

Epigenetic Changes in Barley in Response to Salinity Stress

by

Jason Smith

*Thesis
Submitted to Flinders University
for the degree of*

Doctor of Philosophy
College of Science and Engineering
December 2020

Table of contents

Table of figures.....	v
Table of tables.....	viii
Declaration.....	xi
Abstract.....	xii
Acknowledgements.....	xiv
Abbreviations.....	xvi
1 Introduction.....	1
1.1 Stress in plants.....	2
1.2 Epigenetics in plants.....	3
1.2.1 Small RNA.....	4
1.2.2 DNA methylation.....	7
1.2.3 Detection of DNA methylation.....	9
1.2.4 The RNA directed DNA Methylation (RdDM) pathway.....	14
1.2.5 Silencing of Transposable Elements via the RdDM pathway.....	16
1.2.6 Histone modification and chromatin silencing.....	17
1.2.7 Epigenetic changes in plants under abiotic stress.....	20
1.2.8 Priming plants for abiotic stress.....	22
1.3 Hypotheses.....	23
1.4 Project aims.....	23
2 Barley under salinity stress.....	24
2.1 Introduction.....	24
2.2 Methods.....	25
2.2.1 Seed growth and collection.....	25
2.2.2 Plant growth conditions.....	25
2.2.3 Salinity stress treatment.....	26
2.2.4 Tissue harvesting.....	28
2.3 Results.....	29
2.4 Discussion.....	31
2.4.1 Seed growth and collection.....	31
3 Messenger RNA changes in barley under salinity stress.....	35
3.1 Introduction.....	35

3.2	Methods	38
3.2.1	Messenger RNA library preparation	38
3.2.2	Bioinformatics	39
3.3	Results	40
3.3.1	RNA quality and quantity	40
3.3.2	Differential expression of mRNA	42
3.3.3	Gene Ontology analysis	43
3.4	Discussion	49
3.4.1	Conclusion	59
4	Small RNA changes in barley under salinity stress	60
4.1	Introduction	60
4.2	Methods	63
4.2.1	Small RNA library preparation	63
4.2.2	Bioinformatics	64
4.3	Results	66
4.3.1	Small RNAs guiding DNA methylation	66
4.3.2	Identification of previously described microRNAs	71
4.3.3	Identification of microRNA not previously identified in <i>Hordeum vulgare</i>	73
4.3.4	Differential expression of miRNA	73
4.4	Discussion	82
4.4.1	Identification of previously categorised miRNA	84
4.4.2	Identification of miRNA not previously identified in <i>Hordeum vulgare</i> ..	84
4.4.3	Differential expression of miRNA	85
4.4.4	Conclusion	89
5	DNA methylation changes in barley under salinity stress	90
5.1	Introduction	90
5.1.1	EpiGBS	92
5.2	Methods	95
5.2.1	DNA extraction	95
5.2.2	EpiGBS	96
5.2.3	First bisulphite sequencing run	98
5.2.4	Second bisulphite sequencing run	99

5.2.5	Bioinformatics	100
5.3	Results	104
5.4	Discussion.....	117
5.4.1	DNA extraction.....	117
5.4.2	EpiGBS.....	117
5.4.3	Bioinformatics	120
6	Associations between small RNA, DNA methylation, and mRNA changes in barley under salinity stress	131
6.1	Introduction.....	131
6.1.1	Hypothesis.....	133
6.2	Methods	133
6.3	Results	135
6.3.1	DNA methylation and 23/24 nt sRNA.....	135
6.3.2	Differential DNA methylation and mRNA differential expression.....	142
6.4	Discussion.....	146
6.4.1	23/24 nt sRNA vs DNA methylation.....	146
6.4.2	DNA methylation vs mRNA expression.....	152
6.4.3	Conclusion	154
7	General discussion.....	155
7.1	Novel discoveries and developments from this research.....	159
7.1.1	New and novel micro RNA involved in post-transcriptional repression	159
7.1.2	Development of bioinformatics tools	160
7.1.3	Links between 23/24 nt sRNAs and DNA methylation	160
7.2	Limitations.....	160
7.2.1	Small RNA	160
7.2.2	Multi-mapped methylation sequence reads.....	162
7.2.3	DNA methylation	162
7.3	Future Directions	164
7.3.1	Small RNA	164
7.3.2	DNA methylation	166
7.3.3	Transposable Elements	168
7.4	Conclusion.....	169

8	References	171
9	Appendix A, Supplementary figures and tables	191
10	Appendix B, Custom bioinformatics code.....	245

Table of figures

Figure 1.1. MicroRNA hairpin-loop precursor and miRNA duplex with 5p and 3p strands.	6
Figure 1.2. Bisulphite conversion of an unmethylated cytosine to a uracil, adapted from Tollefsbol (2017).....	12
Figure 1.3. The RNA Directed DNA Methylation pathway.....	15
Figure 1.4. Activity of Transposable Elements enabled by stress induced inhibition of RNA-directed DNA Methylation (RdDM).....	18
Figure 1.5. Initiation and reinforcement of Transposable Element (TE) silencing.....	19
Figure 2.1. Top view diagram of plant positioning in the growth cabinet for the barley salinity stress experiment.	27
Figure 2.2. Plant positional rotation scheme for barley salinity stress experiment.	27
Figure 2.3. Barley plants 21 days after sewing, prior to tissue sampling.	30
Figure 2.4. Comparison of biomass for control and salt treated plants.....	30
Figure 2.5. Evaporation/transpiration loss per day for each pot in the growth cabinet.	31
Figure 3.1. Principle component analysis of mapped messenger RNA reads from salt stressed barley and control plants.....	41
Figure 3.2. Venn diagram of differentially expressed genes in salt treated plants relative to control plants.	43
Figure 3.3. Normalised transcript abundance heat-map, comparing control and salt treated barley plants in (A) third leaf tissue and (B) root tissue.	44
Figure 4.1. Relative abundance of sequenced small RNAs according to read size.....	68
Figure 4.2. Relative abundance of unique small RNAs according to read size.....	69
Figure 4.3. Percentage of significantly differentially expressed 23-24 nt sRNAs targeting genomic features in salinity stressed barley plants.	70
Figure 4.4. Principle component analysis showing separation of barley miRNA mapped to miRBase under 100 mM salinity stress for leaf and root tissue.	71
Figure 4.5. Principle component analysis showing separation of barley miRNAs not represented in miRBase, under 100 mM salinity stress for leaf and root tissue.....	74
Figure 4.6. Sequences and secondary structure of hvu-new-miR396a.	76
Figure 4.7. Hairpin-loop structure of hvu-new-miR477 precursor.	77
Figure 4.8. Sequences and secondary structure of hvu-new-miR528.	78

Figure 4.9. Sequences and secondary structure of hvu-novel-miR90000.....	79
Figure 4.10. Sequences and secondary structure of hvu-novel-miR90001.....	79
Figure 4.11. Sequences and secondary structure of hvu-novel-miR90002.....	80
Figure 4.12. Sequences and secondary structure of hvu-novel-miR90003.....	81
Figure 4.13. Sequences and secondary structures for hvu-novel-miR90004.....	82
Figure 5.1. Sequence of adapters ligated to digested DNA in the epiGBS protocol....	93
Figure 5.2. Stages of the nick repair process to replace one strand of the adapter after ligation in the epiGBS method.....	94
Figure 5.3. Bisulphite treatment, PCR amplification and identification of Watson and Crick strands in the epiGBS method.....	94
Figure 5.4. Capillary electrophoresis results after libraries re-suspended before sequencing.....	106
Figure 5.5. GC content of all sequence reads before and after PhiX removal.....	106
Figure 5.6. Read counts per sample for barcode recognised epiGBS reads of control and salt treated barley plants.....	107
Figure 5.7. Percentage of cytosine methylation in barley grouped by CG, CHG or CHH context for each chromosome using 1,000,000 bp windows.....	108
Figure 5.8. Principle Component Analysis of barley DNA methylation percentage in 100 bp windows for all samples categorised by cytosine contexts.....	110
Figure 5.9. Methylation pattern around Protein Coding Genes (PCGs) in barley.....	111
Figure 5.10. Methylation pattern around Transposable Elements (TEs) in barley.	112
Figure 5.11. Methylation pattern in CHH context cytosines around two selected TE classes in barley.....	113
Figure 5.12. Percentage of differentially methylated sites in barley under 100 mM salinity stress associated with genomic functions based on cytosine context and tissue type.....	115
Figure 5.13. Frequencies of significantly differentially methylation locations relative to the Transcription Start Site (TSS) in barley in salt treated plants vs control.	116
Figure 5.14. Bismark processing of short insert reads by removal of adapters.....	123
Figure 6.1. Quantity of targets of differentially expressed 23 nt miRNA, 24 nt siRNA, and differential methylation CHH sites mapped to 1 Mbp windows in barley under salt stress.....	136
Figure 6.2. Proximity of differentially methylated sites and targets of differentially expressed sRNA in barley under salinity stress.....	138

Figure 6.3. Genomic features of methylated sites where the differential methylation co-locates with differential 23/24 nt sRNA expression within two 100 bp windows.	139
Figure 6.4. Scatter plot of differentially methylated CHH context sites associated with differentially expressed mRNA in barley under salinity stress.	144
Figure 6.5. Scatter plot of differentially methylated CHH context sites associated with differentially expressed mRNA in barley under salinity stress.	145
Figure 7.1. Schematic diagram of important pathways investigated for changes that occurred when barley was exposed to 150 mM salinity stress in soil vs control plants.	158
Figure A.9.1. Agilent Bioanalyser results for leaf RNA extractions.	214
Figure A.9.2. Agilent Bioanalyser results for root RNA extractions.	215
Figure A.9.3. Methylation patterns in barley around different TE classes.	222
Figure A.9.4. <i>Arabidopsis thaliana</i> methylation pattern across each chromosome and cytosine context, reproduced from Cokus et al. (2008).	223
Figure A.9.5. Methylation patterns around different genomic features in <i>Arabidopsis thaliana</i> , reproduced from Cokus et al. (2008).	224
Figure A.9.6. DNA methylation patterns for all cytosine contexts around genes in three different rice varieties, reproduced from Garg et al. (2015).	225
Figure A.9.7. DNA methylation patterns in wild rice for (A) genes and (B) transposable elements, reproduced from Li et al. (2012)	226
Figure A.9.8. DNA methylation patterns in cassava for all cytosine contexts for genes and Transposable Elements (TEs), reproduced from Wang et al. (2015).	227
Figure A.9.9. Quantity of differentially expressed 23 nt miRNA, 24 nt siRNA, and CG context differential methylation mapped to 1 Mbp windows in barley under salt stress.	236
Figure A.9.10. Quantity of differentially expressed 23 nt miRNA, 24 nt siRNA, and CHG context differential methylation mapped to 1 Mbp windows in barley under salt stress.	237

Table of tables

Table 1.1. Function of dicer-like enzymes in plants and the relationship with small RNA function.....	4
Table 1.2. Some of the more common methods for determining methylation states across the genome in genomic DNA.....	10
Table 1.3. Restriction enzymes <i>HpaII</i> and <i>MspI</i> digestion and cytosine methylation.	11
Table 2.1. Watering and salt treatment schedule for salinity stress experiment.	28
Table 3.1. Extracted RNA concentration and RNA Integrity Number.	40
Table 3.2. Total RNA reads per sample.	40
Table 3.3. Shortlist of the ten genes with the greatest increase or decrease in mRNA expression in leaf and root tissue of salt stressed barley relative to control.	45
Table 3.4. Biological process gene ontology terms overrepresented in differential expression of leaf or root tissue in salinity stressed plants.....	47
Table 3.5. Molecular function gene ontology terms overrepresented in differential expression of leaf or root tissue in salinity stressed plants.....	48
Table 3.6. Cellular component gene ontology terms overrepresented in differential expression of leaf or root tissue in salinity stressed plants.....	49
Table 3.7. Comparing GO term results that are common with prior research.	58
Table 3.8. Quantity of GO terms and common terms across different studies.....	59
Table 4.1. Sequence read numbers for small RNA sequencing of barley before and after filtering of reads.	67
Table 4.2. MicroRNA found in barley root or leaf tissue under control or salt treatment conditions that have already been identified in miRBase.	72
Table 4.3. New and novel miRNA found in barley.	73
Table 4.4. Differentially expressed miRNA in leaf and root tissue in barley under salinity stress.....	75
Table 4.5. List of species with an exact match for hvu-new-miR396a when comparing the 5' or 3' strand.....	76
Table 4.6. Selection of blast search results comparing hvu-new-miR477 with miRBase.	77
Table 4.7. Selection of blast search results comparing hvu-new-miR528 with miRBase.	78
Table 4.8. Blast search results comparing hvu-novel-miR90002 with miRBase.	80

Table 5.1. Extracted DNA concentration from barley plant leaf and root tissue samples.....	105
Table 5.2. Total sequence reads and PhiX spike in reads for epiGBS sequencing runs.	105
Table 5.3. Differentially methylated 100 bp windowed reads mapping to sites in the proximity of Protein Coding Genes (PCGs).....	114
Table 5.4. Chi squared analysis of Protein Coding Gene (PCG) count associated with differential methylation against random sampling of the bisulphite-mapped genome to determine overrepresentation.	114
Table 6.1. Number of Protein Coding Gene (PCG) sites targeted by differential methylation co-located with sRNA within two 100 bp genomic windows in barley under salt stress.....	140
Table 6.2. Fold difference of Protein Coding Genes (PCGs) and Transposable Element (TE) superfamilies targeted by differential methylation co-located with sRNA within two 100 bp windows when compared with random sampling.	141
Table 6.3. Correlations between differential methylation and mRNA differential expression in different tissue and cytosine contexts in barley under salt stress. ..	143
Table A.9.1. Differentially expressed genes in salt stressed barley leaf tissue.	191
Table A.9.2. Differentially expressed genes in salt stressed barley root tissue.	199
Table A.9.3. Complete biological process gene ontology hierarchy for leaf and root tissue in salt stressed barley.	208
Table A.9.4. Complete molecular function gene ontology hierarchy for leaf and root tissue in salt stressed barley.	211
Table A.9.5. Complete cellular component gene ontology hierarchy for leaf and root tissue in salt stressed barley.	213
Table A.9.6. Illumina barcodes and corresponding samples for small RNA sequencing.	216
Table A.9.7. Listing of previously undocumented mature miRNA in barley.	217
Table A.9.8. Listing of previously undocumented hairpin-loop precursors for miRNAs in barley	219
Table A.9.9. EpiGBS barcode combinations used with biological samples.	221
Table A.9.10. Differentially methylated CG context genes in barley leaf tissue under salt stress.	228
Table A.9.11. Differentially methylated CHG context genes in barley leaf tissue under salt stress.	229

Table A.9.12. Differentially methylated CHH context genes in barley leaf tissue under salt stress.	230
Table A.9.13. Differentially methylated CG context genes in barley root tissue under salt stress.	231
Table A.9.14. Differentially methylated CHG context genes in barley root tissue under salt stress.....	233
Table A.9.15. Differentially methylated CHH context genes in barley root tissue under salt stress.....	235
Table A.9.16 List of Protein Coding Genes (PCGs) targeted by differential 23/24 nt sRNA expression correlated with CHH context differential DNA methylation in leaf tissue in barley under salt stress.....	238
Table A.9.17 List of Protein Coding Genes (PCGs) targeted by differential 23/24 nt sRNA expression correlated with CHH context differential DNA methylation in root tissue in barley under salt stress.....	242
Table A.9.18. Log ₂ fold difference between Protein Coding Genes (PCGs) and Transposable Element (TE) superfamilies targeted by differential methylation correlated with sRNA within two 100 bp windows when compared with random sampling.....	244

Declaration

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

Jason Lyndon Paul Smith

5 October 2020

Abstract

Epigenetic mechanisms are potentially heritable molecular changes that affect gene expression, leading to differences in phenotype without changing the DNA sequence of the organism. In plants, such mechanisms are involved in the control of a range of processes, including response to stress. DNA methylation is an epigenetic mechanism used by organisms to adapt to changing environmental conditions by altering localised accessibility of the genome to transcription factors, thus ultimately affecting gene expression levels. The sequence and organ specificity of this stress induced de-novo DNA methylation in plants is guided by a class of small RNAs (sRNAs), typically 23 to 24 nt long.

This project uses epiGBS, a reduced representation genome bisulphite sequencing method, coupled with small RNA and whole transcriptome Next Generation Sequencing to investigate interactions between sRNAs, DNA methylation and gene expression in the leaves and roots of barley under salt stress. EpiGBS was chosen as a method for gaining information about the methylation states of the genome since the barley genome is large, at 5.1 Gbp, making whole genome sequencing prohibitive. The method reduces the complexity of the genome by restriction enzyme fragmentation and sequencing based on fragment size selection, with genome coverage scalable by size selection and number of reads. *Hordeum vulgare* (barley) was chosen for this study as it is the fifth most important crop in global agriculture

with good coverage of genome sequence information. It is anticipated that findings in this research should be applicable to other cereal crops such as wheat.

Analysis of small RNA sequence data identified 59 new *H. vulgare* microRNAs (miRNAs) and corresponding precursor hairpin-loop sequences found that have not been previously reported in miRBase. Of these newly discovered miRNAs, 44 did not have sequence similarity to any previously identified miRNAs in other plant species and 15 were similar to known miRNAs in other plant species. Eight of these newly discovered miRNAs correlated with salinity stress and are likely to be involved in stress response. DNA methylation changes were found in response to salinity stress, with 1,210 loci in leaf tissue and 513 loci in root tissue found corresponding to a change in 23/24 nt sRNA expression that targets a protein coding gene. The linkages between sequence information from DNA methylation, small RNA and the transcriptome will lead to a greater understanding of how this crop deals with this important stress, and provide an extremely useful avenue for further research.

Acknowledgements

I would like to acknowledge the contributions of, and thank the AJ & IM Naylon trust for providing the base scholarship for this PhD research, the Grains Research and Development Corporation for providing a top-up scholarship and research funds that made the extensive sequencing possible, and the Playford trust for providing valuable financial support.

My partner Mahnaz has been extremely supportive and accommodating, giving me access to the time I needed to deal with all the bioinformatics obstacles that seemed to be constantly popping up, and the space to write with minimal interruptions, I am enormously grateful. Thanks to my mum and Mahnaz's mum for looking after our young son Samyar, and thanks to Samyar for being a pleasure to be with. Thanks to friends and family for understanding my relative isolation from the world while completing this thesis.

My supervisors have been very useful in the particular roles they took on during the progress of this project and I am very grateful. Peter Anderson provided great feedback, kept a positive outlook when things were not turning out well, and importantly kept the objectives realistic, so that objectives could actually be achieved. Michael Michael provided useful feedback, and valuable advice related to sequencing and some of the early bioinformatics issues I ran into. Yuri Shavrukov

has extensive experience in abiotic stress and cereal crops and was very helpful in early experiments with hydroponic and soil experiments. While most of those experiments are not covered in this thesis, his help was valuable in getting plants growing in a reliable and repeatable fashion. Carlos Rodríguez López was involved later in this project, but his contributions were enormous. His extensive knowledge of epigenetic modifications pointed me in the right direction. He provided the space, time, advice, and connections needed to solve the problems encountered with the wet-lab portion of epiGBS. Thanks also to Carlos for all the extremely useful feedback provided with regard to the thesis. The regular lab group meetings at both Flinders and Waite were useful and I am grateful for the discussions and suggestions coming from lab group members. I appreciate the help and patience that everyone involved has provided.

Abbreviations

bp	base pair
ChIP	Chromatin Immunoprecipitation
CPU	Central Processing Unit
DAS	Days After Sewing
DNA	Deoxyribonucleic acid
GO	Gene Ontology
LTR	Long Terminal Repeat
miRNA	Micro RNA
mRNA	Messenger RNA
MSAP	Methylation Sensitive Amplified Polymorphism
ms-GBS	Methylation Sensitive Genotyping By Sequencing
MW	Molecular Weight
nt	nucleotide
PCA	Principle Component Analysis
PEG	Polyethylene Glycol
PVP	Polyvinylpyrrolidone
siRNA	Small Interfering RNA
sRNA	Small RNA
RdDM	RNA directed DNA Methylation
RIN	RNA Integrity Number
RISC	RNA Induced Silencing Complex
RNA	Ribonucleic acid
ROS	Reactive Oxygen Species
RRBS	Reduced Representation Bisulphite Sequencing
SEM	Standard Error of the Mean
TE	Transposable Element
TES	Transcription End Site
TSS	Transcription Start Site
UMI	Unique Molecular Identifier

1 Introduction

Human food supply is severely affected by abiotic stress in plants, such as drought, salinity and heat. With continuing population growth and a decrease in arable land, there is an increasing need to effectively manage these stresses in the future (Tilman et al. 2002; Godfray et al. 2010). By understanding how plants cope with stress, we can potentially utilise and optimise these stress-coping mechanisms to maintain yield, and therefore food supply.

Research in stress tolerance of plants has demonstrated the important role that epigenetics plays in plant stress response. Epigenetics involves changes in genomic markers without changes to the DNA sequence, which can regulate gene expression levels and manage genome stability, with some epigenetic information being passed onto subsequent generations (Danchin et al. 2011). Environmental factors such as abiotic stress can result in changes in epigenetic states, allowing the plant to better cope with the environmental conditions (Boyko & Kovalchuk 2011).

Barley (*Hordeum vulgare*) is an extremely important crop in global agriculture, ranking fifth in global production (Mayer et al. 2011). Barley is an economically important crop in Australia and is increasingly affected by dryland salinity which decreases crop yield (Patterson et al. 2009). Given this, the research presented here aims to determine the role that epigenetic mechanisms play in the salt stress response

system in barley.

1.1 Stress in plants

Stress in agricultural crops can cause significant loss in yield. It can be defined as biotic (caused by biological disease agents such as fungi, bacteria, viruses and insects), and abiotic (caused by environmental factors). The most widespread abiotic stresses include drought and salinity (Vinocur & Altman 2005), with global climate change expected to increase problems associated with these stress. Other abiotic stresses include heat, cold, flooding and nutrient deficiencies or toxicities. Salinity is a major problem for yield loss, and production of salt tolerant crops is extremely important for global food security in the future (Pitman & Läuchli 2002).

Stress in plants has been shown to stimulate homologous recombination, an effect which can be passed on to progeny which were not exposed to the stress (Pecinka et al. 2009). Tricker et al. (2012) found that application of stress in young *Arabidopsis thaliana* plants resulted in greater tolerance to the same stress in later development, with epigenetic changes involved in the stress response. While some of the stress response mechanisms in plants are known, much of the currently unknown mechanisms may be related to regulation by epigenetic factors.

1.2 Epigenetics in plants

Epigenetic mechanisms have been defined broadly as “the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states” (Bird 2007). Certain molecules attached to specific regions of the genome can change gene expression levels, or genome stability, based on environmental factors, with some of the changes being passed down across multiple generations (Danchin et al. 2011). Environmental factors such as stress can result in changes in epigenetic states, potentially allowing a plant and its progeny to cope with the stress inducing conditions (Boyko & Kovalchuk 2011).

Epigenetic mechanisms are observed in plants, animals and fungi, with some differences observed between the implementation in these three evolutionary branches (Feng & Jacobsen 2011). It has been proposed that the original function of epigenetics was to manage selfish DNA (which could harm the host organism), and adapted to perform other regulatory functions (Slotkin & Martienssen 2007). Modifications to the epigenome can involve DNA methylation, chromatin structure changes and small interfering RNA (siRNA) (Feng & Jacobsen 2011). In plants, 24-nt siRNAs play a significant role in activation and maintenance of epigenetic signalling (Matzke et al. 2007).

1.2.1 Small RNA

There are various small RNAs (sRNAs) in plants that regulate various aspects of gene expression. Some of the sRNAs can be categorised by the molecular machinery that cuts the double-stranded RNA. The dicer-like class of enzymes cleave double-stranded RNA for different functions as shown in Table 1.1. All four of these proteins evolved prior to the evolutionary split between monocotyledonous and dicotyledonous plants (Henderson et al. 2006). According to their sizes, the main classes of small RNAs in plants are microRNAs (miRNAs) and siRNAs.

Table 1.1. Function of dicer-like enzymes in plants and the relationship with small RNA function.

Dicer-like protein	Average size of cleaved RNA	Function of small RNAs
DCL1	21 nt	Post-transcriptional silencing
DCL2	22 nt	Viral resistance
DCL3	24 nt	Guiding DNA methylation
DCL4	21 nt	Post-transcriptional silencing

Some small RNAs have been isolated and sequenced from barley. Schreiber et al. (2011) found small RNA sequences in two barley cultivars (Golden Promise and Pallas), finding that 24 nt siRNA made up about half of the unique short RNA (18 nt to 24 nt) reads. Lv et al. (2012) exposed seedlings of the barley cultivar Clipper to Poly Ethylene Glycol (PEG) and NaCl to simulate drought and salt stress respectively. Small RNAs were extracted from leaf tissue at various stages in treated and non-treated control plants. Some small RNAs were found to be involved in stress tolerance.

The function of sRNAs in plants in relation to DNA methylation control is different to animals (Law & Jacobsen 2010), with 23 nt to 24 nt sRNAs guiding DNA methylation in plants (Matzke et al. 2007). *De novo* methylation is established by the RNA directed DNA methylation (RdDM) pathway and is discussed in section 1.2.4.

MircoRNAs in plants are typically 21 nt in size, but can range in size from 20 nt to 24 nt. They function as a control mechanism in plant gene expression, mostly targeting transcription factors and proteins related to stress response, development, growth and physiology (Rogers & Chen 2013). The miRNA is created through a transcript generated by RNA polymerase II (Pol II). This primary (pri-) miRNA transcript is processed into a stem-loop precursor (pre-) miRNA structure, then a small RNA duplex with 3' overhangs by a complex of proteins including DICER-LIKE 1 (DCL1), HYPONASTIC LEAVES 1 (HYL1), DOUBLE-STRANDED RNA BINDING 1 (DRB1) and SERRATE (SE) (Moro et al. 2018; Wang et al. 2019). The 2' hydroxyl group at the 3' ends are methylated by HUA ENHANCER 1 (HEN1) which decreases the rate of degradation of the miRNA (Yu et al. 2005). Without this methylation modification, uracil bases are added to the 3' end (Li et al. 2005), signalling the miRNA for degradation (Ji & Chen 2012).

One strand of the microRNA duplex is loaded in to the RNA Induced Silencing Complex (RISC), this is often the strand with lower thermodynamic stability at the 5' end (Schwab et al. 2006). The strand that is loaded in to the RISC can be called the

guide strand and the discarded strand the passenger strand (Schwab et al. 2006), but more recently the strands are now defined as the 5p strand and the 3p strand based on the position of the miRNA in the hairpin loop precursor (Figure 1.1), with the 5p strand usually loaded in to the RISC (Kozomara & Griffiths-Jones 2013). The miRNA strand loaded in to the RISC anneals with a partial sequence match from a messenger RNA (mRNA) strand and causes degradation or translational repression of that mRNA strand, decreasing the expression of the targeted gene (Huntzinger & Izaurralde 2011). Generally there is not perfect complementarity between the miRNA strand loaded in to the RISC and the mRNA target. Typically, in plants there is perfect complementarity between bases 2 to 12 and up to 5 mismatches with the target mRNA. There are normally one or two mismatches between the miRNA and the mRNA between bases 17 to 21 to prevent double stranded extension of the mRNA/miRNA pairing by RNA dependent RNA polymerase (Moissiard et al. 2007; Ossowski et al. 2008).

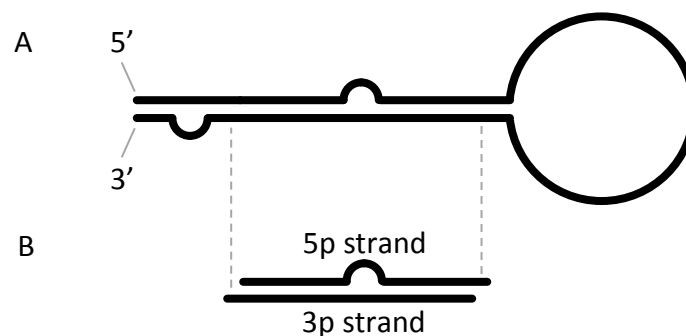


Figure 1.1. MicroRNA hairpin-loop precursor and miRNA duplex with 5p and 3p strands. A, Hairpin-loop miRNA precursor with 5' and 3' ends labelled, is cleaved to form the miRNA duplex. B, The miRNA duplex with the 5p strand and 3p strand definition related to the hairpin-loop precursor that it originated from. Exact sequence complementarity between strands is not necessary for miRNA and can result in small sections of the miRNA duplex that do not pair.

1.2.2 DNA methylation

DNA methylation occurs when a methyl group is added to the 5' carbon in a cytosine nucleotide (Martienssen & Colot 2001). In animals, DNA methylation is typically considered to be important at CG sequences, where the methylation is symmetric at both cytosine bases in the double-stranded DNA. In plants methylation can occur at symmetric CG and CHG contexts (where H is an A, C or T), and also at asymmetric sites with a CHH context; all three contexts are considered important in plants (Mirouze & Paszkowski 2011).

When DNA is replicated, the methylation information will not be present on the newly replicated strand. The DNA METHYLTRANSFERASE 1 (MET1) maintains the methylation state of CG sites in plants during DNA replication by fully methylating hemi-methylated CG sites (Saze et al. 2003; Vanyushin & Ashapkin 2011). The DNA methylation maintenance method for CHG sites is related to histone modifications, which are discussed in further detail in section 1.2.6. Methylation in a CHG context is mainly maintained with CHROMOMETHYLASE 3 (CMT3), but also to a lesser extent with CHROMOMETHYLASE 2 (CMT2) (Zhang et al. 2018). These enzymes cause cytosine methylation at CHG sites based on the methylation state of histone H3 tail at lysine 9 (H3K9) (Stroud et al. 2014; Yaari et al. 2019). The methylation at CHH contexts is predominantly maintained by RdDM (discussed further in section 1.2.4), but can also be maintained with CMT2 in large Transposable Elements (TEs) (Stroud et al. 2014; Kawakatsu et al. 2017).

Demethylation of cytosine bases in plants can occur passively through DNA replication without methylation maintenance, or through active removal of the methylated cytosine base (Viggiano & de Pinto 2017). The DNA glycosylase REPRESSOR OF SILENCING 1 (ROS1), removes the methylated cytosine base, and DNA repair mechanisms replace the excised base with an unmethylated cytosine (Li et al. 2018). The mechanism for targeting a specific region for demethylation is not fully understood. A complex called the Increased DNA Methylation (IDM) complex forms and causes histone acetylation which is believed to recruit ROS1 to target the site (Nie et al. 2019). Given this discussion on how methylation and demethylation occur in plants, it is worth considering its role.

DNA methylation has two main functions in plants; genome stability and regulation of gene expression (Castiglione et al. 2010). Transposable Elements can be silenced by DNA methylation (Teixeira et al. 2009). Epigenetic activity of TEs influences regulation of genes near the TE and can modulate stress response. (McCue et al. 2012).

DNA methylation is related to suppression of gene expression, however the link between DNA methylation and mRNA transcript levels is not simple (Bewick & Schmitz 2017). Many cases have been observed where an increase in DNA methylation of a promoter causes lower production levels of mRNA (Berdasco et al. 2008). However there are many documented instances where the opposite is true;

hundreds of genes associated with tomato fruit ripening increase expression as a result of increased methylation at the promoter (Zhang et al. 2018). An increased level of methylation at another site in a gene body has resulted in decreased production of mRNA in one example (Anastasiadi et al. 2018) and increased production of mRNA in a different example (Shibuya et al. 2009). In most cases an increase in methylation results in silencing of the targeted region (Law & Jacobsen 2010). However there appears to be a greater level of complexity involved than being able to observe methylation level changes and having a simple predictable outcome for the expression of the associated gene.

1.2.3 Detection of DNA methylation

There are a number of ways to detect DNA methylation, with differing levels of resolution, with some of the more common methods listed in Table 1.2. Methylation Sensitive Amplified Polymorphism involves the use of isoschizomer restriction enzymes where the same recognition site is used, but with differing ability in cutting if methylated cytosines are present within the enzyme's recognition site. The restriction enzymes *HpaII* and *MspI* are commonly used in this application, and the methylation states that are cleaved are shown in Table 1.3. Adapters are ligated during digestion and act as primers for PCR amplification of a population of products. The amplified fragments are run on a denaturing acrylamide gel and any unique fragments observed between the *HpaII* and *MspI* digested template are cut

Table 1.2. Some of the more common methods for determining methylation states across the genome in genomic DNA.

Method	Technique overview	Resolution
Whole genome bisulphite sequencing	Conversion of unmethylated cytosine (C) to uracil while methylated cytosine (5mC) remains unchanged. Sequencing of modified DNA.	Methylation states of all cytosines that can be mapped to the genome can be known. Large repetitive elements will not be able to be mapped to the genome.
Third generation direct sequencing	Subtle differences in the way 5mC and C are processed is detected with when sequenced with extremely high coverage.	Methylation state information for mapped cytosines, with potentially better mapping than 2 nd generation sequencing due to longer reads. Base calling accuracy requires improvement before matching whole genome bisulphite sequencing.
Reduced Representation Bisulphite Sequencing (RRBS)	Like whole genome bisulphite sequencing but the complexity of the genome is reduced by use of restriction enzymes and fragment size selection.	Only methylation states of fragments selected in size selection process are obtained. This is somewhat scalable with a trade-off on the sequencing reads required and the breadth of fragment selection.
methylation-sensitive Genotyping-By-Sequencing (ms-GBS)	A variant of Methylation Sensitive Amplified Polymorphism (below) that sequences the amplified fragments rather than observing on a gel.	Only methylation states at restriction digest sites can be determined.
Methylation Sensitive Amplified Polymorphism (MSAP)	The genome is digested by restriction enzymes that cut at the same sequence, but have different abilities in cutting a methylated cytosine. The digested DNA is PCR amplified and run on an acrylamide gel with the presence or absence of bands indicating a difference in methylation state in samples.	Only methylation states at restriction digest sites can be determined.

out of the gel and sequenced (Dong et al. 2006; Akimoto et al. 2007; Zhao et al. 2007; Li et al. 2008; Lister & Ecker 2009; You et al. 2012; Wu et al. 2013; Avramidou et al. 2015; Baránek et al. 2015; Sun et al. 2015). Alternatively, if the specific sequence is not required, information can be gained from the presence or absence of specific fragments. The amplified fragments can be separated by capillary electrophoresis to generate a chromatogram (Yaish et al. 2014). This method is quite limited in that only methylation differences at the restriction site sequence can be observed. A variant of this method is methylation-sensitive Genotyping-By-Sequencing (ms-GBS), where digested fragments are sequenced (Xia et al. 2014). This variant is able to identify differential methylation to a much greater degree than the presence or absence of a band on a gel, but still has the limitation of only providing methylation information at the restriction digest site.

Table 1.3. Restriction enzymes *HpaII* and *MspI* digestion and cytosine methylation. Where C is an unmethylated cytosine and C^m is a methylated cytosine. The restriction enzymes are not perfect; a small amount of digestion will occur when the site would normally remain undigested and vice-versa.

Methylation state of sequence	Restriction enzyme <i>HpaII</i> ability to digest	Restriction enzyme <i>MspI</i> ability to digest
C C G G	digested	digested
C ^m C G G	undigested	digested
C C ^m G G	undigested	undigested
C ^m C ^m G G	undigested	undigested

The bisulphite conversion method modifies an unmethylated cytosine to a uracil as shown in Figure 1.2, while a methylated cytosine will remain unmodified as the sulphonation reaction cannot occur. Control of the pH is important for progression of the equilibrium reactions to convert almost all unmethylated cytosine bases to uracil

(Howkit et al. 2017). The bisulphite converted DNA can then be amplified by PCR, during which uracils are copied as thymines before sequencing. It is possible to determine the methylation state of the genome when compared to a library that is not bisulphite treated or a reference genome. Any C to T conversion indicates an unmethylated cytosine, and an unconverted C indicates a methylated cytosine (Lizardi et al. 2017).

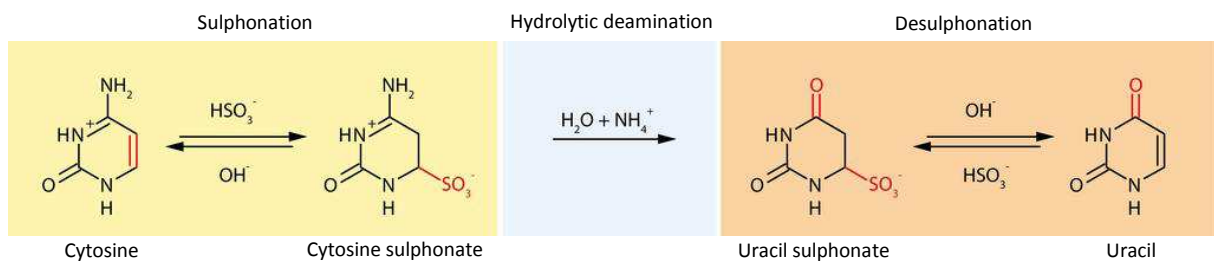


Figure 1.2. Bisulphite conversion of an unmethylated cytosine to a uracil, adapted from Tollefsbol (2017). Cytosine on the left is converted to uracil on the right via two intermediate molecules, cytosine sulphonate and uracil sulphonate. Adjustment of the pH of the reactions results in an equilibrium with uracil as the final product when a low pH in the sulphonation stage and a high pH in the desulphonation stage is used.

The bisulphite treatment can be utilised in multiple ways. Either the methylation state of the entire genome can be determined, that of a reduced representation of the genome, or a melt-curve analysis can be performed for a specific region.

The melt-curve analysis compares the dissociation temperature of PCR amplified bisulphite treated and untreated DNA. It is possible to measure the level of methylation between different samples by bisulphite treatment, PCR and then comparing the dissociation temperature by observing the change in fluorescence

with small increments in temperature at the end of the PCR cycle. A shift in the peak change in fluorescence indicates a difference in dissociation temperature. A lower dissociation temperature indicates a higher AT content, which results from a higher level of unmethylated cytosine (Guldborg et al. 2002).

Reduced Representation Bisulphite Sequencing (RRBS) methods combine the use of non-methylation sensitive restriction enzymes, bisulphite treatment, restriction fragment size selection, and Next Generation Sequencing to reduce the complexity of the data obtained. Reduced representation is achieved as the size selection of restriction products retain a relatively small number of DNA fragments that are shared by all samples (Meissner et al. 2005). This method does not account for single nucleotide polymorphisms, insertions, deletions or TE activity differences between samples at the restriction digestion sites. What this approach does allow is the identification of DNA methylation differences between multiple samples, and at a lower cost to that of complete genome sequencing. The level of representation in the genome is somewhat scalable, as the size selection step can potentially be tuned to deliver a desired level of coverage across the genome (Wang et al. 2012).

More recently, two types of third generation sequencing techniques have become available with direct sequencing potential, without the need for bisulphite treatment of DNA. These methods are currently not as reliable or cost-effective as bisulphite sequencing, but the technology continues to develop and may be a viable alternative

for some situations, particularly if long reads are required. The Oxford Nanopore MinION measures the changes in the ionic current as a single DNA strand is pushed through a tiny pore (Jain et al. 2016). It is able to detect and correctly call a 5-methylcytosine with an accuracy ranging from 83% to 91%. The higher accuracy is achieved when a more strict quality control limit is applied and 32% of the base calls are discarded (Simpson et al. 2017). The PacBio SMRT sequencer can also obtain cytosine methylation information directly from sequencing, however a minimum coverage of 250x is required to detect cytosine methylation (Liu et al. 2020).

Other techniques exist that can determine the methylation states of specific sequences of interest, and are summarised by Šestáková et al. (2019). Global methylation levels can be compared with various techniques detailed by Kurdyukov & Bullock (2016). Such techniques are less relevant for this research as this study is looking to identify and locate various changes in methylation states across the genome.

1.2.4 The RNA directed DNA Methylation (RdDM) pathway

The RNA directed DNA Methylation (RdDM) pathway is the epigenetic pathway in plants where de novo methylation is guided by small RNA (Figure 1.3). This pathway involves two RNA polymerases, Pol IV and Pol V, which are only found in plants (Matzke & Mosher 2014). The RdDM pathway involves production of

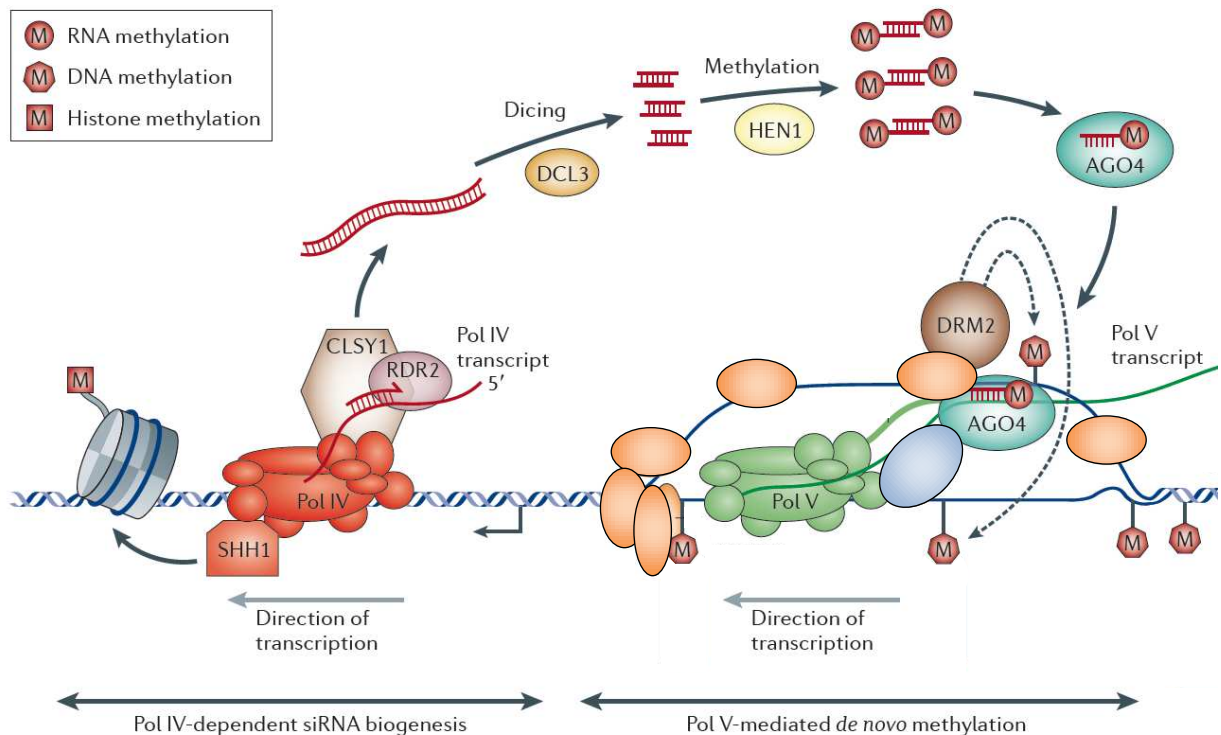


Figure 1.3. The RNA Directed DNA Methylation pathway. RNA polymerase IV (Pol IV) dependent siRNA is generated when SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1) recognises methylation state of the histone 3 tail at lysine 9 (H3k9) and recruits Pol IV for transcription. The Pol IV transcript is converted to double-stranded RNA (dsRNA) with RNA DEPENDENT RNA POLYMERASE 2 (RDR2). DICER-LIKE 3 (DCL3) cleaves the dsRNA in to 24 nt siRNA which is then exported to the cytosol. The 3' ends are methylated by HUA ENHANCER 1 (HEN1) before the siRNA is loaded in to ARGONAUTE 4 (AGO4) and imported back in to the nucleus. RNA polymerase V (Pol V) produces a single-stranded transcript. When the siRNA loaded in to AGO4 matches with the Pol V transcript, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) causes the *de novo* methylation of the DNA. Reproduced with permission: Matzke & Mosher (2014).

transcripts from Pol IV which are made in to double stranded RNA (dsRNA) by RNA DEPENDENT RNA POLYMERASE 2 (RDR2). This dsRNA is then cleaved to 24 nt siRNA with DICER-LIKE 3 (DCL3) (Zhang & Zhu 2011) and exported to the cytoplasm where the 3' ends are methylated by HUA ENHANCER 1 (HEN1) to prevent degradation (Matzke & Mosher 2014). Once a siRNA is loaded into ARGONAUTE 4 (AGO4), it is re-imported to the nucleus where the siRNA pairs

with a complementary section of scaffold transcript produced by Pol V. This pairing results in methylation of cytosine by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Lister et al. 2008). This *de novo* methylation function can be performed in all (CG, CHG and CHH) sequence contexts (Zhang & Zhu 2011; Matzke & Mosher 2014).

1.2.5 Silencing of Transposable Elements via the RdDM pathway

Transposable Elements, found in both prokaryotes and eukaryotes, are self-replicating entities in the genome with the ability to move or copy their sequence throughout the genome (Kim 2017). Historically they had been considered parasitic genomic elements, but more recently have been found to play a significant role in evolution (Makałowski et al. 2019) and also are an important component in plant stress response (Horváth et al. 2017).

There are two top-level classes of TEs, the retrotransposons which utilise RNA transposition and reverse transcription, and the DNA transposons which do not. These are further broken down in classification based on structural patterns that make up the sequence such as terminal repeats, coding regions and non-coding regions (Wicker et al. 2007). During insertion, the TEs can be incorporated in various sections of the genome, which can result in harmful, neutral or beneficial changes (Rebollo et al. 2012). The insertion location of some TEs can target regions close to genes and can have an impact on the regulation of the targeted gene (Galindo-

González et al. 2016). A group of TEs called the Long Terminal Repeat (LTR) retrotransposons have been found to have a preference for insertion within other similar TEs or non-functional TE fragments (Wei et al. 2013; Ma et al. 2019). The LTR retrotransposons are the most abundant TE in plants and have resulted in significant changes to their host genomes (Sabot & Schulman 2006). The ability of TEs to move or replicate throughout the genome is heavily regulated by DNA methylation (Lisch 2009).

The initiation of epigenetic silencing of active TEs is triggered by 21-22-nt siRNA and may be an important part of stress responses in plants (Nuthikattu et al. 2013). Figure 1.4 shows how stress can inhibit RdDM, enabling TE insertion events which can silence or permanently disrupt gene expression (Matzke & Mosher 2014). There is evidence that an un-silenced TE is transcribed by Pol II with RDR6 generating double stranded RNA which is then cut into 21-nt siRNA by DCL4 or 22-nt siRNA by DCL2 (Figure 1.5). The 22-nt siRNA guides DNA methylation, and the methylated DNA causes production of 24-nt siRNA to reinforce methylation of the TE (Nuthikattu et al. 2013).

1.2.6 Histone modification and chromatin silencing

Chromatin is the complexing of genomic DNA with histone proteins related to DNA packaging, protection and gene expression regulation. Histone modification in plants can be initiated by DNA methylation and can modulate the accessibility of genomic

DNA. Chromatin changes occur in response to drought, salinity, heat, and cold stresses in plants (Kim et al. 2015). There are a number of different histone modifications that can increase or decrease expression levels with the most common forms in plants being methylation, acetylation, phosphorylation and ubiquitination (Pfluger & Wagner 2007).

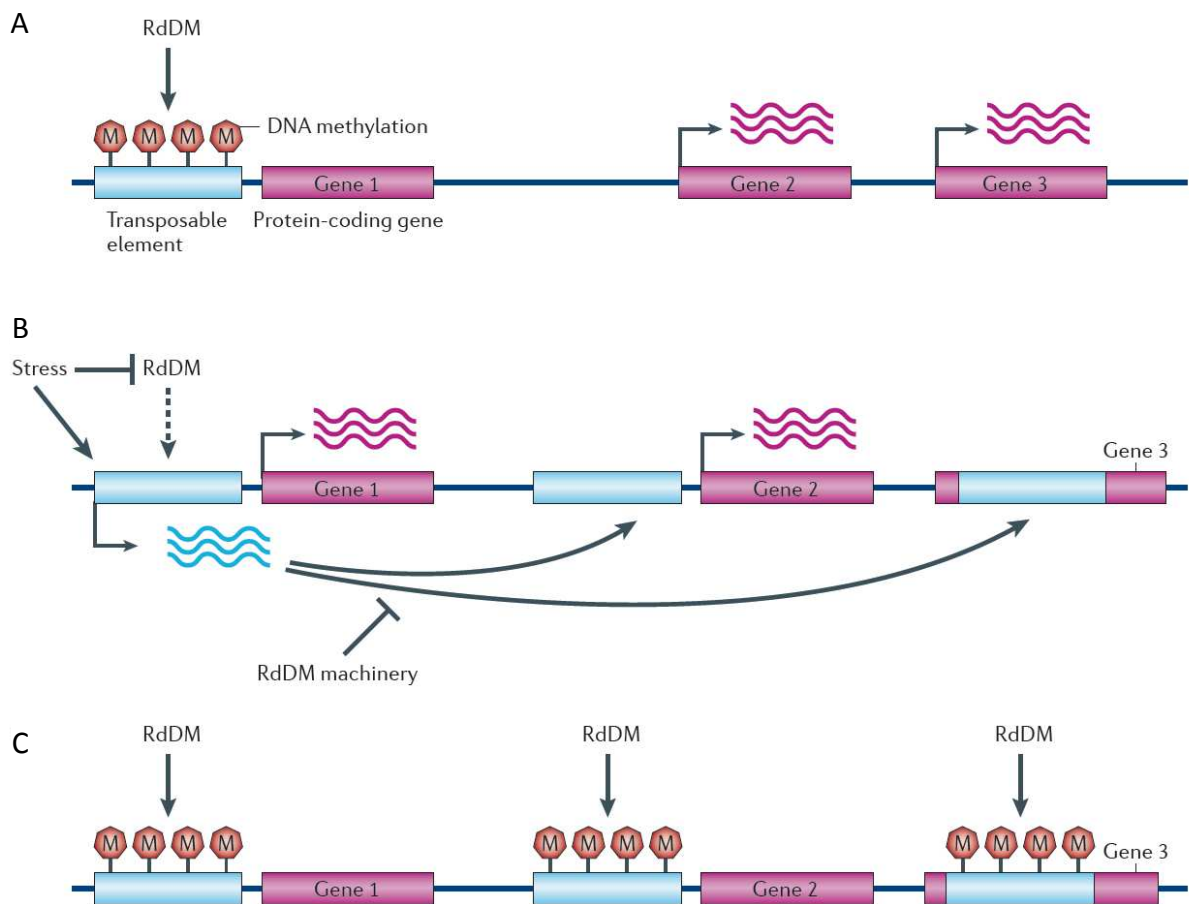


Figure 1.4. Activity of Transposable Elements enabled by stress induced inhibition of RNA-directed DNA Methylation (RdDM). A, A Transposable Element (TE) is prevented from replication by DNA methylation established and maintained by RdDM. B, A stress event prevents RdDM from maintaining methylation at the TE which is able to replicate to certain locations in the genome. C, RdDM re-establishes methylation which suppresses TE replication and gene expression remains altered as a result. Reproduced with permission: Matzke & Moshier (2014).

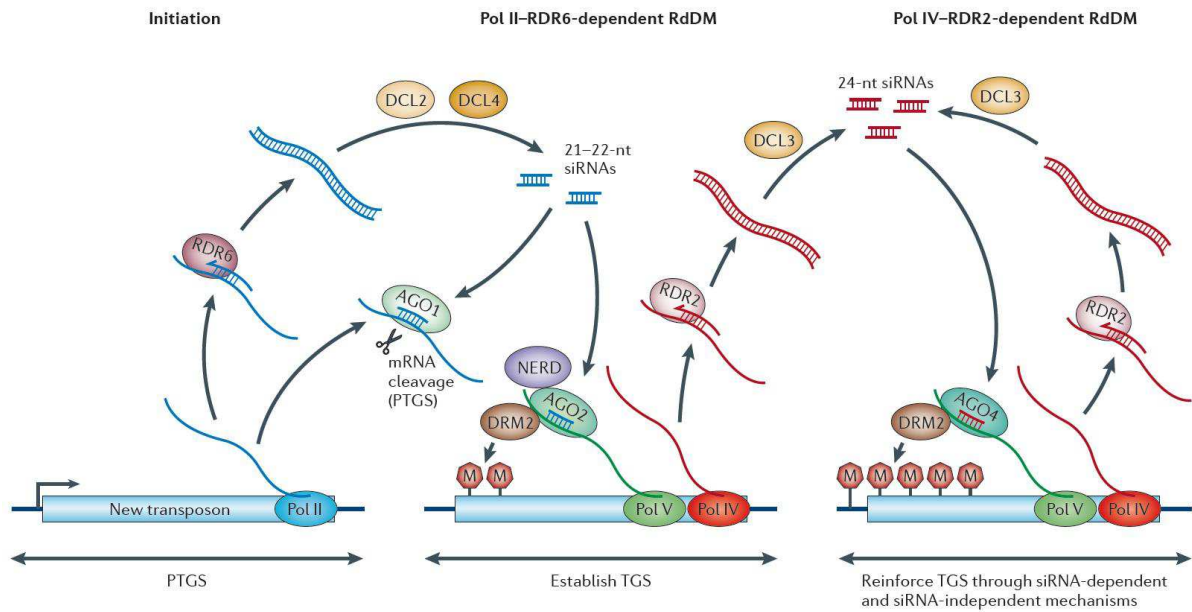


Figure 1.5. Initiation and reinforcement of Transposable Element (TE) silencing. RNA Polymerase II (Pol II) produces a transcript from the TE which is made double-stranded with RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) and cleaved by DICER-LIKE 2 and 4 (DCL2 and DCL4) into 21 nt to 22 nt siRNAs. New Pol II transcripts are then cleaved when the siRNA is loaded into ARGONAUTE 1 (AGO1) and sequence complementarity is found. DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) methylates the TE when the siRNA is loaded into ARGONAUTE 2 (AGO2) and sequence complementarity is found in a Pol V transcript. This methylation is maintained with a feedback loop involving Pol IV generating transcripts that RDR2 converts to double-stranded RNA. DCL3 cleaves this product into 24 nt siRNA which is loaded into ARGONAUTE 4 (AGO4). When a sequence match is found between the siRNA and the Pol V transcript, DRM2 reinforces the methylation of the TE. Reproduced with permission: Matzke & Mosher (2014).

The DNA sequences that specific histone modifications are associated with can be determined by chromatin immunoprecipitation (ChIP). In this technique, the genome is fragmented with the histones still attached, then isolated with antibodies specific to the histone modification of interest. The isolated fragments can then be sequenced to determine the location on the genome and sequence associated with the specific histone modifications of interest (Saleh et al. 2008).

1.2.7 Epigenetic changes in plants under abiotic stress

Retaining the memory of a prior stress event can be very useful to a plant. A form of stress memory that depends on stress related proteins or metabolites would be limited by the half-life of the signalling molecules and only useful over short periods of time. A much longer term memory involves more stable epigenetic modifications, with potential heritability in some cases (Chinnusamy & Zhu 2009).

A variety of experiments have been performed by various research groups to understand how epigenetic changes are related to stress response in plants. Tricker et al. (2012) demonstrated that environmentally induced epigenetic responses were lost in *A. thaliana* mutants lacking *de novo* methylation capability or mutants lacking siRNA production capability. Further research showed that *A. thaliana* exposed to low humidity stress induced epigenetic changes. These changes were passed on to the next generation, resulting in offspring with changed stomata density, and therefore an early increased ability to cope with conditions of low humidity (Tricker et al. 2013).

Under salinity stress, rice varieties that are more salt tolerant have a greater level of flexibility in methylation changes under salinity stress than non-tolerant varieties (Joel 2013). Research in the salinity tolerance of wheat varieties has shown a correlation between high salinity tolerance and high global methylation levels when not under stress. The high tolerance varieties experienced a greater reduction in

methylation during the salinity stress event in a dose-dependent fashion (Zhong et al. 2009). When looking at the methylation characteristics of winter and spring wheat, Sherman & Talbert (2002) found that winter wheat had higher global levels of methylation and that vernalisation (a cold treatment) caused de-methylation in sequences related to floral induction.

Research performed by Pecinka et al. (2010) found evidence for environmental conditions transiently overriding epigenetic states in *A. thaliana*. Several repetitive elements under epigenetic regulation by transcriptional gene silencing became activated by prolonged heat stress. It has also been demonstrated that RNA directed DNA Methylation is essential in basal tolerance against heat stress when comparing the response of mutant plants without RdDM functionality (Kim et al. 2015).

An inverse correlation was found between DNA methylation and gene expression in rice, with the epigenetic changes being heritable. This research also discovered a disease resistance gene was activated by demethylation and expression of this resistance gene was stably inherited (Akimoto et al. 2007). Boyko et al. (2010) verified that the trans-generational response to stress induced changes depended on DNA methylation and small RNA. The progeny of stressed plants had an increase in homologous recombination under non-stress conditions compared with progeny from non-stressed plants.

There has been some prior research in methylation changes in barley under salinity stress. Demirkiran et al. (2013) grew the barley cultivar Tokak in MS media with 0 mM, 50 mM and 100 mM NaCl and used MSAP in combination with specific PCR primers to further reduce the population of fragments. They found that 5 of the 23 amplified bands experienced methylation changes when comparing 100 mM salt stressed plants with control plants. Konate et al. (2018) used ms-GBS to look for differentially methylated markers in five barley varieties in soil exposed to 75 mM, 150 mM and 200 mM NaCl. This led to the discovery of thousands of significantly differentially methylated markers associated with salinity stress. All of these markers were limited to methylation differences at the CCGG restriction enzyme recognition site.

To our knowledge, there has not been any prior research in barley that has attempted to observe this plant under salinity stress and obtain methylation states along with small RNAs and mRNA transcript levels from the same tissue samples.

1.2.8 Priming plants for abiotic stress

Methods have previously been investigated for priming seeds and seedlings to better cope with later stresses. Hydropriming involves imbibing seeds in sterilised water at a specific temperature, and drying back to the original weight of the seed. This method has yielded a 3 to 4 fold increase in the length of roots and shoots in primed plants under low water conditions when compared with control plants (Kaur et al.

2002). Osmopriming is a method where the seeds are imbibed in a solution that possesses a low water potential. This method has been used to improve yields in sugarcane, canola and chickpea under salinity stress (Jisha et al. 2013).

A greater understanding of the interactions between small RNAs, DNA methylation and salinity stress response mechanisms may lead to treatment methods where small RNAs can be introduced to plants to help prime the plant for oncoming stresses.

1.3 Hypotheses

If gene expression in barley plants under salinity stress is partly controlled by epigenetic mechanisms, then exposing the plants to salinity stress will result in changes in sRNA and DNA methylation associated with mRNA expression.

1.4 Project aims

This project aims to identify how different epigenetic mechanisms such as sRNAs and DNA methylation interact and lead to changes in gene expression in barley plants that have been exposed to salt stress. Understanding such interactions at a molecular level will help us understand how barley manages salinity stress, and could potentially lead to novel methods of plant tolerance to stress via breeding programs or via epigenetic priming of planted crops.

2 Barley under salinity stress

2.1 Introduction

Salinity stress is an extremely important issue for grain growing regions in Australia (Rengasamy 2002). A salt stress (step-wise increases in salt concentration), rather than a salt shock (a sudden increase to maximum concentration) more accurately represents natural changes (Shavrukov 2012) and was chosen to be applied to this experiment. Significant differences have been observed previously in the response to salinity stress in barley in soil and in hydroponic systems (Tavakkoli et al. 2010) so the experiment was designed around a salt stress in soil. A salinity stress of 100 mM NaCl was selected, as it produced a significant biomass difference in about a week, and is also in the concentration range most commonly used to impose salt stress in barley in the existing literature (Patterson et al. 2009; Sheldon et al. 2013).

The variety Morex was chosen for the experiment as the genome had been sequenced with the highest degree of coverage (IBGSC 2012) and the transcriptome also has good coverage and annotation (Mascher et al. 2013; Mascher et al. 2017). Prior preliminary growth tests (data not shown) had highlighted the susceptibility of this variety to net blotch and aphid infestations when grown in a glasshouse. These issues were removed by growing plants in a growth chamber. This had the added advantage that the growth conditions were more controlled and repeatable.

Previous observations from growth experiments (data not shown) had indicated that the emergence of the third leaf would occur about 14 Days After Sowing (DAS). The rate of growth of the third leaf would rapidly decrease by around 20 DAS indicating the leaf had reached a relatively mature stage. This was considered important, as the observed difference in expressed genes should be more related to the stress rather than to potential differences in growth stages.

2.2 Methods

2.2.1 Seed growth and collection

Hordeum vulgare variety Morex seeds were germinated to check for viability. Seeds were sterilised using the method described by Tavakkoli et al. (2010) where seeds were placed in a 70% (v/v) ethanol solution and gently agitated for 1 minute, followed by 3% sodium hyperchlorite solution for 5 minutes, then three washes in milliQ water. Three seeds per pot (26 cm diameter, 28 cm high) were grown in a soil mixture made up of 50% (v/v) University of California mix, 35% (v/v) peat mix and 15% (v/v) clay loam soil with pH 6.0. Plants were grown to maturity and seeds collected and stored in envelopes grouped by seed head.

2.2.2 Plant growth conditions

Morex seeds which had been collected from a single head from one healthy barley plant were sterilised using the method described above. Each pot had 1519 g soil (soil mixture as described above) added with an initial water content of 13% (w/w), and

each empty pot and saucer weighed 55 g. Three seeds per pot were placed 2 cm below the surface, each 3 cm apart, covered, and watered to a final soil moisture content of 30% (w/w) making the mass of the pot, saucer and soil 1800 g.

Plants were placed in a Sanyo MLR-352 growth cabinet (Osaka, Japan) set to 20°C with a 16 h day / 8 h night cycle at a light intensity of 380 $\mu\text{mol}/\text{m}^2$. The eight plants were positioned as shown in Figure 2.1 on the middle shelf of the growth cabinet. This shelf was previously found to have the most consistent temperature and even airflow. Plant positions were allocated by random number generation and were rotated daily (Figure 2.2) so that any slight temperature, humidity or lighting bias would not persist over the duration of the experiment. A ninth pot, set up in the same way as described above, was placed on the bottom shelf to act as a reserve if all of the seeds from one pot did not have a successful germination. On the eighth day after sowing, the plants of similar height in each pot were kept such that only one plant per pot remained, and other seedlings were carefully removed.

2.2.3 Salinity stress treatment

Plants were exposed to salt stress using a similar method employed by the University of Adelaide Plant Accelerator Facility for salinity stress experiments (Asif et al. 2018) with some modifications as outlined below. Plants were watered to weight as shown in Table 2.1 with a decreasing water content until reaching 20% (w/w) water content.

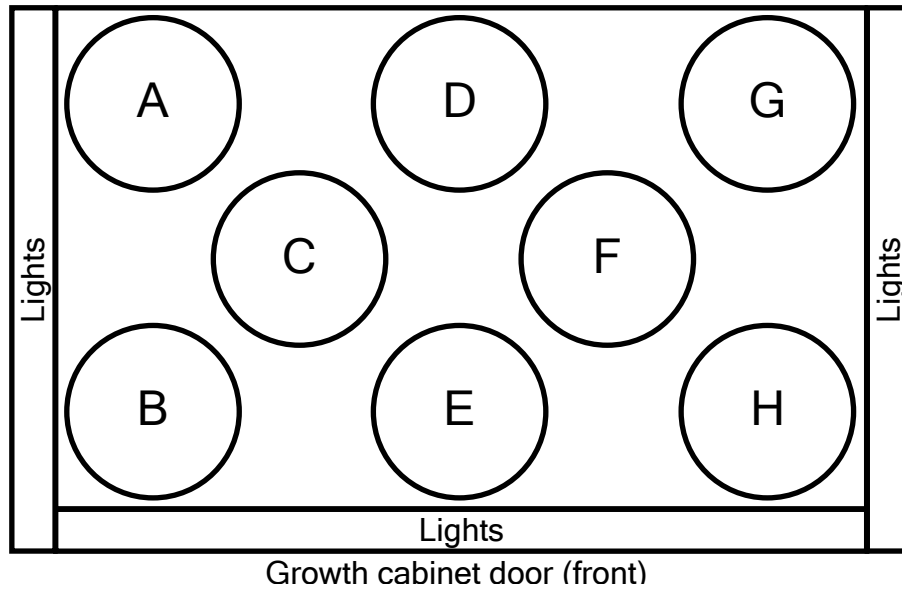


Figure 2.1. Top view diagram of plant positioning in the growth cabinet for the barley salinity stress experiment.

DAS	Position							
	A	B	C	D	E	F	G	H
0	1	7	8	6	3	4	5	2
1	2	1	7	8	6	3	4	5
2	5	2	1	7	8	6	3	4
3	4	5	2	1	7	8	6	3
4	3	4	5	2	1	7	8	6
5	6	3	4	5	2	1	7	8
6	8	6	3	4	5	2	1	7
7	7	8	6	3	4	5	2	1
8	1	7	8	6	3	4	5	2
9	2	1	7	8	6	3	4	5
10	5	2	1	7	8	6	3	4
11	4	5	2	1	7	8	6	3
12	3	4	5	2	1	7	8	6
13	6	3	4	5	2	1	7	8
14	8	6	3	4	5	2	1	7
15	7	8	6	3	4	5	2	1
16	1	7	8	6	3	4	5	2
17	2	1	7	8	6	3	4	5
18	5	2	1	7	8	6	3	4
19	4	5	2	1	7	8	6	3
20	3	4	5	2	1	7	8	6
21	6	3	4	5	2	1	7	8

Figure 2.2. Plant positional rotation scheme for barley salinity stress experiment. Plants 1-4 were control plants and 5-8 were salt treated. DAS = Days After Sowing. The position refers to the label shown in Figure 2.1.

The day before the first salt application, the water content was decreased to 19% (w/w).

On both the 14th and 15th day after sowing, 30 ml of a 224 mM NaCl solution was added to the saucers in which the pots were sitting, once in the morning and again in the evening. The weight of all pots was adjusted by adding water to the saucers to achieve the same mass as shown in Table 2.1. For salt treated plants, this gave a final salinity concentration of 100 mM at 18-21 DAS. The same procedure was repeated on control plants without salt added to the water.

Table 2.1. Watering and salt treatment schedule for salinity stress experiment.

Days after sowing	Mass of soil, pot and saucer watered to (g)	Watered from top or bottom	NaCl added to treated plants (g)	Water content of soil (w/w)	Salinity concentration in soil of treated plants (mM)
0	1800	Top		401 g (30%)	0
1	1785	Top		386 g (29%)	0
2	1765	Top		366 g (27%)	0
3	1745	Top		346 g (26%)	0
4	1725	Top		336 g (25%)	0
5	1705	Top		306 g (23%)	0
6	1685	Top		286 g (21%)	0
7 to 12	1670	Top		271 g (20%)	0
13	1650	Bottom		251 g (19%)	0
14	1690	Bottom	0.393 + 0.393	291 g (22%)	47
15	1730	Bottom	0.393 + 0.393	331 g (25%)	82
16	1705	Bottom		306 g (23%)	89
17	1685	Bottom		286 g (21%)	95
18 to 21	1670	Bottom		271 g (20%)	100

2.2.4 Tissue harvesting

At the point of sampling, the 3rd leaf of control and salt treated plants was removed,

quickly weighed, then snap frozen in liquid nitrogen, and stored at -80°C until DNA and RNA extraction. Immediately after snap freezing the 3rd leaf, the soil was placed in a large bucket of water and the roots gently separated from the soil. Once most of the soil was removed, roots were placed in a fresh bucket of water, rinsed and dried by dabbing with a paper towel. The roots were cut off, weighed, snap frozen in liquid nitrogen, and stored at -80°C until DNA and RNA extraction. The fresh mass of the shoots without the 3rd leaf was then measured, dried for two days at 80°C and the dry mass measured. The total shoot dry mass was estimated to be $(\text{dry mass without the third leaf}) * (\text{3}^{\text{rd}} \text{ leaf fresh mass} + \text{fresh mass without 3}^{\text{rd}} \text{ leaf}) / (\text{fresh mass without 3}^{\text{rd}} \text{ leaf})$. The water content ratio was calculated as $(\text{fresh mass} + \text{dry mass}) / (\text{dry mass})$.

2.3 Results

Images of plants just prior to harvesting tissue are shown in Figure 2.3. Plants 1-4 were control plants and 5-8 received a salt treatment with a final concentration of 100 mM. A small difference in growth between control and salt treated plants can be seen. The fresh mass, dry mass and water content ratio of shoot tissue is shown in Figure 2.4. There is a clear statistically significant difference between control and salt treated plants for each of these measurements, demonstrating that the plants exposed to salt were stressed with a consequential decrease in biomass.

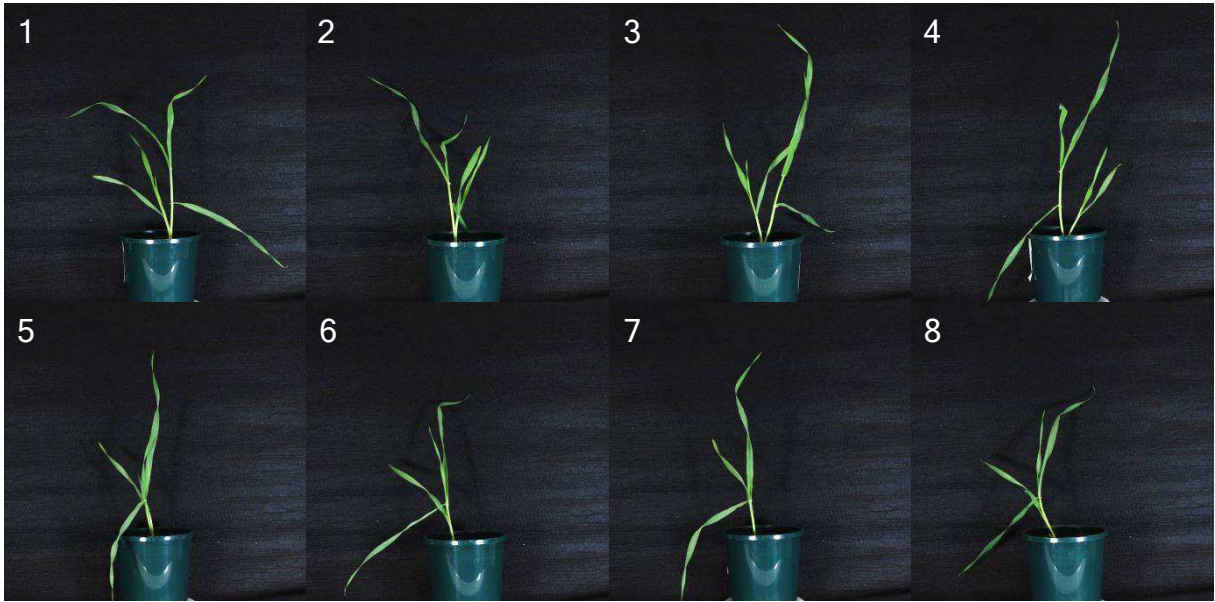


Figure 2.3. Barley plants 21 days after sowing, prior to tissue sampling. Plants 1-4 were control and 5-8 were salt treated to a final salinity concentration of 100 mM.

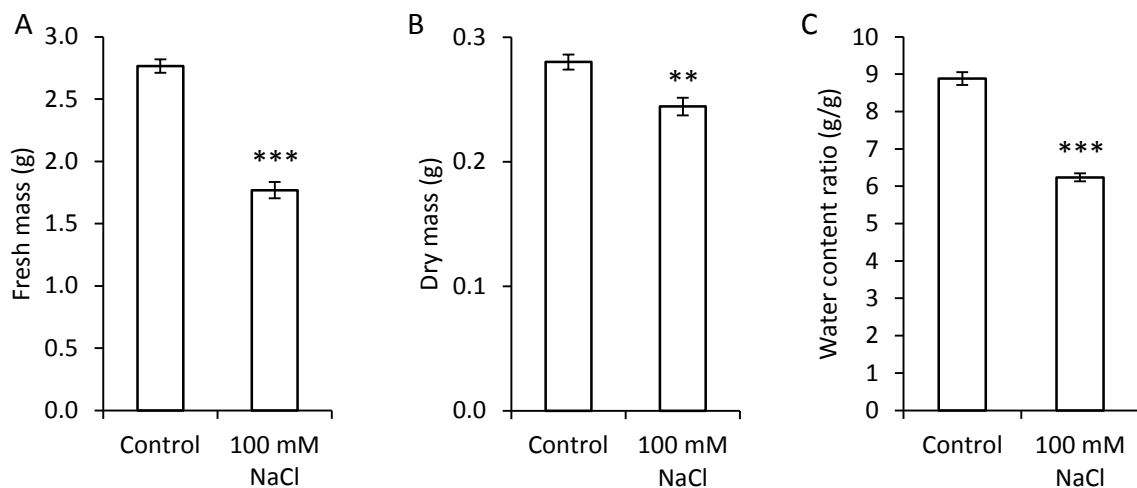


Figure 2.4. Comparison of biomass for control and salt treated plants. A: Fresh mass of shoot tissue. B: Dry mass of shoot tissue. C: Water content ratio of shoot tissue. Error bars represent SEM with ** $p < 0.01$, *** $p < 0.001$.

Evaporation and transpiration rates were measured relative to position in the growth chamber to ensure water loss positional effects were minimal. The evaporation rate over the first six days (before cotyledons emerged from the soil) is shown in Figure 2.5.A. No significant differences in daily water loss were observed across the

different plants. Figure 2.5.B shows evaporation (and transpiration) losses over the first 13 days after sowing, and again no statistically significant differences were found.

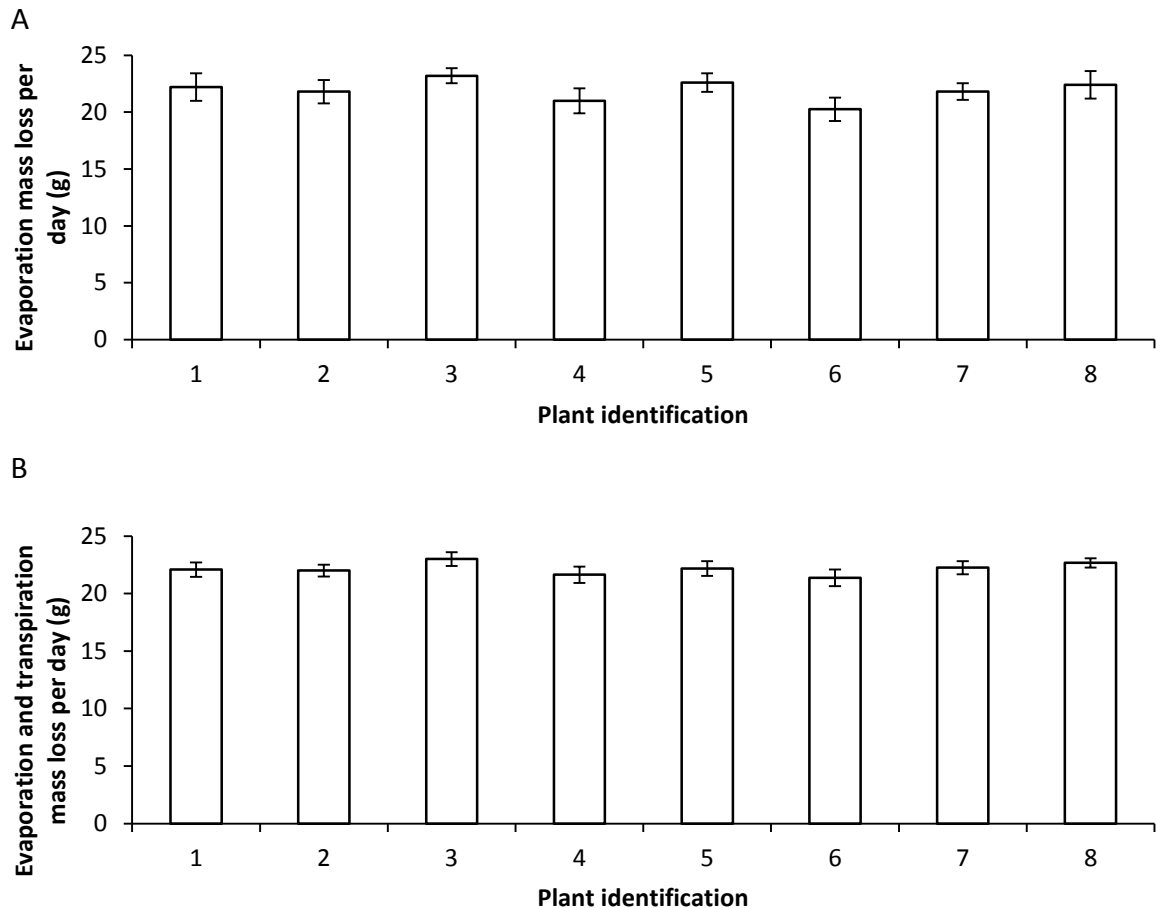


Figure 2.5. Evaporation/transpiration loss per day for each pot in the growth cabinet. Plant identification 1-4 were control plants and 5-8 salt treated plants. A. Average daily evaporation loss over 6 days after sowing for each pot (before cotyledons emerged). B. Average daily evaporation and transpiration loss over the first 13 days after sowing for each pot. Error bars show SEM with $n = 6$ in A and $n = 13$ in B.

2.4 Discussion

2.4.1 Seed growth and collection

Plants were grown to collect seed for use in the experiments. It was thought that seeds from one plant should have a more similar starting epigenetic state than seeds

from different parent plants. It was also expected that seeds from one head of one plant would develop in a shorter time window than all seeds from different heads, and are therefore likely to have a more similar epigenetic starting condition than different heads from the one plant. These expectations were not tested due to the cost and time required to verify these assumptions.

In this experiment it was important to maintain equal growing conditions between control and salt treated plants. To achieve this, the pot positions were rotated and weighed daily in an attempt to minimise any confounding variables such as highly localised airflow, lighting, temperature and humidity. Such issues had been observed in earlier trials investigating drought stress (data not shown), where plants in certain locations in the growth chamber lost water at noticeably different rates when left to dry over a two week period. These issues with growth chamber spatial variability have been noted in prior research (Potvin et al. 1990; Liu et al. 2000). The daily watering to maintain identical weights of pots and rotation in the growth chamber appears to have minimised any positional effects with relation to transpiration and evaporation as seen by the lack of any statistically significant differences in Figure 2.5.

There were some variations in the measurement of light levels at different locations in the growth cabinet and at different light sensor orientations, with a light level reading of 352 to 451 $\mu\text{mol}/\text{m}^2$ at position B, and 199 to 353 $\mu\text{mol}/\text{m}^2$ at position D.

Possible impacts from this difference on subsequent data (Chapters 3-6) cannot be ruled out. Prior tests had indicated that the middle shelf of the growth cabinet was the best choice for minimal variations resulting from the position. The middle and bottom shelf was found to be best for minimal differences in soil evaporation rates which was presumably related to air flow coming from the top of the growth cabinet. The middle shelf was found to have minimal variation in light levels between positions and the best with regard to temperature stability.

Salt stressed plants clearly showed a decrease in biomass (Figure 2.4). This indicates that the salinity exposure did effectively stress the plants. This decrease in fresh mass, dry mass and water content ratio has been observed previously in experiments and is consistent with other experiments involving transcriptome changes in response to salt stress in barley (Long et al. 2013; Shen et al. 2016).

The salt stress may also have had an effect on leaf growth stages as the salt stressed plants were smaller. It is possible that some observed gene expression or epigenetic differences may be related to growth cycle stages rather than directly related to salinity stress. The timing had been chosen so that the third leaf was relatively mature and a similar size in both control and salt treated plants based on prior tests. At the time of harvesting, the salt treated total shoot fresh tissue biomass was 36% less than the biomass from control plants, however the 3rd leaf experienced a 12% lower biomass in salt treated vs control. This suggests that the 3rd leaf had reached a

period of less rapid growth compared with other newer leaves on the plant at the time of harvesting tissue.

The measurements collected from these control and salt-treated plants indicate that the treatment imposed here was effective in eliciting a stress response and consistent with other literature reports on salt-stressed barley plants. The following chapters describe the subsequent DNA and RNA extraction, messenger RNA and small RNA sequencing, and DNA methylation sequencing performed on these plants to identify overlapping regions of the genome that respond to salinity stress in barley.

3 Messenger RNA changes in barley under salinity stress

3.1 Introduction

Transcriptome analysis is an extremely useful tool for observing changes in mRNA transcripts under different conditions. Comparing mRNA transcript levels under different treatment conditions can highlight genes that are important to and potentially related to dealing with the applied treatment. The rapid decrease in costs related to sequencing has resulted in a huge increase in transcriptome-related experiments (McCarthy et al. 2012). There are two major methods for RNA library preparation for transcriptome analysis. The first is poly A capture, which generates sequence data from mRNA alone; while the second is more comprehensive, employing ribosome depletion, to include other non-coding long RNAs in addition to mRNA reads (Zhao et al. 2014).

Prior research by Ziemann et al. (2013) had been performed using transcriptome analysis of salt stressed barley using poly A capture. They applied a 150 mM salt shock to *Hordeum vulgare* in soil and recorded mRNA changes 12 hours later. The limitations in read count and the incomplete assembly of the barley transcriptome at the time meant that only 1.65 million reads per sample could be mapped to an mRNA transcript. These limitations resulted in only 110 genes that were found to be differentially expressed.

In another study, Bahieldin et al. (2015) applied a 500 mM salt shock to *Hordeum spontaneum* (the ancestor to cultivated barley) after two weeks growth in soil and observed transcriptome changes in leaf tissue at four time points. They found that a high proportion of genes that mapped to the wild barley transcriptome were differentially expressed. Some genes were found to increase production of enzymes to protect cells from the damaging effects of the Reactive Oxygen Species (ROS) produced under stress conditions. These included antioxidants, iron ion binding proteins, and dehydrins which are all known to interact with ROS. Some signalling molecules, serine threonine-protein kinases and mitogen-activated protein kinases were found to have expression related to the salt shock. The involvement of ethylene production in barley stress response has been known (Dey & Vlot 2015), and ACC oxidase, an enzyme with increased expression in the stressed plants, is required for ethylene production. A transcription factor related to ethylene detection (ethylene-responsive element binding factor) showed increased expression in stressed plants. They also found that genes encoding refolding proteins were differentially expressed when the seedlings were exposed to the salt shock.

Hill et al. (2016) applied a 100 mM salt stress to germinating *H. vulgare* seeds and observed changes in the transcripts in different root zones three days after germination. The gene with the highest increase in expression under salt stress was a dehydrin. Carbohydrate-binding glycoproteins were found to be under increased expression in the elongation region of roots under salinity stress, and the production

and transportation of sugars was also increased. An ethylene responsive transcription factor BABY BOOM (BBM) had increased expression under salt stress. A number of the differentially expressed genes and molecular functions correlated with prior research.

Gene Ontology (GO) analysis has become a useful tool for distilling gene expression information to gain an insight in to the changes in processes, molecular functions, and cellular compartments. It can help with understanding the overall systems that are changing rather than looking at specific up and down regulation of individual genes, assisting in dealing with and interpreting the large quantity of data generated by transcriptome analysis (Doniger et al. 2003).

This chapter investigates the transcriptome changes occurring in barley grown in soil under 100 mM salt stress when compared with control plants. Differential expression patterns of genes are explored, used for gene ontology analysis and compared with prior literature. Observing similarities between the transcriptome changes and GO terms in this experiment with prior research provides validity to the use of transcriptome analysis in this experiment. It indicates that the tissue under stress reacted in the expected fashion, which is also useful for methylation and small RNA sequencing of the same tissue.

The hypothesis tested in this chapter is: if gene expression in barley is related to stress, then exposing barley to a salt stress will result in changes in the transcriptome. These changes should be relatable to prior research.

3.2 Methods

3.2.1 Messenger RNA library preparation

Snap frozen tissue described in section 2.2.4 was ground under liquid nitrogen. Either 100 mg of leaf tissue or 200 mg of root tissue was used for RNA extraction and further ground using a 1.5 ml tube and micropestle in liquid nitrogen until it was an extremely fine powder. The root tissue extractions had 20 mg of PVP 40 (40,000 MW) added. Then 1 ml of TRIzol (Life Technologies, NY, USA) was added and extraction proceeded as per the manufacturer's instructions with two exceptions. The supernatant was poured off after pelleting RNA, and two washes in cold 75% ethanol were performed as earlier trial experiments attained better yield, better RNA integrity and lower phenol contamination. The final pellet was dissolved in 25 µl of DEPC-treated water, with the concentration measured using the Nanodrop spectrophotometer (Thermo Fisher, DE, USA) and RNA was stored at -80°C. RNA quality and concentration was determined using the Agilent 2100 Bioanalyser with a RNA 6000 kit (Agilent Technologies, CA, USA) following the manufacturer's instructions with samples diluted to be in optimal instrument range based on the previous Nanodrop results. A messenger RNA library was generated by the Flinders Genomics Facility (Adelaide, Australia) using the TruSeq messenger RNA library kit

(Illumina Incorporated, CA, USA) as per the manufacturer's instructions. Sequencing was performed by SAHMRI (Adelaide, Australia) using Illumina NextSeq with 76 bp + 76 bp paired-end reads.

3.2.2 Bioinformatics

Read quality was checked with fastqc (Andrews 2017) and the splice-aware sequencing aligner STAR (Dobin et al. 2013) was used to map mRNA reads against the ensemble annotated genome for *Hordeum vulgare* (Kersey et al. 2017). A custom R script mRNA_STAR_PCA.R shown in section 10.1 was written to perform principle component analysis and observe separation between groups. Differential expression between control and salt treated groups was determined by a custom R script mRNA_DE.R shown in section 10.2, which used the false discovery rate threshold of $p_{adj} < 0.01$ and fold change threshold of $|\log_2(\text{fold change})| > \log_2(1.5)$. The scripts used the additional R libraries DESeq2 (Love et al. 2014), edgeR (Robinson et al. 2010), ggplot2 (Wickham 2016a), ggfortify (Tang et al. 2016) and Rsamtools (Morgan et al. 2016). PANTHER (Mi et al. 2019) using the GO ontology database released 8 October 2019 was used to find gene ontology terms for biological, molecular and cellular function using the PANTHER overrepresentation test with Fisher's exact test and a false discovery rate correction.

3.3 Results

3.3.1 RNA quality and quantity

Extracted RNA quality results are shown in Table 3.1 with further information in Appendix A (Figure A.9.1 and Figure A.9.2). The RNA Integrity Number (RIN), which is an algorithm that calculates RNA degradation, based on a variety of parameters including the ratios of the 18S and 25S ribosomal RNAs, was determined to be acceptable for all extracted RNA. The RIN for leaf tissue RNA was less than root due to the method for calculation not accounting for extra peaks from plastid ribosomal RNA. The quality of all RNA preparations was considered acceptable for both mRNA and small RNA sequencing. The number of reads obtained by sequencing each mRNA library is shown in Table 3.2.

Table 3.1. Extracted RNA concentration and RNA Integrity Number. Tissue samples 1-4 were from control barley plants and 5-8 were from 100 mM salt treated barley plants.

Leaf tissue sample	Concentration (ng/ul)	RIN	Root tissue	Concentration (ng/ul)	RIN
1	1336	7.1	1	818	9.2
2	1614	7.1	2	628	9.4
3	1142	7.2	3	604	9.1
4	1172	6.4	4	902	9.0
5	1326	7.3	5	506	9.2
6	1432	7.2	6	798	8.9
7	1274	7.3	7	1062	9.1
8	1796	7.2	8	1305	9.1

Table 3.2. Total RNA reads per sample. Tissue samples 1-4 were from control barley plants and 5-8 were from 100 mM salt treated barley plants.

Tissue	1 Control	2 Control	3 Control	4 Control	5 Salt	6 Salt	7 Salt	8 Salt
Leaf	26,233,942	22,465,779	28,866,034	23,663,568	28,871,394	29,670,450	30,413,398	27,476,863
Root	27,663,203	29,049,610	28,728,790	36,855,037	26,724,253	29,814,578	27,510,581	26,708,516

Sequence reads were mapped against the reference transcriptome and a Principle Component Analysis (PCA) was performed on the counts per million (CPM) of mapped reads (Figure 3.1). The PCA plot shows clear separation between leaf and root tissue along with separation between control and salt treated plants. The separation of tissue type shows a greater contrast than for the salt treatment, indicating the difference in gene expression across all samples is primarily from tissue type and secondarily from exposure to salinity stress.

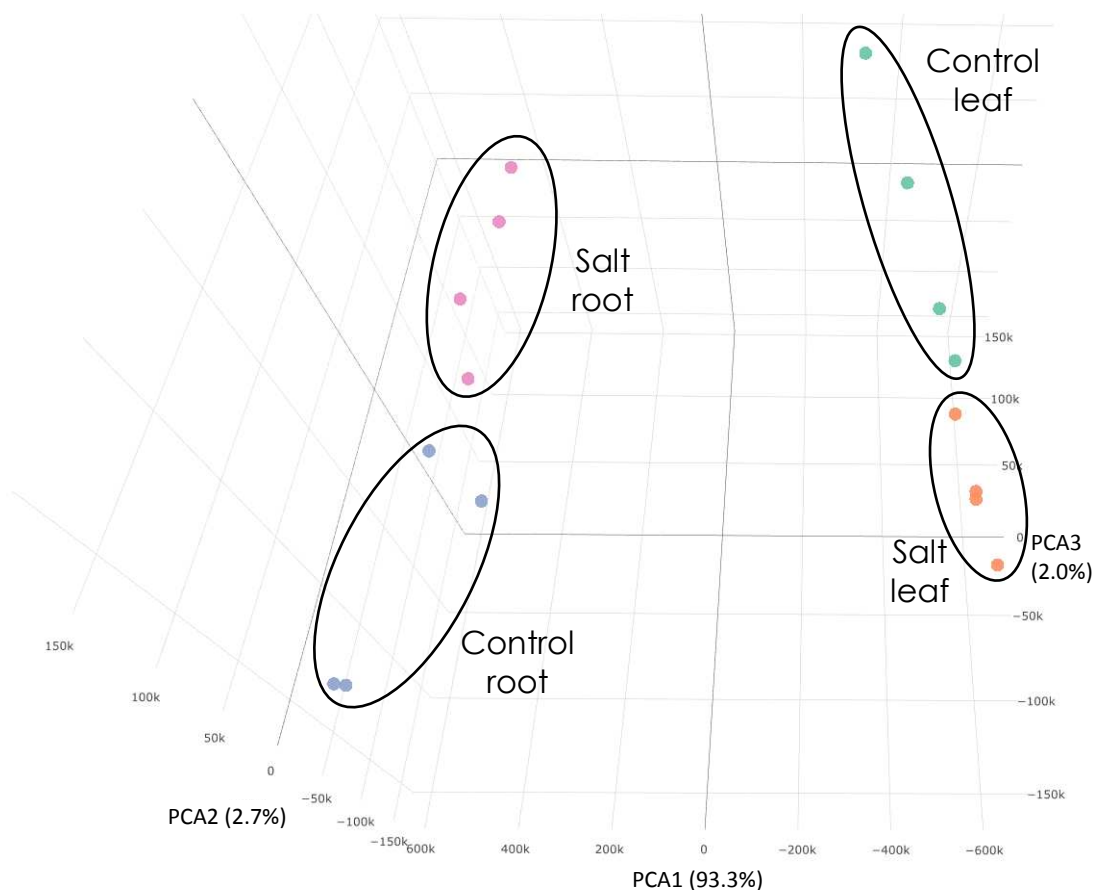


Figure 3.1. Principle component analysis of mapped messenger RNA reads from salt stressed barley and control plants. Leaf control tissue in green, leaf salt treated tissue in orange, root control tissue in blue and root salt treated tissue in purple. Principle components 1, 2 and 3 account for 93.3%, 2.7% and 2.0% of the total variance respectively. Barley plants were grown in soil and exposed to 100 mM NaCl stress or a control.

3.3.2 Differential expression of mRNA

Differentially expressed genes between salt treated and control plants were found and these are summarised in Figure 3.2 with normalised relative expression levels shown graphically in Figure 3.3. There was an increase in expression of 762 or 625 genes in leaf and root tissue respectively, with a decrease in expression of 404 or 591 genes in leaf and root tissue respectively ($p_{\text{adj}} < 0.01$, $|\log_2(\text{fold change})| > \log_2(1.5)$). A complete list of genes found to be differentially expressed is also included in Appendix A (Table A.9.1 and Table A.9.2). A summary of the ten genes with the most increased or decreased transcript expression for leaf and root tissue is shown in Table 3.3.

The most highly up-regulated transcript found in barley leaf tissue under salinity stress was an RNase S-like protein which also had 78% protein sequence identity with the third most up-regulated transcript. Some other up-regulated transcripts include a polyamine oxidase, an unknown protein containing a BURP domain, a xylanase inhibitor, a dehydrin, an invertase inhibitor, a purine-uracil permease, and a papain-like cysteine proteinase. Some of the most heavily down-regulated transcripts include a metacaspase, a dirigent protein, and some signalling related proteins.

In root tissue some of the transcripts which had the greatest increase in expression under salinity stress were a glycosyltransferase, an agglutinin isolectin, a sucrose synthase, an endo-1,3;1,4-beta-D-galactanase, an apoplastic invertase and a defence

response protein. Some of the transcripts that were heavily down-regulated include a peroxidase, an S-adenosylmethionine decarboxylase, a polygalacturonase, a SAUR50-like auxin-responsive protein and a calcineurin B-like protein.

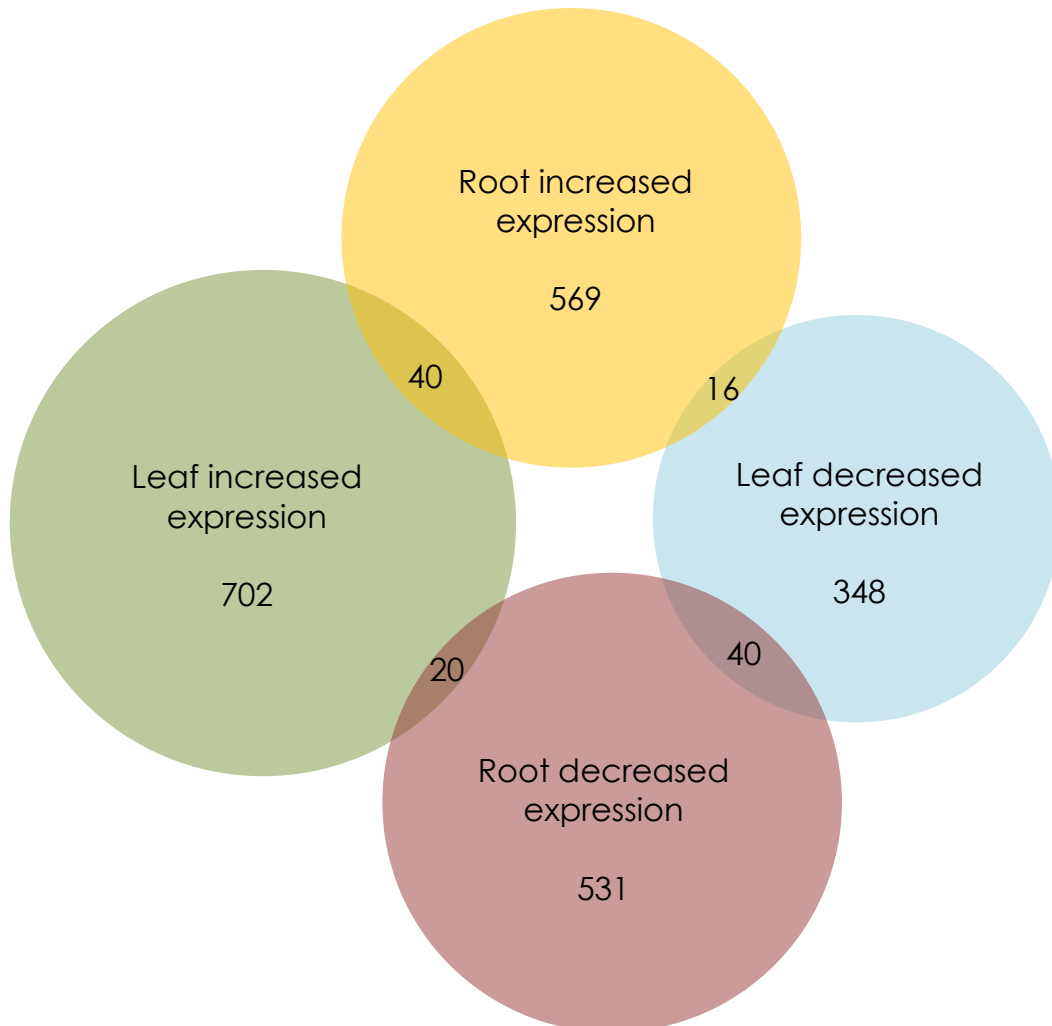


Figure 3.2. Venn diagram of differentially expressed genes in salt treated plants relative to control plants. Differentially expressed genes determined by use of DESeq2 with adjusted p-value < 0.01 and $|\log_2(\text{fold change})| > \log_2(1.5)$.

3.3.3 Gene Ontology analysis

The differentially expressed genes were used to find significant changes in GO terms using the PANTHER online database. High level biological, molecular and cellular functions that were overrepresented in salt treated plants are shown in Table 3.4,

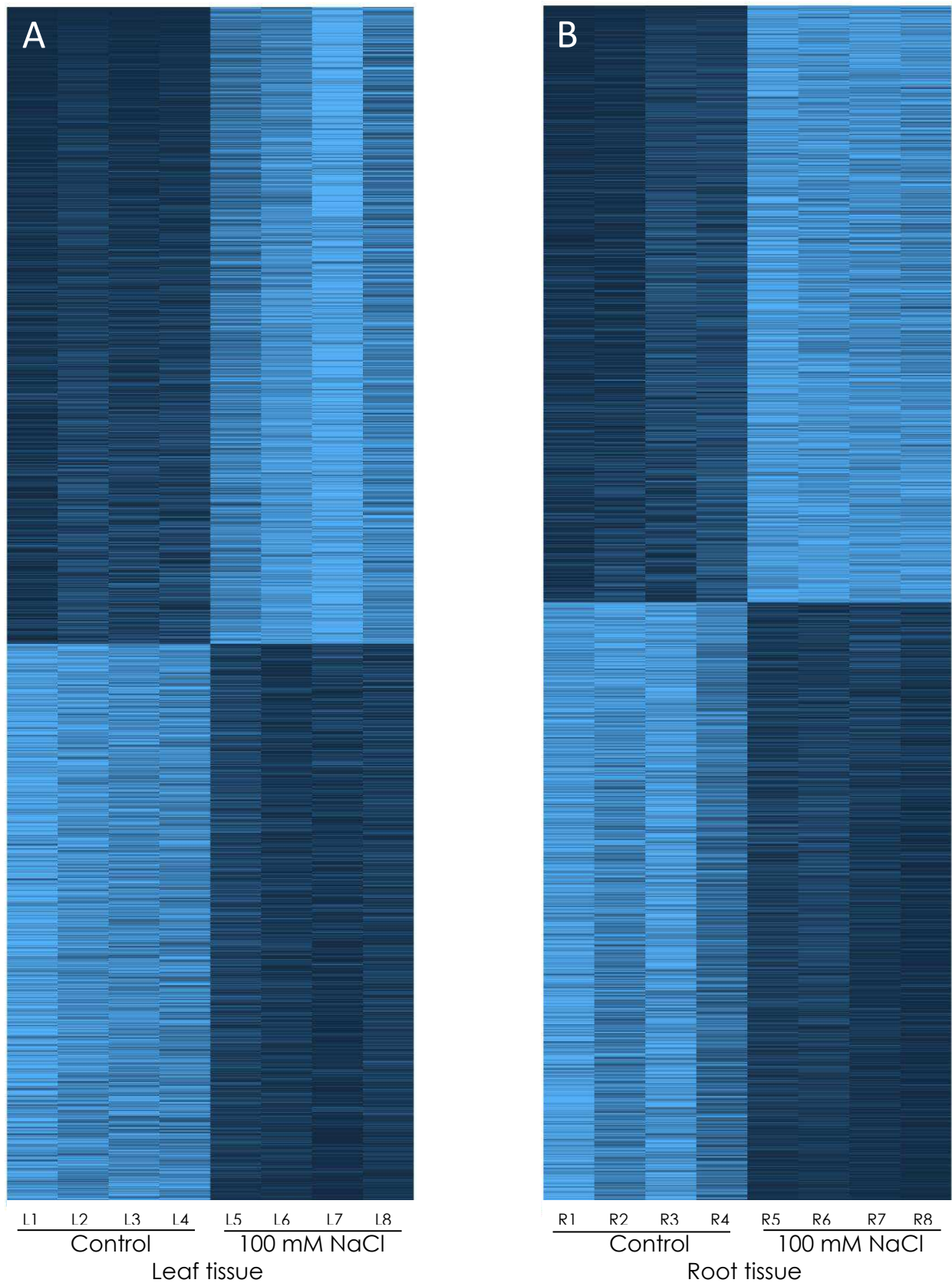


Figure 3.3. Normalised transcript abundance heat-map, comparing control and salt treated barley plants in (A) third leaf tissue and (B) root tissue. Dark blue is maximum normalised expression and light blue minimum normalised expression. Only transcripts with DESeq2 adjusted p-values < 0.01 and $|\log_2(\text{fold change})| > \log_2(1.5)$ are shown.

Table 3.3. Shortlist of the ten genes with the greatest increase or decrease in mRNA expression in leaf and root tissue of salt stressed barley relative to control. Differential expression is represented by $\log_2(\text{fold change})$ and the \log_{10} of the adjusted p-value is shown due to extremely small values. Plants were grown in soil and salt treated plants were exposed to a 100 mM NaCl stress. The annotation is shown, and if unknown in the Ensembl database a protein BLAST was performed with the highest match with a known protein shown and the identity of the match in brackets.

Tissue	Gene	Annotation or function of similar BLASTP result (identity of BLASTP in brackets if applicable)	$\log_2(\text{fold change})$	$\log_{10}(\text{adj-p})$	
Leaf	HORVU1Hr1G093780	RNAse S-like protein	5.96	-102.3	
	HORVU7Hr1G090410	Polyamine oxidase	3.94	-32.4	
	HORVU1Hr1G093660	RNAse S-like protein (78% identity)	3.83	-30.4	
	HORVU3Hr1G069650	BURP domain containing protein 3-like	3.30	-53.3	
	HORVU2Hr1G043890	Xylanase inhibitor	3.19	-48.9	
	HORVU6Hr1G084070	Dehydrin	3.12	-20.2	
	HORVU7Hr1G008260	Flavonoid O-methyltransferase-like protein (92% identity)	3.01	-23.7	
	HORVU2Hr1G103150	Invertase Inhibitor	2.96	-20.1	
	HORVU7Hr1G073640	Purine-Uracil permease	2.92	-40.6	
	HORVU7Hr1G120060	Papain-like cysteine proteinase	2.90	-54.4	
	HORVU4Hr1G090860	Metacaspase	-3.00	-40.9	
	HORVU5Hr1G057090	Alcohol dehydrogenase	-2.73	-15.9	
	HORVU7Hr1G047910	MYB-related pretein (transcription factor)	-2.44	-32.4	
	HORVU2Hr1G028780	Protease inhibitor	-2.37	-13.1	
	HORVU4Hr1G000040	Unknown (no matches)	-2.35	-12.9	
	HORVU6Hr1G004440	Wall-associated receptor kinase	-2.32	-12.1	
	HORVU0Hr1G000910	Putative Glucan 1,3-beta-glucosidase	-2.23	-13.7	
	HORVU4Hr1G000030	Putative LRR receptor-like serine/threonine-protein kinase	-2.19	-26.8	
	HORVU7Hr1G036960	Unknown (no matches)	-2.16	-9.9	
	HORVU1Hr1G000920	Dirigent protein	-2.03	-7.6	
	Root	HORVU3Hr1G065420	Glycosyltransferase	3.10	-48.9
		HORVU7Hr1G106900	Agglutinin isolectin (91.1% identity)	2.49	-34.7
		HORVU2Hr1G030870	Sucrose synthase	2.40	-37.6
		HORVU3Hr1G009360	Non-specific lipid-transfer protein	2.20	-14.3
HORVU4Hr1G087870		Unknown (no matches)	2.19	-16.0	
HORVU3Hr1G093170		endo-1,3;1,4-beta-D-glucanase (91.4% identity)	2.10	-12.8	
HORVU2Hr1G073210		Apoplastic invertase	2.07	-10.6	
HORVU1Hr1G092310		Glucan endo-1,3-beta-glucosidase	1.98	-9.0	
HORVU2Hr1G011550		Unknown (no matches)	1.97	-8.0	
HORVU4Hr1G071300		Defence response protein	1.90	-8.3	
HORVU3Hr1G066090		High affinity nitrate transporter	-2.91	-18.2	
HORVU1Hr1G080790		Calcineurin B-like protein 4 (92.9% identity)	-2.84	-20.9	
HORVU5Hr1G076740		SAUR50-like auxin-responsive protein (87.6% identity)	-2.80	-26.8	
HORVU2Hr1G117610		Serine/threonine protein kinase	-2.69	-21.1	
HORVU5Hr1G114000		Transmembrane transporter	-2.59	-29.6	
HORVU5Hr1G109980		Polygalacturonase (90.8% identity)	-2.51	-12.3	
HORVU2Hr1G007510		Unknown (no matches)	-2.28	-11.6	
HORVU5Hr1G064040		Unknown (no matches)	-2.14	-13.3	
HORVU3Hr1G071800		Unknown (no matches)	-2.12	-12.7	
HORVU2Hr1G127480		Peroxidase 47	-2.09	-10.6	

Table 3.5 and Table 3.6 respectively, with lower level GO terms shown in Appendix A (Table A.9.3, Table A.9.4 and Table A.9.5 respectively). From the biological GO terms, only two high-level GO terms were enriched in both leaf and root tissue: GO:1901607, alpha-amino acid biosynthetic process; and GO:0055114 oxidation-reduction process. In enriched high-level GO terms in molecular processes, two were overrepresented in both leaf and root tissue: GO:0005506, iron ion binding; and GO:0020037, heme binding. The one high-level cellular function GO term that was enriched in both leaf and root tissue of salt treated plants was located in the cell wall.

Table 3.4. Biological process gene ontology terms overrepresented in differential expression of leaf or root tissue in salinity stressed plants.

GO term	GO term process	Fold over-represented	
		Leaf	Root
0009834	plant-type secondary cell wall biogenesis	29.3	
0006782	protoporphyrinogen IX biosynthetic process	24.5	
0031408	oxylipin biosynthetic process	22.6	
0030244	cellulose biosynthetic process	13.8	
0005985	sucrose metabolic process	10.9	
0000302	response to reactive oxygen species	8.7	
0046677	response to antibiotic	8.2	
0044247	cellular polysaccharide catabolic process	8.2	
0009251	glucan catabolic process	8.2	
0071555	cell wall organization	5.1	
0019318	hexose metabolic process	5.1	
0044036	cell wall macromolecule metabolic process	5.0	
1901607	alpha-amino acid biosynthetic process	4.4	4.4
0010035	response to inorganic substance	4.2	
0007017	microtubule-based process	3.9	
0001101	response to acid chemical	3.6	
0098869	cellular oxidant detoxification	3.0	
0055114	oxidation-reduction process	1.9	2.5
0009627	systemic acquired resistance		23.7
0006002	fructose 6-phosphate metabolic process		20.2
0006568	tryptophan metabolic process		14.4
0042401	cellular biogenic amine biosynthetic process		12.7
0009664	plant-type cell wall organization		8.1
0008299	isoprenoid biosynthetic process		6.0
0048544	recognition of pollen		5.0
0032269	negative regulation of cellular protein metabolic process		4.6
0042737	drug catabolic process		4.0
0017001	antibiotic catabolic process		3.6
0005975	carbohydrate metabolic process		2.3
0006468	protein phosphorylation		1.9
0055085	transmembrane transport		1.9

Table 3.5. Molecular function gene ontology terms overrepresented in differential expression of leaf or root tissue in salinity stressed plants.

GO term	GO term function	Fold over-represented	
		Leaf	Root
0004575	sucrose alpha-glucosidase activity	21.0	
0016157	sucrose synthase activity	16.3	
0016760	cellulose synthase (UDP-forming) activity	12.6	
0005200	structural constituent of cytoskeleton	12.2	
0033897	ribonuclease T2 activity	12.2	
0051087	chaperone binding	10.9	
0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	5.8	
0008017	microtubule binding	4.0	
0016209	antioxidant activity	3.0	
0005506	iron ion binding	2.5	5.0
0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	2.3	
0020037	heme binding	2.1	3.9
0000234	phosphoethanolamine N-methyltransferase activity		20.5
0004350	glutamate-5-semialdehyde dehydrogenase activity		20.5
0004349	glutamate 5-kinase activity		20.5
0005504	fatty acid binding		18.7
0005315	inorganic phosphate transmembrane transporter activity		16.5
0003872	6-phosphofructokinase activity		15.1
0030410	nicotianamine synthase activity		15.1
0030598	rRNA N-glycosylase activity		8.6
0016831	carboxy-lyase activity		7.5
0030170	pyridoxal phosphate binding		6.3
0030145	manganese ion binding		5.3
0004497	monooxygenase activity		4.8
0042626	ATPase-coupled transmembrane transporter activity		4.7
0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen		4.4
0030246	carbohydrate binding		3.3
0016758	transferase activity, transferring hexosyl groups		2.6
0004553	hydrolase activity, hydrolyzing O-glycosyl compounds		2.4
0004672	protein kinase activity		1.9
0005524	ATP binding		1.6

Table 3.6. Cellular component gene ontology terms overrepresented in differential expression of leaf or root tissue in salinity stressed plants.

GO term	GO term component	Fold over-represented	
		Leaf	Root
0042644	chloroplast nucleoid	18.3	
0005874	microtubule	6.1	
0005618	cell wall	3.5	6.4
0005576	extracellular region	2.4	
0048046	apoplast		5.4
0016021	integral component of membrane		1.5

3.4 Discussion

The RNA extracted from root and leaf tissue was of sufficient quality for use in both small RNA and mRNA transcriptome sequencing. The transcriptome data used for PCA (Figure 3.1) demonstrated clear separation between groups. When observing the separation of clusters in the PCA, the separation between root and leaf tissue is very clear. This indicates that the largest components of variability in sample groups are related to the tissue type. This is expected as different tissue groups specialise in different functions and will utilise various transcripts to a greater or lesser extent. The separation between salt and control tissue is not as great as the tissue type, but can clearly be seen. This indicates that the salt stress had an effect on gene expression and that the plants under stress had a similar gene expression response for each tissue type.

A number of differentially expressed genes were found in salt treated plants relative to control plants, as shown in Figure 3.2. Reports in the literature describe various approaches for selection of the most biologically important genes by setting cut-off

values for adjusted p-value and fold change. Some researchers have used low stringency threshold with adjusted p-value cut-off of 0.2 and a fold change greater than 1.3 (McCarthy & Smyth 2009). In doing this, the aim is to capture as much biologically significant gene expression data as possible with less concern for the collection of false-positive results. Some researchers have suggested the use of more strict selection criteria, with adjusted p-values of 0.01 and fold change limits of 1.5. This came after comparing results from different platforms and results using p-value thresholds of 0.01 and 0.05, and fold change thresholds of 1.5, 2 and 4 (Patterson et al. 2006). Many researchers choose thresholds for an adjusted p-value of 0.05 and a fold change of 2 (or $|\log_2(\text{fold change})| > 1$) (Xiao et al. 2012); which can potentially retain some genes that are statistically not relevant, while also discarding other genes that have some biological relevance. Genes with a fold change of less than 2 can still have biological relevance, subject to the appropriate statistical significance test (Patterson et al. 2006). In the research presented here, a compromise was used with an adjusted p-value < 0.01 (a relatively strict cut-off for statistical significance) and relatively relaxed biological significance threshold with a fold change of 1.5 (or $|\log_2(\text{fold change})| > \log_2(1.5)$). Using the R library DESeq2, the error correction for low read count genes has already been incorporated in to the adjusted p-value (Love et al. 2014).

Of the transcripts that were found to be significantly differentially expressed between salt-treated barley and control plants, a few of the most differentially

expressed were compared with prior literature. The gene with the greatest increase in transcript expression in barley under salt stress was an RNase S-like protein. The protein sequence from this transcript had 100% sequence identity with a rice RNase S-like protein that had been found to have the greatest increase in protein abundance in a proteomics study of drought stressed rice (Salekdeh et al. 2002). This protein is unable to operate as an RNase as it lacks histidine residues at the required sites (Parry et al. 1997). Using a protein BLAST search, this protein also has a 78% identity with the third most increased barley transcript under salinity stress. The transcript in salt stressed leaf tissue that experienced the second highest level of increased expression was a polyamine oxidase. This enzyme produces hydrogen peroxide through polyamine catabolism, which enables hypocotyl growth in soybean under salinity stress and is an important factor for salinity tolerance in soybean (Campestre et al. 2011). The fourth greatest increase in expression in leaf tissue was from a transcript labelled BURP domain containing protein 3-like, which has a 95% sequence identity with a transcript found to be related to drought stress in *Triticum aestivum* (bread wheat) (Han et al. 2016). The BURP domain is typically found in proteins embedded in the cell wall in plants (Batchelor et al. 2002), indicating a probable location for this particular protein. Dehydrin had previously been reported as highly increased under salt stress in barley and is known for interactions with ROS (Bahieldin et al. 2015; Hill et al. 2016). This transcript was also highly up-regulated in our salt stressed plant leaves. Other transcripts with highly increased expression include a xylenase inhibitor previously found to be required in wheat and

Arabidopsis thaliana for salt and drought stress tolerance (Xin et al. 2014); a flavonoid O-methyltransferase-like protein, previously shown to increase expression in salt and cold stressed *Hibiscus cannabinus* (Ghosh et al. 2012); an invertase inhibitor, which is known to play an important role in development and stress response (Castrillon-Arbelaez & Delano-Frier 2011); a purine-uracil permease, previously found to have increased expression in *Saccharum spontaneum* under drought stress (Kai-chao et al. 2018); and a papain-like cysteine proteinase which are known for their involvement in biotic and abiotic stress response (Grudkowska & Zagdańska 2004)

Of the transcripts that experienced the greatest decrease in expression, many were found to be related to stress in prior research. A metacaspase transcript experienced the greatest down-regulation in salt-stressed barley. Metacaspases are proteases whose function is related to stress, programmed cell death and cell proliferation (Tsiatsiani et al. 2011). Prior research has found a positive correlation between expression of a metacaspase and exposure to biotic and abiotic stresses in rice, and that different metacaspases respond differently to a variety of stresses (Fagundes et al. 2015). This seemingly opposing result may be explained by research in *Arabidopsis* that has shown that two different metacaspases were regulating the hypersensitive response with one acting as a positive regulator and the other a negative regulator (Tsiatsiani et al. 2011). In a proteome analysis of soybean, alcohol dehydrogenase was found to be at elevated levels in the hypocotyls of salt stressed plants (Sobhanian et al. 2010). An increase in alcohol dehydrogenase transcripts has been observed in the

leaves of salt stressed barley plants (Lee et al. 2009). The result from our experiment does not match earlier findings as alcohol dehydrogenase was found to have significantly decreased levels in leaf tissue in the current study. The reason for this is unknown. An MYB-related protein is a transcription factor and was under decreased expression in the salt stressed leaf tissue in this experiment. Ziemann et al. (2013) observed two different MYB-related transcripts that were under decreased expression in leaf tissue in barley under salinity stress, and Lee et al. (2007) found that an MYB transcription factor was under decreased expression when exposed to a waterlogging stress and increased expression under a low-oxygen stress. An MYB transcription factor has also been found under increased expression in sugarcane under drought and salt stress (Prabu & Prasad 2012). A proteome analysis of soybean found that a protease inhibitor was found at decreased levels in salt stressed plants (Aghaei et al. 2009), which is in agreement with the findings of our experiment. A wall-associated receptor kinase experienced decreased expression in leaf tissue, which has also been observed in drought stressed barley, in both the leaf and root tissues (Ozturk et al. 2002), and also salt stressed barley root tissue (Marakli & Gozukirmizi 2018). A dirigent protein was also found at decreased expression levels under salt stress in leaf tissue. The dirigent proteins have been linked to biotic and abiotic stress response (Paniagua et al. 2017) and have been observed under increased expression in sugarcane under a polyethylene glycol stress (a drought simulant) (Jin-long et al. 2012), and decreased expression in alfalfa under heat and cold stress (Behr et al. 2015). Overall the transcripts with the greatest decrease in

expression in leaf tissue in barley under salinity stress were mostly in concordance with prior literature in other plant species, but with lower concordance than shown from the transcripts under increased expression in leaf tissue.

In root tissue under salinity stress, a glycosyltransferase had the greatest increase in expression when compared with control plants. Glycosyltransferases modify compounds by glycosylation which is the transfer of sugars from donor to acceptor molecules. They have been observed under increased expression in *A. thaliana* under salinity stress and have improved the salt tolerance of a tobacco plant modified to overexpress an *A. thaliana* glycosyltransferase (Sun et al. 2013). The root transcript with the second greatest increase in expression was not known in barley, but had a 91.1% identity with an agglutinin isolectin from *Aegilops taushii*. These proteins are responsive to salt and drought stress in rice (Lannoo & Van Damme 2010). A sucrose synthase increased expression in root tissue under salt stress in our experiment. Prior research has shown increased expression of sucrose synthase transcripts in barley under hypoxia, drought and salt stress treatments (Barrero-Sicilia et al. 2011). A transcript which was unannotated in barley, but had 91.4% translated sequence identity with endo-1,3;1,4-beta-D-glucanase protein from *A. taushii* experienced increased expression in root tissue of salt stressed barley. This transcript has previously been observed with increased expression in rice when exposed to ethylene, wounding, fungal infection, cytokinin or salicylic acid (Simmons et al. 1992). Ethylene production has been previously established as a signalling molecule

for abiotic and biotic stress response (Fujita et al. 2006), which would make our result match with this finding. The transcripts of an apoplastic invertase had increased expression in the root tissue of salt stressed barley plants in this experiment. Research by Fukushima et al. (2001) resulted in a transgenic tobacco plant with improved salt stress response by overexpression of a yeast apoplastic invertase. A glucan endo-1,3-beta-glucosidase was under increased expression in the roots of salt stressed plants in our experiment. This protein has been observed with higher levels of expression during drought stress in a wheat variety with higher drought tolerance when compared against a variety with lower tolerance (Faghani et al. 2015). The transcripts with the greatest increases in expression in root tissue under salt stress were in concordance with prior research.

The transcripts with the greatest decrease in expression in barley root tissue under salt stress include a high affinity nitrate transporter. In prior research in tomato roots, a nitrate transporter was found under significantly decreased expression under salinity stress (Yao et al. 2008). An unannotated transcript in barley, which when translated had 92.9% identity with a Calcineurin B-like protein 4 from *A. taushii* had the second most decreased transcript abundance in salt treated root tissue. This is a somewhat unexpected result as this protein has previously been shown to be involved in stress response, with *A. thaliana* mutants lacking this gene being more sensitive to drought and salt stresses (Luan et al. 2002), and expression is induced in salt and heat stressed foxtail millet seedlings (Zhang et al. 2017). Another

unannotated gene in barley under decreased expression had 87.6% translated sequence identity with a SAUR50-like auxin-responsive protein from *A. taushii*. This protein, which is related to growth and development, was previously found to be downregulated in gerbera flower stems under low-water stress (Ge et al. 2019). An unannotated gene with decreased expression in barley root tissue had a translated identity of 90.8% with polygalacturonase from *Triticum uratu*. Liu et al. (2014) found that overexpression of polygalacturonase in rice resulted in increased abiotic sensitivity and decreased pectin content, which aligns with the decreased expression we observed. A second somewhat unexpected transcript decrease in root tissue was a peroxidase. Peroxidases are known for ROS scavenging and increased expression under stress events (Zipor & Oren-Shamir 2013). However, Ziemann et al. (2013) discovered a peroxidase was under decreased expression in the leaves of salt stressed barley plants, so this is not the first time this seemingly unusual result has been observed. The transcripts with the greatest level of increase and decrease in root and leaf tissue from salt stressed barley generally align with the prior literature, which supports the validity of these data.

With a considerable number of differentially expressed genes, it is useful to observe groupings of gene functions rather than attempting to look at the function of every individual gene. Gene Ontology analysis was used to observe a higher level perspective of changes that occurred. PANTHER was chosen for GO analysis due to the monthly update and continual improvement of GO annotation terms (Carbon et

al. 2017; Mi et al. 2019). In the biological process GO terms (Table 3.4), some highly enriched terms related to Reactive Oxygen Species (ROS). ROS is a destructive by-product of metabolism that needs to be carefully managed to prevent toxicity. This class of molecules is also used for signalling and initiating a stress response (Choudhury et al. 2017). High-level leaf GO terms related to ROS included; GO:0000302, response to reactive oxygen species; GO: 0098869, cellular oxidant detoxification; and GO:0055114 oxidation-reduction process. In the root tissue, the high-level GO term related to ROS was GO:0055114, oxidation-reduction process.

Prior research of Bahieldin et al. (2015) and Hill et al. (2016) have reported biological process GO terms related to salinity stress and these are summarised and compared to the findings of this study in Table 3.7 and Table 3.8. Bahieldin et al. (2015) studied leaf tissue transcriptome changes in wild barley *Hordeum spontaneum* under a 500 mM NaCl salt shock after two weeks growth in soil. Hill et al. (2016) investigated the transcriptome of root tissue of *H. vulgare* germinated in a control or 100 mM NaCl nutrient medium and harvested three days after germination. It is important to note that recent advances in gene ontology completeness (Carbon et al. 2017) will mean that prior research will be likely to yield fewer GO terms. Different enrichment tools will use differing background comparisons and will also result in some differences. Gene Ontology information and annotations are updated frequently, with changes to how information is mapped between GO terms and genes. This can result in

variations in enrichment of GO terms between different releases of the same database or between separate databases (Rhee et al. 2008).

Table 3.7. Comparing GO term results that are common with prior research. Comparing the gene ontology results from this study with prior research by Bahieldin et al. (2015) in leaf tissue and Hill et al. (2016) in root tissue.

GO term	GO term function	Enriched term			
		Present study Leaf	Present study Root	Bahieldin 2015 (leaf)	Hill 2016 (root)
0071554	cell wall organization or biogenesis	X		X	
0009987	cellular process	X		X	X
0071840	cellular component organization or biogenesis	X		X	
0044237	cellular metabolic process	X		X	X
0008152	metabolic process	X		X	X
0071704	organic substance metabolic process	X		X	
0044249	cellular biosynthetic process	X			X
0009058	biosynthetic process	X		X	X
0044281	small molecule metabolic process	X		X	
0006629	lipid metabolic process	X			X
0044238	primary metabolic process	X		X	X
0009059	macromolecule biosynthetic process	X			X
0005975	carbohydrate metabolic process	X			X
0042221	response to chemical	X		X	
0050896	response to stimulus	X		X	X
0016043	cellular component organization	X		X	X
0006520	cellular amino acid metabolic process	X	X		X
0007017	microtubule-based process	X		X	
0051716	cellular response to stimulus	X		X	
0055114	oxidation-reduction process	X	X	X	
0006950	response to stress		X	X	X
0051707	response to other organism		X	X	
0009605	response to external stimulus		X	X	X
0009607	response to biotic stimulus		X	X	X
0051704	multi-organism process		X	X	X
0006955	immune response		X	X	
0002376	immune system process		X	X	
0008152	metabolic process		X	X	X
0071704	organic substance metabolic process		X	X	
0044281	small molecule metabolic process		X	X	
0071554	cell wall organization or biogenesis		X	X	
0009875	pollen-pistil interaction		X		X
0008037	cell recognition		X	X	
0009056	catabolic process		X	X	X
0005975	carbohydrate metabolic process		X		X

Table 3.8. Quantity of GO terms and common terms across different studies. Comparing the overlap in gene ontology terms found from this study with prior research by Bahieldin et al. (2015) in leaf tissue and Hill et al. (2016) in root tissue.

	Present study leaf	Present study root	Bahieldin 2015 (leaf)	Hill 2016 (root)
Present study leaf	71	11	15	12
Present study root		86	1	8
Bahieldin 2015 (leaf)			176	39
Hill 2016 (root)				80

3.4.1 Conclusion

The information provided in this chapter indicates that the transcriptome changes that occurred in treated barley plants were in response to the applied salt stress. A comparison was performed between this present study and prior research with reasonable concordance found in the most significantly modified transcript expression levels and enriched gene ontology terms. Perfect matching between the transcriptome profile of different experiments is not possible, as many factors including timing of sampling, age of plants, growth conditions, stress conditions and tissue type all have an impact on the expression profile. However, the correlations with prior studies provide confidence that the changes in transcriptome were related to the salinity stress. This supports the use of these tissue samples for further molecular analysis of small RNA and DNA methylation in chapters 4 and 5 respectively.

4 Small RNA changes in barley under salinity stress

4.1 Introduction

Small RNAs (sRNAs) are non-coding RNA molecules less than 200 nucleotides (nt) in length, which can be further broken into two classes, microRNAs (miRNAs) and small interfering RNAs (siRNAs) (Mallory & Vaucheret 2006; Chen 2012). In plants they perform regulatory functions including post-transcriptional regulation and the guiding of DNA methylation.

MicroRNAs in plants are generated from cleavage of hairpin-loop RNA structures by RNase III and DICER-LIKE 1 (DCL1) (Liu et al. 2017). The miRNA duplex has two imperfectly paired strands, typically around 21 nt long with overhangs at the 3' ends (Jones-Rhoades et al. 2006). A naming convention distinguishes the strands based on the proximity to the ends of the hairpin-loop precursor. Mature miRNA strands processed from the 5' end of the RNA hairpin-loop sequence are referred to as '5p', and strands processed from the 3' end are named the '3p' strands (Kozomara & Griffiths-Jones 2013). Typically the 5p strand is loaded in to ARGONAUTE 1 (AGO1) forming the RNA-Induced Silencing Complex (RISC), however there are instances where the 3p strand performs this function (Wang et al. 2019). When the miRNA strand loaded into the RISC anneals with a mRNA strand from a partial sequence match, the mRNA strand is either degraded or translational repression occurs (Huntzinger & Izaurralde 2011).

Endogenous siRNA in plants are produced from double-stranded RNA (dsRNA) rather than hairpin-loop structures (Mallory & Vaucheret 2006; Zhang et al. 2012). Small interfering RNA can be broken in to the sub-categories of trans-acting siRNA (ta-siRNA), natural antisense transcript siRNA (nat-siRNA), and heterochromatic siRNA (hc-siRNA) (Phillips et al. 2007).

Both ta-siRNAs and nat-siRNAs in plants have a similar mature form and function to miRNA, but differ in their method of synthesis. Trans-acting siRNAs are generated from a single stranded transcript which is targeted by a miRNA and after cleavage becomes double-stranded by the action of RNA Dependant RNA Polymerase 6. It is cut into 21 nt long segments with DICER-LIKE 4. Natural antisense transcript siRNAs are formed when two complementary sections of different mRNA strands generate a dsRNA which is then cut into 21 nt or 24 nt long segments (Phillips et al. 2007).

The hc-siRNAs in plants are typically 24 nt long and they are involved in RNA-directed DNA Methylation (RdDM). They are produced by RNA polymerase IV, where the product is then made double-stranded by RNA dependant RNA polymerase 2. This dsRNA is cleaved by DICER-LIKE 3, methylated by HEN1 and loaded in to ARGONAUTE 4 (AGO4). DOMAINS REARRANGED METHYLTRANSFERASE 2 (DMR2) is induced to methylate nearby genome cytosines when base pairing between the siRNA loaded in AGO4 and an RNA

transcript created by RNA polymerase V occurs (Matzke & Mosher 2014; Xie & Yu 2015).

While 24 nt siRNA targets DNA methylation, there is also a microRNA (miRNA) class which can range in size from 23 to 27 nt which also targets DNA methylation. These small RNAs (sRNA) were discovered relatively recently and are classified as miRNA due to the production involving a hairpin-loop structure forming from a single stranded RNA, which is then processed to form a miRNA (Teotia et al. 2017). These miRNA are significantly less abundant in quantity than 24 nt siRNA (Jia et al. 2011), and in some prior research the 23 and 24 nt sRNAs had been grouped together (Qi et al. 2006; Kurihara et al. 2008; Groszmann et al. 2011; Blevins et al. 2015) as they both caused DNA methylation via the RdDM pathway (Jia et al. 2011). In this chapter, the 23 nt sRNA is referred to as miRNA, 24 nt sRNA is referred to as siRNA since siRNA makes up a significant majority of 24 nt sRNA reads, and the grouping of the two is referred to as 23/24 nt sRNA. Changes in the expression of 23/24 nt sRNAs has previously been linked to DNA methylation, which targets genes and Transposable Elements (TEs) and modulates expression of these targets (Fultz et al. 2015).

Small RNAs have been found to be involved with abiotic stress response. A number of miRNAs have been identified that are involved in plant management of various

abiotic stresses including drought, cold, salinity, nutrient deficiency and radiation (Lu & Huang 2008; Khraiwesh et al. 2012).

The objective of this section of the study was to identify sRNAs that were involved in salinity stress response. Fifty-nine previously unknown miRNAs in barley were found, and six of these previously unknown miRNAs appear to be related to salinity stress response. A number of genomic features were targeted for methylation by 23/24 nt sRNA, these targets are further explored in Chapter 6 when incorporated with DNA methylation information from Chapter 5.

4.2 Methods

4.2.1 Small RNA library preparation

The small RNA sequencing libraries were prepared from the RNA extraction described in Chapter 3. The Flinders Genomics Facility (Adelaide, Australia) generated the libraries with size selection of 16 base pairs (bp) to 34 bp performed using the Pippin Prep (Sage Science, MA, USA). The library was generated using the Illumina Truseq small RNA library kit (Illumina Incorporated, CA, USA). The barcode assignment for samples is shown in Appendix A (Table A.9.6). Sequencing was performed by SAHMRI (Adelaide, Australia) using Illumina NextSeq with 76 bp single end reads.

4.2.2 Bioinformatics

Read quality was examined with fastqc (Andrews 2017) to check that the raw sequencing reads were as expected. Adapters were then trimmed, sequences filtered for quality of 30 or greater and size filtered for reads between 18 and 26 nt with cutadapt (Martin 2011). Reads were aligned using Bowtie2 (Langmead & Salzberg 2012) against *Hordeum vulgare* var Morex ribosomal RNA, snRNA and snoRNA sequences (Mascher et al. 2017) with matching reads discarded. Relative abundances of read sizes and unique reads were calculated using a custom python script Count_siRNA.py shown in section 10.3. Unique reads are only counted once for each sequence; if a specific sequence is read multiple times, it only counts once as a unique read. Reads were matched against the miRBase database entries for *H. vulgare*. A custom script Barley_SmallRNA_PCA.R, shown in section 10.4 was written in R to perform Principle Component Analysis (PCA) on the relative read abundance. The custom R script used the additional packages Rsamtools (Morgan et al. 2016), ggplot2 (Wickham 2016a), plyr (Wickham 2016b), and edgeR (Robinson et al. 2010).

Previously unknown small RNAs were inferred using sRNAbench (Aparicio-Puerta et al. 2019). PCA was performed with this new set of discovered miRNA using the custom R script Barley_SmallRNA_PCA.R, shown in section 10.4 with the variable miRBase set to FALSE. The online tool mfold (Zuker 2003) was used to determine the secondary structure of the hairpin-loop miRNA precursor. Differential expression

and adjusted p-values of miRNA between treatment groups was determined by a custom R script `Barley_SmallRNA_DE.R`, shown in section 10.6.

Small RNAs were grouped by size using the custom python script `SmallRNA_size_grouping.py` shown in section 10.6. SmallRNA targeting DNA methylation was analysed by mapping 23-24 nt sRNA reads against the *H. vulgare* genome using Bowtie2. Reads were normalised as counts per million reads for each sample and were binned in to 100 bp windows using the custom python program `smallRNA_windowing.py`, shown in section 10.7. Reads that were mapped to multiple locations were fractionally assigned to the relevant bins. For example, one sRNA that mapped to ten different 100 bp bins only added 0.1 reads to each relevant bin (and were later scaled by counts per million reads). The mapped reads for each sample were compared by grouping control and treatment and determining the adjusted p-value with the two R scripts: `windowed_smallRNA_ttests.R` and `ttests_to_adjusted_p.R`, shown in section 10.8 and section 10.9 respectively. Genomic locations with a statistically significant difference between the control and treatment were compared with known genomic features such as Protein Coding Genes (PCGs) and TEs using the custom R script `Annotate_sRNA.R`, shown in section 10.10. A modified version of this script `Annotate_sRNA_random_sampling.R` (section 10.11) was used to determine random sampling of the 100 bp windows for comparison of sRNA changes targeting certain genomic features.

New miRNAs that have a high sequence similarity with existing plant miRNA are named hvu-new-miRxxx where xxx is the miR number of the closest existing known miRNA. Where a novel miRNA has been found with no similarities to other plants, the name is of the form hvu-novel-miRxxxxxx where xxxxx is a number assigned from 90000 so that it does not clash with any known miRNA. These designations are temporary placeholders until a paper is published and fixed designators are defined by miRBase with the new and novel miRNA that has been found through this research.

4.3 Results

The total reads after filtering out low quality reads and reads that mapped to ribosomal RNA, snRNA or snoRNA are shown in Table 4.1. Analysis of the relative abundance of sRNA species grouped by size showed prominent peaks at sizes of 21 and 24 nt in leaf samples (Figure 4.1, unique reads in Figure 4.2). Root tissue small RNA did not show the same clear peaks at 21 and 24 nt, but rather a continuous spectrum of sizes sloping away from a 24 nt size peak. No statistically significant differences (t-tests, $p > 0.15$) were observed when comparing small RNA size read populations between control and salt treated plants.

4.3.1 Small RNAs guiding DNA methylation

Small RNAs that guide DNA methylation are 23/24 nt long and are involved in the RNA directed DNA methylation pathway. The 23/24 nt sRNAs were mapped against

the reference genome and assigned to 100 bp windows. Control and treatment relative count numbers were compared and the genomic features nearby were determined. A 23/24 nt sRNA was considered to be targeting a PCG if the differential expression occurred within 1.5 kbp of the PCG. For all other genomic features, it was considered to be related if the differential sRNA expression fell within the feature. Figure 4.3 shows the breakdown of significantly differentially expressed 23/24 nt sRNA and the associated genomic features. The differential expression of 23/24 nt sRNA targeted PCGs (with or without TEs) at a noticeably higher level than when compared against random sampling of the genome (chi-squared for 23/24 nt sRNA differential expression targeting PCGs against random sampling, $p < 0.001$). Changes in targeting of PCGs is at a higher level in 23 nt miRNA than 24 nt siRNA in both leaf and root tissue. DNA transposons were targeted by changes in 23/24 nt sRNA differential expression to a lesser extent than random sampling and retrotransposons were targeted to a greater level than random sampling.

Table 4.1. Sequence read numbers for small RNA sequencing of barley before and after filtering of reads. Reads were filtered for quality, size of 18 to 26 nt and reads that matched rRNA, snRNA or snoRNA were removed.

Small RNA sequence	Reads
Raw reads	389,086,591
Total reads after size, quality and contamination filtering	150,318,401
Average reads per sample for leaf control	7,945,356
Average reads per sample for leaf salt treated	8,095,103
Average reads per sample for root control	11,187,958
Average reads per sample for root salt treated	10,351,184

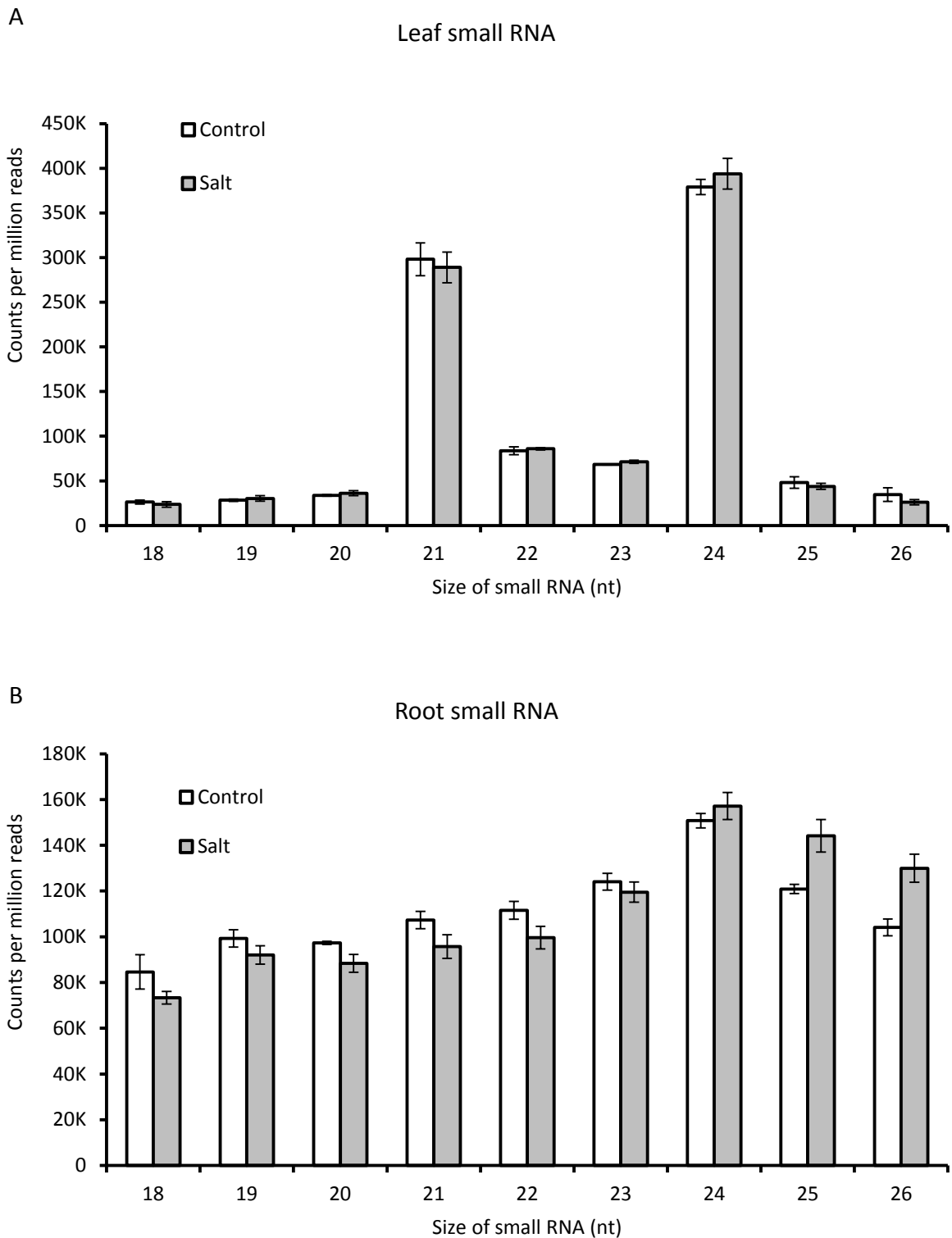


Figure 4.1. Relative abundance of sequenced small RNAs according to read size. Vertical bars show the relative abundance (counts per million) of small RNA molecules ranging from 18 to 26 nucleotides in length in (A) leaf and (B) root tissue collected from plants grown under control and salt stress (100 mM NaCl), $n = 4$. Error bars show SEM.

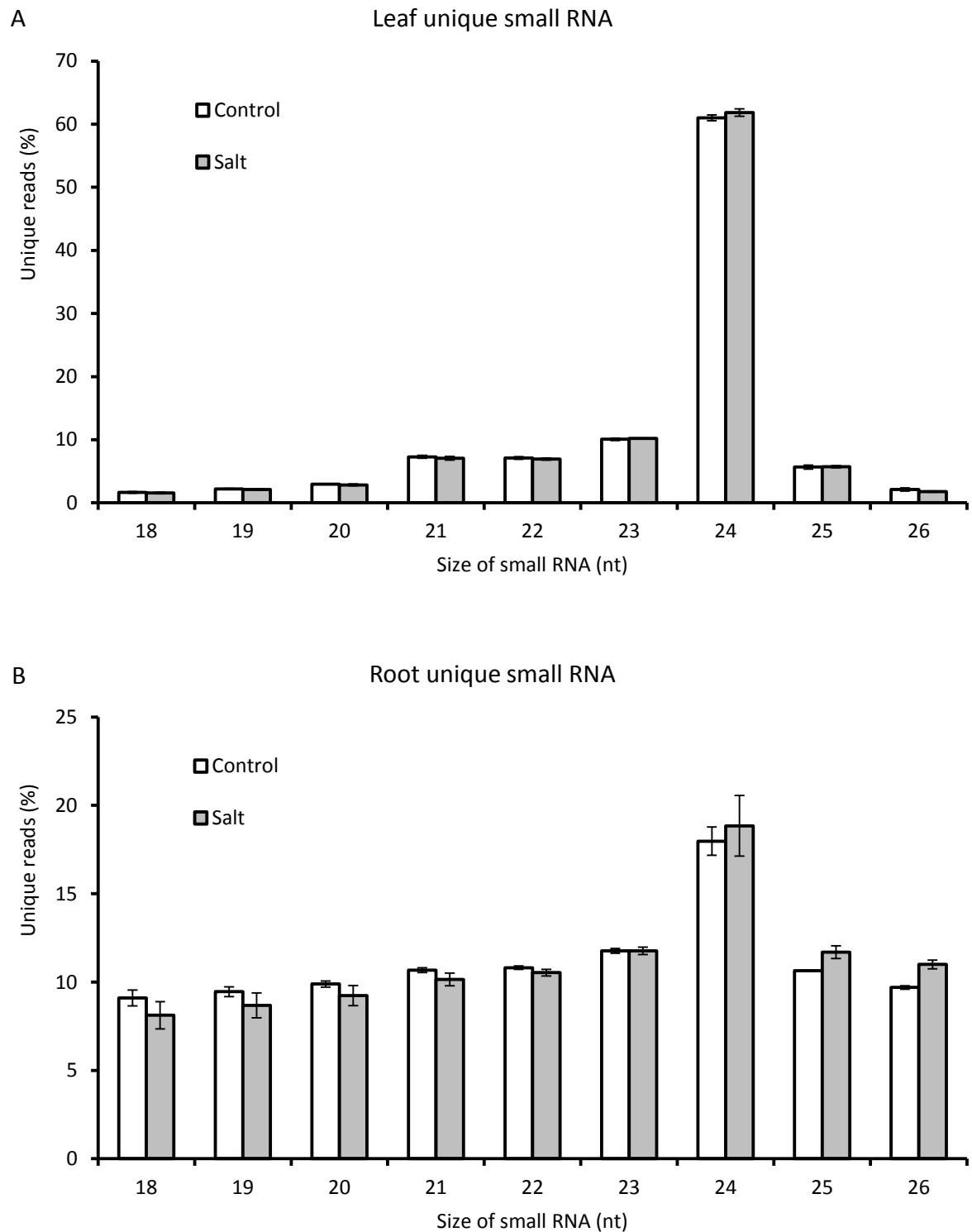


Figure 4.2. Relative abundance of unique small RNAs according to read size. Multiple reads of the same sequence count as a single unique read. Vertical bars show the relative abundance (%) of unique small RNA reads ranging from 18 to 26 nucleotides in length in (A) leaf and (B) root tissue collected from plants grown under control and salt stress (100 mM NaCl), n = 4. Error bars show SEM.

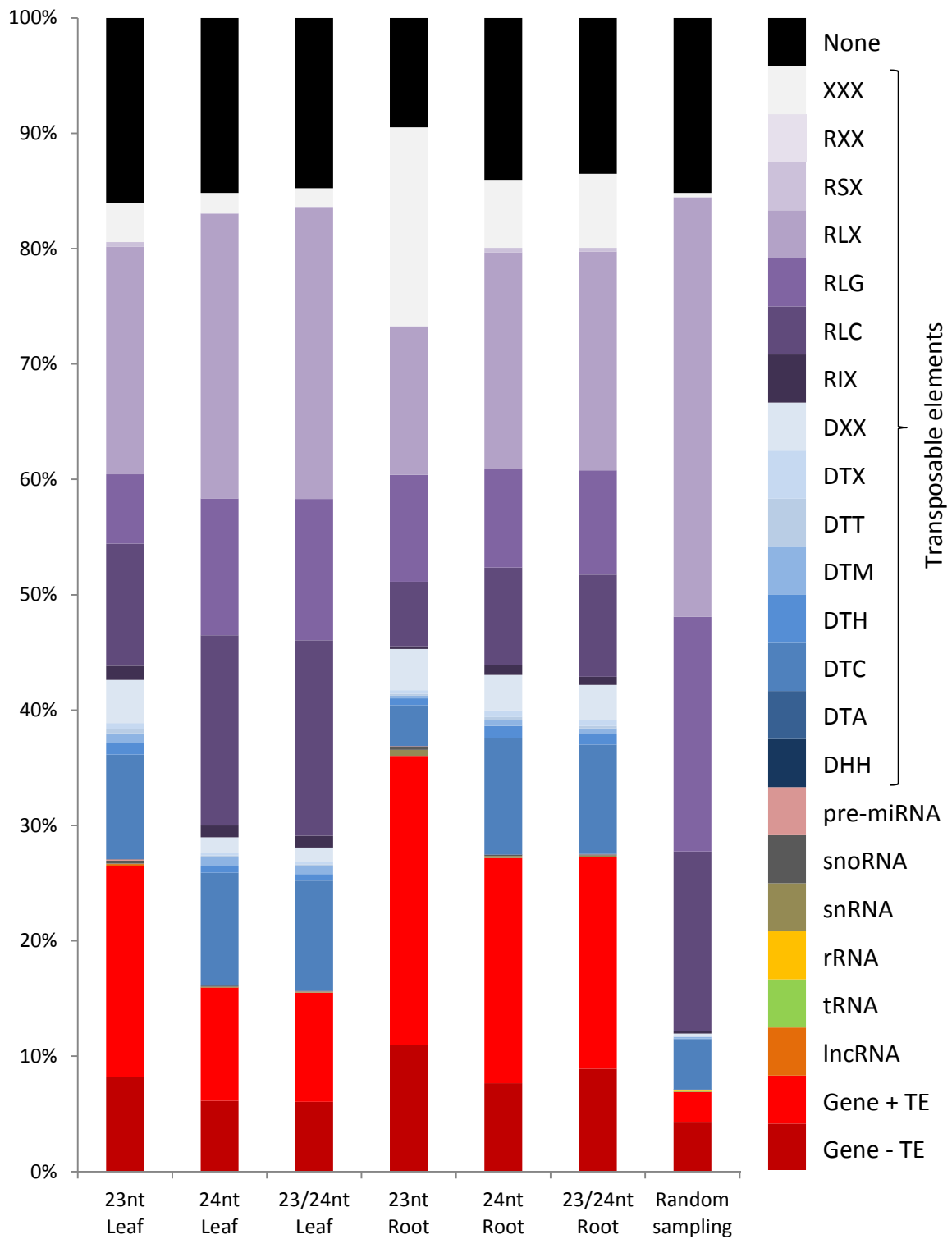


Figure 4.3. Percentage of significantly differentially expressed 23-24 nt sRNAs targeting genomic features in salinity stressed barley plants. SmallRNAs were mapped to the genome and grouped in 100 bp windows, differential expression of sRNAs was determined across the 100 bp window in 100 mM salinity stressed plants when compared with control plants. Transposable Elements are classified according to Wicker et al. (2007). Random sampling of 100 bp windows is included for comparison.

4.3.2 Identification of previously described microRNAs

Previously identified miRNAs were categorised by comparison between the small RNA library and miRBase 22.1 (Griffiths-Jones 2006) and is summarised in Table 4.2.

PCA of the relative abundance of miRNA reads identified in miRBase shows that the largest fraction of the variability (91.0%) in miRNA expression is explained by the tissue of origin (leaf vs root) (Figure 4.4). The second main component of the variability is differences in miRNA expression between root samples from plants grown under control and salt conditions. Finally, no clear differences between treatments in leaf samples were observed.

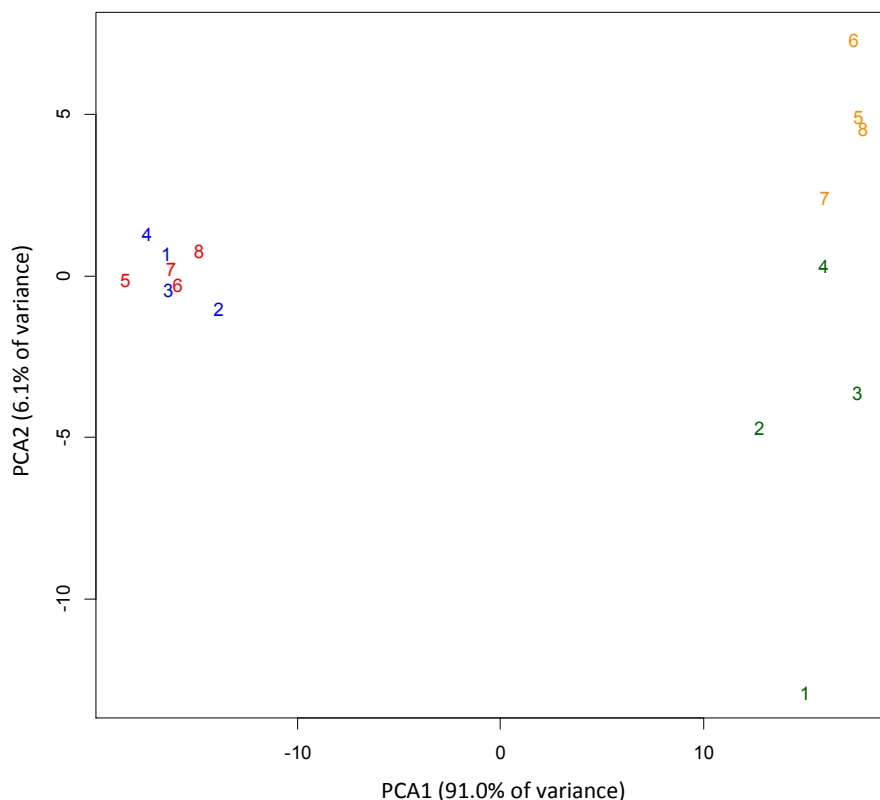


Figure 4.4. Principle component analysis showing separation of barley miRNA mapped to miRBase under 100 mM salinity stress for leaf and root tissue. Leaf control tissue in blue, leaf salt treated tissue in red, root control tissue in green and root salt treated tissue in orange. Principle component 1 accounts for 91.0% of the total variance, principle component 2 accounts for 6.1% of the total variance. PCA of miRNA was performed using count per million reads on the small RNA sequences that mapped to 40 different known barley miRNA on miRBase.

Table 4.2. MicroRNA found in barley root or leaf tissue under control or salt treatment conditions that have already been identified in miRBase. X indicates a minimum of 10 reads of the sequence across a treatment/tissue group.

miRBase identifier	Leaf		Root	
	Control	Salt treated	Control	Salt treated
miR156a	X	X	X	X
miR156b	X	X	X	X
miR159a	X	X	X	X
miR159b	X	X	X	X
miR166a	X	X	X	X
miR166b	X	X	X	X
miR166c	X	X	X	X
miR168	X	X	X	X
miR171	X	X		
miR397a		X		
miR397b	X	X		
miR444b	X	X	X	X
miR1120	X	X	X	X
miR1130	X	X		
miR1436	X	X	X	
miR5048a	X	X	X	X
miR5048b	X	X	X	X
miR5049a	X	X		
miR5049b	X	X	X	X
miR5049c	X	X	X	
miR5049e	X	X	X	X
miR5049f	X	X		
miR5051	X	X	X	X
miR5052	X	X		
miR6177			X	X
miR6181			X	X
miR6184			X	X
miR6186			X	
miR6187				X
miR6191	X	X		
miR6195			X	
miR6196	X	X		X
miR6200	X	X		
miR6201	X	X	X	
miR6204			X	X
miR6205	X	X		
miR6206			X	X
miR6207			X	X
miR6209			X	X
miR6213	X	X	X	

4.3.3 Identification of microRNA not previously identified in *Hordeum vulgare*

A total of 59 previously unidentified miRNAs in *H. vulgare* were discovered. A summary of newly discovered sequences that have sequence similarity with existing known plant miRNAs, and novel miRNAs that have not been observed in other plant species is shown in Table 4.3. Newly discovered mature *H. vulgare* miRNA and hairpin-loop sequences are found in Appendix A in Table A.9.7 and Table A.9.8 respectively. A PCA plot of the relative abundance of these newly identified miRNAs (Figure 4.5) also indicates a prominent separation in tissue type.

Table 4.3. New and novel miRNA found in barley. New sequences are defined as sequences that are not already known to exist in barley, but have high sequence similarity to other plant miRNA. Novel sequences are defined as being previously unknown in barley without having sequence similarity to known miRNA in other plant species.

Sequence information	Number of new sequences that have high sequence similarity with existing known miRNA in other plant species	Number of novel sequences that do not have similarity with any known plant miRNA
5p miRNA sequence only	1	11
3p miRNA sequence only	1	13
Both 5p and 3p miRNA sequences	13	20

4.3.4 Differential expression of miRNA

The miRNA identified as having statistically significant differential expression between control and salt treated groups is shown in Table 4.4. Only one of these miRNA had been classified in *H. vulgare* in miRBase. Only two miRNA, labelled hvu-new-miR477a and hvu-novel-miR90000 had significant expression differences in both leaf and root tissue. The miRNA designated hvu-new-miR477a had decreased

expression in both root and leaf tissue in salt treated plants relative to control. Conversely, hvu-novel-miR90000 was found to have decreased expression in leaf tissue and increased expression in root tissue.

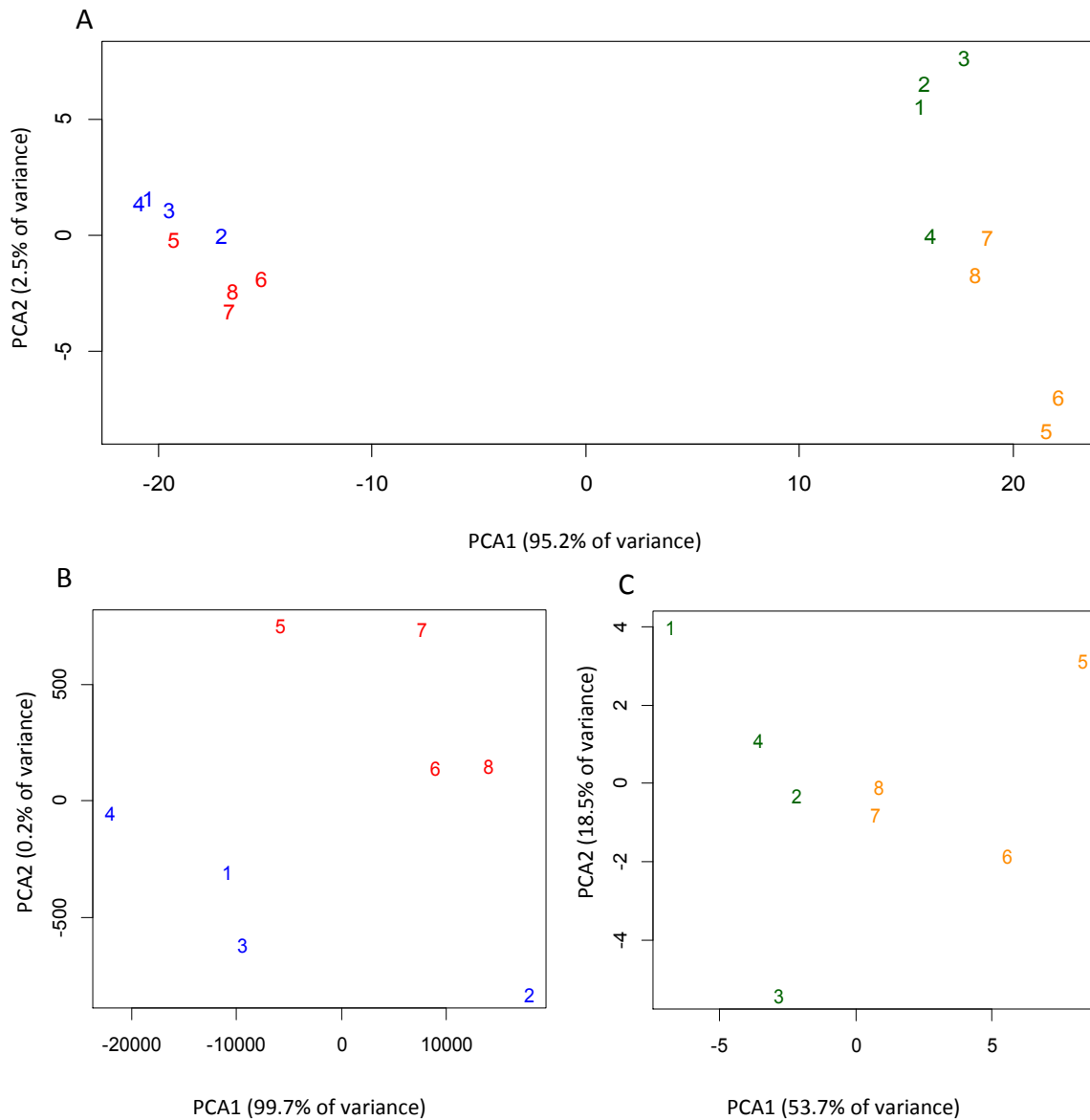


Figure 4.5. Principle component analysis showing separation of barley miRNAs not represented in miRBase, under 100 mM salinity stress for leaf and root tissue. Leaf control tissue in blue, leaf salt treated tissue in red, root control tissue in green and root salt treated tissue in orange. Principle component variance is shown in brackets for each axis. PCA of small RNA was performed using count per million reads on the miRNA sequences that mapped to the miRNA strands previously unknown in barley. A, Root and leaf. B, Leaf only. C, Root only.

Table 4.4. Differentially expressed miRNA in leaf and root tissue in barley under salinity stress. Differential expression of a given small RNA in log₂ fold change of 100 mM salt treated plants relative to control.

MicroRNA label	Leaf (salt/control)		Root (salt/control)	
	Log ₂ fold change	p-adjusted	Log ₂ fold change	p-adjusted
hvu-new-miR528	1.88	< 0.0001		
hvu-novel-miR90001	0.31	0.041		
hvu-miR5051	-0.37	0.033		
hvu-novel-miR90002	-0.50	0.0005		
hvu-novel-miR90003	-0.91	0.003		
hvu-novel-miR90000	-1.01	0.043	0.83	0.0005
hvu-new-miR396a	-1.04	0.040		
hvu-new-miR477	-1.32	0.0009	-1.42	0.0003
hvu-novel-miR90004	-1.71	0.0001		

4.3.4.1 hvu-miR5051

Only one miRNA under differential expression in barley under salt stress was previously recorded in miRBase. This miRNA, hvu-miR5051, decreased expression in leaf tissue under salt stress conditions.

4.3.4.2 hvu-new-miR396a

The miRNA designated hvu-new-miR396a (Figure 4.6) is extremely highly conserved in various plants with 52 exact matches across 36 plant species when compared with the 5' strand and six exact matches across four monocot species when compared with the 3' strand (Table 4.5).

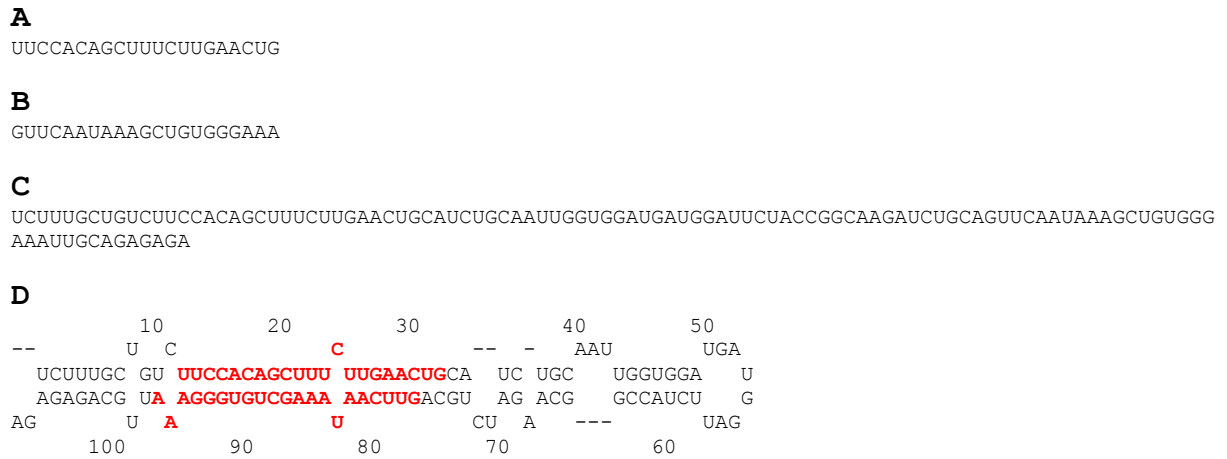


Figure 4.6. Sequences and secondary structure of hvu-new-miR396a. A: 5' strand sequence. B: 3' strand sequence. C: Hairpin-loop sequence. D: Secondary structure of hairpin-loop.

Table 4.5. List of species with an exact match for hvu-new-miR396a when comparing the 5' or 3' strand.

Identical 5' strand		Identical 3' strand		
<i>Acacia auriculiformis</i>	<i>Bruguiera gymnorhiza</i>	<i>Gossypium hirsutum</i>	<i>Populus trichocarpa</i>	<i>Aegilops tauschii</i>
<i>Acacia mangium</i>	<i>Camelina sativa</i>	<i>Hevea brasiliensis</i>	<i>Saccharum officinarum</i>	<i>Brachypodium distachyon</i>
<i>Aegilops tauschii</i>	<i>Carica papaya</i>	<i>Linum usitatissimum</i>	<i>Saccharum sp.</i>	<i>Vriesea carinata</i>
<i>Aquilegia caerulea</i>	<i>Citrus sinensis</i>	<i>Lotus japonicus</i>	<i>Salvia sclarea</i>	<i>Zea mays</i>
<i>Arabidopsis lyrata</i>	<i>Cucumis melo</i>	<i>Malus domestica</i>	<i>Solanum lycopersicum</i>	
<i>Arabidopsis thaliana</i>	<i>Digitalis purpurea</i>	<i>Manihot esculenta</i>	<i>Sorghum bicolor</i>	
<i>Asparagus officinalis</i>	<i>Eugenia uniflora</i>	<i>Medicago truncatula</i>	<i>Theobroma cacao</i>	
<i>Brachypodium distachyon</i>	<i>Fragaria vesca</i>	<i>Nicotiana tabacum</i>	<i>Vitis vinifera</i>	
<i>Bruguiera cylindrica</i>	<i>Glycine max</i>	<i>Oryza sativa</i>	<i>Zea mays</i>	

4.3.4.3 hvu-new-miR477

The only small RNA to have significant differential expression in both root and leaf tissue in the same regulatory direction was the small RNA labelled hvu-new-miR477.

This small RNA was most similar to miR477 found in a number of plants. Table 4.6 shows the alignment comparison with miR477 from various plant species.

The analysis with sRNAbench produced a mature small RNA sequence along with the hairpin-loop structure sequence. The lowest free-energy state of the hairpin-loop

secondary structure is shown in Figure 4.7. The suggested hairpin-loop sequence and structure is consistent with a class of small RNA called 451-like miRNA.

Table 4.6. Selection of blast search results comparing hvu-new-miR477 with miRBase.

Species	Designation	Sequence	Alignment score
<i>Hordeum vulgare</i>	hvu-new-miR477	UUCUCCUCUCAAAGGCCUCCAACG	
<i>Vitis vinifera</i>	miR477	AUCUCCUCUCAAAGGCCUCCA...	91
<i>Physcomitrella patens</i>	miR477c	CUCUCCUCUCAAAGGCCUCCA...	86
<i>Physcomitrella patens</i>	miR477f	...UCCUCUCAAAGGCCUCCAACAA	86
<i>Cucumis melo</i>	miR477a	ACCUCCUCUCAAAGGCCUCCA...	86
<i>Manihot esculenta</i>	miR477f	AUCUCCUCUCAAAGGCCUCCA...	86
<i>Populus trichocarpa</i>	miR477a	AUCUCCUCAGAGGCCUCCA...	82
<i>Asparagus officinalis</i>	miR477a	ACUCUCCUCUCAAAGGCCUCCG...	72

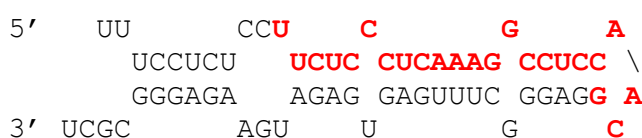


Figure 4.7. Hairpin-loop structure of hvu-new-miR477 precursor. The red highlighted bases are the mature miRNA sequence.

4.3.4.4 hvu-new-miR528

The microRNA hvu-new-miR528 (Figure 4.8) is not reported in miRBase for barley. This miRNA had the largest expression difference of all detected miRNAs between control and salt treated plants in leaf tissue. This sequence appears to be perfectly conserved in a selection of monocots as shown in Table 4.7.

4.3.4.5 hvu-novel-miR90000

The miRNA hvu-novel-miR90000 (Figure 4.9) showed down-regulation in leaf tissue in plants under salt stress, but up-regulation in roots of the same plants. There were no notable sequence similarities with other reported plant miRNAs. The differential

expression in leaf tissue was largely from the 3' strand reads, but mainly from 5' strand reads in root tissue.

Table 4.7. Selection of blast search results comparing hvu-new-miR528 with miRBase.

Species	Designation	Sequence	Alignment score
<i>Hordeum vulgare</i>	hvu-new-miR528	UGGAAGGGGCAUGCAGAGGAG	
<i>Oryza sativa</i>	osa-miR528	UGGAAGGGGCAUGCAGAGGAG	105
<i>Zea mays</i>	zma-miR528a	UGGAAGGGGCAUGCAGAGGAG	105
<i>Sorghum bicolor</i>	sbi-miR528	UGGAAGGGGCAUGCAGAGGAG	105
<i>Saccharum sp.</i>	ssp-miR528	UGGAAGGGGCAUGCAGAGGAG	105
<i>Brachypodium distachyon</i>	bdi-miR528	UGGAAGGGGCAUGCAGAGGAG	105
<i>Aegilops tauschii</i>	ata-miR528	UGGAAGGGGCAUGCAGAGGAG	105
<i>Vriesea carinata</i>	vca-miR528	UGGAAGGGGCAUGCAGAGGAG	105

A

UGGAAGGGGCAUGCAGAGGAG

B

CCUGUGCCUGCCUCUCCAUAU

C

CCGGAGCAGCAGCGUGGAAGGGGCAUGCAGAGGAGCGGCCAUGC AUGGGAGCUUUGCUUUGCUUUGCCUCUCCUGCUCUGGGCUCUAGCUCUCU
CCUGUGCCUGCCUCUCCAUAUCCUGCCGCUAA

D

```

      10      20      30      40      50
5' CCGG  A  CG          U  ---  -  ---  AU--  U  -  UUU  U
    AGC GCAG  GUGGAAGGGGCA GCA  GAG  GAGC  GGCC  GCA GGA GC  GC \
    UCG CGUC  UACCUUCUCCGU CGU  CUC CUCG  UCGG  CGU CUCU CG  CG U
3'  AA  C  CU          C  GUC  U  AUC  GUCU  C  C  UU-  U
      120      110      100      90      80      70

```

Figure 4.8. Sequences and secondary structure of hvu-new-miR528. A: 5' strand sequence. B: 3' strand sequence. C: Hairpin-loop sequence. D: Secondary structure of hairpin-loop with mature miRNA shown in red.

4.3.4.6 hvu-novel-miR90001

The miRNA designated hvu-novel-miR90001 did not have any close matches when performing a BLAST search against miRBase. A modest increase in expression occurred in leaf tissue in plants exposed to salinity stress. The sequence and proposed structure of the hairpin-loop precursor is shown in Figure 4.10.

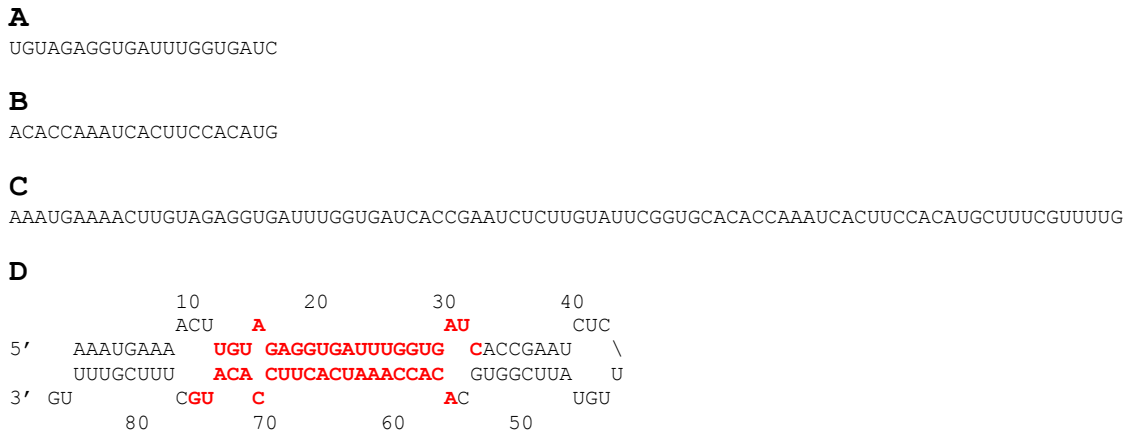


Figure 4.9. Sequences and secondary structure of hvu-novel-miR90000. A: 5' strand sequence. B: 3' strand sequence. C: Hairpin-loop sequence. D: Secondary structure of hairpin-loop.

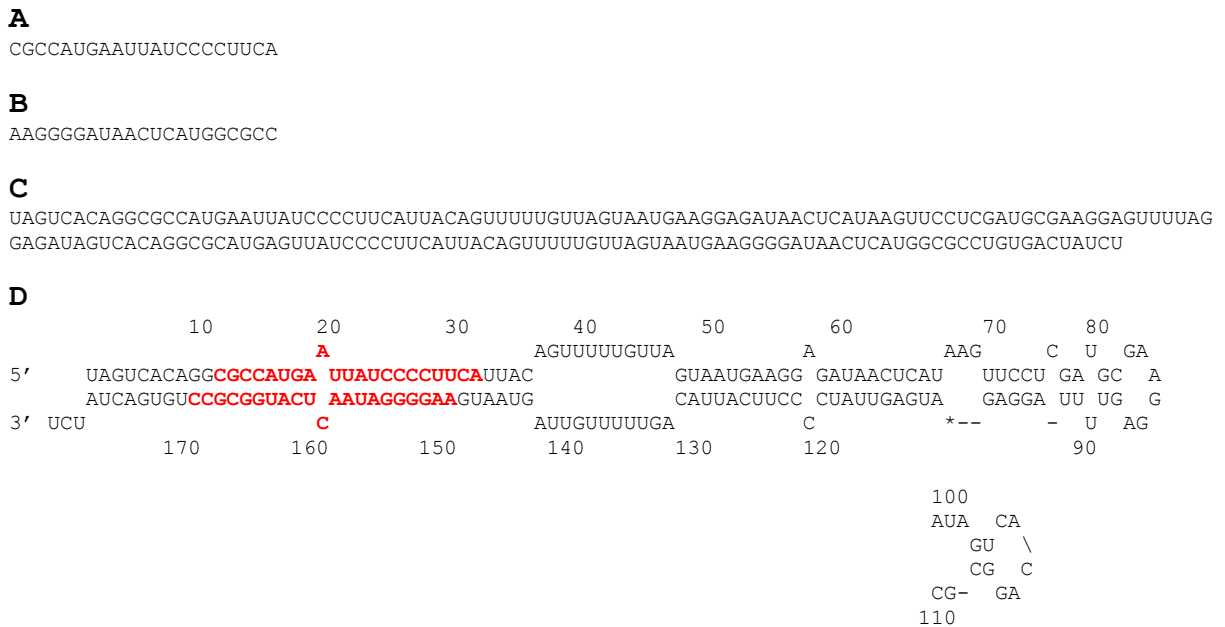


Figure 4.10. Sequences and secondary structure of hvu-novel-miR90001. A: 5' strand sequence. B: 3' strand sequence. C: Hairpin-loop sequence. D: Secondary structure of hairpin-loop with insertion point of side-loops shown with *.

4.3.4.7 hvu-novel-miR90002

The miRNA designated hvu-novel-miR90002 (Figure 4.11) was down regulated under salt stress in barley leaf tissue. One plant BLAST match was found with miR9655 in *Triticum aestivum* (wheat), however the alignment was poor (Table 4.8).

When comparing the hairpin-loop sequences with NEEDLE there was only a 24.2% similarity.

Table 4.8. Blast search results comparing hvu-novel-miR90002 with miRBase.

Species	Designation	Sequence	Alignment score
<i>Hordeum vulgare</i>	hvu-novel-miR90002	UGGGCCAAGGGAAGGAAAUA	
<i>Triticum aestivum</i>	tae-miR9655 CAAGGGAAGGAAGUAGCCAAC	66

A

UGGGCCAAGGGAAGGAAAUA

B

AUUUUCUCCCUUGGUCCGUU

C

GUCAAAGCUAAUGGGCCAAGGGAAGGAAAUAUGAUGACAUGAUCCAGCAGGGGAACAGUGGGCUACAUGCAUGGCUCUUGAGUGAGUAUAGGA
GUGUGGGAAAUACAUUGACGAAUUAGUUAUGUAAGAAUCUCUUUCAUUACUAUUUUCUCCCUUGGUCCGUUAACUUUGAUAG

D

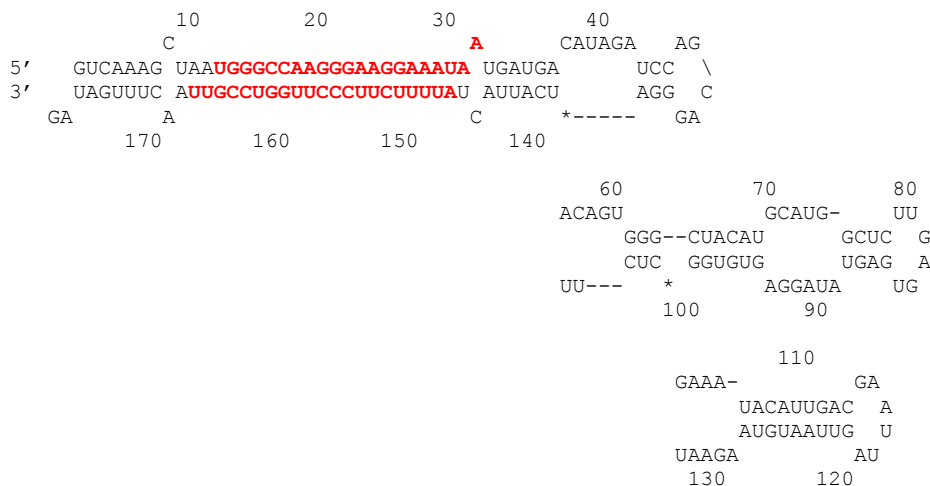


Figure 4.11. Sequences and secondary structure of hvu-novel-miR90002. A: 5' strand sequence. B: 3' strand sequence. C: Hairpin-loop sequence. D: Secondary structure of hairpin-loop with insertion point of side-loops shown with *.

4.3.4.8 hvu-novel-miR90003

The miRNA designated hvu-novel-miR90003 (Figure 4.12) does not return any BLAST results in plants against miRBase and therefore appears to be novel in barley.

The expression of this miRNA decreased by 47% in leaf tissue in plants exposed to salinity stress.

A

GAGAACGGGAUGCAGCCAAGG

B

UUGGCUACACCUAGUUCUCUU

C

CGCUAAGGGGCAGAGAACGGGAUGCAGCCAAGGAUGACUUGCCGGCUUCUGGUGUUGGGAGUUCGUAGAGCCUUAAGAAUUAGCCGGCAAGCUG
UCCUUGGCUACACCUAGUUCUCUUCUUCUGGUGUU

D

	10	20	30	40	50	60
	A C	- A C		GA	UCUGGUG	A G
5'	CGCUA GGGG A	GAGAAC GGG UG	AGCCAAGGAU	CUUGCCGGCU	UUGGG GUUC \	
	GUGGU CUUC	UCUCUUG UCC AC	UCGGUUCUG	GAACGGCCGA	AAUUC CGAG U	
3'	UU -	U A - A		UC	UUAAG--	- A
	120	110	100	90	80	70

Figure 4.12. Sequences and secondary structure of hvu-novel-miR90003. A: 5' strand sequence. B: 3' strand sequence. C: Hairpin-loop sequence. D: Secondary structure of hairpin-loop.

4.3.4.9 hvu-novel-miR90004

The microRNA, hvu-novel-miR90004 has three different hairpin loop structures which generate the same miRNA sequence in the 5' strand, but a different sequence in the 3' strand (Figure 4.13). There were no close matches when performing a BLASTN on the 5' strand or any of the 3' strands against miRBase, so this appears to be either unique to barley or currently undetected in other species.

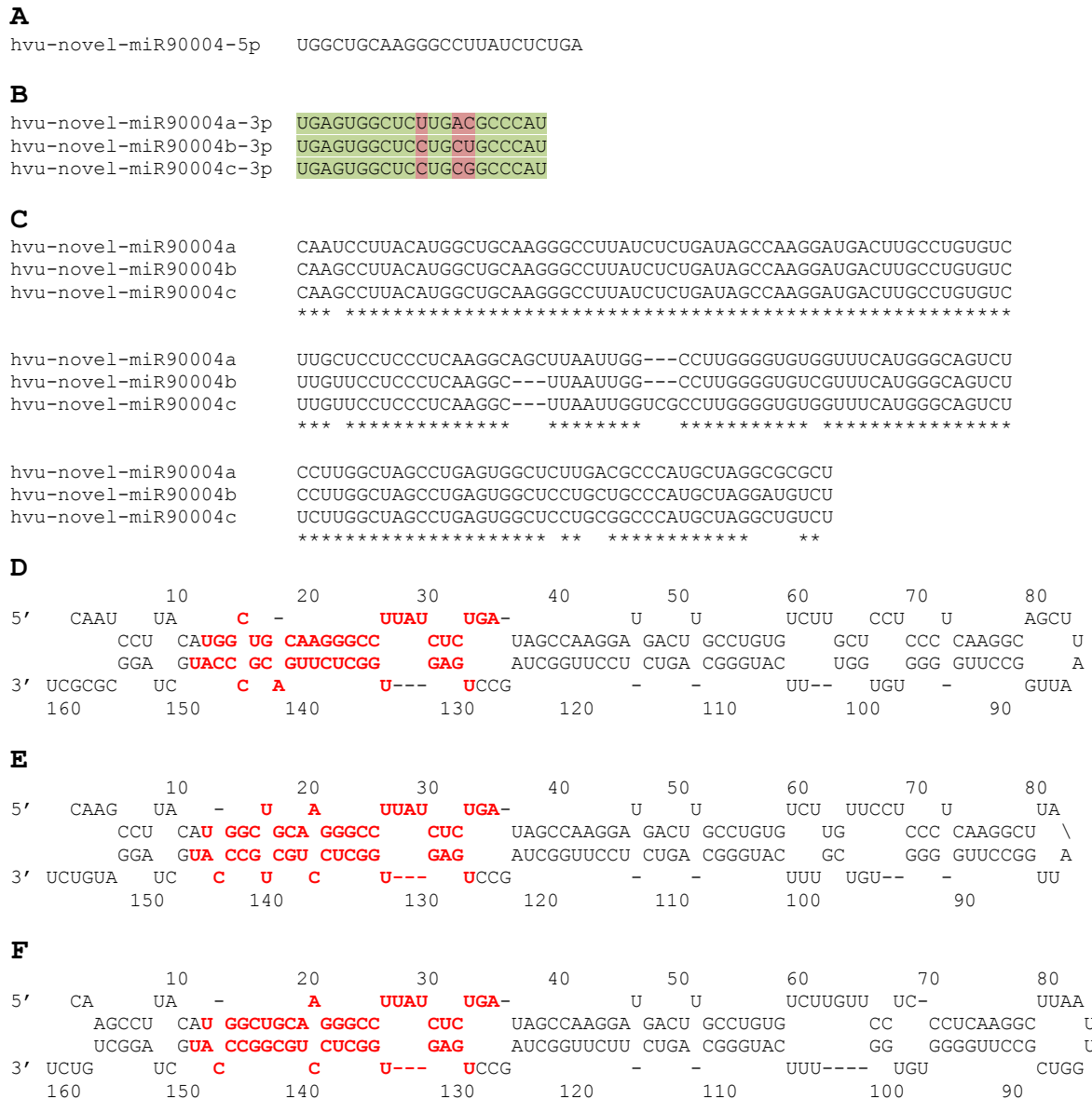


Figure 4.13. Sequences and secondary structures for hvu-novel-miR90004. A: 5' strand sequence. B: Alignment of the three different 3' strand sequences. C: Sequence alignment of the hairpin-loop precursors. D-F: The lowest free-energy secondary structures of the three hairpin-loop sequences hvu-novel-miR90004a, hvu-novel-miR90004b and hvu-novel-miR90004c respectively.

4.4 Discussion

The sRNA fragment size population (Figure 4.1) in leaf tissue is similar to prior research in barley, with 21-nt and 24-nt small RNAs making up the majority of reads (Schreiber et al. 2011). The root tissue did not show the expected size population.

Potential contaminants have been considered such as degraded barley mRNA, bacterial RNA and fungal RNA. Degraded barley mRNA was ruled out as a source of contamination by alignment of the sRNA library with the barley transcriptome. Microbiome analysis was considered, but the short length of reads is not compatible with the existing microbiome analysis tools available.

The number of unique reads (Figure 4.2) for each small RNA size in leaf tissue was in concordance with prior research (Schreiber et al. 2011), with the vast majority of unique reads coming from 24 nt small RNAs. While there was a significantly higher number of 24 nt unique reads in the root tissue, the abundance of other sizes indicates some form of contamination.

While the differences in read size populations for control and salt treated libraries were not statistically significant, there appears to be a slight bias to larger small RNAs in the salt treated root compared with control root samples. The reason for this is unknown, it is possible that the difference in salt concentration created a slight size selection bias during the RNA extraction process. A search of the literature did not find any relevant research that might explain this apparent bias.

Changes in 23/24 nt sRNA expression were mapped to genomic features and compared with random sampling of the genome for reference (Figure 4.3). In both leaf and root tissue in barley under salt stress, the changes in 23/24 nt sRNA

expression was targeting PCGs to a much higher level than random sampling. The 23 nt miRNAs appeared to target PCG more than 24 nt siRNA. There was a general overrepresentation of 23/24 nt sRNAs targeting DNA transposons and an underrepresentation of 23/24 nt sRNAs targeting RNA transposons.

4.4.1 Identification of previously categorised miRNA

While the separation between control and salt treated plants for miRBase mapped miRNA was clear in the principle component analysis for root tissue, this was not the case for leaf tissue (Figure 3.1). This may be due to the limited number of 71 unique miRNAs categorised in miRBase for *H. vulgare*. Only 40 of these known small RNAs (Table 4.2) were expressed at high enough levels to be useful for analysis.

4.4.2 Identification of miRNA not previously identified in *Hordeum vulgare*

The principle component analysis of miRNA that had not been previously categorised (Figure 4.5) showed clear separation between root and leaf tissue. When looking at only root or only leaf tissue, the control miRNA expression separated well from the salt treated group, indicating that exposure to salinity stress has an effect on at least some of these miRNAs.

Fifty-nine unique miRNAs were discovered that had not previously been classified in miRBase. All of these unique miRNAs were expressed and useful for analysis. This greater diversity of miRNAs could explain the clear separation in PCA of leaf tissue

between control and salt treated groups in the new small RNAs compared with previously categorised small RNAs which did not experience a clear separation. At the time of writing, there were 71 *H. vulgare* miRNAs in miRBase. When the newly found miRNAs from this study are published and incorporated in miRBase, this will represent an increase of 83% in *H. vulgare* miRBase records.

4.4.3 Differential expression of miRNA

The observed differences in sRNA size profiles between root and leaf tissue along with the small number of differentially expressed miRNA in root tissue may indicate problems with the sRNA library obtained for root tissue. The potential issues with root tissue sRNA may have resulted in significantly fewer reads of *H. vulgare* miRNA compared with other sources.

4.4.3.1 hvu-miR5051

The results from differential expression analysis highlighted a single previously categorised small RNA in *H. vulgare* hvu-miR5051. This miRNA was discovered by Schreiber et al. (2011). Ozhuner et al. (2013) investigated miRNA changes in barley exposed to excess boron stress. They found that there was a 2-fold up-regulation of miR5051 in the roots of barley under 1 mM boron toxicity stress compared with control plants, with no difference observed in leaves. This is different to the log₂ fold change of -0.37 (p = 0.033) we observed in leaves under salinity stress relative to

control plants. Research by Alpaslan & Gunes (2001) investigating the interaction between boron and salinity stress in cucumbers found that a higher dose of boron resulted in higher levels of sodium accumulation in the plant. While there may be interactions between boron and salinity stress, it is difficult to draw comparisons with research in this area. Bai et al. (2017) indicated that miR5051 targets a serine/threonine protein kinase by use of degradome sequencing and analysis, however this research was focussed on barley seed development and germination. Other potential targets for miR5051 may not have been present in the abundance required for degradome analysis during seed development and germination stages.

4.4.3.2 hvu-new-miR396a

In *Zea mays* (maize), miR396 had a 48% decreased expression in leaves but no significant changes in roots under salinity stress (Fu et al. 2017). This result compares well with the decrease of 50% in miR396 expression we observed in *H. vulgare* leaf tissue under salinity stress and no significant changes in root tissue. Although it should be noted that the salinity stress conditions were different and the salinity tolerance between species would also be different. Under drought stress in *Z. mays*, an increase in the expression of miR396 has been reported (Aravind et al. 2017).

In *Brachypodium distachyon* (a close relative to *H. vulgare*) the function of miR396 has been observed with an increase in expression under cold stress. This miRNA targets

a serine/threonine-protein kinase, a PTII-like protein and seven uncharacterised proteins (Zhang et al. 2009; Zhang et al. 2013).

4.4.3.3 hvu-new-miR477

The microRNA miR477 was first discovered in *Populus trichocarpa* (poplar) and showed differential expression related to tension and compression stress (Lu et al. 2005). The class of 451-like miRNA are different from other miRNAs by not producing a double-stranded mature miRNA, and are produced when a short hairpin structure is directly loaded into and cleaved by AGO2 without any Dicer processing (Yang et al. 2010). In *Vitis vinifera* (grape), miR477 was found to be expressed at significant levels in the immature berry, veraison and mature berry but not in the leaf, inflorescence or root (Mica et al. 2009). In *Physcomitrella patens* (moss), mature miR477 was found in the protonema stage and an incompletely processed precursor in gametophores (Fattash et al. 2007). In *Manihot esculenta* (cassava), miR477 was found to be highly expressed in male flower tissue (Khatabi et al. 2016), with the GRF transcription factor which is related to leaf and cotyledon growth being a target (Chen et al. 2015). *Asparagus officinalis* (asparagus) expresses miR477 (Harkess et al. 2017). This is the closest relative to barley that has been shown to express this miRNA, but the function of miR477 in asparagus is unknown.

4.4.3.4 hvu-new-miR528

In *Saccharum sp.* (sugarcane) the expression of miR528 was found to fluctuate heavily in shoot tissue under a strong salinity shock of 340 mM, with a sharp increase in expression at 1 h, and decreasing at both 6 h and 24 h, but remaining above control levels (Bottino et al. 2013). This increase in miR528 is consistent with our findings for salinity stress.

4.4.3.5 hvu-novel-miR90000

The miRNA designated hvu-novel-miR90000 showed down-regulation in leaf tissue in plants under salt stress, but up-regulation in roots of the same plants. It should be noted that the adjusted p-value of the leaf differential expression is only marginal at $p = 0.043$. The differential expression in leaf tissue was largely from the 3' strand reads, but mainly from 5' strand reads in root tissue. The reason for this is unknown and may be worthy of future investigation.

4.4.3.6 hvu-novel-miR90002

While a BLAST search yields one match of miR9655 from *T. aestivum* for hvu-novel-miR90002, the alignment is poor. The two sequences are not similar enough to confidently say that they are related. Two known targets for miR9655 in *T. aestivum* by degradome analysis were an RPM1-like disease resistance protein and an unknown protein labelled TC402663 (Li et al. 2015).

4.4.3.7 hvu-novel-miR90004

Interestingly, hvu-novel-miR90004 has three different hairpin loop structures which generate the same sequence in the 5' strand, but a different sequence in the 3' strand (Figure 4.13). There were no close matches when performing a BLASTN on the 5' strand or any of the 3' strands against miRBase, so this appears to be either unique to barley or currently undetected in other species. The \log_2 fold change in leaf tissue was -1.71. It was not possible to determine which of the three hairpin-loop isoform precursors had the greatest effect on the differential expression due to the lower number of reads in the 3' strands.

4.4.4 Conclusion

Fifty-nine previously unknown barley miRNAs have been identified, with six of these miRNAs displaying a significant change in expression in response to salinity stress. Genomic features targeted by 23/24 nt sRNA in response to salinity stress were discovered with PCGs being targeted at a much higher rate than random sampling of the genome. In chapter 6, the targets for 23/24 nt sRNAs are compared with DNA methylation changes (Chapter 5) to observe how these changes in sRNA expression and DNA methylation come together.

5 DNA methylation changes in barley under salinity stress

5.1 Introduction

DNA methylation plays an important role in a range of functions including gene control, development and cellular differentiation, management of Transposable Elements (TEs), and stress response (Downen et al. 2012; Elhamamsy 2016). A cytosine base can be modified to 5-methylcytosine by the addition of a methyl group (Saze et al. 2012). The cytosine methylation pathway is related to the context of the cytosine, and is grouped by CG, CHG and CHH where H can be A, C or T (He et al. 2011).

There are various techniques for determining methylation changes in a genome. Some, like Methylation Sensitive Amplified Polymorphism (MSAP) (Reyna-Lopez et al. 1997) or Methylation Sensitive Genotyping By Sequencing (ms-GBS) (Kitimu et al. 2015), use methylation sensitive restriction enzymes to generate fragments, which can then be analysed by comparing bands following polyacrylamide/capillary electrophoresis or next generation sequencing. The methylation status of the generated fragments is then inferred from presence/absence or relative abundance (Rodríguez-López et al. 2012). However, this approach generates information from a very limited number of methylation sites (Rauluseviciute et al. 2019).

A method commonly referred to as methyl-capture involves fragmenting the genome, capturing methylated cytosines by immunoprecipitation and then

sequencing the captured fragments (Down et al. 2008). This method does not provide base-pair resolution information of methylation, only methylation states of genomic regions (Teh et al. 2016). Two third-generation sequencing techniques, The Oxford Nanopore MinIon and PacBio SMTR are able to gain methylation information at base-pair resolution by detecting the methylation state of a cytosine base without the need for chemically altering bases (Weirather et al. 2017). There are limitations with these methods, the MinION accuracy is 91% for correctly calling a 5-methylcytosine base when a strict filtering is applied that discards 32% of reads (Simpson et al. 2017). The PacBIO SMRT sequencer needs greater than 250 times coverage to accurately call a 5-methylcytosine base (Liu et al. 2020).

A bisulphite treatment converts unmethylated cytosines to uracil which are then converted to thymine after PCR amplification. Methylated cytosines remain as cytosines after the treatment. By comparing the bisulphite treated sequence with either untreated DNA, a reference genome, or complementary bisulphite treated strands highlights those cytosines that are methylated and those that are not. The entire genome bisulphite state can be determined by whole genome bisulphite sequencing (Cokus et al. 2008).

The reduced representation bisulphite sequencing method is a compromise between full information for each cytosine base and lower resolution methods. It can be useful in cases of very large genomes where the information derived from MSAP is

technically demanding and very limited in the information obtained, and in the case of whole genome bisulphite sequencing, prohibitively expensive.

5.1.1 EpiGBS

The method used in this study to obtain methylation information was epiGBS (epigenetic Genotyping By Sequencing), which is a reduced representation bisulphite sequencing method developed by van Gurp et al. (2016). To distinguish the epiGBS “wet-lab” protocol from the bioinformatics tools the former will be referred to as the epiGBS protocol, and the latter the epiGBS bioinformatics pipeline.

EpiGBS is able to extract bisulphite sequencing information from a reduced portion of the genome by restriction digest and size selection. The sites selected for sequencing should be the same for each sample, enabling a significantly reduced cost compared to that of whole genome bisulphite sequencing. Of course, there is a trade-off in the number of sites that will be sequenced. In brief, genomic DNA is digested with two different Type II restriction enzymes. Then adapters designed with co-adhesive ends complementary to those generated by the restriction enzymes (Ba and Co shown in Figure 5.1), are ligated to the 3' and the 5' ends of the restriction products. The 5' end of the adapter is not joined to the 3' end of the genomic DNA as the adapter is lacking a 5' phosphate group. That section of the adapter is replaced through a process referred to as ‘nick repair’, using A, G, T and 5-methylcytosine with a DNA polymerase that has 5'-3' exonuclease activity (Figure 5.2). The three

base Unique Molecular Identifier (UMI) (N's in Figure 5.1 and Figure 5.2) in the adapter were not complementary before the nick repair process, but are complementary after the process. Following nick repair, the entire adapter section has 5-methylcytosine at all cytosine locations except for one base at the 3' end of the barcode. This cytosine is used as a control to measure the conversion rate of the bisulphite treatment, and to identify the strand that was bisulphite treated. Conversion of all other adapter cytosines to 5-methylcytosine is to maintain the sequence after bisulphite treatment.

Ba adapter

```
5' AXAXTXTTXXXTAXAXGAXGXTXTTXXGATXTNNNBBBBC 3'
3' TGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGANNNBBBBGAT 5'
```

Co adapter

```
5' XTXGGXATTXXTGXTGAAXXGXTXTTXXGATXTNNNBBBCTGXA 3'
3' GAGCCGTAAGGACGACTTGGCGAGAAGGCTAGANNNBBBBG 5'
```

Figure 5.1. Sequence of adapters ligated to digested DNA in the epiGBS protocol. B represents the barcode sequence used to determine sample identity. N is a random base consisting of A, G, T or 5-methylcytosine to make up the Unique Molecular Identifier, and is used to find and remove PCR duplicates with a total of 4096 different possible combinations when both adapters used. X represents 5-methylcytosine which is used to prevent sequence change during the bisulphite treatment process.

The two DNA strands are separated and bisulphite treated, converting unmethylated cytosines to uracil. The strands are then PCR amplified as shown in Figure 5.3. The strand that is bisulphite treated has the cytosine in the adapter designated as the Watson/Crick identifier modified to a uracil then a thymine. The complementary strand after PCR replication remains as a cytosine. If both paired reads have a cytosine in this location, then this indicates the bisulphite treatment failed to convert

that base and can be used to determine the conversion rate.

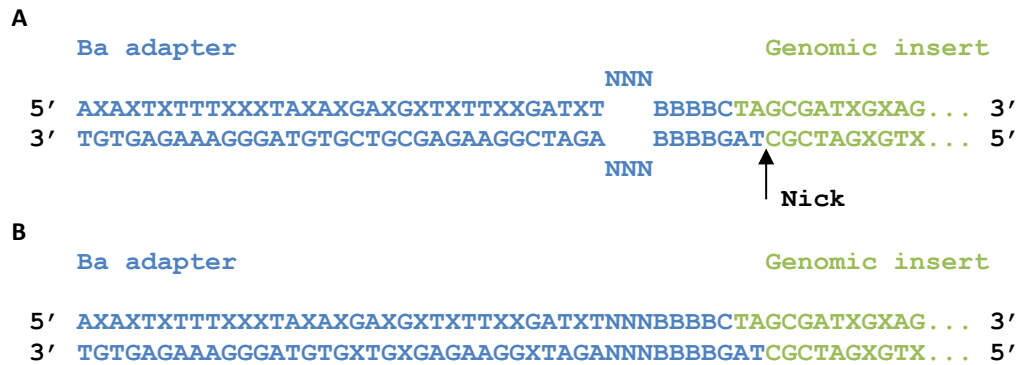


Figure 5.2. Stages of the nick repair process to replace one strand of the adapter after ligation in the epiGBS method. This figure only shows one adapter, the same process occurs in both Ba and Co adapters. A. After ligation the adapter and genomic insert is not ligated at the site shown by the nick label, the UMI site (Ns) does not have sequence complementarity. B. After the nick repair process which uses A, G, T and 5-methylcytosine with a DNA polymerase that has 5'-3' exonuclease activity the cytosines on the original Ba adapter have been replaced with 5-methylcytosine (shown as X) and the UMI has complementarity between adapter strands.



Figure 5.3. Bisulphite treatment, PCR amplification and identification of Watson and Crick strands in the epiGBS method. A, Double stranded DNA after ligation of adapters and nick repair. B, Strand 1 from A after bisulphite treatment. C, Double stranded DNA after PCR replication of the single DNA strand from B. D, Strand 2 from A after bisulphite treatment. E, Double stranded DNA after PCR replication of the single DNA strand from D. The Watson/Crick identifier is shown in light orange with a T indicating that strand had received the bisulphite treatment (Watson strand) and a C indicating it was a complementary strand to the bisulphite treated strand (Crick strand).

The four base barcode in the adaptors is used to identify samples, and the three base UMI at either end is used to identify PCR duplicates. Reading the sequence multiple times from PCR duplicates does not add any useful information and so these reads need to be identified and removed. The total of six UMI bases per sample enables identification of 4096 different combinations. If two sequences have an almost identical insert sequence, and the same barcode and UMI base combination, then there is a very high chance that the read is a PCR duplicate and only the highest quality read should be kept for further analysis.

Once PCR duplicates are identified and removed, all the sequences are then mapped against the reference genome and the percentage methylation at each mapped cytosine is determined for each sample. The differences in percentage methylation can then be investigated between control and treatment samples or different tissue types. In this study, barley plants grown in saline conditions were compared with those in control conditions, and nucleic acid samples collected from roots and third leaf of the treated and control plants.

5.2 Methods

5.2.1 DNA extraction

DNA was extracted from leaf and root tissue that had been collected as described in chapter 2 using the Qiagen DNeasy Kit (CA, USA). Tissue had been ground under liquid nitrogen with a mortar and pestle and aliquoted for storage at -80°C for DNA

and RNA extractions. Approximately 100 mg of ground leaf or root tissue was digested with 400 µl of buffer AP1 and 4 µl of RNase A, mixed with a vortex and a micropestle was used to crush any clumped material. The DNeasy kit was then used as per the manufacturer's instructions and the final elution used 50 µl of Buffer AE. A Nanodrop spectrophotometer (Thermo Fisher, DE, USA) was used to quantify DNA concentration and purity.

5.2.2 EpiGBS

The epiGBS method is outlined by van Gurp et al. (2016), however modifications were needed to obtain useable results. The method with modification is as follows. Genomic DNA concentration was checked with a Qubit (Life Technologies, NY, USA) and diluted to 20 ng/µl. A 40 µl restriction enzyme digestion reaction containing 400 ng of DNA, 4 µl Thermo Fast Digest Buffer (10x) (MA, USA), 6.25 µl of 20 mg/ml BSA (NEB B9000S) (MA, USA), 1 µl of 20 units/µl *Nsi*I-HF (NEB R3127S) (MA, USA), 2 µl of *Csp*6I restriction enzyme (Thermo FD0214) (MA, USA) was performed over 16 hours at 37°C then held at 4°C. Sequencing adaptors were ligated to restriction products in a ligation reaction containing: 40 µl of product from the digestion reaction, 2 µl of 5 ng/µl Ba adapter, 2 µl of 6 ng/µl Co adapter, 6 µl of T4 ligase buffer and 10 µl of T4 ligase (NEB M0202L) (MA, USA) with the reaction held at 22°C for 3 hours then 4°C for 16 hours. Adapter design and sequence are shown in Figure 5.1 where X is 5-methylcytosine, N is a random UMI base used to later

determine PCR duplicates, and B is the barcode sequence as shown in Appendix A Table A.9.9.

The ligated products were pooled and split across four Qiaquick spin columns (used as per the manufacturers protocol), and each column eluted with 60 μ l water. Size selection was performed using AMPure XP beads (Beckman Coulter) with 47.2 μ l beads added to each 59 μ l of DNA product from the previous stage and pipette mixed 10 times. The bead solution was incubated at room temperature for 5 mins then placed on a magnetic rack for 2 mins. Supernatant was pipetted off and discarded, then 200 μ l of 70% ethanol was added for 2 minutes and the supernatant discarded. This step was repeated once more then the beads were allowed to dry in air for 4 mins. Tubes were removed from the magnetic rack with 12 μ l water added, pipette mixed 10x and left to incubate at room temperature away from the magnetic rack for 2 mins. The tubes were then transferred to the magnetic rack and left for 2 mins before 11 μ l of the supernatant was collected by pipette for the nick repair process.

Four nick repair reactions containing the following reagents were incubated at 15°C for 1 hour: 9 μ l of size selected DNA from the previous step, 0.625 μ l water, 1.25 μ l of 10 mM 5-methylcotisine dNTP mix, 1.25 μ l of NEB B7002S Buffer2, 0.375 μ l of NEB M0209S DNA polymerase. A PCR was used to check the success of the previous stages and the products observed by electrophoresis on a 1% agarose gel. The PCR

contained 1 μl of the nick repaired products, 3.4 μl water, 5 μl Kappa HiFi HotStart + Uracil x2 (Kappa Biosystems, MA, USA), 0.3 μl of 10 mM Illumina PE forward primer, 0.3 μl of 10 mM Illumina PE forward primer (CA, USA). The PCR cycling conditions consisted of 95°C for 3 min then 18 cycles of 98°C for 10s, 65°C for 15s, 72°C for 15s then a final 72°C for 5 min before a 4°C hold.

The bisulphite treatment was performed on the nick repaired products using the Zymo EZ DNA Methylation-Lightning Kit (CA, USA) as per the manufacturer's instructions. A PCR was performed immediately after the bisulphite treatment as the bisulphite treatment would degrade DNA over time. This PCR consisted of 4 separate reactions each containing 1 μl of the bisulphite treated DNA, 3.4 μl water, 5 μl Kappa HiFi HotStart + Uracil x2 (MA, USA), 0.3 μl of 10 mM Illumina PE forward primer, 0.3 μl of 10 mM Illumina PE forward primer (CA, USA). The PCR cycling conditions consisted of 95°C for 3 min then 18 cycles of 98°C for 10s, 65°C for 15s, 72°C for 15s then a final 72°C for 5 min before a 4°C hold. The products from this PCR were checked by electrophoresis on a 1% agarose gel, and later using an Agilent 2100 Bioanalyser with a High Sensitivity DNA kit (Agilent Technologies, CA, USA).

5.2.3 First bisulphite sequencing run

An in-silico digestion (DigestBarleyDouble.R shown in section 10.12) was performed on the *H. vulgare* Morex genome in an attempt to determine the best size selection for good sequence coverage. This simulation suggested optimal coverage would be

achieved with a selection of fragments ranging from 380 bp (base-pairs) to 560 bp (including the adapters). A Pippin prep (Sage Science, MA, USA) was used as per the manufacturers protocol to select DNA fragments from 380 bp to 560 bp. The selected fragments were dried for shipping to NovoGene (Hong Kong, China) using a DNA Speed Vac (Savant, NY, USA). Upon arrival, the DNA was re-suspended and checked for quality. The library was spiked with 30% PhiX and sequenced using a single lane of an Illumina HiSeq (CA, USA) with 150 + 150 bp paired end reads.

5.2.4 Second bisulphite sequencing run

The quantity of PCR duplicate reads from the first sequence run indicated a much wider size selection should have been used. The library constructed as described in section 5.2.2 was used for the size selection process, with 20 µl of DNA library mixed with 10 µl of AMPure XP beads and pipette mixed 10 times. This mixture was left to incubate at room temperature for 5 minutes and then placed on the magnetic rack for 2 minutes. The supernatant contained the smaller fragments and was collected, and the beads bound the large fragments and were discarded. The supernatant had 0.8x AMPure XP beads added and was pipette mixed 10 times. The beads were allowed to incubate at room temperature for 5 minutes then were placed on a magnetic rack for 2 mins. Supernatant was pipetted off and discarded, then 200 µl of 70% ethanol was added for 2 minutes then the supernatant discarded. This step was repeated once more then the beads were allowed to dry in air for 4 mins. Tubes were removed from the magnetic rack with 20 µl water added, pipette mixed 10x and left to incubate at

room temperature away from the magnetic rack for 2 mins. The tubes were then transferred to the magnetic rack and left for 2 mins before 19 μ l of the supernatant was collected by pipette. A Bioanalyser with a High Sensitivity DNA Kit (Agilent Technologies, CA, USA) was used to check the concentration and size distribution of fragments. The DNA library was dried, re-suspended, checked, spiked with 20% PhiX and sequenced by NovoGene as described in section 5.2.3.

5.2.5 Bioinformatics

Sequence reads were quality checked with fastqc (Andrews 2017) and PhiX reads were removed by alignment with the PhiX genome using Bowtie2 (Langmead & Salzberg 2012) and discarding matching reads. The remaining reads were checked again using fastqc to ensure the PhiX reads had been removed. The distribution of reads by GC content indicated that PhiX had been removed with the absence of the previously visible peak at around 46% GC content.

The epiGBS bioinformatics pipeline provided in the github repository (<https://github.com/thomasvangurp/epiGBS>) contains multiple errors which prevent operation without significant modification to the poorly documented code. A modified version of this pipeline was used to generate differential methylation information, but with the errors and lack of feedback from the author, I was not confident with the analysis. Given this, I created a completely different bioinformatics pipeline to perform the task.

A python program `fastqfilter.py` was written to perform the filtering of the paired-end reads and this can be found in section 10.13. To determine the quality thresholds that would be most appropriate for the sequence reads, a sample of one million reads was used to test the rate of discarded reads at different Phred qualities. The Phred sequencing quality is related to probability of accurately calling a base (P), where the Phred quality $Q = -10 \log_{10}(P)$. A strict quality filtering was applied with any read containing a base sequenced with a Phred quality lower than 5 or an average Phred quality over the whole paired-end read lower than 35 being discarded.

There were a significant number of PCR duplicates and these reads needed to be removed to reduce any resulting bias. Reads were split so that the bioinformatic processes could be performed faster in parallel. These split files were further separated into files for each barcode combination (each sample) using the custom python script `Barcodes.py` shown in section 10.14. Each sample was split based on the six base UMI combination using the two custom programs `WobbleIndexingToPickle.py` (section 10.15) and `WobblePickleToFiles02.py` (section 10.16), which generated 4096 files per sample corresponding to each UMI combination. These files were then processed with the custom program `dedup.py` (section 10.17) which used multithreading and the custom Cython program `dedupcy.pyx` (section 10.18) to compare reads and remove any duplicates, only keeping the highest read quality of the duplicates. After PCR duplicates were removed, the 4096 UMI reads were merged again using the linux command `cat`, and

FastQC (Andrews 2017) was used again to check the quality of the remaining reads.

PEAR (Paired-End reAd merger) (Zhang et al. 2014) converted each paired end sample in to either a merged read or left as two unmerged paired reads if there was no significant overlap. The custom program `epiGBS2Bismark06.py` (section 10.19) prepared reads for use with Bismark (Krueger & Andrews 2011) by performing reverse complements on all Crick assembled reads, all Crick unassembled second strand reads and all Watson unassembled second strand reads. This was necessary as PEAR reverse complements all of the second strand reads which needed to be converted back for Bismark and the assembled reads were required to all be in Watson orientation for Bismark.

A genome index for Bismark was created using the ensemble genome for *Hordeum vulgare*, IBSC v2.45 (Mascher et al. 2017). Bismark was used twice, once with the assembled reads and again with the unassembled paired end reads. Bismark command-line options for the unassembled reads include `-p 8 --maxins 2000 --unmapped --ambiguous` while options for the assembled reads include `-p 8 --unmapped --ambiguous`. Then `bismark_methylation_extractor.pl` ran with the options `--bedGraph --CX_context` to generate files with methylation counts for each cytosine read.

The `.cov` files generated by Bismark for each sample were separated for each

chromosome using the custom pypy program `splitcov.py` (section 10.20). Then samples were merged, but keeping separate files for each chromosome using the custom pypy program `mergecov.py` (section 10.21). The context for each cytosine was determined by mapping against the reference genome using the custom pypy program `covcontext.py` (section 10.22) and files were split by CG, CHG, or CHH context with the custom pypy program `cov2R.py` (section 10.23) to be able to import the data set into custom scripts developed for the R programming language (Team 2019).

The R programming language was used for data analysis along with the libraries, `Genomation` (Akalin et al. 2015), `GenomicFeatures` (Lawrence et al. 2013), `methylKit` (Akalin et al. 2012), `Rsamtools` (Morgan et al. 2016). Principle Component Analysis using percentage methylation levels was performed with the custom R script `PCA_methylation.R` (section 10.24). Plots of global methylation levels in 1 million bp windows were performed for each cytosine context with the custom R script `plot_chromosome_methylation.R` (section 10.25). The methylation pattern around genes and transposable elements for each cytosine context were determined with the custom pypy programs `methylation_levels_genes.py` (section 10.26) and `methylation_levels_TEs.py` (section 10.27) respectively. The custom R script `find_differentially_methylated.R` (section 10.28) grouped the methylation data in 100 bp windows and by cytosine context and found significantly differentially methylated regions using a false discovery rate threshold of $p_{adj} < 0.01$. These

differentially methylated regions were epiGBS fragments mapped to 100 bp genomic window sites with significantly differential methylation across the window, with the results being used in multiple custom R scripts. One generated histograms of the distance between the differentially methylated site and TSS (Transcription Start Site) `dist_to_TSS.R` (section 10.29), another looked for location of differential methylation and locality with annotated regions such as genes, non-coding RNAs and transposable elements (TEs) `diff_meth_annotation.R` (section 10.30). The enrichment of differentially methylated regions with annotated regions was compared by generating a random sampling of the bisulphite reads that were mapped to the genome, and running `diff_meth_annotation.R` as though these randomly selected were differentially methylated regions using the custom R script `random_genome_sampling.R` (section 10.31).

5.3 Results

The extracted DNA from root and leaf tissue had concentrations shown in Table 5.1. Capillary electrophoresis electropherograms of bisulphite treated DNA libraries used for sequencing are shown in Figure 5.4 with the first library being a narrow size selection and the second having a much wider size selection profile.

The sequence read numbers are shown in Table 5.2 and the GC content of reads is shown in Figure 5.5. The two peaks are from the combination of the bisulphite treated barley DNA library at around 28% GC content and the PhiX spike in at

around 46% GC content. Only the one broad peak remained after discarding PhiX reads, indicating the PhiX reads were successfully removed. The fastq files from the two sequence runs were merged, filtered, and samples were determined by matching up to the associated barcode as shown in Appendix A Table A.9.9, with Watson and Crick reads identified along with unconverted reads as shown in Figure 5.6. The unconverted reads were used along with the Watson and Crick reads to determine the bisulphite conversion rate was 94.2%.

Table 5.1. Extracted DNA concentration from barley plant leaf and root tissue samples. Tissue samples 1-4 were from control plants and 5-8 were from salt treated plants.

Leaf tissue sample	Concentration (ng/ul)	Root tissue sample	Concentration (ng/ul)
1	41.4	1	38.3
2	55.8	2	36.7
3	43.8	3	53.6
4	54.5	4	51.4
5	97.3	5	40.0
6	80.0	6	54.2
7	45.0	7	54.1
8	44.5	8	53.9

Table 5.2. Total sequence reads and PhiX spike in reads for epiGBS sequencing runs.

Sequencing run	Total raw reads	Reads after PhiX removed	PhiX % of reads
1	492,629,410	285,503,368	42%
2	446,094,532	349,415,026	22%

Figure 5.6 shows the reads assigned to each sample by barcode matching. There was a noticeable decrease in read quantity in L4, L8, R4 and R8 compared with all other reads. These samples all used the Ba4 barcode.

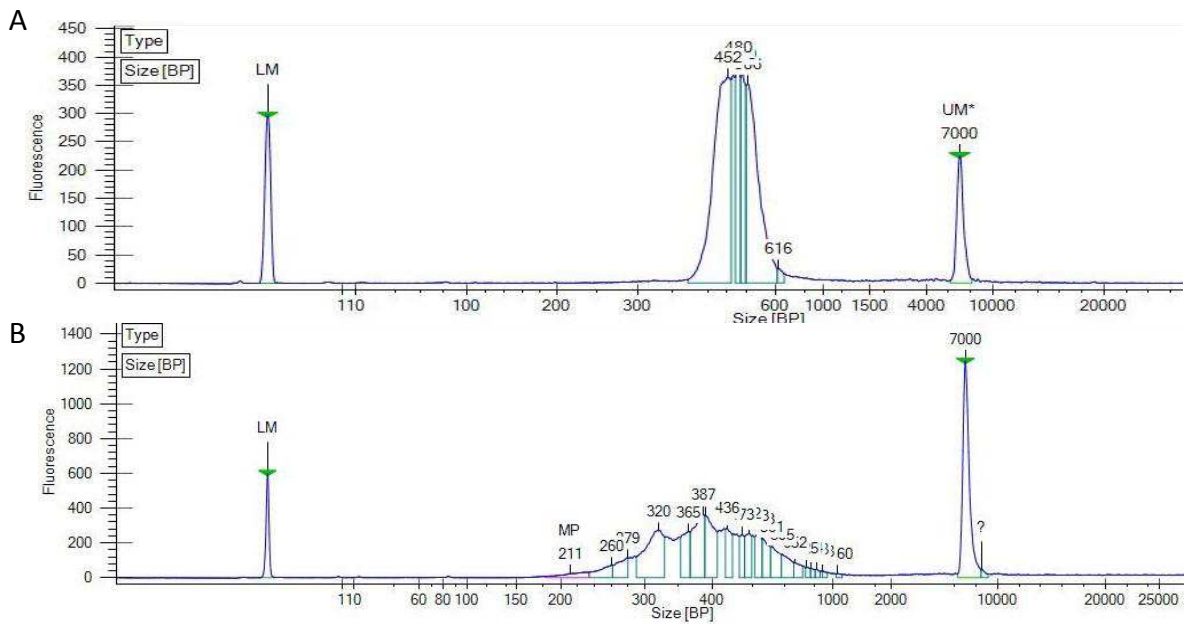


Figure 5.4. Capillary electrophoresis results after libraries re-suspended before sequencing. A, The first library for sequencing had a narrow size selection using the Pippin Prep. B, The second library had a much wider size range with magnetic beads used for the size selection profile. These two capillary electrophoresis runs were performed by NovoGene.

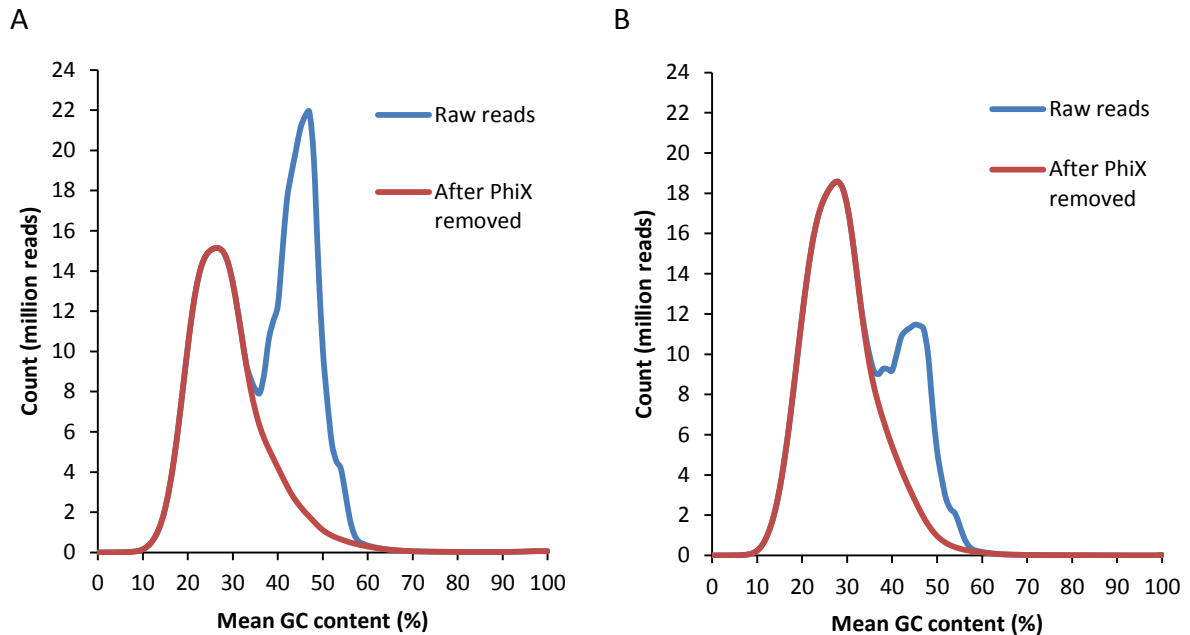


Figure 5.5. GC content of all sequence reads before and after PhiX removal. A, First sequence run. B, Second sequence run.

The methylation levels across each chromosome were grouped in to 1 million base

pair windows and plotted for average methylation levels for each cytosine context as shown in Figure 5.7. Methylation levels in CG contexts were highest globally, peaking broadly around the centre of chromosomes and sharply decreasing at the ends, with an average level of 91.0%. A similar pattern was seen in CHG contexts with a lower average methylation level of 59.6%. The CHH context methylation levels were the lowest, averaging 4.1%.

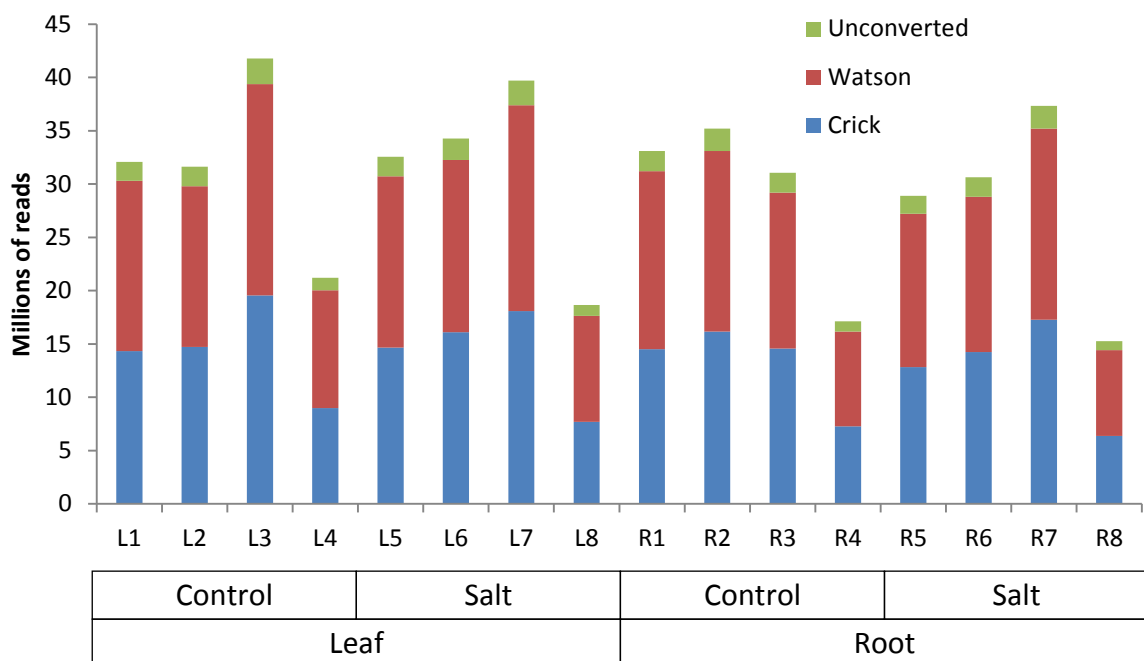


Figure 5.6. Read counts per sample for barcode recognised epiGBS reads of control and salt treated barley plants. Watson and Crick reads are defined based on the strand that was bisulphite treated, with a single unmethylated cytosine from each adapter that signals which it is. If both unmethylated cytosines from adapters remain unconverted then this indicates the non-conversion rate of bisulphite treatment.

Principle Component Analysis (PCA) was performed on the methylation percentage in 100 bp windows grouped by cytosine context as shown in Figure 5.8. All cytosine contexts show separation between leaf and root tissue.

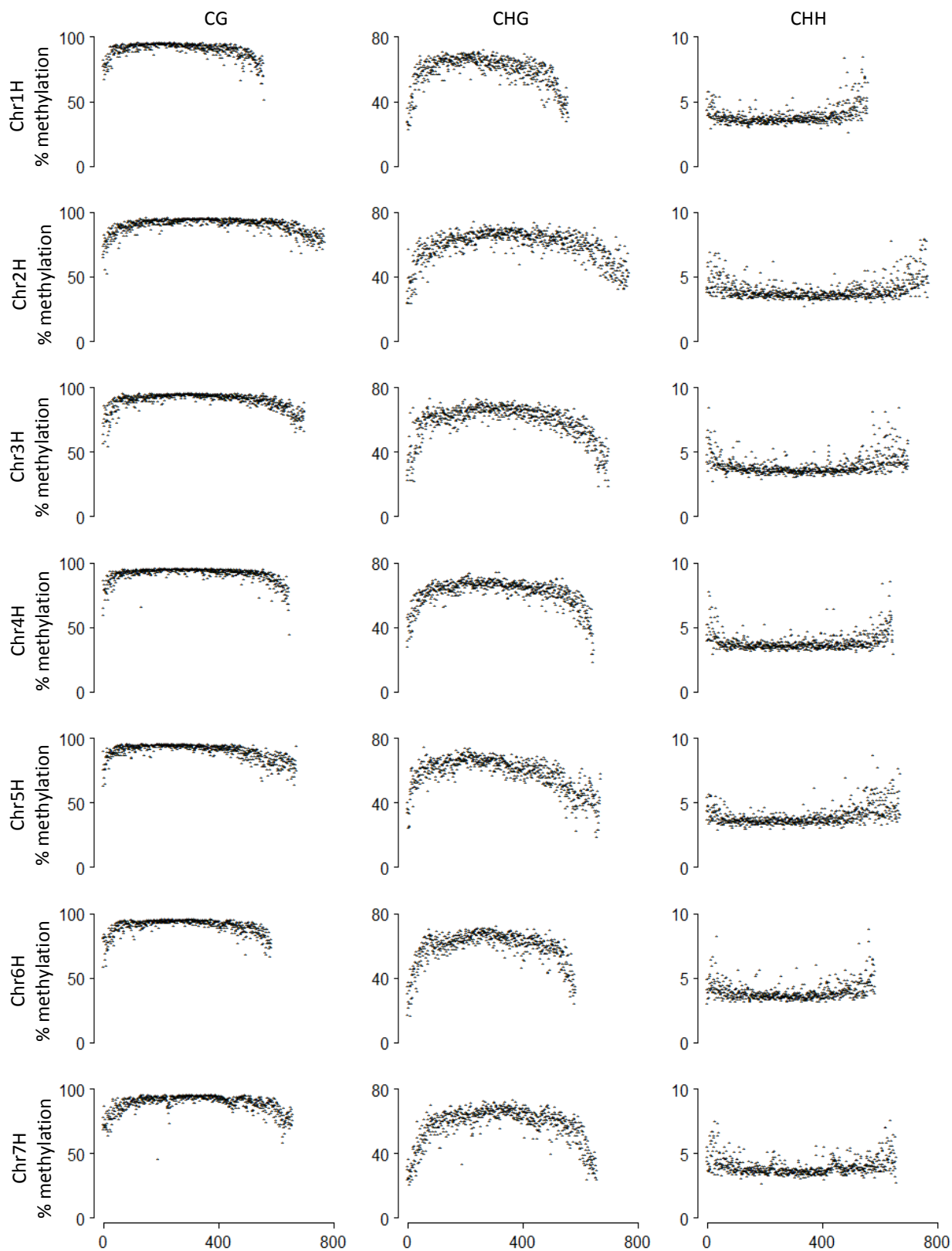


Figure 5.7. Percentage of cytosine methylation in barley grouped by CG, CHG or CHH context for each chromosome using 1,000,000 bp windows. The x-axis is the position on the chromosome in Mbp, with the y-axis being the cytosine methylation percentage. The y-axis is scaled differently for CG, CHG and CHH contexts to show detail.

Methylation patterns around Protein Coding Genes (PCGs) were determined and are shown in Figure 5.9. The methylation level in a CG context cytosine is high in the upstream and downstream flanking sequences of the PCG as well as in the middle of the gene body, with the lowest value near the Transcription Start Site (TSS) and a low level also observed at the Transcription End Site (TES). Under the CHH context, a different pattern is observed with methylation ramping up to a peak about 200 bp before the TSS and dropping to a low level in the PCG.

The methylation patterns around TEs are shown in Figure 5.10. A similar pattern is seen in CG and CHG contexts with a sharp increase in methylation at the boundary of the TE albeit at different average methylation levels. The level of methylation just outside of the retrotransposon TE is higher and more consistent than the DNA transposon, which decreases to much lower levels until the TE boundary. The retrotransposons dominate reads that were mapped to TEs and so the combination of all TEs more closely follows the pattern of retrotransposons than DNA transposons. In a CHH context, the methylation in retrotransposons takes on a slightly lower level than outside the TE, but methylation around the DNA transposons increases to higher levels within the TE.

The different TE classes were also plotted for methylation patterns and are shown in Appendix A, Figure A.9.3. Two of the patterns that stood out were from methylation in a CHH context and is shown in Figure 5.11. The TE grouping DTX (class: DNA

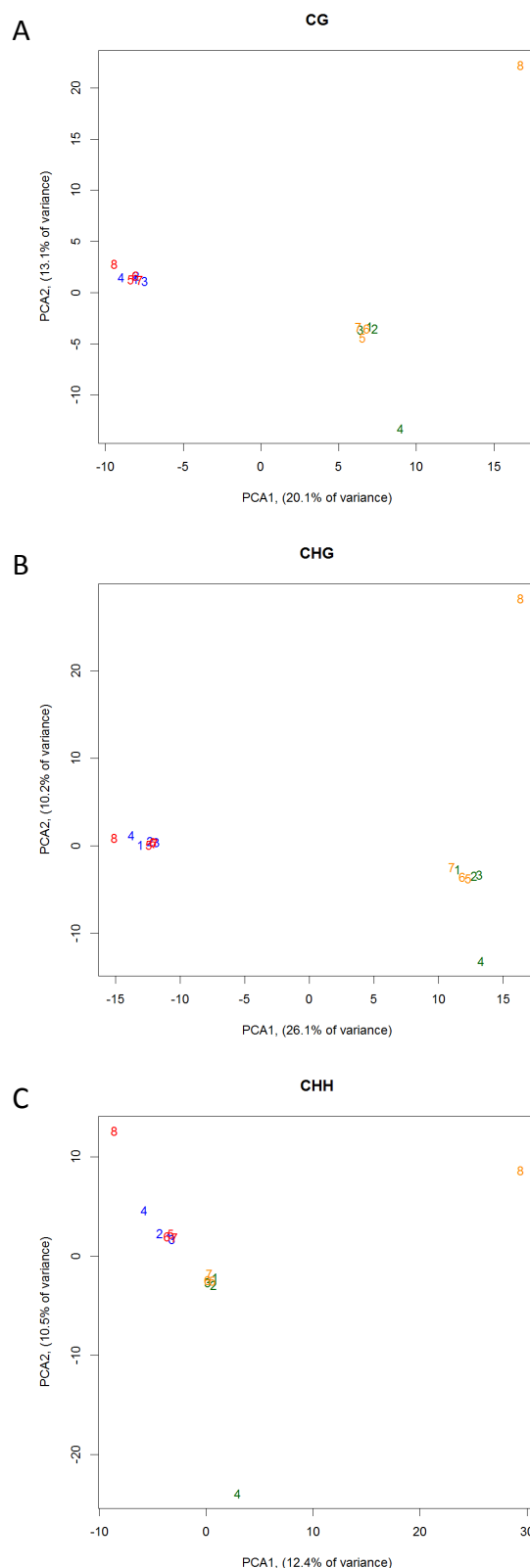


Figure 5.8. Principle Component Analysis of barley DNA methylation percentage in 100 bp windows for all samples categorised by cytosine contexts. A, Cytosine context CG. B, Cytosine context CHG. C, Cytosine context CHH. Samples shown are blue: control leaf tissue, red: salt treated leaf tissue, green: control root tissue, orange: salt treated root tissue. Salinity exposure was 100 mM NaCl in soil.

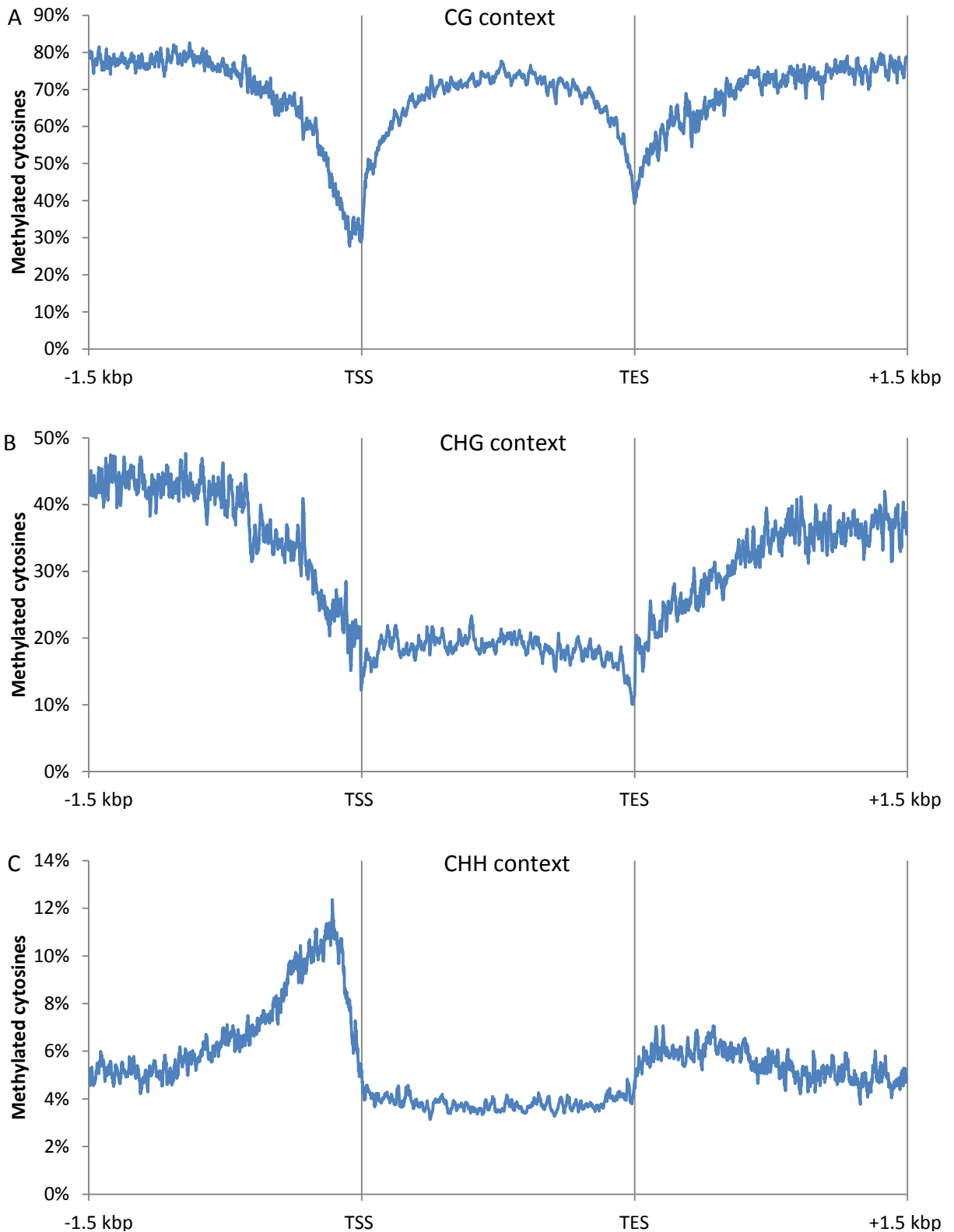


Figure 5.9. Methylation pattern around Protein Coding Genes (PCGs) in barley. The methylation patterns for the three different cytosine contexts A: CG methylation, B: CHG methylation and C: CHH methylation. All cytosine reads for the relevant context that were within 1.5 kbp of a PCG were included. The PCG body from Transcription Start Site (TSS) to Transcription End Site (TES) was scaled for each PCG to fit in the same width.

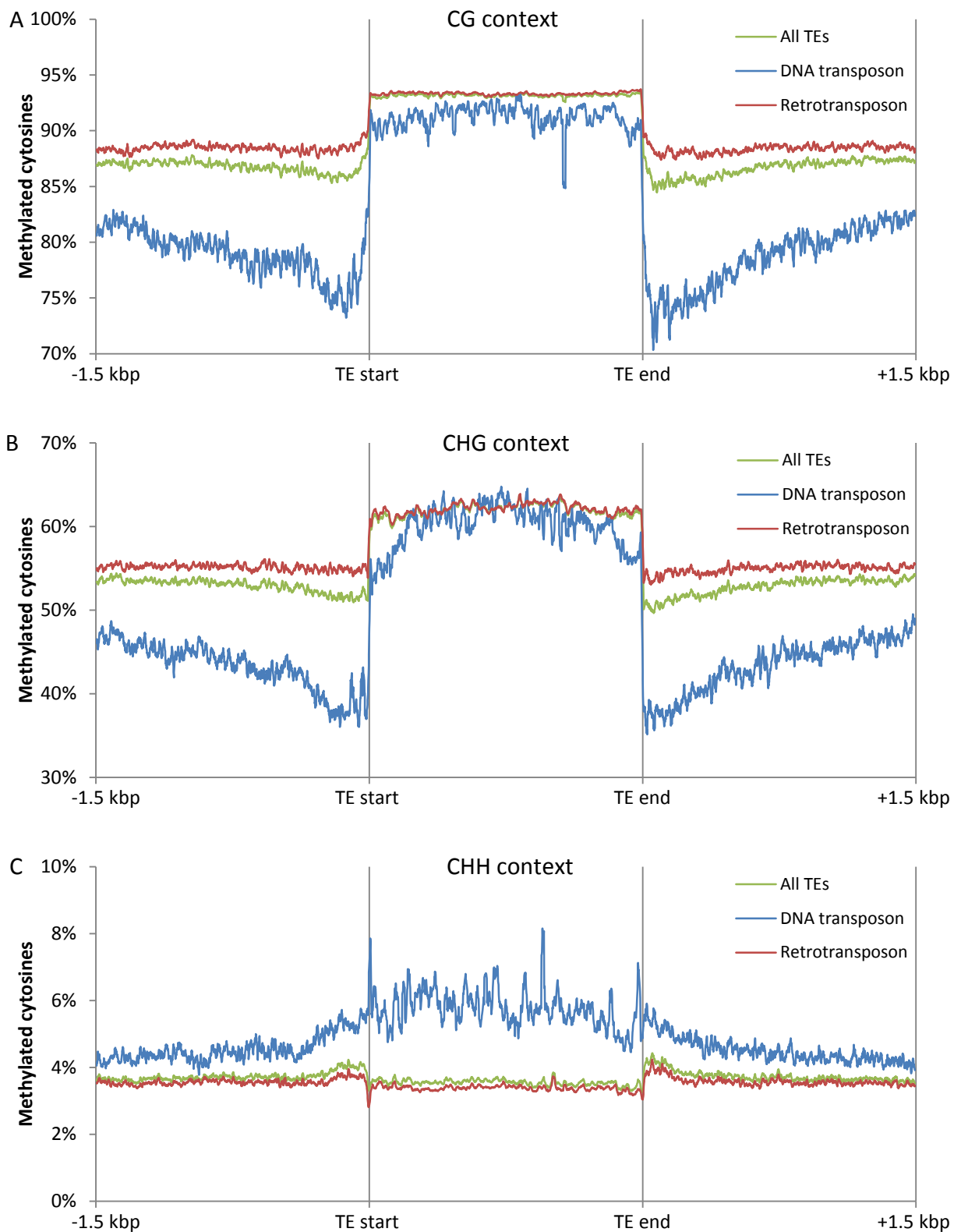


Figure 5.10. Methylation pattern around Transposable Elements (TEs) in barley. The methylation patterns for the three different cytosine contexts A: CG methylation, B: CHG methylation and C: CHH methylation. All cytosine reads for the relevant context that were within 1.5 kbp of a TE were included. The TE body from TE start to TE end was scaled for each TE to fit in the same width.

transposon, order: terminal inverted repeat, superfamily: unknown) and RSX (class: retrotransposon, order: non-long terminal repeat, superfamily: unknown). These two groupings displayed significant spikes in the methylation levels in the body of the TE, even with a reasonable number of average samples per base of 3864 and 964 for DTX and RSX respectively.

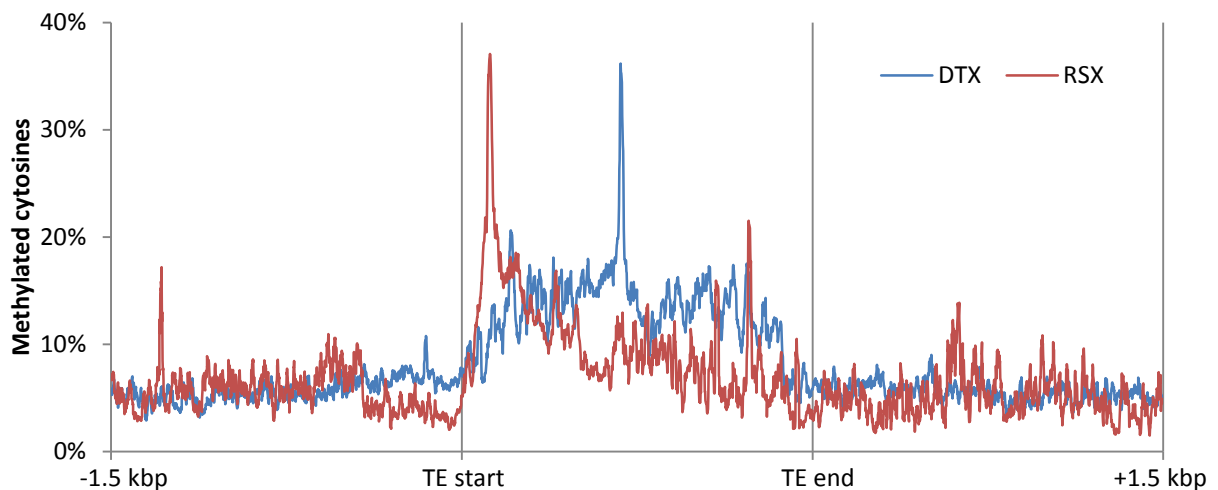


Figure 5.11. Methylation pattern in CHH context cytosines around two selected TE classes in barley. The two TE classes were selected for the highly localised, relatively high concentration of methylation.

Individual methylation sites were grouped by cytosine context and 100 bp windows throughout the genome. Statistically significant differentially methylated 100 bp windows (p -adjusted < 0.05 , $|\text{methylation difference}| > 10\%$) were found and are summarised in Table 5.3. These differentially methylated sites were compared with the annotated genome and transposable elements to find the percentage of genomic features associated with differentially methylated sites as shown in Figure 5.12, with all differentially methylated genes found being listed in Appendix A (Table A.9.10, Table A.9.11, Table A.9.12, Table A.9.13, Table A.9.14 and Table A.9.15). The expected

representation of different genomic features, by random sampling of 100,000 sites that had been included in the mapped reduced-representation bisulphite reads was also determined. Chi squared analysis was performed comparing the observed count of genes associated with differential methylation against the expected count from random sampling of mapped regions. The result is shown in Table 5.4 with PCG being overrepresented in differentially methylated sites in all leaf tissue and only in CHH context in root tissue.

Table 5.3. Differentially methylated 100 bp windowed reads mapping to sites in the proximity of Protein Coding Genes (PCGs).

	Leaf			Root		
	CG	CHG	CHH	CG	CHG	CHH
Number of 100 bp windows differentially methylated	331	639	54,708	3,416	1,823	60,805
Number of 100 bp windows differentially methylated in the region of PCGs	84	69	5,587	247	143	6,295
Number of unique PCGs associated with differential methylation	78	67	4,337	238	134	4,704

Table 5.4. Chi squared analysis of Protein Coding Gene (PCG) count associated with differential methylation against random sampling of the bisulphite-mapped genome to determine overrepresentation.

Tissue	Leaf			Root		
	CG	CHG	CHH	CG	CHG	CHH
PCG overrepresentation p-value	< 0.001	< 0.001	< 0.001	0.26	0.058	< 0.001

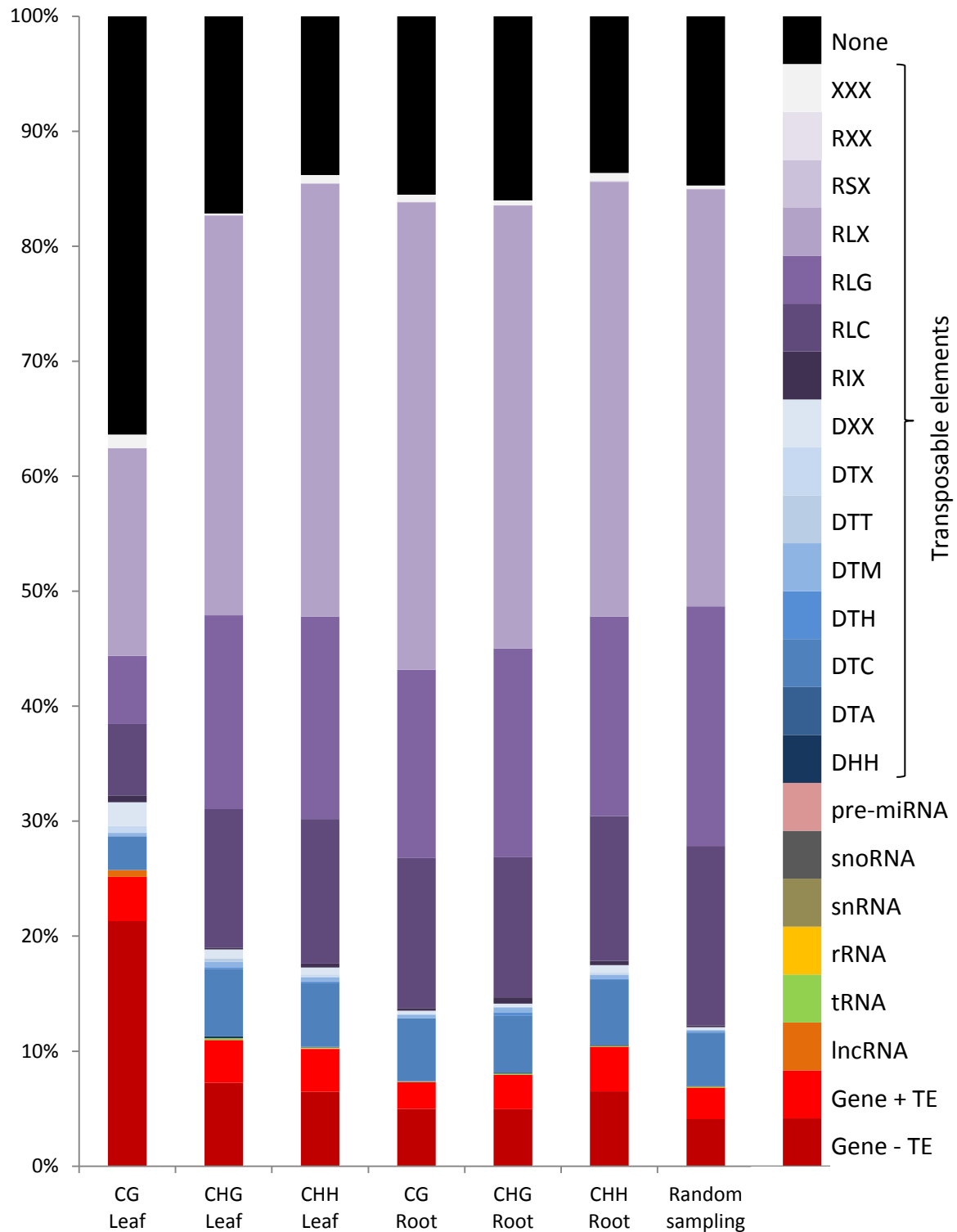


Figure 5.12. Percentage of differentially methylated sites in barley under 100 mM salinity stress associated with genomic functions based on cytosine context and tissue type. Cytosine methylation grouped by CG, CHG and CHH context. Transposable elements were categorised according to the standard established by Wicker et al. (2007). The same annotation was performed with random sampling of mapped locations of bisulphite reads to compare targeted genomic features. Gene + TE implies both a gene and a TE were targeted by this differentially methylated site, Gene - TE implies only a gene was targeted.

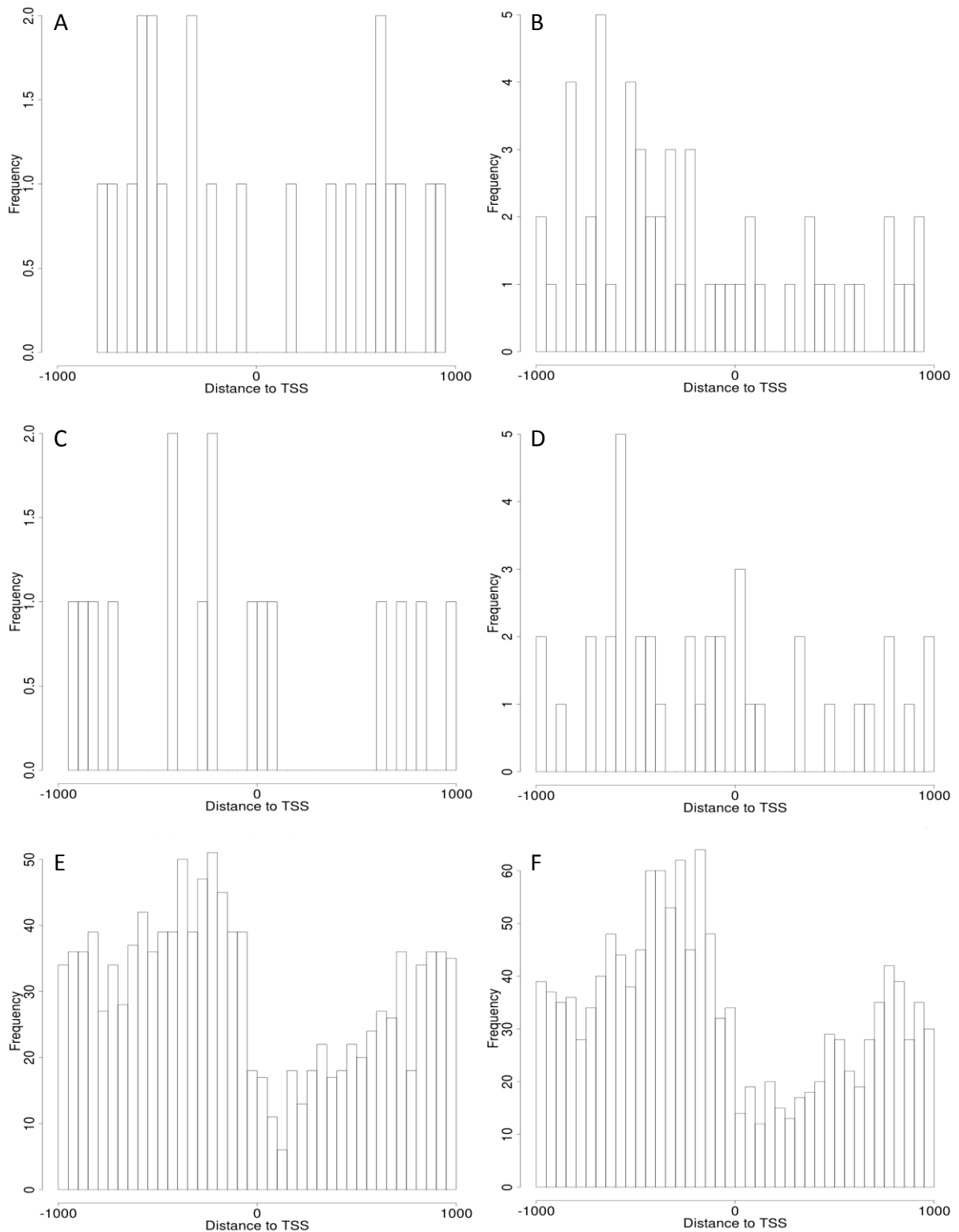


Figure 5.13. Frequencies of significantly differentially methylation locations relative to the Transcription Start Site (TSS) in barley in salt treated plants vs control. The TSS is at 0 on the x-axis with negative numbers being before the TSS and positive within the transcribed area. A, Leaf tissue, CG context. B, Root tissue, CG context. C, Leaf tissue, CHG context. D, Root tissue, CHG context. E, Leaf tissue, CHH context. F, Root tissue, CHH context.

5.4 Discussion

5.4.1 DNA extraction

Multiple methods for DNA extraction were explored (data not shown) before settling on the Qiagen DNeasy kit with some customised improvements for the specific plant tissue, described in section 5.2.1. The other methods that were explored included a CTAB protocol (Porebski et al. 1997), a phenol/chloroform/isoamyl alcohol protocol (Weining & Langridge 1991), and a Trizol method that was able to extract and separate DNA and RNA (as per the manufacturer's instructions). The Trizol protocol resulted in denatured DNA which was not useful for downstream restriction digestion, and the other protocols did not produce DNA yields as high as the method described in section 5.2.1.

5.4.2 EpiGBS

The most significant and important deviation from the published epiGBS method was the concentration of adapters for ligation. Experimentation demonstrated that the amount of adapters for the ligation stage needed to be 12 ng of each Ba and Co adapter, while the method from van Gurp et al. (2016) required 1.2 ng. It is possible that the difference in the DNA extraction technique may have had some impact on the ligation stage. The van Gurp et al. (2016) paper used the Macherey-Nagel Nucleospin Plant II kit while the Qiagen DNeasy kit was used by us in this experiment. The epiGBS protocol had been further developed by van Gurp et al. (2016), after release of the paper, to use restriction enzymes *Nsi*I and *Csp*6I with the

appropriately adjusted adapters. Apart from the extraction process, other reagents and kits were from the same manufacturers as van Gurp et al. (2016). Many individual changes were systematically made to resolve the adapter concentration issue, and along the way minor improvements were also made to the clean-up and size selection processes.

An in-silico digest was performed to estimate the coverage with different size selection ranges. A size selection of epiGBS fragments from 380 bp to 560 bp which included the adapter sequence was chosen based on this in-silico digest. Most of the reads from the first sequence run were PCR duplicates with an average of each unique sequence read being represented about three times, as determined by the combination of barcodes, sequence and the six UMI bases. This suggests that the in-silico digestion modelling was not particularly useful in determining the optimal epiGBS fragment size range. In the second sequence run with a wider size selection from approximately 250 bp to 1000 bp, the number of PCR duplicate reads decreased such that each unique sequence was represented 1.9 times on average.

The epiGBS library was sent to NovoGene Co. Ltd. for sequencing. For the first sequence run the requested PhiX content was 30% based on the advice from NovoGene based upon previous bisulphite sequencing they had performed. The actual PhiX content in reads was 42%, however the large number of PCR duplicate reads meant that a lower PhiX spike-in would not have resulted in many more useful

reads as PCR duplicates are only useful once in the bioinformatics pipeline. For the second sequence run a PhiX content of 20% was requested and the result was 22% of reads mapping to PhiX. The lower PhiX concentration was used after private correspondence with Neils Wagemaker stating that they had acceptable read quality when PhiX was as low as 8% (Wagemaker 2018). PhiX is added to prevent an excessive imbalance in C and G sensing chemistry relative to A & T in the sequencing machine, which would cause read quality degradation at the end of reads (Kircher et al. 2009). A novel alternative developed by (Suzuki et al. 2018) uses a known library with a high GC content to balance out the reads, with fewer reads being assigned to the spike-in library, however NovoGene did not have access to such a library.

There was a noticeable decrease in reads that used the Ba4 ligated adapter barcode (for barcode designations see Table A.9.9). The cause for this is unknown, the concentration of the adapter may have been incorrect or there may have been an issue with this particular adapter that caused it to function less efficiently.

The bisulphite conversion rate was 94.2% which is noticeably less than the 99.5% claimed conversion rate of the EZ DNA bisulphite conversion kit (Zymo 2020) and other common expected conversion rates of 98.7% to 99.9% (Holmes et al. 2014). The reason for this low conversion rate is not known. This would have an impact on the accuracy of the methylation information. It causes a small over-reporting in global methylation, under-reporting in global non-methylation, and increasing the noise in

individual un-methylated reads. The error rate of methylated cytosines being incorrectly converted is not determined by the epiGBS protocol, but this error will also be a source of noise as it has been found to be 2.7% in the EZ DNA bisulphite conversion kit (Holmes et al. 2014).

5.4.3 Bioinformatics

The bioinformatics pipeline for the epiGBS method provided in the github repository (<https://github.com/thomasvangurp/epiGBS>) contains multiple errors. Some of the errors could be corrected with modifications to the poorly documented code. The most significant error in the pipeline was that PCR duplicates were found but not flagged or removed before further processing. Consequently, the library generated using this pipeline contained many PCR duplicates. They needed to be removed before running the epiGBS bioinformatics pipeline, and a custom written cython & python code performed this function.

Errors were found in the epiGBS bioinformatics program demultiplex.py that caused almost all Crick reads to be discarded unless read errors were permitted in the barcode and restriction enzyme portion of the read. The errors in the epiGBS pipeline, combined with a lack of assistance from the author for overcoming many issues raised on the github repository from multiple sources, created distrust for the validity of information from that epiGBS bioinformatics pipeline. The current

literature suggests that this is not a commonly used bioinformatics pipeline, despite being cited in review articles.

A well established and frequently used program for mapping bisulphite sequences and performing methylation calls is Bismark (Krueger & Andrews 2011). The Bismark program is not able to accept reads produced by the epiGBS protocol. To use in this study, modifications to the reads were required before using Bismark. The sequence fastq files needed to be modified and processed in two batches with aligned (overlap between the two read strands) and unaligned reads (no overlap) separately. The Bismark bioinformatics protocol is well cited and utilised. This provided the necessary confidence to continue with the analysis of the sequence data obtained in this research.

The epiGBS pipeline uses the splice aware aligner STAR (Dobin et al. 2013). It is possible to choose variables such that read splicing is decreased, but splicing is not completely removed, and some reads will be mapped as though there is a splice site. Splicing is only relevant to mRNA and not DNA methylation information. This only affected 54,002 reads out of the 126,252,514 reads that were mapped (0.04%). This error occurs due to the use of an inappropriate aligner, but the error is extremely small. Bismark uses the Bowtie2 sequence aligner which does not generate this error. While the Bowtie2 sequence aligner is slower than the STAR splice-aware sequence aligner (when comparing with the same number of CPU (Central Processing Unit)

cores), the entire Bismark pipeline is considerably faster than the epiGBS pipeline. Bismark is able to make use of multiple CPUs through the entire process. Much of the epiGBS pipeline runs in python or pypy with only a single CPU in use and as a result is comparatively slow. In the end, the entire bioinformatics pipeline was re-written such that no part of the epiGBS pipeline was used in the generation of the final results.

Bismark was not able to import reads generated using the epiGBS protocol, and so custom code was written to modify the data set so it could be used in the Bismark bioinformatics pipeline. Bismark has a PCR duplicate removal function, however this is not compatible with the epiGBS method which uses UMI bases for identification of duplicates. The PCR duplicates were removed by the custom written program `dedup.py` (section 10.17).

The Bismark pipeline is not capable of appropriately recognising short-insert reads from the epiGBS protocol (Figure 5.14) so PEAR (Paired-End reAd mergeR) (Zhang et al. 2014) was used to convert overlapping paired reads into merged reads and leave reads that do not have overlap unaltered. Bismark could then be used to perform two rounds of bisulphite sequence information extraction, with merged reads and un-merged reads, and with the methylation calling information being combined later.

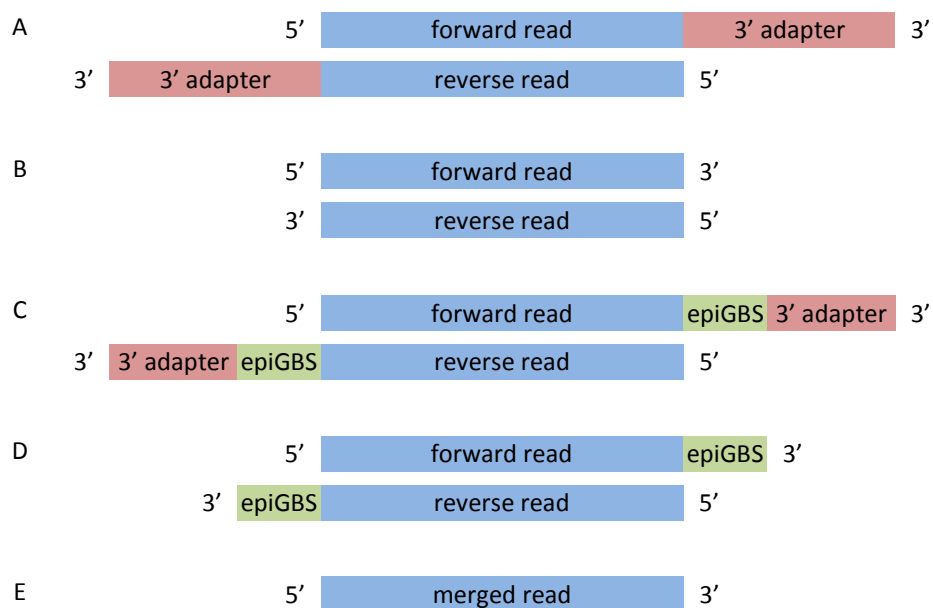


Figure 5.14. Bismark processing of short insert reads by removal of adapters. A. An example read expected by Bismark which can be dealt with by strict removal of adapters. B. The read after the strict adapter removal. C. An example read from the epiGBS protocol where the green epiGBS section represents the single unmethylated cytosine, 3 base UMI, and barcode between the sequence read and the adapter. D. The read from C after adapter removal by Bismark showing why pre-processing before the use of Bismark is required. E. Read merging with the use of PEAR (Paired End reAd mergeR), which can then be used by Bismark as a single ended read.

Cytosine methylation information is categorised by the context CG, CHG or CHH which exhibit different average levels throughout the genome (Law & Jacobsen 2010). The pattern of cytosine methylation in Figure 5.7 shows the relatively high levels of CG methylation, moderate levels of CHG methylation and low levels of CHH methylation. The pattern with higher methylation levels near centromeric regions of chromosomes and lower levels toward the chromosome ends, is also clearly visible in CG and CHG contexts. Cytosines in CHH contexts had very low levels of methylation. These patterns have been observed previously in various plant species. *Arabidopsis thaliana* has broad similarities to this data set with CG contexts showing the highest average methylation, followed by CHG then CHH showing the

lowest levels (Cokus et al. 2008). While *A. thaliana* also shows higher methylation levels around the centromere and lower at the chromosome ends, the decrease in methylation around the centromere was more pronounced than was observed in barley. The high average methylation levels in barley in CG and CHG contexts compared with *A. thaliana* are most likely related to higher quantities of TEs in barley, which has previously been shown to correlate with higher methylation levels (Hirsch & Springer 2017).

The PCA plots for 100 bp windowed methylation levels in different contexts shown in Figure 5.8 shows clear separation between leaf and root tissue in CG and CHG contexts, similar to reported methylation differentiation patterns in barley leaf and root tissue (Konate et al. 2020). The CHH context PCA separation is not as distinct, which could be related to the lower levels of methylation associated with this context. A greater number of reads per cytosine may be required to separate the tissue groups more clearly when comparing small differences in methylation percentage. The fourth sample from each treatment or tissue group had noticeably lower read numbers (Figure 5.6) and similarly may have contributed to the divergence in PCA from others in their group. This does not appear to have a noticeable impact on downstream processing. All bioinformatics tools used were also run with only three replicates per sample, excluding the fourth sample in each group which had lower read quantities, with similar outcomes to using all four samples per group. Overall the PCAs are similar to what would be expected based on

methylation differences in different tissue groups.

The methylation pattern around genes as shown in Figure 5.9 can be compared with prior research in different plants such as *Arabidopsis thaliana* (Figure A.9.5), *Oryza sativa* (rice) (Figure A.9.6, Figure A.9.7) and *Manihot esculenta* (cassava) (Figure A.9.8) (Cokus et al. 2008; Li et al. 2012; Garg et al. 2015; Wang et al. 2015). Of the mentioned species for comparison, the closest relative to barley is *O. sativa*. The pattern of methylation around the gene in a CG context has most similarities with rice with the lowest methylation level at the TSS and second lowest point at the TES, and relatively high levels in other regions. The average levels of methylation in CG context around the gene in barley were higher than rice, presumably due to the higher percentage of transposable elements in the barley genome. A generally similar pattern is observed in *A. thaliana* and cassava.

The CHG context methylation pattern around genes in barley is also quite similar to rice, with a slightly lower level at the TES than the TSS in barley, but a lower level at the TSS than the TES in rice. Again average methylation levels are higher in barley. There are some similarities here too with *A. thaliana* and cassava. Methylation in the CHH context from this research produced a pattern where the methylation rose from upstream toward the TSS before dropping to a low level in the gene body. This same pattern is observed in rice, but is quite different to patterns observed in *A. thaliana* and cassava which look more similar to a low level CHG methylation than CHH

methylation found in barley. It is possible that there was an evolutionary split in the function of CHH methylation around genes. Barley and rice are both monocot grasses while *A. thaliana* and cassava are dicots. Much more information would be needed to establish if this methylation pattern is related to the monocot/dicot evolutionary divergence, or perhaps it is something unusual to grasses, or any other evolutionary split. As the methylomes and annotations of more genomes become available, this question should be able to be answered.

The methylation pattern around transposable elements in barley is shown in Figure 5.10. The patterns of methylation from CG and CHG context cytosines are similar to patterns in rice, *A. thaliana* and cassava, but with different average methylation levels. The notable difference is observed in CHH context methylation around transposable elements. In barley the CHH context methylation increases in DNA transposons, but in retrotransposons there is a slight rise either side of the TE and slightly lower levels in the TE than outside of the TE. The retrotransposons dominate the TEs in barley; so when looking at all TEs, the pattern is similar to the retrotransposon pattern. Rice, *A. thaliana* and cassava all adopt the pattern similar to DNA transposons in barley in CHH context, but different to retrotransposons in barley in the CHH context.

The confirmation of the methylation patterns with existing research indicates the success of the epiGBS method along with the custom made Bismark-based

bioinformatics pipeline. If the bioinformatics pipeline had not operated correctly, then the emergence of these patterns would be extremely unlikely.

The observed level of occurrence of differentially methylated regions corresponding to PCGs (Figure 5.12 and Table 5.4) indicates that all cytosine contexts are targeted for epigenetic modification in leaves under salt stress, while only in the CHH context in root tissue. Leaf tissue CG context differentially methylated regions were targeting PCGs at 3.7 times the rate that would be expected of random sampling of bisulphite mapped reads. However, root tissue CG context was not significantly different from the random sampling when targeting PCGs. It is not clear why there would be such a difference between root and leaf tissue. Possible explanations are that the reduced representation nature of epiGBS protocol means that not all PCGs are represented, and possibly by chance the PCGs important for leaf tissue methylation changes happened to be over-represented in restriction digest and size selection. It is possible that many PCGs need to be controlled via methylation changes in leaf tissue and only relatively few in root tissue. There were only 331 leaf tissue CG context differentially methylated sites observed compared to much higher numbers for other cytosine contexts (Table 5.3) which may be a factor in explaining this observation, where a smaller sample size could create a distortion. Prior research has shown that environmental stress results in more methylation changes in CG contexts than non-CG contexts (Dubin et al. 2015).

The position of differential methylation relative to the Transcription Start Site (TSS) shown in Figure 5.13 shows a very weak relationship between the TSS and location of differentially methylated sites in CG and CHG contexts. There is a much clearer concentration of differentially methylated sites about 200 bp upstream of the TSS in CHH methylation and fewer changes in methylation near the start and downstream of the transcribed region. The pattern may be more clearly visible in CHH than other contexts because of the much larger number of differentially methylated sites that could contribute to the frequency analysis plot. Prior research by Wicker et al. (2017) found that CHH methylation experienced the greatest methylation levels compared with elsewhere in the genome within 1.5 kbp upstream of the TSS, and CG and CHG experienced smaller elevated levels. These methylation levels were associated with TEs in promoter regions.

Some of the differentially methylated sites identified were associated with PCGs without TEs, and some associated both with PCGs and TEs. The selection criterion for being classified as associated with a PCG was the differential methylation needed to be overlapping or be within 1.5 kbp of the PCG, but TEs needed to be directly overlapping the differentially methylated window.

It should not be surprising that most of the differential methylation in salt treated plants is related to TEs. Mobile elements and repeat structures make up 84% of the barley genome (IBGSC 2012). Prior research has indicated the importance of DNA

methylation and transposable elements in stress response (McCue et al. 2012; Mao et al. 2015; Zhang et al. 2018), with TE families targeting specific functions rather than previous concepts of TEs ending up in locations where the minimum disruptions to genomic structure occur (Wicker et al. 2017).

Possibly the most useful future direction for the research presented in this chapter would be to further develop the bioinformatics pipeline so that other researchers can make use of the epiGBS protocol without the need to deal with the pitfalls in the existing epiGBS bioinformatics pipeline. This would require modifications to the code to allow different samples, barcodes and restriction enzymes. The many individual custom made programs would need to be simplified so that the users only need to interact with one or two scripts which then pass commands to the appropriate programs. Good documentation and code comments are essential for being able to be utilised by other researchers. The current documentation for the epiGBS bioinformatics pipeline is almost non-existent, and combined with the lack of support via the github repository when issues are discovered, does not make this method useful. There are benefits to this protocol over other methods, with the UMI bases able to distinguish PCR duplicates and the unmethylated cytosine in the adapter providing useful information on the conversion rate. It is unfortunate that this useful protocol was let down by the epiGBS bioinformatics pipeline. A new pipeline has been created, and will be of use to future research that makes use of this

particular reduced representation bisulphite sequencing method for DNA methylation studies.

6 Associations between small RNA, DNA methylation, and mRNA changes in barley under salinity stress

6.1 Introduction

The regulation of gene expression is complex, with many counteracting signals modulating messenger RNA (mRNA) production, resulting in a unique population of transcripts in any given cell type. DNA methylation has been recognised as a method for suppression of mRNA production in eukaryotes, but the link between the two is not trivial (Bewick & Schmitz 2017). While there are instances where an increase in DNA methylation of a promoter region results in decreased production of the associated mRNA (Berdasco et al. 2008), there are also instances where increased methylation of other sites in the gene body results in similar decreased production of mRNA (Anastasiadi et al. 2018) but also instances of increased production of mRNA (Shibuya et al. 2009).

The RNA directed DNA Methylation (RdDM) pathway uses complementarity between a 24 nt small interfering RNA (siRNA) and a single stranded RNA transcript to cause DNA methylation of a cytosine in the region (Matzke & Mosher 2014). This pathway controls de-novo methylation of all sequence contexts (CG, CHG, CHH as defined in section 5.1) (Williams & Gehring 2017). Maintenance of the symmetric CG context methylation after DNA replication is performed by DNA METHYLTRANSFERASE 1 (MET1) (Bewick & Schmitz 2017). Maintenance of CHG

and CHH sequence context methylation is performed by CHROMOMETHYLASE 3 (CMT3) and CHROMOMETHYLASE 2 (CMT2) respectively, in conjunction with DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), where the feedback loop is based on the state of histone methylation at H3K9 (Zemach et al. 2013; Stroud et al. 2014). These processes were described in greater detail in section 1.2.4.

While 24 nt siRNA targets DNA methylation, there is also a microRNA (miRNA) class which can range in size from 23 to 27 nt which also targets DNA methylation. These small RNAs (sRNA) were discovered relatively recently and are classified as miRNA due to the production involving a hairpin-loop structure forming from a single stranded RNA, which is then processed to form a miRNA (Teotia et al. 2017). These miRNA are significantly less abundant in quantity than 24 nt siRNA (Jia et al. 2011), and in some prior research the 23 and 24 nt sRNAs had been grouped together (Qi et al. 2006; Kurihara et al. 2008; Groszmann et al. 2011; Blevins et al. 2015) as they both caused DNA methylation via the RdDM pathway (Jia et al. 2011). In this chapter, the 23 nt sRNA is referred to as miRNA, 24 nt sRNA is referred to as siRNA since siRNA makes up a significant majority of 24 nt sRNA reads, and the grouping of the two is referred to as 23/24 nt sRNA. Changes in the expression of 23/24 nt sRNAs has previously been linked to DNA methylation, which targets genes and Transposable Elements (TEs) and modulates expression of these targets (Fultz et al. 2015).

Transposable elements are known to play an important role in stress response (Negi et al. 2016) and are capable of moving or copying their sequence and inserting in adjacent or distal regions of the genome (Slotkin et al. 2012). There are two main classes of TEs, the DNA transposon which either moves or replicates as a DNA intermediate, and the retrotransposon which replicates via an RNA intermediate (Wicker et al. 2007). Transposable Elements make up a sizable portion of the barley genome, with 84% of the genome sequence being TEs or non-functional sequences related to TEs (IBGSC 2012). Transposable Elements played an important role in this research with changes in methylation targeting TEs being one of the main foci of discussion.

6.1.1 Hypothesis

The hypothesis under test is that if gene expression in barley under salinity stress is at least partially controlled by epigenetic mechanisms, then exposing the plants to salinity stress will result in changes in sRNA and DNA methylation associated with gene expression. With the expectation being that the change in 23/24 nt sRNA expression would correlate with changes in DNA methylation and that in turn would inversely correlate with mRNA expression.

6.2 Methods

The short RNA sequences that were found and described in section 4.3 were separated by size with the custom python script `SmallRNA_size_grouping.py`

(section 10.6). Bowtie2 (Langmead & Salzberg 2012) was used to map 23 nt miRNA and 24 nt siRNA reads against the ensemble reference genome for *Hordeum vulgare*, IBSC v2.45 (Mascher et al. 2017). Information from multimapped reads was preserved. Mapped reads were assigned to 100 bp windows throughout the genome with the custom python script `sRNA_windowing.py` (section 10.7). If a read was mapped to multiple locations then the single read was added to the relevant windows as a fraction of the number of locations it mapped to. For example, if a read was mapped to five locations then 0.2 would be added to each 100 bp window that the read mapped to. This was done separately for 23 nt miRNA, 24 nt siRNA, and also by combining 23 and 24 nt sRNA reads (referred to as 23/24 nt sRNA). A custom R script `windowed_sRNA_ttests.R` (section 10.8), performed t-tests for each windowed location that had at least three biological samples with non-zero reads comparing control to salt treated plants. Adjusted p-values were determined using the custom R script `ttests_to_adjusted_p.R` (section 10.9), which utilises the R library `qvalue` (Dabney et al. 2010). Statistically significant differential expression in 23 nt miRNA, 24 nt siRNA and differential methylation was binned in 1 Mbp windows to generate a map comparing changes in expression of the two data sets across the genome with the R script `chromosome_mapping_siRNA_meth.R` (section 10.33). The custom python script `sRNA_methylation_co-location.py` (section 10.34), found co-locations of statistically significant differences in sRNA mapped to 100 bp windows of the genome with the differential methylation 100 bp window mapped data set

which had been created in section 5.3. Co-locations of differentially methylated sites and differential sRNA expression were determined by the custom R script `plot_barley_sRNA_meth_co-locations.R` (section 10.35). The genomic functions of the differentially methylated sites had already been found in section 5.3 and the sRNA sites co-located with methylation were recorded. Annotation was performed by the custom R script `annotate_co-located_sRNA_methylation.R` (section 10.36). The plots comparing significantly different methylation levels with significantly different mRNA expression levels were created with the custom R script `methylation_co-location_mRNA.R` (section 10.37). This script also categorised the methylation targets of protein coding genes (e.g. upstream, first intron, etc.), along with information about TEs that may have also been targeted by the change in methylation.

6.3 Results

6.3.1 DNA methylation and 23/24 nt sRNA

Both methylation and 23/24 nt sRNA were mapped to 100 bp windows across the barley genome. Statistically significant differences between control and salt treated plants were classified as having an adjusted p-value of less than 0.05. A graphical representation of the quantity of significant differentially expressed 23/24 nt sRNA and differential methylation in CHH, CG and CHG contexts across the whole genome is shown in Figure 6.1, Figure A.9.9 and Figure A.9.10 respectively.

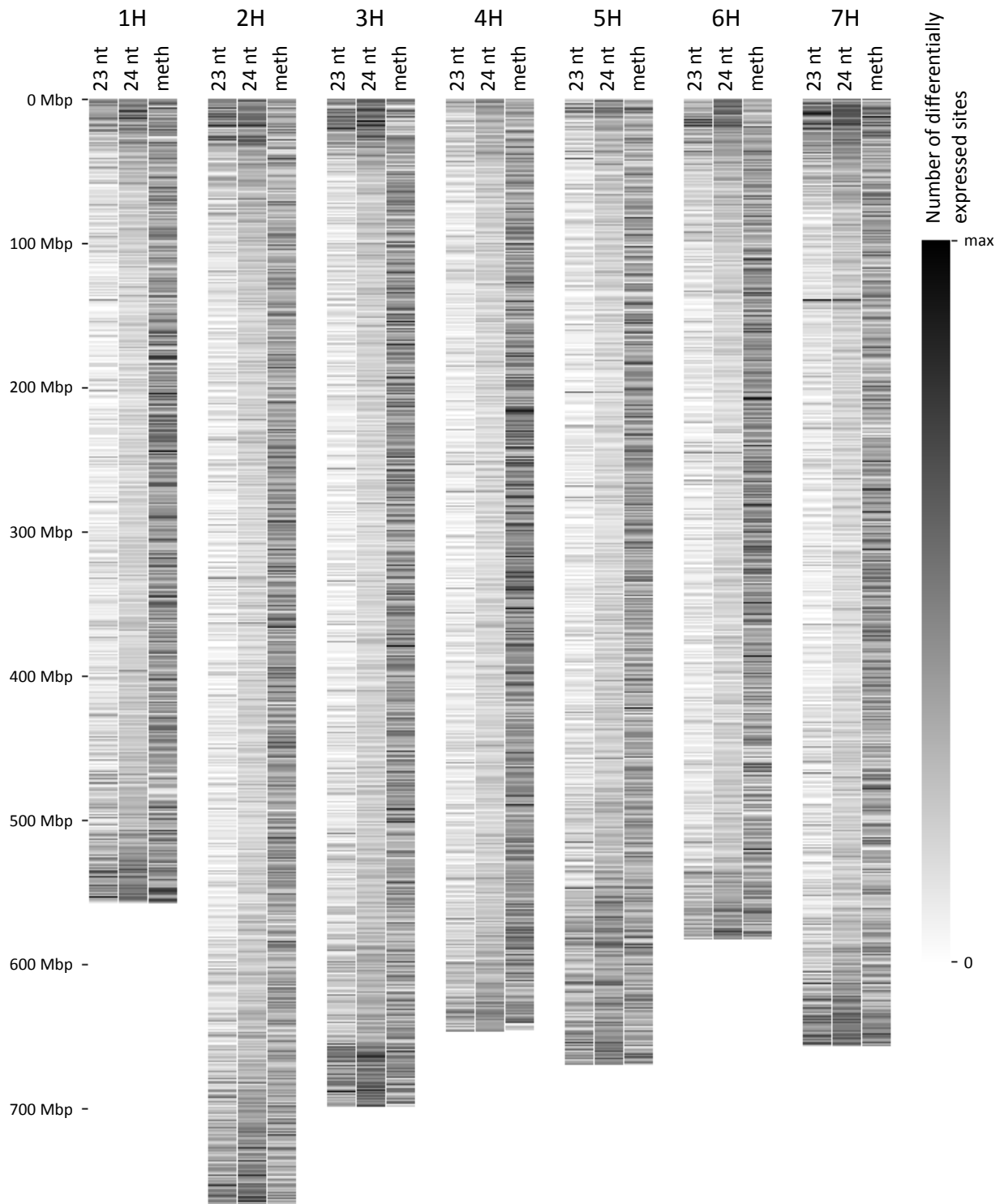
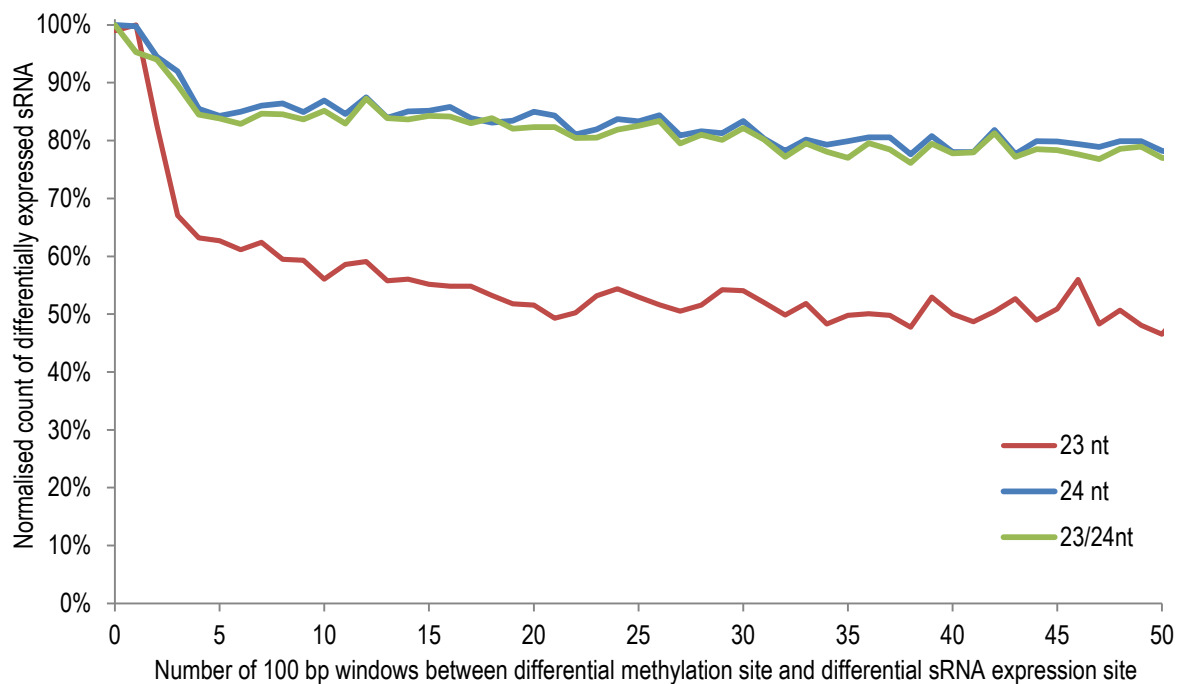


Figure 6.1. Quantity of targets of differentially expressed 23 nt miRNA, 24 nt siRNA, and differential methylation CHH sites mapped to 1 Mbp windows in barley under salt stress. Differential expression of 23 nt miRNA, 24 nt siRNA, and CHH context differential methylation was determined by mapping reads to 100 bp windows in the barley reference genome and finding differences between salt treated and control with adjusted p-values of less than 0.05. The number of differentially expressed 23/24 nt siRNA and CHH context differentially methylated sites were counted for each 1 Mbp window in the genome. Maximum count values for 23 nt miRNA was 24, for 24 nt siRNA was 90 and DNA methylation was 30 differentially methylated 100 bp windows per 1 Mbp genomic window.

The differential methylation and sRNA data sets were compared for co-locations. Only the methylation data set for the CHH context contained enough data points useful for analysis. Figure 6.2 shows the distance between a differentially methylated 100 bp window and a differentially expressed set of 23/24 nt sRNAs, and how the differential expression profile changes with proximity. The number of locations where the differential methylation was within two 100 bp windows of the differentially expressed 23/24 nt sRNA is noticeably greater than larger distances in both leaf and root tissue. The close proximity of differentially methylated sites and differential expression of 23/24 sRNA throughout the genome indicates that 23/24 nt sRNA and DNA methylation data sets are related.

The genomic functions associated with sites where the differential methylation co-located with differential 23/24 nt sRNA expression within two 100 bp windows is shown in Figure 6.3. A list of all PCGs at sites associated with the co-location of 23/24 nt sRNA and DNA methylation is shown in Table A.9.16 and Table A.9.17, for leaf and root tissue respectively. There is a noticeable increase of 23/24 nt sRNAs co-located with methylation in PCGs when compared with random sampling of epiGBS read mapping locations as shown in Table 6.1. Chi squared analysis showed that genes were significantly over-represented relative to random sampling of mapped methylation reads with each combination in Table 6.1 resulting in p-values of less than 0.001.

A. Leaf tissue



B. Root tissue

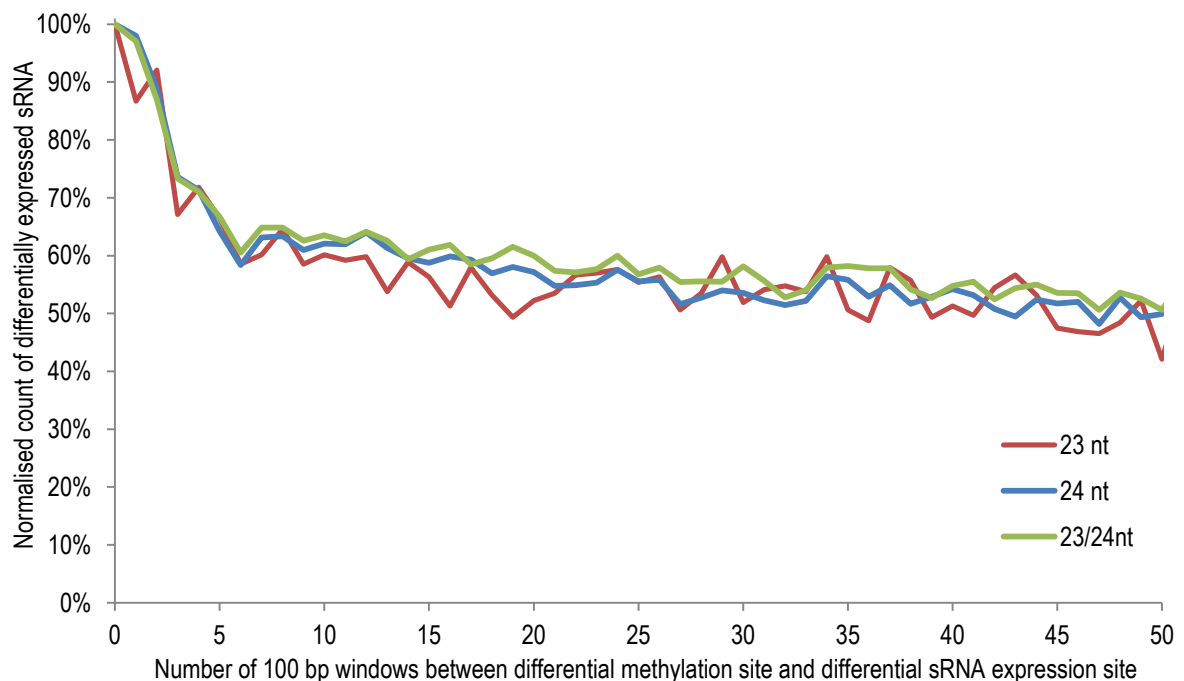


Figure 6.2. Proximity of differentially methylated sites and targets of differentially expressed sRNA in barley under salinity stress. A, Leaf tissue; B, Root tissue. Count of differentially expressed siRNA in 100 bp genomic windows is relative to 100 bp windowed differentially methylated sites. Counts are normalised relative to the maximum count of differentially expressed siRNA per window distance. Small interfering RNA is broken in to the categories of 23 nt miRNA, 24 nt siRNA, and 23/24 nt sRNA which includes both 23 and 24 nt sRNAs.

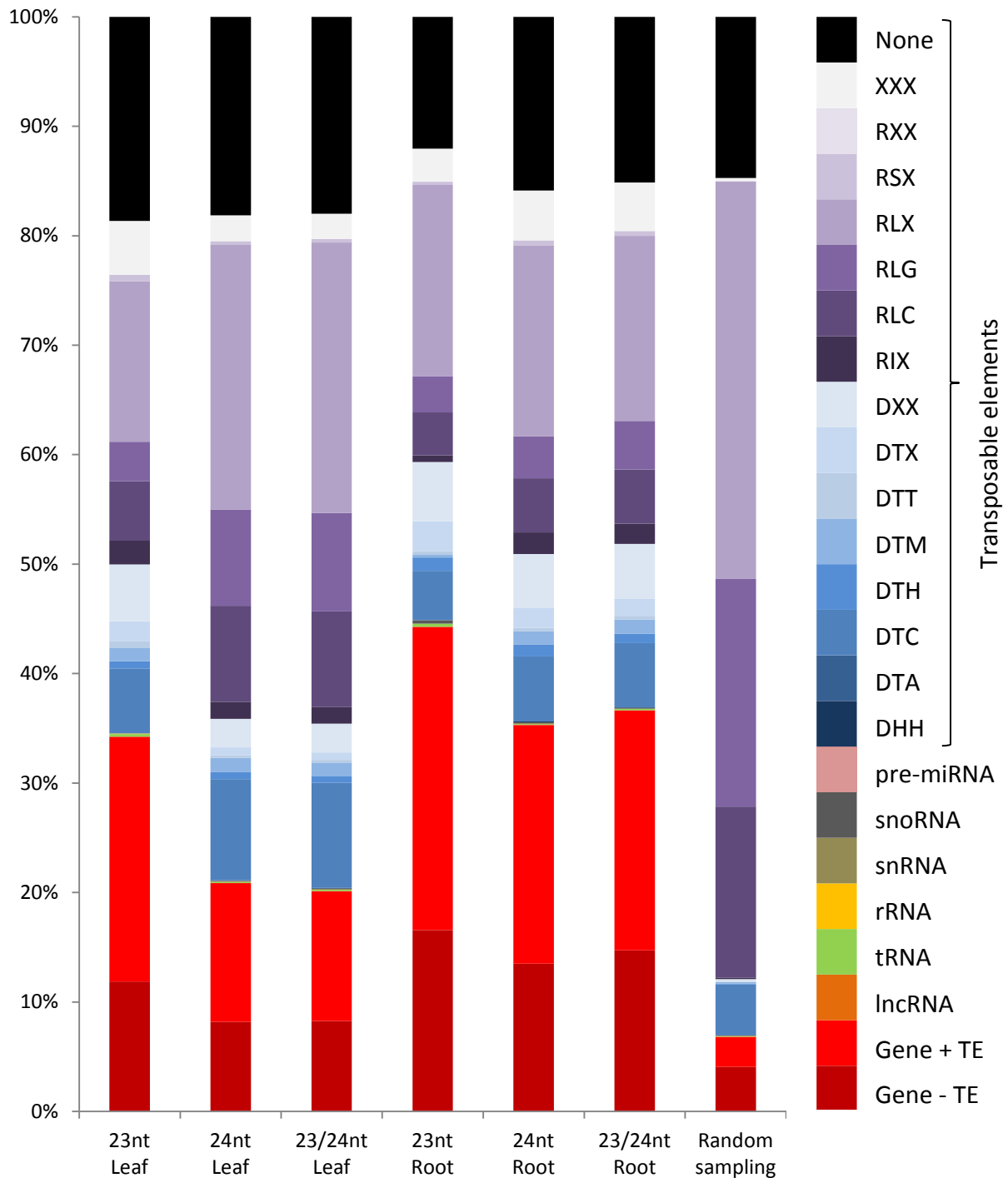


Figure 6.3. Genomic features of methylated sites where the differential methylation colocalizes with differential 23/24 nt sRNA expression within two 100 bp windows. Transposable elements (TEs) were categorised according to the standard established by Wicker et al. (2007). The random sampling method described in section 5.2.5 is shown for comparison. Methylation is classified as targeting a Protein Coding Gene (PCG) if it is between 1.5 kbp upstream of the transcription start site and 1.5 kbp downstream of the transcription end site of a PCG. Methylation is classified as targeting a TE if the differential methylation is inside the transcribed region of the TE. Gene + TE is where both a PCG and a TE were targeted, Gene - TE is where a PCG was targeted without a TE being targeted, and None is where none of the items listed in the legend were targeted.

Table 6.1. Number of Protein Coding Gene (PCG) sites targeted by differential methylation co-located with sRNA within two 100 bp genomic windows in barley under salt stress. Methylation is classified as targeting a PCG if it is between 1.5 kbp upstream of the transcription start site and 1.5 kbp downstream of the transcription end site. Methylation is classified as also targeting a TE if the differential methylation is between the start and end of the TE. Each cell contains the total count of PCG sites targeted by sRNAs (a single PCG may have more than one site targeted), the percentage of sRNA target sites associated with a PCG in parenthesis, and the fold change that targets are over-represented, relative to random sampling of the epiGBS mapped sites.

Tissue and sRNA length	sRNA targeting CHH methylation of PCGs and TEs	sRNA targeting CHH methylation of PCGs, not TEs	sRNA targeting CHH methylation of any PCG	Total sRNA targeting CHH methylation
Leaf, 23 nt miRNA	445 (22.3%) [x 2.9]	236 (11.8%) [x 8.2]	681 (34.1%) [x 5.0]	1995
Leaf, 24 nt siRNA	717 (12.6%) [x 2.0]	463 (8.2%) [x 4.6]	1180 (20.8%) [x 3.1]	5668
Leaf, 23/24nt sRNA	713 (11.8%) [x 2.0]	497 (8.3%) [x 4.3]	1210 (20.1%) [x 3.0]	6020
Root, 23 nt miRNA	92 (27.5%) [x 4.0]	55 (16.5%) [x 10.1]	147 (44.0%) [x 6.5]	334
Root, 24 nt siRNA	268 (21.8%) [x 3.3]	166 (13.5%) [x 8.0]	434 (35.2%) [x 5.2]	1232
Root, 23/24nt sRNA	307 (21.9%) [x 3.6]	206 (14.7%) [x 8.0]	513 (36.6%) [x 5.4]	1403
Random sampling	(2.7%) [x 1.0]	(4.1%) [x 1.0]	(6.8%) [x 1.0]	(100%)

Some transposable elements were highly over-represented when compared with random sampling of methylation mapped sites, others notably under-represented as shown in Table 6.2. Some of the highest levels of enrichment of TE targets were found in root tissue, with 23 nt miRNA differential expression co-located with methylation that corresponded to the genomic features of PCGs with a TE classified as DTH (class: DNA transposon, order: terminal inverted repeat; superfamily: *Harbinger*), or without a gene but a TE classified as DTX (class: DNA transposon, order: terminal inverted repeat; superfamily: unknown) with 91.2 and 72.0 fold increase above random sampling respectively. The targeting of most TEs was similar in the presence or absence of a PCG, with the exception of RLC, RLG and RLX (class: retrotransposon; order: long terminal repeat; superfamily: *Copia*, *Gypsy* or unknown, respectively). These groups were underrepresented when outside a PCG region and overrepresented when inside a PCG region.

Table 6.2. Fold difference of Protein Coding Genes (PCGs) and Transposable Element (TE) superfamilies targeted by differential methylation co-located with siRNA within two 100 bp windows when compared with random sampling. Methylation is classified as targeting a TE if the differential methylation is inside the TE, and targeting a PCG if it is between 1.5 kbp upstream of the transcription start site and 1.5 kbp downstream of the transcription end site. Classification of TEs is done according to the standard established by Wicker et al. (2007). Only TE classifications where random sampling was greater than 0.01% are shown.

Differential methylation target of siRNA	Fold difference relative to random sampling					
	Leaf			Root		
	23 nt	24 nt	23/24 nt	23 nt	24 nt	23/24 nt
PCG without TE	2.9	2.0	2.0	4.0	3.3	3.6
PCG with TE	8.2	4.6	4.3	10.1	8.0	8.0
PCG with or without TE	5.0	3.1	3.0	6.5	5.2	5.4
DTC without PCG	1.3	2.0	2.1	1.0	1.3	1.3
DTH without PCG	8.7	8.2	8.0	16.0	14.1	11.4
DTM without PCG	9.2	9.7	9.3	2.2	9.0	9.4
DTT without PCG	27.8	10.6	11.5	13.8	15.0	16.4
DTX without PCG	48.2	19.3	18.2	72.0	47.7	41.9
DXX without PCG	25.4	12.8	12.9	26.6	24.4	24.6
RIX without PCG	13.7	9.8	9.5	3.8	12.4	11.8
RLC without PCG	0.3	0.6	0.6	0.2	0.3	0.3
RLG without PCG	0.2	0.4	0.4	0.2	0.2	0.2
RLX without PCG	0.4	0.7	0.7	0.5	0.5	0.5
RSX without PCG	33.9	19.9	16.9	16.9	27.5	24.1
XXX without PCG	16.3	7.8	7.7	9.9	15.1	14.7
DTC with PCG	4.5	4.6	4.2	8.9	6.5	5.0
DTH with PCG	30.5	10.7	11.0	91.2	33.0	32.6
DTM with PCG	4.5	6.3	6.0	8.9	24.2	21.3
DTT with PCG	40.7	44.8	38.8	30.4	49.4	43.4
DTX with PCG	61.8	32.0	31.3	21.7	47.1	41.3
DXX with PCG	43.8	20.0	17.6	42.4	43.1	37.0
RIX with PCG	20.5	10.3	8.8	40.9	30.1	30.6
RLC with PCG	3.1	2.4	2.3	0.9	2.4	3.3
RLG with PCG	3.6	1.8	1.8	4.9	2.8	3.8
RLX with PCG	7.9	4.0	3.9	10.9	6.4	6.6
RSX with PCG	44.5	17.9	19.0	38.0	61.8	63.3
XXX with PCG	21.9	10.7	9.8	23.8	22.6	22.7
DTC with or without PCG	1.5	2.1	2.2	1.4	1.6	1.5
DTH with or without PCG	13.2	8.8	8.6	31.7	18.0	15.8
DTM with or without PCG	8.3	9.1	8.6	3.5	12.0	11.8
DTT with or without PCG	31.8	21.3	20.0	19.0	25.7	24.9
DTX with or without PCG	51.9	22.7	21.7	58.4	47.5	41.7
DXX with or without PCG	30.8	14.9	14.3	31.2	29.9	28.2
RIX with or without PCG	15.4	10.0	9.3	12.9	16.7	16.4
RLC with or without PCG	0.5	0.6	0.6	0.3	0.4	0.4
RLG with or without PCG	0.2	0.5	0.5	0.3	0.2	0.3
RLX with or without PCG	0.6	0.8	0.8	0.8	0.6	0.6
RSX with or without PCG	37.2	19.3	17.5	23.4	38.0	36.2
XXX with or without PCG	17.7	8.5	8.2	13.4	17.0	16.7
None	1.3	1.2	1.2	0.8	1.1	1.0

6.3.2 Differential DNA methylation and mRNA differential expression

Correlations between co-located differential DNA methylation and mRNA differential expression were determined and summarised in Table 6.3. In roots, 17 differentially methylated sites in a CG context were associated with PCGs (both with and without TEs). Of these, 69% of differentially methylated sites across PCGs without TEs were inversely correlated with a change in gene expression (i.e. an increase in CG methylation corresponding with a decrease in gene expression or vice-versa). Seventy five percent of the differentially CG methylated sites spanning both PCGs and TEs, showed changes in gene expression correlated with differential methylation. Forty seven percent of the differentially methylated CHH sites in leaf tissue had differentially methylated sites that were inversely correlated with gene expression. No differences were observed between differentially CHH methylated genes with or without TEs. In roots, 55% of the differentially methylated CHH sites showed an inverse correlation between DNA methylation and gene expression levels (61% and 52% of differentially methylated CHH sites spanning only genes or genes and TEs respectively).

Scatter plots were used to visualise the changes in mRNA expression and differential methylation, and are shown in Figure 6.4 and Figure 6.5. There were no significant correlations found when performing Pearson correlation tests on the various subsets shown in Figure 6.4 and Figure 6.5. All correlation p-values were greater than 0.21.

Table 6.3. Correlations between differential methylation and mRNA differential expression in different tissue and cytosine contexts in barley under salt stress. Differential methylation determined by finding the average methylation state across 100 bp windows throughout the genome for each cytosine context and finding significant differences between control and salt treated plants. Methylation is defined as targeting Protein Coding Genes (PCGs) if the differential methylation is within 1.5 kbp of the transcribed region of the gene and targeting a TE if the differential methylation is within the TE.

Methylation context and tissue	Methylation targeting	Methylation correlated with mRNA	Methylation inversely correlated with mRNA	Total
CG Leaf	Any PCG	5 (71%)	2 (29%)	7
	PCG, not TEs	4 (67%)	2 (33%)	6
	PCG and TEs	1 (100%)	0 (0%)	1
CHG Leaf	Any PCG	0 (0%)	1 (100%)	1
	PCG, not TEs	0	0	0
	PCG and TEs	0 (0%)	1 (100%)	1
CHH Leaf	Any PCG	199 (53%)	176 (47%)	375
	PCG, not TEs	122 (53%)	107 (47%)	229
	PCG and TEs	77 (53%)	69 (47%)	146
CG Root	Any PCG	7 (41%)	10 (59%)	17
	PCG, not TEs	4 (31%)	9 (69%)	13
	PCG and TEs	3 (75%)	1 (25%)	4
CHG Root	Any PCG	2 (33%)	4 (67%)	6
	PCG, not TEs	2 (40%)	3 (60%)	5
	PCG and TEs	0 (0%)	1 (100%)	1
CHH Root	Any PCG	33 (45%)	41 (55%)	74
	PCG, not TEs	22 (48%)	24 (52%)	46
	PCG and TEs	11 (39%)	17 (61%)	28

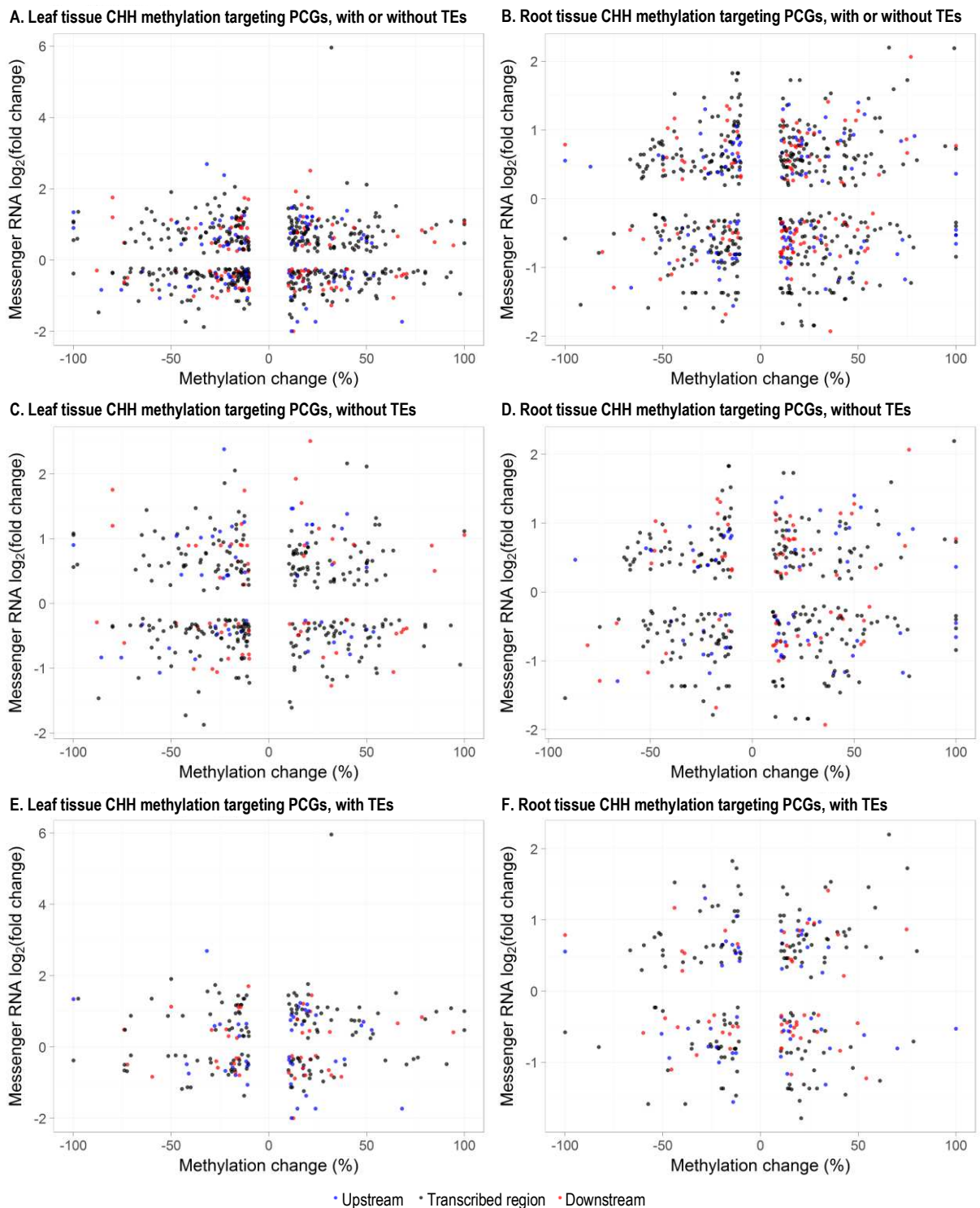


Figure 6.4. Scatter plot of differentially methylated CHH context sites associated with differentially expressed mRNA in barley under salinity stress. Tissue type and presence or absence of TE combinations: A. Leaf tissue, Differentially Expressed (DE) Protein Coding Genes (PCGs) with or without TEs; B. Root tissue, DE PCGs with or without TEs; C. Leaf tissue, DE PCGs without TEs; D. Root tissue, DE PCGs without TEs; E. Leaf tissue, DE PCGs and TEs targeted; F. Root tissue, DE PCGs and TEs targeted. PCG targeting is broken into upstream (blue, 1.5 kbp upstream to TSS), transcribed region (black, TSS to TES), and downstream (red, TES to 1.5 kbp downstream). Overlap between the differential methylation site and a TE is required to be classified as targeting a TE.

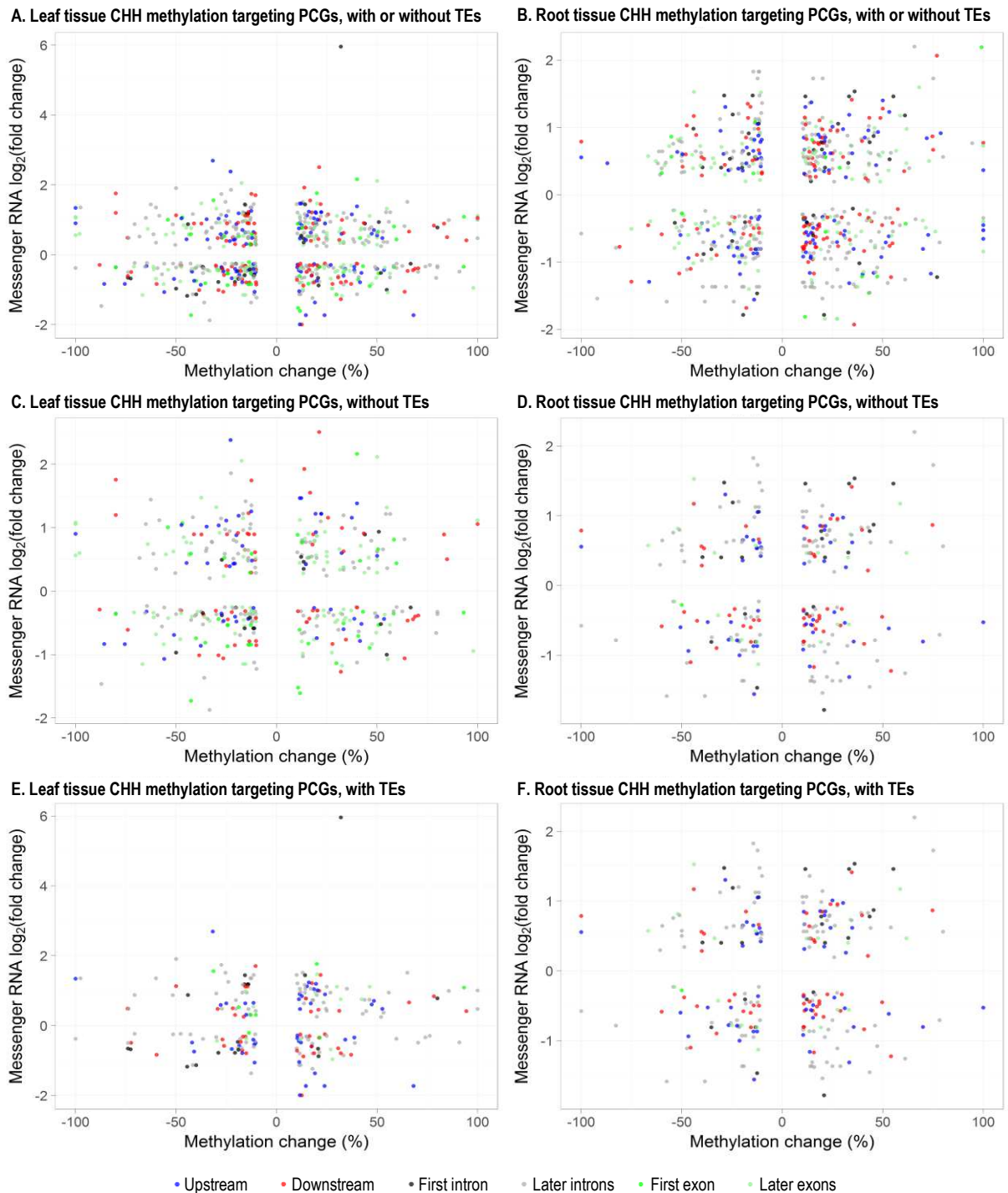


Figure 6.5. Scatter plot of differentially methylated CHH context sites associated with differentially expressed mRNA in barley under salinity stress. Tissue type and presence or absence of TE combinations: A. Leaf tissue, all Differentially Expressed (DE) Protein Coding Genes (PCGs) with or without TEs; B. Root tissue, all DE PCGs with or without TEs; C. Leaf tissue, DE PCGs without TEs; D. Root tissue, DE PCGs without TEs; E. Leaf tissue, DE PCGs and TEs targeted; F. Root tissue, DE PCGs and TEs targeted. PCG targeting is broken in to upstream (blue, 1.5 kbp upstream to TSS), first intron (dark grey), later introns (light grey), first exon (bright green), later exons (pale green) and downstream (red, TES to 1.5 kbp downstream). Direct overlap between the differential methylation site and a TE is required to be classified as targeting a TE.

6.4 Discussion

6.4.1 23/24 nt sRNA vs DNA methylation

Patterns of differential expression in 23/24 nt sRNA and changes in DNA methylation were not obvious when looking at a genome wide scale (Figure 6.1). This was most likely due to the limited sampling of the genome for methylation state information which excluded a large number of sites, along with the limitations in multimapped reads which would generally be associated with TEs, further discussed in section 7.2.2. The density of differential expression in 23/24 nt sRNAs increased at the ends of chromosomes, but no similar pattern was observed in changes in DNA methylation at the chromosomal scale. A smaller scale observation is required to search for differential 23/24 nt sRNA expression around regions where methylation read information was available to observe correlations in these two sequence sets.

Differentially expressed 23/24 nt sRNAs in barley under salinity stress were frequently found within 200 bp of a differentially methylated site as seen in Figure 6.2. This indicates that the methylation and 23/24 nt sRNA were related by proximity. This was observed in both root and leaf tissue in CHH context methylation. Other methylation contexts also hinted at similar patterns, however the small amount of statistically significant differentially methylated sites in CG and CHG contexts was of limited analytical use. In root tissue the patterns for the proximal relationship between differential methylation and siRNA all followed a similar pattern for 23 nt miRNA, 24 nt siRNA and the combination of the two (23/24nt sRNA), where the

normalised count of differential methylation in combination with differentially expressed siRNA dropped rapidly from 100% when aligned to about 65% over a 500 bp distance, and then decreased slowly to about 50% over a 5000 bp distance. This same pattern was observed in the differential expression of 23 nt miRNA with differential methylation in leaf tissue. The pattern for 24 nt siRNA and the combination of 23 nt and 24 nt sRNA was noticeably different in leaf tissue with much higher normalised counts. This may indicate that the 24 nt siRNA was involved in clustered targeting of genomic sites than the other sRNAs. It is possible that multiple 24 nt siRNAs in leaf tissue targeted multiple sequences over a region of thousands of base pairs to a higher degree than in root tissue. Changes in DNA methylation can be clustered, with changes frequently occurring over a region rather than individual cytosine bases (Gehring & Henikoff 2007). Alternatively it could be the result of the reduced representation bisulphite sequencing method creating biases in the sections of the genome that are sequenced in combination with different responses from different tissue. The amount of 23 nt miRNA reads was considerably less than 24 nt siRNA reads (Figure 4.1) and so the combination of 23 and 24 nt sRNA had a closer pattern to 24 nt siRNA alone. It is also possible that the lower number of 23 nt miRNA reads than 24 nt siRNA reads, in combination with extra noise in root sRNA (discussed in section 4.4.3) meant that fewer statistically significant sRNA expression sites were observable than in leaf 24 nt siRNA. Some combination of these possibilities might be able to explain the differences seen when comparing the patterns of 23 nt and 24 nt sRNA in proximity to differential methylation in root and

leaf tissue.

The genomic features of the methylated sites correlating with changes in 23/24 nt sRNA, shown in Figure 6.3, indicate a noticeable increase in siRNAs sharing homology with differentially methylated PCGs when compared with random sampling of the available sites. Both PCGs without TEs and PCGs with TEs were targeted by sRNAs much more than random sampling, with 23 nt miRNA targeting methylation of PCGs 5 times and 6.5 times the expectation from random sampling in leaf and root tissue respectively. The 24 nt siRNA also had targeted methylation of PCGs at a greater level than expected from random sampling, with 3.1 times and 5.2 times in leaf and root tissue respectively. It is of interest that 23 nt miRNAs appear to target differentially methylated PCGs more strongly than 24 nt siRNA. Transposable elements appear to have been a greater target in 24 nt siRNA than 23 nt miRNA. This indicates that the 23 nt miRNA and 24 nt siRNA have different functions with relation to targeting PCGs and TEs.

Transposable elements have been shown to play an important role in stress response in plants (Grandbastien 1998; Le et al. 2014; Makarevitch et al. 2015). The over-representation of siRNA targeting PCGs with TEs as shown in Figure 6.3 is quite striking when compared with random sampling of methylation reads mapped to the genome. These differences are described in more detail with Table 6.2 and Table A.9.18. When siRNA targets PCGs and TEs at the same time, those sites are over-

represented at a higher rate than siRNA targeting PCGs and not targeting TEs.

In roots under salt stress, the TE classes most highly targeted by siRNAs were DTH (class: DNA transposon, order: terminal inverted repeat; superfamily: *Harbinger*) when it was also targeted with a PCG, and DTX (class: DNA transposon, order: terminal inverted repeat; superfamily: unknown) when it was targeted without a PCG (Table 6.2 and Table A.9.18). Both of these had the highest fold difference relative to random in the 23 nt miRNA targeted sites, with lower but still elevated levels in 24 nt siRNA targeted sites. A similar pattern for these TE classes was seen in leaf tissue with higher levels in 23 nt miRNA targeted sites than 24 nt siRNA targeted sites. This provides some confirmation of the difference in function between 23 nt miRNA and 24 nt siRNA, as previously reported (Teotia et al. 2017).

The *Harbinger* superfamily of TEs is known to have autonomous TEs with all of the genes required for replication and movement throughout the genome, while also having a large number of Miniature Inverted repeat Transposable Elements (MITEs) in grasses (Bureau & Wessler 1994; Wicker et al. 2017). MITEs are much smaller than full TEs, and rely on some or all genes from functional TEs to mobilise. There are 10,634 copies of autonomous and non-autonomous *Harbinger* TEs in the reference barley genome, with 25.7% of these being found within 5 kbp of genes, and 40% within 10 kbp of genes (Wicker et al. 2017). Many of the methylation reads that were mapped to *Harbinger* TEs would actually have been related to these MITEs due to the

large number of MITEs relative to fully functional TEs, and a significant proportion of these were in close proximity to genes as seen in Table 6.2.

Most of the TEs were over-represented relative to random sampling in barley under salt stress with the clear exception of RLC (class: retrotransposon, order: long terminal repeat; superfamily: *Copia*), RLG (class: retrotransposon, order: long terminal repeat; superfamily: *Gypsy*) and RLX (class: retrotransposon, order: long terminal repeat; superfamily: *unknown*). These three groupings stood out with an underrepresentation when not also associated with a gene and mostly an overrepresentation when also associated with a gene.

The *Copia* superfamily contains the most abundant family of TEs in the barley genome known as *BARE1* (Wicker et al. 2017). This family of TEs is known to form virus-like particles (VLPs) (Chang et al. 2013). The transcripts from *BARE1* can be differentially spliced which causes the formation of the VLPs and is known to be involved with stress response (Jääskeläinen et al. 2013). The *Gypsy* superfamily contains the four next most abundant families of TEs in the barley genome. Both *Gypsy* and *Copia* TEs are relatively evenly spread throughout the genome, but individual families of TEs have differing distribution patterns (Wicker et al. 2017). Research in *Zea mays* has shown the importance of RLG (*Gypsy*), RLC (*Copia*) and RLX (*unknown*) in stress response of cold, heat, salt and UV stress, where stress response activated genes were related to these TEs (Makarevitch et al. 2015).

Non-functional sequences associated with TEs account for 84% of the barley genome (IBGSC 2012). With such a huge portion of the genome allocated to TEs it is reasonable to expect that TEs provide some benefit. If TEs caused more harm than benefit to the host genome then individuals with mutations preventing the functioning of TEs would have an advantage over those without. Over evolutionary timescales, the TEs would decrease in number under those circumstances. Transposable Elements are known to play a beneficial role in plant stress response (McCue et al. 2012; Mao et al. 2015; Zhang et al. 2018), and with the high proportion of the genome related to TEs it would be reasonable to suggest that TEs are more beneficial than they are harmful to the survival and reproduction of barley.

The Bismark bisulphite methylation calling program (Krueger & Andrews 2011) has limitations in that a read must be mapped to a single location in the genome for analysis. Multi-mapped reads cannot be processed without significant extra work which is discussed later in section 7.2.2. Transposable Elements are replicated in multiple locations throughout the genome. Combined with the fact that epiGBS sequence read sizes frequently spanned less than 1 kbp meant that many TEs could not be analysed. Of the epiGBS reads that were not filtered out for low quality or PCR duplication and could be mapped to the genome, only 42.7% were mapped to a single location, with 57.3% mapping to multiple locations and being discarded by Bismark. Generally, only reads of TEs that spanned across a TE boundary could be mapped to a single location. With a multi-mapping modification to Bismark, this

would massively increase the TE data set and enable the gathering of further information about the role that TEs play.

The chromosomal pattern of changes in expressed 23 nt miRNA, 24 nt siRNAs and methylation in barley under salt stress is shown in Figure 6.1. A noticeable increase in differentially expressed 23/24 sRNAs is visible within about 30 Mbp of the ends of chromosomes. The same pattern is not observed in the differential methylation data set. Research by Mascher et al. (2017) demonstrated that genes, Miniature Inverted repeat Transposable Elements (MITEs) and Long Interspersed Nuclear Elements (LINEs) also were heavily concentrated at the ends of chromosomes. It is possible that many of the bisulphite reads of MITEs and/or LINEs were discarded due to multi-mapping and hence would not have shown up clearly in the mapped bisulphite reads in Figure 6.1. Improvements in how Bismark handles multi-mapped reads and how it copes with TEs could provide further information on why the methylation pattern does not also increase at the ends of chromosomes.

6.4.2 DNA methylation vs mRNA expression

Figure 6.4 and Figure 6.5 plots the differential methylation level and differential mRNA expression of salt treated plants relative to control plants. There are not any clear patterns observed when breaking the sites in to categories of upstream, downstream, first intron, other introns, first exon or other exons. Breaking down methylation that targets genes but not TEs or targeting both genes and TEs does not

help to show any clear patterns either. The grouping of 5' Untranslated Region (UTR) and 3' UTR was also tested, but did not yield any significant patterns (data not shown). Correlations were tested for different classes of TEs and how they were associated with differential methylation and mRNA differential expression. No statistically significant correlations were found in either leaf or root tissue.

Prior research has shown that the link between DNA methylation and mRNA expression is not straightforward (Bewick & Schmitz 2017). There are instances where increased methylation in the promoter region was correlated with a decrease in associated transcript production (Berdasco et al. 2008), and other instances where this simple model does not adequately explain the interaction between methylation and gene expression (Wang et al. 2015). Some have suggested that gene body methylation can be a better indicator of the changes in mRNA production than methylation of the promoter (Lou et al. 2014). Anastasiadi et al. (2018) found a clear relationship between increased methylation of the first intron and decreased gene expression in various vertebrates. Shibuya et al. (2009) found an example where increased methylation of the second intron resulted in increased gene expression in the pMADS3 gene in *Petunia hybrida*. Much of this prior research has focussed on CG methylation, while the findings in the present research indicates that other cytosine contexts also do not present a clear link between methylation and gene expression in barley. Clearly there is scope for future work to investigate how DNA methylation around genes relates to gene expression in barley as the dataset presented does not

provide clear or simple correlations.

6.4.3 Conclusion

The hypothesis was partially supported. Exposure to salinity stress in barley resulted in demonstrable changes in the expression profile of 23/24 nt sRNA, changes in DNA methylation and changes in mRNA expression. The link between 23/24 nt sRNA and DNA methylation is correlated, however the expected inverse correlation between DNA methylation and mRNA expression was not observed. There does not appear to be a direct link between genomic methylation of protein coding genes and expression of those genes. This could be a result of some unknown intermediary complicating the issue. Transposable elements play a greater role in the barley genome than many other plants that have been studied, with TEs making up 84% of the barley genome (IBGSC 2012). Further work investigating the methylation changes that occur in TEs may shed some light on the indirect relationships between DNA methylation and gene expression in barley under salt stress. Further work can also be performed to gain a greater understanding of the newly discovered 23/24 nt sRNAs that guide DNA methylation in barley under salinity stress.

7 General discussion

Understanding and dealing with salinity stress is an important factor for growing grain crops in Australia, with a large overlap in regions where grain crops are grown in Australia and soil salinity issues (Patterson et al. 2009). Barley and wheat play an extremely important role in global agriculture, ranking fifth and first respectively in global food production (Mayer et al. 2011). As barley is a close relative of wheat (Middleton et al. 2014), research on barley can have a significant impact on global food security, as research in barley is likely to be adaptable to wheat and other grasses such as rice and maize.

Experiments were performed that investigated the changes in small RNA (sRNA), DNA methylation and mRNA expression that occurred when barley plants were exposed to a salinity stress. The aim was to overlay these data sets to identify sequences that were shared, with the expectation that small RNAs guide methylases to genes that are regulated under conditions of salinity stress. The ultimate aim was to gain a greater understanding of the plant response to salinity, so as to assist in the breeding and/or transgenic approaches to improve salinity and other stress tolerance in the future.

The mRNA, sRNA and DNA methylation data sets analysed were overall in concordance with prior research. The Protein Coding Genes (PCGs) with the greatest

increase and decrease in barley under salinity stress mostly correlated with mRNA expression changes from prior research. The number of known sRNAs in barley is limited, but correlations between known sRNAs in other plants with high sequence similarity indicated the validity of this dataset. Methylation levels at the chromosomal scale, and in the vicinity of PCGs and TEs found in this study matched those with other plants, and were similar to earlier reports in more closely related species. These findings validate the datasets for sRNAs, mRNAs and DNA methylation reported here.

As discovered in this study, the interactions between sRNA, DNA methylation and mRNA expression are not straightforward in barley under salinity stress. Changes in expression of sRNAs that were 23 nt to 24 nt long were found to be well correlated with changes in DNA methylation, however these shared sequences did not overlap with PCGs that were differentially expressed under salinity stress. Prior research has shown that these sRNAs guide DNA methylation through the RNA directed DNA Methylation (RdDM) pathway (Matzke & Mosher 2014). Data presented here show DNA methylation targeted various genomic features, with PCGs and some classes of Transposable Elements (TEs) being highly over-represented when compared with random targeting of the locations where methylation sequence data were available. The proximal correlations between sRNA differential expression and differential methylation give an indication that the two are linked and change in response to salt stress in both leaf and root tissue in barley. There were however no clear indications

that the changes in DNA methylation in a specific location relative to a PCG resulted in consistent expression changes of those genes. Figure 7.1 shows a schematic representation of the pathways investigated. While some of these pathways have been explored in barley previously, at the time of writing there does not appear to be any prior research that integrates sRNAs, DNA methylation and mRNA sequence information in barley under salt stress.

The hypothesis under test was that if gene expression in barley under salinity stress is at least partly controlled by epigenetic mechanisms, then exposing the plants to salinity stress will result in changes in sRNA and DNA methylation associated with mRNA expression. This hypothesis was partially accepted and partially rejected. The plants exposed to salinity stress clearly had changes in sRNA expression, DNA methylation, and mRNA expression. Links between sRNA expression and DNA methylation were observed, however the link between DNA methylation changes and mRNA expression was not clear. It is possible that some intermediary such as TEs or histone modifications were involved in the association between DNA methylation changes and mRNA expression in barley.

Further research into TEs, with possible improvements in further analysing the collected dataset are suggested in section 7.3.3. The 23/24 nt sRNAs found to be differentially expressed in salt treated plants relative to control plants are likely to be due to re-establishing silencing of TEs that had recently been enabled. These TE

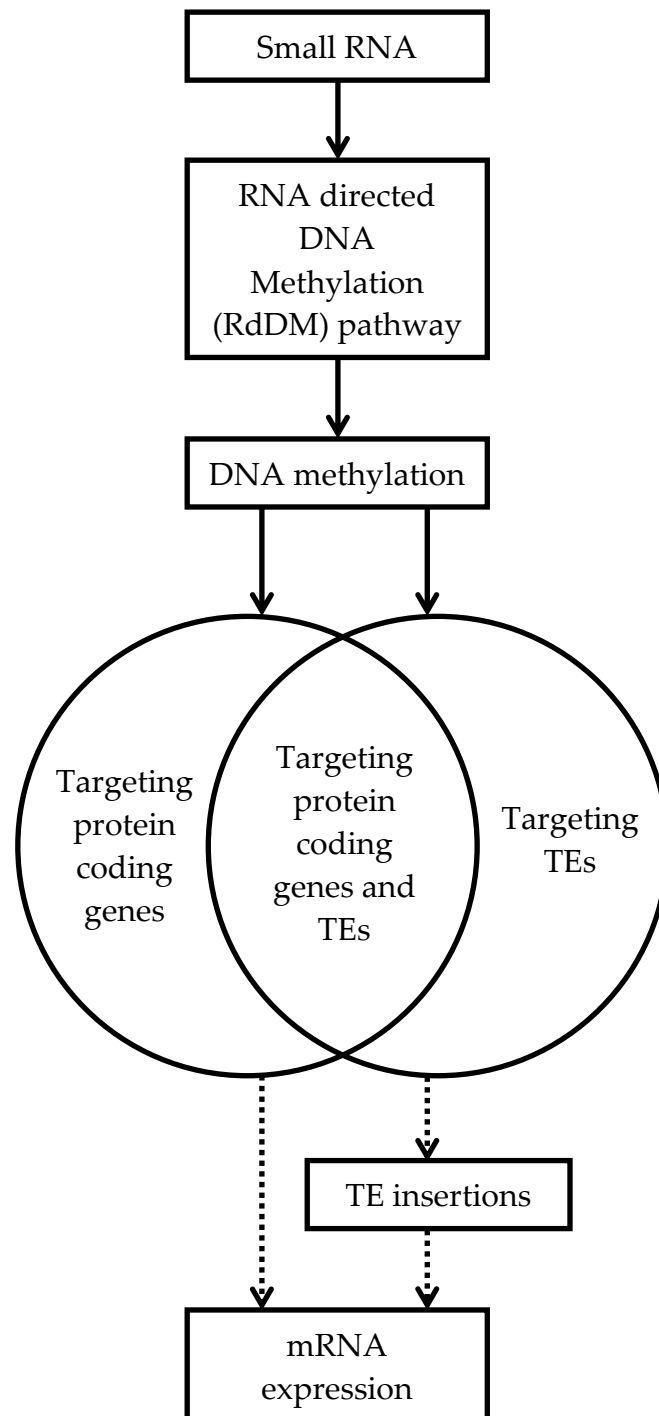


Figure 7.1. Schematic diagram of important pathways investigated for changes that occurred when barley was exposed to 150 mM salinity stress in soil vs control plants. Solid arrows indicate connections that were observable in this study and dashed arrows indicate expected connections which were not verified in this study.

insertions could result in increased or decreased expression of a targeted mRNA based on where the TE inserted relative to gene features such as enhancers, repressors or promoters. There is some precedence for this, with some TEs found to target certain genomic features, with an elevated propensity in some TEs for insertion at or near existing non-functional TE fragments, thus influencing the expression of neighbouring genes (Wei et al. 2013; Ma et al. 2019).

7.1 Novel discoveries and developments from this research

7.1.1 New and novel micro RNA involved in post-transcriptional repression

New and novel microRNAs (miRNA) were found in this study by comparing small RNA sequences with genomic sequences that could generate an RNA hairpin-loop structure corresponding with the sequenced sRNA data set. There were 59 unique miRNAs that had not previously been recorded for barley in miRBase. Of these newly discovered sequences, 44 were classified as novel as they were not similar to any known plant miRNA, and 15 were classified as new since similar sequences were known in other plants, but had not been previously recorded in barley. Five of the novel miRNAs and three of the new miRNAs were found to be differentially expressed in leaf tissue in barley under salt stress, and one novel miRNA and one new miRNA were differentially expressed in root tissue. The discovery of these miRNAs in barley will contribute to the knowledge of miRNAs in barley once published, as only 71 *Hordeum vulgare* miRNAs were reported in the miRBase at the time of writing.

7.1.2 Development of bioinformatics tools

A bioinformatics pipeline was developed that could analyse reads generated by the epiGBS method and generate methylation calls by using the Bismark aligning and methylation calling program. This development required considerable time, but with a comparatively small amount of extra effort, will become a useful tool for other researchers who wish to use the epiGBS protocol.

7.1.3 Links between 23/24 nt sRNAs and DNA methylation

Links between changes in expression of 23/24 nt sRNAs and DNA methylation in barley under salinity stress were discovered. The correlations between differential 23/24 nt sRNA expression and changes in DNA methylation confirmed that the changes were linked. These correlated changes were found to cluster with PCGs and certain TE classes at a high level indicating that PCGs and some TE classes were targeted by 23/24 nt sRNAs and DNA methylation.

7.2 Limitations

7.2.1 Small RNA

There was a problem with the sequence data set for sRNAs from root tissue. The distribution of sRNAs by size (Figure 4.1) extracted from leaf tissue was as expected, but this was not the case for root tissue. The leaf tissue sRNA profile had a clear bimodal distribution centred around sRNAs 21 nt and 24 nt long which had been observed in prior research (Schreiber et al. 2011), but the root tissue sRNAs had a

single peak at 24 nt with a wide span. The considered explanations for the unexpected distribution in root sRNA sequences are fragmented barley mRNA, or contaminants of bacterial and fungal RNA. Removing reads that mapped to the known barley transcriptome removed the same proportion of reads when applied to sRNA extracted from root tissue and leaf tissue. Therefore the range of sRNA sizes found specifically in root tissue cannot be attributed to endogenous fragmented barley root mRNA.

Most microbiome analysis techniques use 16S ribosomal RNA (rRNA) matches to determine the make-up of the microbiome (Yuan et al. 2012). Specific regions of the 16S rRNA are targeted for sequencing where the primers bind in regions of low variability, whereas the region between the primers have relatively high variability in the 16S rRNA sequence across microbe species (Claesson et al. 2010). In soil samples, a correct classification at the family level can be obtained approximately 70% of the time when using average read lengths of 215 bp or more (Vasileiadis et al. 2012). The read lengths obtained in the small RNA sequence were size limited in the Pippin-prep size selection stage to a maximum of 34 bp. This limitation in fragment size means that the small RNA sequence dataset cannot be used with existing microbiome tools. It is not possible to determine the cause of the contamination in root sRNAs without a significant of further research. The most likely source of this contamination is of bacterial or fungal origin. This would explain why it is in root tissue and not found in leaf tissue. While microbes will be present on or in leaves,

their abundance in and around root tissue is potentially much higher. While efforts were made to remove soil from the roots, there was also a need to prevent damaging the roots with vigorous cleaning. Plant mRNA expression has been shown to change rapidly under mechanical wounding. Research by Koo et al. (2009) demonstrated that a gene coding for the protein JA-IIe in the jasmonate signalling pathway had been transcribed, translated and exited leaves within two minutes of wounding.

7.2.2 Multi-mapped methylation sequence reads

The Bismark bisulphite read mapper was unable to deal with reads that map to multiple locations. Of the epiGBS reads that were able to be mapped to a sequence on the genome, 42.7% mapped to a single location in the genome while 57.3% mapped to multiple locations and were thus discarded by Bismark. Many of the reads associated with TEs would map to multiple locations as the same TE sequence occurs multiple times throughout the genome. Many of the reads that Bismark did successfully map to TEs probably either crossed the boundary of the TE, anchoring the read to a specific site in the genome, or were somehow different to other TEs, perhaps with a unique insertion/deletion or other sequence change. Further work could be done to enable the use of larger numbers of multi-mapped reads, and a potential method for this is described in section 7.3.3.

7.2.3 DNA methylation

The epiGBS technique is a reduced representation bisulphite sequencing method,

and as a result, only small sampled sections of the genome were assessed for DNA methylation. This method was used instead of entire genome sequencing due to cost limitations. While it is a very effective tool, the limitations of the epiGBS technique prevent a more detailed picture than that attained by entire genome bisulphite sequencing. However, there would still be sections that could not be mapped to a single location in the genome, typically those related to TEs. Sequencing techniques that are able to gain information from long reads such as the Third Generation Sequencing technologies offered by PacBio and Oxford Nanopore Technologies (Weirather et al. 2017) may be accessible in the future, as costs continue to come down and sequence reliability increases. The current bisulphite treatment method will often cause damage to large DNA fragments (Yang et al. 2015), but the Oxford Nanopore MinION and PacBio SMRT sequencers do not require bisulphite treatment to obtain methylation information (Clarke et al. 2009; Flusberg et al. 2010). These techniques are not currently as reliable or cost effective as bisulphite sequencing. The Oxford Nanopore MinION accuracy varies between 83% and 91% for correctly calling a 5-methylcytosine base, but the higher accuracy rate comes at the expense of 32% of base calls being discarded (Simpson et al. 2017). The PacBio SMRT sequencer requires a minimum of 250x coverage to detect cytosine methylation (Liu et al. 2020). At current costs, the sequencing of the whole barley genome to identify 5-methylcytosine at single base resolution would cost about AUD26,000, based on a 15x coverage recommended by Ziller et al. (2015) and not including library preparation costs. For a project of this size, these costs are prohibitive.

Many of the sequence reads related to the epiGBS protocol were PCR duplicates. Only one copy of a PCR duplicate is useful for analysis. The only time a PCR duplicate is of use is when one or more copies have a low read quality score and one PCR duplicate has a high read quality score, otherwise they are wasted reads. Using more starting material and fewer PCR cycles should reduce the number of reads wasted on PCR duplicates. Given the chance to repeat this research, this would be a recommended change. Furthermore, the bisulphite conversion rate (94.2%) was lower than the manufacturers claimed conversion rate of 99.5% (Zymo 2020). A lower conversion rate would have increased the noise in methylation base calling. It is not clear why this poor conversion rate occurred, but it would be worth attempting to attain a higher bisulphite conversion rate in future experiments.

7.3 Future Directions

7.3.1 Small RNA

Existing microbiome tools expect a certain read size, and reads from a specific location in the 16S rRNA are required to determine the makeup of the microbiome. It may be possible to adapt these tools, or at least use the database of 16S rRNA, to reference against the small RNA reads and determine the source of small RNA contamination in the root sRNA data found here. A preliminary search against a 16S rRNA microbiome database with the Bowtie sequence aligner found that approximately 10% of sequences that were 26 nt long in root small RNA reads aligned with microbiome 16S rRNA. However, this was after removal of sequences

that aligned to the barley rRNA, mitochondria and plastid sequences. The barley 18S rRNA would have some similarities to the prokaryotic 16S rRNA, and mitochondria and plastid rRNA would also have similarities to parts of other prokaryotic 16S rRNA. Only the sections that had not aligned to barley rRNA or rRNA of mitochondrial or plastid origin would have remained to align with other microbiome 16S rRNAs. This indicates that further investigation is warranted to be able to understand the microbiome interactions and potentially observe microbiome differences between salt and control root tissue. With a more thorough analysis of the microbiome it may be possible to 'clean up' the root small RNA reads and get a more detailed picture of what is going on with sRNAs in root tissue.

The sRNAs that were involved with DNA methylation warrant further investigation. These sRNAs are expected to be related to changes in DNA methylation that corresponds with increased stress tolerance. Further understanding of the pathway between the DNA methylation change and the specific stress tolerance in plants could yield useful tools. It may be possible to treat plants with an sRNA or a collection of sRNAs that change the methylation state of the desired regions of the genome to increase stress tolerance of the plant. Doing this prior to the onset of the prolonged stress may result in greater crop yields. Possibilities for application of the treatment include spraying or soaking seeds before sowing, or spraying the leaves of established plants with a sRNA solution which could potentially enter the plant and modify methylation states. It seems to be reasonable to expect that different

collections of sRNAs could be developed to prime plants for different stresses. An example of a possible use is when a long range forecast predicts drought conditions; a farmer could treat their plants with a set of sRNAs that prime the plants for a drought stress, enabling a greater yield than without the sRNA treatment.

Further work could also be done with the miRNA data set. This data set was used to find new and novel barley miRNAs, and further work could determine the targets of these miRNAs. A degradome analysis finds mRNA that has been degraded through miRNA induced cleavage (German et al. 2009). By comparing sites of mRNA cleavage and sequence similarity of miRNA, it would be possible to understand what genes are targeted by specific miRNA. It may also be useful to perform sequencing on tissue at multiple time points to see the changes that occur at different times after the application of the stress.

7.3.2 DNA methylation

Enabling the management of reads mapped to multiple locations in the genome is another consideration for future research. Bismark is an open source tool with the PERL source code easily accessible and with a reasonable level of documentation. It would be possible to modify the Bismark code to deal with multi-mapped reads. Bismark calls the sequence aligner Bowtie2 which is capable of analysing reads that map to multiple locations. There appear to be two ways to increase the number of reads mapped to TEs, either significant modification of Bismark to be able to deal

with multi-mapped reads, or generate a new reference genome.

Bismark could be used with the normal reference genome. The option of saving multi-mapped reads to a different location exists for Bowtie2, so very minor changes to Bismark would enable the creation of the multi-mapped files. A reference genome of short paired-end reads could be generated from these normally discarded reads. Bismark could then be used again with the saved multi-mapped reads mapping to the newly generated reference genome. This reference genome could then be mapped back to the original genome to find the genomic features associated with these multi-mapped reads. It would be expected that a large proportion of these reads would be associated with TEs. While it would not be possible to determine a single location in the genome, and these methylation calls would be of limited use when comparing with the small RNA and mRNA data sets, it would still contain useful information about the methylation states of some TEs that may perform functions in response to salinity stress. This would be useful future work as it would make good use of the 57.3% of mapped reads that were discarded.

A significant number of the total epiGBS reads (37.1%) did not map to any location in the genome. A similar operation to that discussed above could be performed with the unmapped reads to see if these were found across all samples. It is possible that more useful information could be gained by further analysis of the reads that were discarded because they did not match the reference genome.

It is not clear why the CG differential methylation pattern in leaf tissue is so concentrated around PCGs while other cytosine contexts and other tissue appears to be more similar to each other (see Figure 5.12). Further work could be done to investigate the specific PCGs and TEs that are targeted and compare with other contexts and tissue types to see if any significant patterns emerge.

While a bioinformatics pipeline was developed that was able to extract useful information from the epiGBS sequence reads, more work needs to be done to make this code useful for other users. In many cases the code was written for the specific barcode set and a specific number of samples used in this experiment. Some parts of the code would need significant changes to be useful for other situations which use different sets of barcodes, and different sample designations for control and treatment groups. To create useful tools for other users, the documentation needs to be detailed enough, and easy enough to understand that a user with some bioinformatics knowledge is able to install all the dependencies, and successfully run the program on a variety of relevant Linux-based platforms.

7.3.3 Transposable Elements

Further work could be done to analyse the specific TEs that are targeted by DNA methylation and 23/24 nt sRNA. The research suggested above to deal with multi-mapped reads would also be useful for gaining further information about TEs. The analysis performed in this research looked at classes of TEs, but further work to drill

down to specific TEs and their interactions with PCGs is warranted. This could involve looking at methylation changes in TEs and the proximity of the same TEs or non-functional TE fragments at different genomic locations near or within PCGs. This would be of particular interest with TEs that are already known to insert close to existing TE fragments (Wei et al. 2013; Ma et al. 2019), as changes in methylation in those TEs are likely to involve TE insertions at certain sites. Some interesting patterns emerged when looking at the methylation levels in two classes of TEs shown in Figure 5.11. Understanding why there are spikes in CHH methylation at specific locations in DTX (class: DNA transposon, order: terminal inverted repeat; superfamily: unknown), and RSX (class: retrotransposon, order: non-long terminal repeat, superfamily: unknown) TEs may be worthy of further investigation. Transposable Elements are known to play an important role in stress response (Horváth et al. 2017) and this project has found some interesting TE groups worthy of further investigation.

7.4 Conclusion

This research resulted in new discoveries in the response of barley to salt stress. New miRNAs (previously unknown in barley but with sequence similarity to other plants) and novel miRNAs (no sequence similarity with any known miRNA) were found. In total, 59 different miRNAs that had not been recorded in the microRNA database miRbase are reported here. Of these newly discovered sequences, 44 were novel and 15 were new. Five of the novel miRNAs and three of the new miRNAs were

differentially expressed in barley under salt stress. The latest miRBase release at the time of writing had only 71 miRNAs listed for *H. vulgare*, and once the newly discovered miRNAs are published, it would result in an 83% increase in known miRNAs for *H. vulgare* in miRBase.

Correlations between differential expression of 23/24 nt sRNAs and differential methylation were observed in barley under salt stress. The sites discovered with 23/24 nt sRNAs and methylation changes were significantly more enriched at PCGs and TEs than random sampling. This indicates that the PCGs and TEs were specifically targeted. Some (or all) of these sites targeted by the 23/24 sRNAs would be related to changing mRNA expression in relation to the salt stress applied, however a direct link between the two was not established. It is possible that further research into the targeted TEs would help in understanding how genomic methylation changes result in altered mRNA expression in barley under salt stress. A greater understanding of how these TE translocations elicit a stress response will also help plant breeders develop crops more resilient to the stressful environments that are likely in the future.

8 References

- Aghaei, K., Ehsanpour, A.A., Shah, A. & Komatsu, S. 2009, 'Proteome analysis of soybean hypocotyl and root under salt stress', *Amino acids*, vol. 36, no. 1, pp. 91-8.
- Akalin, A., Franke, V., Vlahoviček, K., Mason, C.E. & Schübeler, D. 2015, 'Genomation: a toolkit to summarize, annotate and visualize genomic intervals', *Bioinformatics*, vol. 31, no. 7, pp. 1127-9.
- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A. & Mason, C.E. 2012, 'methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles', *Genome biology*, vol. 13, no. 10, p. R87.
- Akimoto, K., Katakami, H., Kim, H.-J., Ogawa, E., Sano, C.M., Wada, Y. & Sano, H. 2007, 'Epigenetic inheritance in rice plants', *Annals of botany*, vol. 100, no. 2, pp. 205-17.
- Alpaslan, M. & Gunes, A. 2001, 'Interactive effects of boron and salinity stress on the growth, membrane permeability and mineral composition of tomato and cucumber plants', *Plant and Soil*, vol. 236, no. 1, pp. 123-8.
- Anastasiadi, D., Esteve-Codina, A. & Piferrer, F. 2018, 'Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species', *Epigenetics & chromatin*, vol. 11, no. 1, p. 37.
- Andrews, S. 2017, 'FastQC: a quality control tool for high throughput sequence data. 2010'.
- Aparicio-Puerta, E., Lebrón, R., Rueda, A., Gómez-Martín, C., Giannoukakos, S., Jaspez, D., Medina, J.M., Zubkovic, A., Jurak, I. & Fromm, B. 2019, 'sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression', *Nucleic acids research*.
- Aravind, J., Rinku, S., Pooja, B., Shikha, M., Kaliyugam, S., Mallikarjuna, M.G., Kumar, A., Rao, A.R. & Nepolean, T. 2017, 'Identification, characterization, and functional validation of drought-responsive microRNAs in subtropical maize inbreds', *Frontiers in plant science*, vol. 8, p. 941.
- Asif, M.A., Schilling, R.K., Tilbrook, J., Brien, C., Dowling, K., Rabie, H., Short, L., Trittermann, C., Garcia, A. & Barrett-Lennard, E.G. 2018, 'Mapping of novel salt tolerance QTL in an Excalibur× Kukri doubled haploid wheat population', *Theoretical and applied genetics*, vol. 131, no. 10, pp. 2179-96.
- Avramidou, E., Kapazoglou, A., Aravanopoulos, F.A., Xanthopoulou, A., Ganopoulos, I., Tsaballa, A., Madesis, P., Doulis, A.G. & Tsaftaris, A. 2015, 'Global DNA methylation changes in *Cucurbitaceae* inter-species grafting', *Crop Breeding and Applied Biotechnology*, vol. 15, no. 2, pp. 112-6.
- Bahieldin, A., Atef, A., Sabir, J.S., Gadalla, N.O., Edris, S., Alzohairy, A.M., Radhwan, N.A., Baeshen, M.N., Ramadan, A.M. & Eissa, H.F. 2015, 'RNA-Seq analysis of

- the wild barley (*H. spontaneum*) leaf transcriptome under salt stress', *Comptes rendus biologiques*, vol. 338, no. 5, pp. 285-97.
- Bai, B., Shi, B., Hou, N., Cao, Y., Meng, Y., Bian, H., Zhu, M. & Han, N. 2017, 'microRNAs participate in gene expression regulation and phytohormone cross-talk in barley embryo during seed development and germination', *BMC plant biology*, vol. 17, no. 1, p. 150.
- Baránek, M., Čechová, J., Raddová, J., Holleínová, V., Ondrušíková, E. & Pidra, M. 2015, 'Dynamics and Reversibility of the DNA Methylation Landscape of Grapevine Plants (*Vitis vinifera*) Stressed by In Vitro Cultivation and Thermo-therapy', *PLOS one*.
- Barrero-Sicilia, C., Hernando-Amado, S., González-Melendi, P. & Carbonero, P. 2011, 'Structure, expression profile and subcellular localisation of four different sucrose synthase genes from barley', *Planta*, vol. 234, no. 2, pp. 391-403.
- Batchelor, A.K., Boutilier, K., Miller, S.S., Hattori, J., Bowman, L., Hu, M., Lantin, S., Johnson, D.A. & Miki, B.L. 2002, 'SCB1, a BURP-domain protein gene, from developing soybean seed coats', *Planta*, vol. 215, no. 4, pp. 523-32.
- Behr, M., Legay, S., Hausman, J.-F. & Guerriero, G. 2015, 'Analysis of cell wall-related genes in organs of *Medicago sativa* L. under different abiotic stresses', *International journal of molecular sciences*, vol. 16, no. 7, pp. 16104-24.
- Berdasco, M., Alcázar, R., García-Ortiz, M.V., Ballestar, E., Fernández, A.F., Roldán-Arjona, T., Tiburcio, A.F., Altabella, T., Buisine, N. & Quesneville, H. 2008, 'Promoter DNA hypermethylation and gene repression in undifferentiated *Arabidopsis* cells', *PloS one*, vol. 3, no. 10.
- Bewick, A.J. & Schmitz, R.J. 2017, 'Gene body DNA methylation in plants', *Current opinion in plant biology*, vol. 36, pp. 103-10.
- Bird, A. 2007, 'Perceptions of epigenetics', *Nature*, vol. 447, no. 7143, pp. 396-8.
- Blevins, T., Podicheti, R., Mishra, V., Marasco, M., Wang, J., Rusch, D., Tang, H. & Pikaard, C.S. 2015, 'Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*', *Elife*, vol. 4, p. e09591.
- Bottino, M.C., Rosario, S., Grativol, C., Thiebaut, F., Rojas, C.A., Farrineli, L., Hemerly, A.S. & Ferreira, P.C.G. 2013, 'High-throughput sequencing of small RNA transcriptome reveals salt stress regulated microRNAs in sugarcane', *PloS one*, vol. 8, no. 3, p. e59423.
- Boyko, A., Blevins, T., Yao, Y., Golubov, A., Bilichak, A., Ilnytsky, Y., Hollander, J., Meins Jr, F. & Kovalchuk, I. 2010, 'Transgenerational adaptation of *Arabidopsis* to stress requires DNA methylation and the function of Dicer-like proteins', *PLOS one*, vol. 5, no. 3, p. e9514.
- Boyko, A. & Kovalchuk, I. 2011, 'Genome instability and epigenetic modification—heritable responses to environmental stress?', *Current opinion in plant biology*, vol. 14, no. 3, pp. 260-6.

- Bureau, T.E. & Wessler, S.R. 1994, 'Mobile inverted-repeat elements of the Tourist family are associated with the genes of many cereal grasses', *Proceedings of the National Academy of Sciences*, vol. 91, no. 4, pp. 1411-5.
- Campestre, M.P., Bordenave, C.D., Origone, A.C., Menéndez, A.B., Ruiz, O.A., Rodríguez, A.A. & Maiale, S.J. 2011, 'Polyamine catabolism is involved in response to salt stress in soybean hypocotyls', *Journal of plant physiology*, vol. 168, no. 11, pp. 1234-40.
- Carbon, S., Chan, J., Kishore, R., Lee, R., Muller, H.-M., Raciti, D., Van Auken, K. & Sternberg, P. 2017, 'Expansion of the Gene Ontology knowledgebase and resources', *Nucleic acids research*, vol. 45, no. D1, pp. D331-D8.
- Castiglione, M.R., Venora, G., Ravalli, C., Gecheff, K., Stoilov, L. & Cremonini, R. 2010, 'DNA methylation pattern in a barley reconstructed karyotype with deleted ribosomal gene cluster of chromosome 6H', *Protoplasma*, vol. 242, no. 1-4, pp. 13-8.
- Castrillon-Arbelaez, P.A. & Delano-Frier, J.P. 2011, 'The sweet side of inhibition: invertase inhibitors and their importance in plant development and stress responses', *Current Enzyme Inhibition*, vol. 7, no. 3, pp. 169-77.
- Chang, W., Jääskeläinen, M., Li, S.-p. & Schulman, A.H. 2013, 'BARE retrotransposons are translated and replicated via distinct RNA pools', *PLoS One*, vol. 8, no. 8.
- Chen, X. 2012, 'Small RNAs in development—insights from plants', *Current opinion in genetics & development*, vol. 22, no. 4, pp. 361-7.
- Chen, X., Xia, J., Xia, Z., Zhang, H., Zeng, C., Lu, C., Zhang, W. & Wang, W. 2015, 'Potential functions of microRNAs in starch metabolism and development revealed by miRNA transcriptome profiling of cassava cultivars and their wild progenitor', *BMC plant biology*, vol. 15, no. 1, p. 33.
- Chinnusamy, V. & Zhu, J.-K. 2009, 'Epigenetic regulation of stress responses in plants', *Current opinion in plant biology*, vol. 12, no. 2, pp. 133-9.
- Choudhury, F.K., Rivero, R.M., Blumwald, E. & Mittler, R. 2017, 'Reactive oxygen species, abiotic stress and stress combination', *The Plant Journal*, vol. 90, no. 5, pp. 856-67.
- Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P. & O'Toole, P.W. 2010, 'Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions', *Nucleic acids research*, vol. 38, no. 22, pp. e200-e.
- Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S. & Bayley, H. 2009, 'Continuous base identification for single-molecule nanopore DNA sequencing', *Nature nanotechnology*, vol. 4, no. 4, p. 265.
- Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M. & Jacobsen, S.E. 2008, 'Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning', *Nature*, vol. 452, no. 7184, pp. 215-9.

- Dabney, A., Storey, J.D. & Warnes, G. 2010, 'qvalue: Q-value estimation for false discovery rate control', *R package version*, vol. 1, no. 0.
- Danchin, É., Charmantier, A., Champagne, F.A., Mesoudi, A., Pujol, B. & Blanchet, S. 2011, 'Beyond DNA: integrating inclusive inheritance into an extended theory of evolution', *Nature Reviews Genetics*, vol. 12, no. 7, pp. 475-86.
- Demirkiran, A., Marakli, S., Temel, A. & Gozukirmizi, N. 2013, 'Genetic and epigenetic effects of salinity on in vitro growth of barley', *Genetics and molecular biology*, vol. 36, no. 4, pp. 566-70.
- Dey, S. & Vlot, A.C. 2015, 'Ethylene responsive factors in the orchestration of stress responses in monocotyledonous plants', *Frontiers in Plant Science*, vol. 6, p. 640.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T.R. 2013, 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, vol. 29, no. 1, pp. 15-21.
- Dong, Z., Wang, Y., Zhang, Z., Shen, Y., Lin, X., Ou, X., Han, F. & Liu, B. 2006, 'Extent and pattern of DNA methylation alteration in rice lines derived from introgressive hybridization of rice and *Zizania latifolia* Griseb', *Theoretical and Applied Genetics*, vol. 113, no. 2, pp. 196-205.
- Doniger, S.W., Salomonis, N., Dahlquist, K.D., Vranizan, K., Lawlor, S.C. & Conklin, B.R. 2003, 'MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data', *Genome biology*, vol. 4, no. 1, p. R7.
- Downen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Downen, J.M., Nery, J.R., Dixon, J.E. & Ecker, J.R. 2012, 'Widespread dynamic DNA methylation in response to biotic stress', *Proceedings of the National Academy of Sciences*, vol. 109, no. 32, pp. E2183-E91.
- Down, T.A., Rakyán, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Graef, S., Johnson, N., Herrero, J. & Tomazou, E.M. 2008, 'A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis', *Nature biotechnology*, vol. 26, no. 7, pp. 779-85.
- Dubin, M.J., Zhang, P., Meng, D., Remigereau, M.-S., Osborne, E.J., Casale, F.P., Drewe, P., Kahles, A., Jean, G. & Vilhjálmsson, B. 2015, 'DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation', *elife*, vol. 4, p. e05255.
- Elhamamsy, A.R. 2016, 'DNA methylation dynamics in plants and mammals: overview of regulation and dysregulation', *Cell biochemistry and function*, vol. 34, no. 5, pp. 289-98.
- Faghani, E., Gharechahi, J., Komatsu, S., Mirzaei, M., Khavarinejad, R.A., Najafi, F., Farsad, L.K. & Salekdeh, G.H. 2015, 'Comparative physiology and proteomic analysis of two wheat genotypes contrasting in drought tolerance', *Journal of proteomics*, vol. 114, pp. 1-15.
- Fagundes, D., Bohn, B., Cabreira, C., Leipelt, F., Dias, N., Bodanese-Zanettini, M.H. & Cagliari, A. 2015, 'Caspases in plants: metacaspase gene family in plant stress responses', *Functional & integrative genomics*, vol. 15, no. 6, pp. 639-49.

- Fattash, I., Voß, B., Reski, R., Hess, W.R. & Frank, W. 2007, 'Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution', *BMC Plant Biology*, vol. 7, no. 1, p. 13.
- Feng, S. & Jacobsen, S.E. 2011, 'Epigenetic modifications in plants: an evolutionary perspective', *Current opinion in plant biology*, vol. 14, no. 2, pp. 179-86.
- Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J. & Turner, S.W. 2010, 'Direct detection of DNA methylation during single-molecule, real-time sequencing', *Nature methods*, vol. 7, no. 6, p. 461.
- Fu, R., Zhang, M., Zhao, Y., He, X., Ding, C., Wang, S., Feng, Y., Song, X., Li, P. & Wang, B. 2017, 'Identification of salt tolerance-related microRNAs and their targets in maize (*Zea mays* L.) using high-throughput sequencing and degradome analysis', *Frontiers in plant science*, vol. 8, p. 864.
- Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K. & Shinozaki, K. 2006, 'Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks', *Current opinion in plant biology*, vol. 9, no. 4, pp. 436-42.
- Fukushima, E., Arata, Y., Endo, T., Sonnewald, U. & Sato, F. 2001, 'Improved salt tolerance of transgenic tobacco expressing apoplastic yeast-derived invertase', *Plant and Cell Physiology*, vol. 42, no. 2, pp. 245-9.
- Fultz, D., Choudury, S.G. & Slotkin, R.K. 2015, 'Silencing of active transposable elements in plants', *Current opinion in plant biology*, vol. 27, pp. 67-76.
- Galindo-González, L., Mhiri, C., Grandbastien, M.-A. & Deyholos, M.K. 2016, 'Ty1-copia elements reveal diverse insertion sites linked to polymorphisms among flax (*Linum usitatissimum* L.) accessions', *BMC genomics*, vol. 17, no. 1, p. 1002.
- Garg, R., Chevala, V.N., Shankar, R. & Jain, M. 2015, 'Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response', *Scientific reports*, vol. 5, p. 14922.
- Ge, Y., Lai, Q., Luo, P., Liu, X. & Chen, W. 2019, 'Transcriptome profiling of *Gerbera hybrida* reveals that stem bending is caused by water stress and regulation of abscisic acid', *BMC genomics*, vol. 20, no. 1, p. 600.
- Gehring, M. & Henikoff, S. 2007, 'DNA methylation dynamics in plant genomes', *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, vol. 1769, no. 5-6, pp. 276-86.
- German, M.A., Luo, S., Schroth, G., Meyers, B.C. & Green, P.J. 2009, 'Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome', *Nature protocols*, vol. 4, no. 3, p. 356.
- Ghosh, R., Choi, B.S., Jeong, M.-J., Bae, D.W., Shin, S.C., Park, S.U., Lim, H.-S., Kim, J. & Bae, H. 2012, 'Comparative transcriptional analysis of caffeoyl-coenzyme A 3-O-methyltransferase from '*Hibiscus cannabinus*' L., during developmental stages in various tissues and stress regulation', *Plant Omics*, vol. 5, no. 2, p. 184.

- Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M. & Toulmin, C. 2010, 'Food security: the challenge of feeding 9 billion people', *science*, vol. 327, no. 5967, pp. 812-8.
- Grandbastien, M.-A. 1998, 'Activation of plant retrotransposons under stress conditions', *Trends in plant science*, vol. 3, no. 5, pp. 181-7.
- Griffiths-Jones, S. 2006, 'miRBase: the microRNA sequence database', *MicroRNA Protocols*, Springer, pp. 129-38.
- Groszmann, M., Greaves, I.K., Albertyn, Z.I., Scofield, G.N., Peacock, W.J. & Dennis, E.S. 2011, 'Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor', *Proceedings of the National Academy of Sciences*, vol. 108, no. 6, pp. 2617-22.
- Grudkowska, M. & Zagdańska, B. 2004, 'Multifunctional role of plant cysteine proteinases', *Acta Biochimica Polonica*, vol. 51, no. 3, pp. 609-24.
- Guldborg, P., Worm, J. & Grønbaek, K. 2002, 'Profiling DNA methylation by melting analysis', *Methods*, vol. 27, no. 2, pp. 121-7.
- Han, J., Zhang, W., Sun, L., Cui, F., Ji, J. & Li, J. 2016, 'Aquaporin TaPIP1A mediates salt and drought tolerance via interaction with TaRD28 in Wheat and Arabidopsis', *Chinese Academy of Sciences*, vol. 2016.
- Harkess, A., Zhou, J., Xu, C., Bowers, J.E., Van der Hulst, R., Ayyampalayam, S., Mercati, F., Riccardi, P., McKain, M.R. & Kakrana, A. 2017, 'The asparagus genome sheds light on the origin and evolution of a young Y chromosome', *Nature communications*, vol. 8, no. 1, p. 1279.
- He, X.-J., Chen, T. & Zhu, J.-K. 2011, 'Regulation and function of DNA methylation in plants and animals', *Cell research*, vol. 21, no. 3, pp. 442-65.
- Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J. & Jacobsen, S.E. 2006, 'Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning', *Nature genetics*, vol. 38, no. 6, pp. 721-5.
- Hill, C.B., Cassin, A., Keeble-Gagnère, G., Doblin, M.S., Bacic, A. & Roessner, U. 2016, 'De novo transcriptome assembly and analysis of differentially expressed genes of two barley genotypes reveal root-zone-specific responses to salt exposure', *Scientific reports*, vol. 6, p. 31558.
- Hirsch, C.D. & Springer, N.M. 2017, 'Transposable element influences on gene expression in plants', *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1860, no. 1, pp. 157-65.
- Holmes, E.E., Jung, M., Meller, S., Leisse, A., Sailer, V., Zech, J., Mengdehl, M., Garbe, L.-A., Uhl, B. & Kristiansen, G. 2014, 'Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine', *PloS one*, vol. 9, no. 4, p. e93933.
- Horváth, V., Merenciano, M. & González, J. 2017, 'Revisiting the relationship between transposable elements and the eukaryotic stress response', *Trends in Genetics*, vol. 33, no. 11, pp. 832-41.

- Howkit, A., Teyssier, E., Deleuze, J.-F. & Gallusci, P. 2017, 'Locus-Specific DNA Methylation Analysis and Applications to Plants', *Plant Epigenetics*, Springer, pp. 303-27.
- Huntzinger, E. & Izaurralde, E. 2011, 'Gene silencing by microRNAs: contributions of translational repression and mRNA decay', *Nature Reviews Genetics*, vol. 12, no. 2, p. 99.
- IBGSC 2012, 'A physical, genetic and functional sequence assembly of the barley genome', *Nature*, vol. 491, no. 7426, p. 711.
- Jääskeläinen, M., Chang, W., Moisy, C. & Schulman, A.H. 2013, 'Retrotransposon BARE displays strong tissue-specific differences in expression', *New Phytologist*, vol. 200, no. 4, pp. 1000-8.
- Jain, M., Olsen, H.E., Paten, B. & Akeson, M. 2016, 'The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community', *Genome biology*, vol. 17, no. 1, p. 239.
- Ji, L. & Chen, X. 2012, 'Regulation of small RNA stability: methylation and beyond', *Cell research*, vol. 22, no. 4, pp. 624-36.
- Jia, X., Yan, J. & Tang, G. 2011, 'MicroRNA-mediated DNA methylation in plants', *Frontiers in Biology*, vol. 6, no. 2, pp. 133-9.
- Jin-long, G., Li-ping, X., Jing-ping, F., Ya-chun, S., Hua-ying, F., You-xiong, Q. & Jing-sheng, X. 2012, 'A novel dirigent protein gene with highly stem-specific expression from sugarcane, response to drought, salt and oxidative stresses', *Plant cell reports*, vol. 31, no. 10, pp. 1801-12.
- Jisha, K., Vijayakumari, K. & Puthur, J.T. 2013, 'Seed priming for abiotic stress tolerance: an overview', *Acta Physiologiae Plantarum*, vol. 35, no. 5, pp. 1381-96.
- Joel, A.J. 2013, 'Epigenetic responses to drought stress in rice (*Oryza sativa* L.)', *Physiology and molecular biology of plants*, vol. 19, no. 3, pp. 379-87.
- Jones-Rhoades, M.W., Bartel, D.P. & Bartel, B. 2006, 'MicroRNAs and their regulatory roles in plants', *Annu. Rev. Plant Biol.*, vol. 57, pp. 19-53.
- Kai-chao, W., Li-ping, W., Lin, X., Shi-yun, T., Yuan-wen, W., Cheng-mei, H., Hui-qing, C., Hai-bin, L., Sheng-li, J. & Zhi-nian, D. 2018, 'Differentially expressed gene analysis by cDNA-SCoT in *Saccharum spontaneum* under drought stress', *Journal of Southern Agriculture*, vol. 49, no. 2.
- Kaur, S., Gupta, A.K. & Kaur, N. 2002, 'Effect of osmo- and hydropriming of chickpea seeds on seedling growth and carbohydrate metabolism under water deficit stress', *Plant growth regulation*, vol. 37, no. 1, pp. 17-22.
- Kawakatsu, T., Nery, J.R., Castanon, R. & Ecker, J.R. 2017, 'Dynamic DNA methylation reconfiguration during seed development and germination', *Genome biology*, vol. 18, no. 1, p. 171.
- Kersey, P.J., Allen, J.E., Allot, A., Barba, M., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P. & Grabmueller, C. 2017, 'Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species', *Nucleic acids research*, vol. 46, no. D1, pp. D802-D8.

- Khatabi, B., Arikat, S., Xia, R., Winter, S., Oumar, D., Mongomake, K., Meyers, B.C. & Fondong, V.N. 2016, 'High-resolution identification and abundance profiling of cassava (*Manihot esculenta* Crantz) microRNAs', *BMC genomics*, vol. 17, no. 1, p. 85.
- Khraiwesh, B., Zhu, J.-K. & Zhu, J. 2012, 'Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants', *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1819, no. 2, pp. 137-48.
- Kim, J.-M., Sasaki, T., Ueda, M., Sako, K. & Seki, M. 2015, 'Chromatin changes in response to drought, salinity, heat, and cold stresses in plants', *Frontiers in plant science*, vol. 6.
- Kim, N.-S. 2017, 'The genomes and transposable elements in plants: are they friends or foes?', *Genes & Genomics*, vol. 39, no. 4, pp. 359-70.
- Kircher, M., Stenzel, U. & Kelso, J. 2009, 'Improved base calling for the Illumina Genome Analyzer using machine learning strategies', *Genome biology*, vol. 10, no. 8, p. R83.
- Kitimu, S.R., Taylor, J., March, T.J., Tairo, F., Wilkinson, M.J. & Rodríguez López, C.M. 2015, 'Meristem micropropagation of cassava (*Manihot esculenta*) evokes genome-wide changes in DNA methylation', *Frontiers in Plant Science*, vol. 6, p. 590.
- Konate, M., Wilkinson, M.J., Mayne, B.T., Pederson, S.M., Scott, E.S., Berger, B. & Rodríguez Lopez, C.M. 2018, 'Salt stress induces non-CG methylation in coding regions of barley seedlings (*Hordeum vulgare*)', *Epigenomes*, vol. 2, no. 2, p. 12.
- Konate, M., Wilkinson, M.J., Mayne, B.T., Scott, E.S., Berger, B. & Lopez, C.M.R. 2020, 'Atlas of Age-and Tissue-specific DNA Methylation during Early Development of Barley (*Hordeum vulgare*)', *DNA Methylation Mechanism*, IntechOpen.
- Koo, A.J., Gao, X., Daniel Jones, A. & Howe, G.A. 2009, 'A rapid wound signal activates the systemic synthesis of bioactive jasmonates in *Arabidopsis*', *The Plant Journal*, vol. 59, no. 6, pp. 974-86.
- Kozomara, A. & Griffiths-Jones, S. 2013, 'miRBase: annotating high confidence microRNAs using deep sequencing data', *Nucleic acids research*, vol. 42, no. D1, pp. D68-D73.
- Krueger, F. & Andrews, S.R. 2011, 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications', *bioinformatics*, vol. 27, no. 11, pp. 1571-2.
- Kurdyukov, S. & Bullock, M. 2016, 'DNA methylation analysis: choosing the right method', *Biology*, vol. 5, no. 1, p. 3.
- Kurihara, Y., Matsui, A., Kawashima, M., Kaminuma, E., Ishida, J., Morosawa, T., Mochizuki, Y., Kobayashi, N., Toyoda, T. & Shinozaki, K. 2008, 'Identification of the candidate genes regulated by RNA-directed DNA methylation in *Arabidopsis*', *Biochemical and biophysical research communications*, vol. 376, no. 3, pp. 553-7.

- Langmead, B. & Salzberg, S.L. 2012, 'Fast gapped-read alignment with Bowtie 2', *Nature methods*, vol. 9, no. 4, p. 357.
- Lannoo, N. & Van Damme, E.J. 2010, 'Nucleocytoplasmic plant lectins', *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1800, no. 2, pp. 190-201.
- Law, J.A. & Jacobsen, S.E. 2010, 'Establishing, maintaining and modifying DNA methylation patterns in plants and animals', *Nature Reviews Genetics*, vol. 11, no. 3, pp. 204-20.
- Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T. & Carey, V.J. 2013, 'Software for computing and annotating genomic ranges', *PLoS computational biology*, vol. 9, no. 8.
- Le, T.-N., Schumann, U., Smith, N.A., Tiwari, S., Au, P.C.K., Zhu, Q.-H., Taylor, J.M., Kazan, K., Llewellyn, D.J. & Zhang, R. 2014, 'DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis', *Genome biology*, vol. 15, no. 9, p. 458.
- Lee, S., Lee, K., Kim, K., Choi, G.J., Yoon, S.H., Ji, H.C., Seo, S., Lim, Y.C. & Ahsan, N. 2009, 'Identification of salt-stress induced differentially expressed genes in barley leaves using the annealing-control-primer-based GeneFishing technique', *African Journal of Biotechnology*, vol. 8, no. 7.
- Lee, T.G., Jang, C.S., Kim, J.Y., Kim, D.S., Park, J.H., Kim, D.Y. & Seo, Y.W. 2007, 'A Myb transcription factor (TaMyb1) from wheat roots is expressed during hypoxia: roles in response to the oxygen concentration in root environment and abiotic stresses', *Physiologia Plantarum*, vol. 129, no. 2, pp. 375-85.
- Li, J., Yang, Z., Yu, B., Liu, J. & Chen, X. 2005, 'Methylation Protects miRNAs and siRNAs from a 3'-End Uridylation Activity in Arabidopsis', *Current biology*, vol. 15, no. 16, pp. 1501-7.
- Li, T., Ma, L., Geng, Y., Hao, C., Chen, X. & Zhang, X. 2015, 'Small RNA and degradome sequencing reveal complex roles of miRNAs and their targets in developing wheat grains', *PLoS One*, vol. 10, no. 10, p. e0139658.
- Li, X., Zhu, J., Hu, F., Ge, S., Ye, M., Xiang, H., Zhang, G., Zheng, X., Zhang, H. & Zhang, S. 2012, 'Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression', *BMC genomics*, vol. 13, no. 1, p. 300.
- Li, Y., Kumar, S. & Qian, W. 2018, 'Active DNA demethylation: mechanism and role in plant development', *Plant cell reports*, vol. 37, no. 1, pp. 77-85.
- Li, Y., Shan, X., Liu, X., Hu, L., Guo, W. & Liu, B. 2008, 'Utility of the methylation-sensitive amplified polymorphism (MSAP) marker for detection of DNA methylation polymorphism and epigenetic population structure in a wild barley species (*Hordeum brevisubulatum*)', *Ecological research*, vol. 23, no. 5, pp. 927-30.
- Lisch, D. 2009, 'Epigenetic regulation of transposable elements in plants', *Annual review of plant biology*, vol. 60, pp. 43-66.
- Lister, R. & Ecker, J.R. 2009, 'Finding the fifth base: genome-wide sequencing of cytosine methylation', *Genome research*, vol. 19, no. 6, pp. 959-66.

- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H. & Ecker, J.R. 2008, 'Highly integrated single-base resolution maps of the epigenome in Arabidopsis', *Cell*, vol. 133, no. 3, pp. 523-36.
- Liu, H., Ma, Y., Chen, N., Guo, S., Liu, H., Guo, X., Chong, K. & Xu, Y. 2014, 'Overexpression of stress-inducible OsBURP16, the β subunit of polygalacturonase 1, decreases pectin content and cell adhesion and increases abiotic stress sensitivity in rice', *Plant, cell & environment*, vol. 37, no. 5, pp. 1144-58.
- Liu, L., Hoogenboom, G. & Ingram, K. 2000, 'Controlled-environment sunlit plant growth chambers', *Critical reviews in plant sciences*, vol. 19, no. 4, pp. 347-75.
- Liu, W.-w., Meng, J., Cui, J. & Luan, Y.-s. 2017, 'Characterization and Function of MicroRNA* s in Plants', *Frontiers in plant science*, vol. 8, p. 2200.
- Liu, Y., Cheng, J., Siejka-Zielińska, P., Weldon, C., Roberts, H., Lopopolo, M., Magri, A., D'Arienzo, V., Harris, J.M. & McKeating, J.A. 2020, 'Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS', *Genome Biology*, vol. 21, no. 1, pp. 1-9.
- Lizardi, P.M., Yan, Q. & Wajapeyee, N. 2017, 'DNA bisulfite sequencing for single-nucleotide-resolution DNA methylation detection', *Cold Spring Harbor Protocols*, vol. 2017, no. 11, p. pdb. prot094839.
- Long, N.V., Dolstra, O., Malosetti, M., Kilian, B., Graner, A., Visser, R.G. & van der Linden, C.G. 2013, 'Association mapping of salt tolerance in barley (*Hordeum vulgare* L.)', *Theoretical and applied genetics*, vol. 126, no. 9, pp. 2335-51.
- Lou, S., Lee, H.-M., Qin, H., Li, J.-W., Gao, Z., Liu, X., Chan, L.L., Lam, V.K., So, W.-Y. & Wang, Y. 2014, 'Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation', *Genome biology*, vol. 15, no. 7, pp. 1-21.
- Love, M.I., Huber, W. & Anders, S. 2014, 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome biology*, vol. 15, no. 12, p. 550.
- Lu, S., Sun, Y.-H., Shi, R., Clark, C., Li, L. & Chiang, V.L. 2005, 'Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*', *The Plant Cell*, vol. 17, no. 8, pp. 2186-203.
- Lu, X.-Y. & Huang, X.-L. 2008, 'Plant miRNAs and abiotic stress responses', *Biochemical and biophysical research communications*, vol. 368, no. 3, pp. 458-62.
- Luan, S., Kudla, J., Rodriguez-Concepcion, M., Yalovsky, S. & Gruissem, W. 2002, 'Calmodulins and calcineurin B-like proteins: calcium sensors for specific signal response coupling in plants', *The Plant Cell*, vol. 14, no. suppl 1, pp. S389-S400.
- Lv, S., Nie, X., Wang, L., Du, X., Biradar, S.S., Jia, X. & Weining, S. 2012, 'Identification and characterization of MicroRNAs from barley (*Hordeum vulgare* L.) by high-throughput sequencing', *International journal of molecular sciences*, vol. 13, no. 3, pp. 2973-84.

- Ma, B., Xin, Y., Kuang, L. & He, N. 2019, 'Distribution and Characteristics of Transposable Elements in the Mulberry Genome', *The plant genome*, vol. 12, no. 2.
- Makałowski, W., Gotea, V., Pande, A. & Makałowska, I. 2019, 'Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics', *Evolutionary Genomics*, Springer, pp. 177-207.
- Makarevitch, I., Waters, A.J., West, P.T., Stitzer, M., Hirsch, C.N., Ross-Ibarra, J. & Springer, N.M. 2015, 'Transposable elements contribute to activation of maize genes in response to abiotic stress', *PLoS genetics*, vol. 11, no. 1.
- Mallory, A.C. & Vaucheret, H. 2006, 'Functions of microRNAs and related small RNAs in plants', *Nature genetics*, vol. 38, no. 6s, p. S31.
- Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Tran, L.-S.P. & Qin, F. 2015, 'A transposable element in a NAC gene is associated with drought tolerance in maize seedlings', *Nature Communications*, vol. 6, no. 1, pp. 1-13.
- Marakli, S. & Gozukirmizi, N. 2018, 'Analyses of abiotic stress and brassinosteroid-related some genes in barley roots grown under salinity stress and HBR treatments: Expression profiles and phylogeny', *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*, vol. 152, no. 3, pp. 324-32.
- Martienssen, R.A. & Colot, V. 2001, 'DNA methylation and epigenetic inheritance in plants and filamentous fungi', *Science*, vol. 293, no. 5532, pp. 1070-4.
- Martin, M. 2011, 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet. journal*, vol. 17, no. 1, pp. 10-2.
- Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E. & Russell, J. 2017, 'A chromosome conformation capture ordered sequence of the barley genome', *Nature*, vol. 544, no. 7651, p. 427.
- Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., Ayling, S., Steuernagel, B., Pfeifer, M. & D'Ascenzo, M. 2013, 'Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond', *The Plant Journal*, vol. 76, no. 3, pp. 494-505.
- Matzke, M., Kanno, T., Huettel, B., Daxinger, L. & Matzke, A.J. 2007, 'Targets of RNA-directed DNA methylation', *Current opinion in plant biology*, vol. 10, no. 5, pp. 512-9.
- Matzke, M.A. & Mosher, R.A. 2014, 'RNA-directed DNA methylation: an epigenetic pathway of increasing complexity', *Nature Reviews Genetics*, vol. 15, no. 6, pp. 394-408.
- Mayer, K.F., Martis, M., Hedley, P.E., Šimková, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S. & Gundlach, H. 2011, 'Unlocking the barley genome by chromosomal and comparative genomics', *The Plant Cell*, vol. 23, no. 4, pp. 1249-63.

- McCarthy, D.J., Chen, Y. & Smyth, G.K. 2012, 'Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation', *Nucleic acids research*, vol. 40, no. 10, pp. 4288-97.
- McCarthy, D.J. & Smyth, G.K. 2009, 'Testing significance relative to a fold-change threshold is a TREAT', *Bioinformatics*, vol. 25, no. 6, pp. 765-71.
- McCue, A.D., Nuthikattu, S., Reeder, S.H. & Slotkin, R.K. 2012, 'Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA', *PLoS genetics*, vol. 8, no. 2, p. e1002474.
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S. & Jaenisch, R. 2005, 'Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis', *Nucleic acids research*, vol. 33, no. 18, pp. 5868-77.
- Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X. & Thomas, P.D. 2019, 'Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v. 14.0)', *Nature protocols*, vol. 14, no. 3, p. 703.
- Mica, E., Piccolo, V., Delledonne, M., Ferrarini, A., Pezzotti, M., Casati, C., Del Fabbro, C., Valle, G., Policriti, A. & Morgante, M. 2009, 'High throughput approaches reveal splicing of primary microRNA transcripts and tissue specific expression of mature microRNAs in *Vitis vinifera*', *BMC genomics*, vol. 10, no. 1, p. 558.
- Middleton, C.P., Senerchia, N., Stein, N., Akhunov, E.D., Keller, B., Wicker, T. & Kilian, B. 2014, 'Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe', *PLoS One*, vol. 9, no. 3.
- Mirouze, M. & Paszkowski, J. 2011, 'Epigenetic contribution to stress adaptation in plants', *Current opinion in plant biology*, vol. 14, no. 3, pp. 267-74.
- Moissiard, G., Parizotto, E.A., Himber, C. & Voinnet, O. 2007, 'Transitivity in Arabidopsis can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins', *Rna*, vol. 13, no. 8, pp. 1268-78.
- Morgan, M., Pages, H., Obenchain, V. & Hayden, N. 2016, 'Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import', *R package version*, vol. 1, no. 0, pp. 677-89.
- Moro, B., Chorostecki, U., Arikat, S., Suarez, I.P., Höbartner, C., Rasia, R.M., Meyers, B.C. & Palatnik, J.F. 2018, 'Efficiency and precision of microRNA biogenesis modes in plants', *Nucleic acids research*, vol. 46, no. 20, pp. 10709-23.
- Negi, P., Rai, A.N. & Suprasanna, P. 2016, 'Moving through the stressed genome: emerging regulatory roles for transposons in plant stress response', *Frontiers in plant science*, vol. 7, p. 1448.
- Nie, W.-F., Lei, M., Zhang, M., Tang, K., Huang, H., Zhang, C., Miki, D., Liu, P., Yang, Y. & Wang, X. 2019, 'Histone acetylation recruits the SWR1 complex to regulate active DNA demethylation in Arabidopsis', *Proceedings of the National Academy of Sciences*, vol. 116, no. 33, pp. 16641-50.

- Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N. & Slotkin, R.K. 2013, 'The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs', *Plant physiology*, vol. 162, no. 1, pp. 116-31.
- Ossowski, S., Schwab, R. & Weigel, D. 2008, 'Gene silencing in plants using artificial microRNAs and other small RNAs', *The Plant Journal*, vol. 53, no. 4, pp. 674-90.
- Ozhuner, E., Eldem, V., Ipek, A., Okay, S., Sakcali, S., Zhang, B., Boke, H. & Unver, T. 2013, 'Boron stress responsive microRNAs and their targets in barley', *PLoS one*, vol. 8, no. 3, p. e59543.
- Ozturk, Z.N., Talamé, V., Deyholos, M., Michalowski, C.B., Galbraith, D.W., Gozukirmizi, N., Tuberosa, R. & Bohnert, H.J. 2002, 'Monitoring large-scale changes in transcript abundance in drought-and salt-stressed barley', *Plant molecular biology*, vol. 48, no. 5-6, pp. 551-73.
- Paniagua, C., Bilkova, A., Jackson, P., Dabravolski, S., Riber, W., Didi, V., Houser, J., Gigli-Bisceglia, N., Wimmerova, M. & Budínská, E. 2017, 'Dirigent proteins in plants: modulating cell wall metabolism during abiotic and biotic stress exposure', *Journal of experimental botany*, vol. 68, no. 13, pp. 3287-301.
- Parry, S., Newbigin, E., Currie, G., Bacic, A. & Oxley, D. 1997, 'Identification of active-site histidine residues of a self-incompatibility ribonuclease from a wild tomato', *Plant physiology*, vol. 115, no. 4, pp. 1421-9.
- Patterson, J.H., Newbigin, E., Tester, M., Bacic, A. & Roessner, U. 2009, 'Metabolic responses to salt stress of barley (*Hordeum vulgare* L.) cultivars, Sahara and Clipper, which differ in salinity tolerance', *Journal of Experimental Botany*, vol. 60, no. 14, pp. 4089-103.
- Patterson, T.A., Lobenhofer, E.K., Fulmer-Smentek, S.B., Collins, P.J., Chu, T.-M., Bao, W., Fang, H., Kawasaki, E.S., Hager, J. & Tikhonova, I.R. 2006, 'Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project', *Nature biotechnology*, vol. 24, no. 9, p. 1140.
- Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N. & Scheid, O.M. 2010, 'Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*', *The Plant Cell Online*, vol. 22, no. 9, pp. 3118-29.
- Pecinka, A., Rosa, M., Schikora, A., Berlinger, M., Hirt, H., Luschnig, C. & Scheid, O.M. 2009, 'Transgenerational stress memory is not a general response in *Arabidopsis*', *PLoS One*, vol. 4, no. 4, p. e5202.
- Pfluger, J. & Wagner, D. 2007, 'Histone modifications and dynamic regulation of genome accessibility in plants', *Current opinion in plant biology*, vol. 10, no. 6, pp. 645-52.
- Phillips, J.R., Dalmay, T. & Bartels, D. 2007, 'The role of small RNAs in abiotic stress', *FEBS letters*, vol. 581, no. 19, pp. 3592-7.
- Pitman, M.G. & Läuchli, A. 2002, 'Global impact of salinity and agricultural ecosystems', *Salinity: environment-plants-molecules*, Springer, pp. 3-20.

- Porebski, S., Bailey, L.G. & Baum, B.R. 1997, 'Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components', *Plant molecular biology reporter*, vol. 15, no. 1, pp. 8-15.
- Potvin, C., Lechowicz, M., Bell, G. & Schoen, D. 1990, 'Spatial, temporal, and species-specific patterns of heterogeneity in growth chamber experiments', *Canadian journal of botany*, vol. 68, no. 3, pp. 499-504.
- Prabu, G. & Prasad, D.T. 2012, 'Functional characterization of sugarcane MYB transcription factor gene promoter (PScMYBAS1) in response to abiotic stresses and hormones', *Plant cell reports*, vol. 31, no. 4, pp. 661-9.
- Qi, Y., He, X., Wang, X.-J., Kohany, O., Jurka, J. & Hannon, G.J. 2006, 'Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation', *Nature*, vol. 443, no. 7114, pp. 1008-12.
- Rauluseviciute, I., Drabløs, F. & Rye, M.B. 2019, 'DNA methylation data by sequencing: experimental approaches and recommendations for tools and pipelines for data analysis', *Clinical Epigenetics*, vol. 11, no. 1, pp. 1-13.
- Rebollo, R., Romanish, M.T. & Mager, D.L. 2012, 'Transposable elements: an abundant and natural source of regulatory sequences for host genes', *Annual review of genetics*, vol. 46, pp. 21-42.
- Rengasamy, P. 2002, 'Transient salinity and subsoil constraints to dryland farming in Australian sodic soils: an overview', *Australian Journal of Experimental Agriculture*, vol. 42, no. 3, pp. 351-61.
- Reyna-Lopez, G., Simpson, J. & Ruiz-Herrera, J. 1997, 'Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms', *Molecular and General Genetics MGG*, vol. 253, no. 6, pp. 703-10.
- Rhee, S.Y., Wood, V., Dolinski, K. & Draghici, S. 2008, 'Use and misuse of the gene ontology annotations', *Nature Reviews Genetics*, vol. 9, no. 7, pp. 509-15.
- Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010, 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, vol. 26, no. 1, pp. 139-40.
- Rodríguez-López, C.M., Morán, P., Lago, F., Espiñeira, M., Beckmann, M. & Consuegra, S. 2012, 'Detection and quantification of tissue of origin in salmon and veal products using methylation sensitive AFLPs', *Food chemistry*, vol. 131, no. 4, pp. 1493-8.
- Rogers, K. & Chen, X. 2013, 'Biogenesis, turnover, and mode of action of plant microRNAs', *The Plant Cell*, vol. 25, no. 7, pp. 2383-99.
- Sabot, F. & Schulman, A. 2006, 'Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome', *heredity*, vol. 97, no. 6, pp. 381-8.
- Saleh, A., Alvarez-Venegas, R. & Avramova, Z. 2008, 'An efficient chromatin immunoprecipitation (ChIP) protocol for studying histone modifications in Arabidopsis plants', *Nature Protocols*, vol. 3, no. 6, pp. 1018-25.

- Salekdeh, G.H., Siopongco, J., Wade, L.J., Ghareyazie, B. & Bennett, J. 2002, 'Proteomic analysis of rice leaves during drought stress and recovery', *PROTEOMICS: International Edition*, vol. 2, no. 9, pp. 1131-45.
- Saze, H., Scheid, O.M. & Paszkowski, J. 2003, 'Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis', *Nature genetics*, vol. 34, no. 1, pp. 65-9.
- Saze, H., Tsugane, K., Kanno, T. & Nishimura, T. 2012, 'DNA methylation in plants: relationship to small RNAs and histone modifications, and functions in transposon inactivation', *Plant and Cell Physiology*, vol. 53, no. 5, pp. 766-84.
- Schreiber, A.W., Shi, B.-J., Huang, C.-Y., Langridge, P. & Baumann, U. 2011, 'Discovery of barley miRNAs through deep sequencing of short reads', *BMC genomics*, vol. 12, no. 1, p. 129.
- Schwab, R., Ossowski, S., Riester, M., Warthmann, N. & Weigel, D. 2006, 'Highly specific gene silencing by artificial microRNAs in Arabidopsis', *The Plant Cell Online*, vol. 18, no. 5, pp. 1121-33.
- Šestáková, Š., Šálek, C. & Remešová, H. 2019, 'DNA Methylation Validation Methods: a Coherent Review with Practical Comparison', *Biological procedures online*, vol. 21, no. 1, p. 19.
- Shavrukov, Y. 2012, 'Salt stress or salt shock: which genes are we studying?', *Journal of Experimental Botany*, vol. 64, no. 1, pp. 119-27.
- Shelden, M.C., Roessner, U., Sharp, R.E., Tester, M. & Bacic, A. 2013, 'Genetic variation in the root growth response of barley genotypes to salinity stress', *Functional Plant Biology*, vol. 40, no. 5, pp. 516-30.
- Shen, Q., Fu, L., Dai, F., Jiang, L., Zhang, G. & Wu, D. 2016, 'Multi-omics analysis reveals molecular mechanisms of shoot adaption to salt stress in Tibetan wild barley', *BMC genomics*, vol. 17, no. 1, p. 889.
- Sherman, J.D. & Talbert, L.E. 2002, 'Vernalization-induced changes of the DNA methylation pattern in winter wheat', *Genome*, vol. 45, no. 2, pp. 253-60.
- Shibuya, K., Fukushima, S. & Takatsuji, H. 2009, 'RNA-directed DNA methylation induces transcriptional activation in plants', *Proceedings of the National Academy of Sciences*, vol. 106, no. 5, pp. 1660-5.
- Simmons, C.R., Litts, J.C., Huang, N. & Rodriguez, R.L. 1992, 'Structure of a rice β -glucanase gene regulated by ethylene, cytokinin, wounding, salicylic acid and fungal elicitors', *Plant molecular biology*, vol. 18, no. 1, pp. 33-45.
- Simpson, J.T., Workman, R.E., Zuzarte, P., David, M., Dursi, L. & Timp, W. 2017, 'Detecting DNA cytosine methylation using nanopore sequencing', *Nature methods*, vol. 14, no. 4, p. 407.
- Slotkin, R.K. & Martienssen, R. 2007, 'Transposable elements and the epigenetic regulation of the genome', *Nature Reviews Genetics*, vol. 8, no. 4, pp. 272-85.
- Slotkin, R.K., Nuthikattu, S. & Jiang, N. 2012, 'The impact of transposable elements on gene and genome evolution', *Plant Genome Diversity Volume 1*, Springer, pp. 35-58.

- Sobhanian, H., Razavizadeh, R., Nanjo, Y., Ehsanpour, A.A., Jazii, F.R., Motamed, N. & Komatsu, S. 2010, 'Proteome analysis of soybean leaves, hypocotyls and roots under salt stress', *Proteome science*, vol. 8, no. 1, p. 19.
- Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J. & Jacobsen, S.E. 2014, 'Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis', *Nature structural & molecular biology*, vol. 21, no. 1, p. 64.
- Sun, L.-F., Liu, T.-J., Shan, X.-H., Su, S.-Z., Li, S.-P., Yuan, Y.-P. & Zhang, J. 2015, 'Analysis of DNA cytosine methylation patterns in maize hybrids and their parents', *Biologia Plantarum*, vol. 59, no. 2, pp. 266-72.
- Sun, Y.-G., Wang, B., Jin, S.-H., Qu, X.-X., Li, Y.-J. & Hou, B.-K. 2013, 'Ectopic expression of Arabidopsis glycosyltransferase UGT85A5 enhances salt stress tolerance in tobacco', *PLoS one*, vol. 8, no. 3.
- Suzuki, M., Liao, W., Wos, F., Johnston, A.D., DeGrazia, J., Ishii, J., Bloom, T., Zody, M.C., Germer, S. & Grealley, J.M. 2018, 'Whole-genome bisulfite sequencing with improved accuracy and cost', *Genome research*, vol. 28, no. 9, pp. 1364-71.
- Tang, Y., Horikoshi, M. & Li, W. 2016, 'ggfortify: unified interface to visualize statistical results of popular R packages', *The R Journal*, vol. 8, no. 2, pp. 474-89.
- Tavakkoli, E., Rengasamy, P. & McDonald, G.K. 2010, 'The response of barley to salinity stress differs between hydroponic and soil systems', *Functional Plant Biology*, vol. 37, no. 7, pp. 621-33.
- Team, R.C. 2019, 'R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012', URL <http://r-project.org>.
- Teh, A.L., Pan, H., Lin, X., Lim, Y.I., Patro, C.P.K., Cheong, C.Y., Gong, M., MacIsaac, J.L., Kwoh, C.-K. & Meaney, M.J. 2016, 'Comparison of methyl-capture sequencing vs. Infinium 450K methylation array for methylome analysis in clinical samples', *Epigenetics*, vol. 11, no. 1, pp. 36-48.
- Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M. & Fraga, M.F. 2009, 'A role for RNAi in the selective correction of DNA methylation defects', *Science*, vol. 323, no. 5921, pp. 1600-4.
- Teotia, S., Singh, D. & Tang, G. 2017, 'DNA Methylation in Plants by microRNAs', *Plant Epigenetics*, Springer, pp. 247-62.
- Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R. & Polasky, S. 2002, 'Agricultural sustainability and intensive production practices', *Nature*, vol. 418, no. 6898, pp. 671-7.
- Tollefsbol, T. 2017, *Handbook of epigenetics: the new molecular and medical genetics*, Academic Press.
- Tricker, P.J., Gibbings, J.G., López, C.M.R., Hadley, P. & Wilkinson, M.J. 2012, 'Low relative humidity triggers RNA-directed de novo DNA methylation and suppression of genes controlling stomatal development', *Journal of experimental botany*, vol. 63, no. 10, pp. 3799-813.

- Tricker, P.J., López, C.M.R., Gibbings, G., Hadley, P. & Wilkinson, M.J. 2013, 'Transgenerational, dynamic methylation of stomata genes in response to low relative humidity', *International journal of molecular sciences*, vol. 14, no. 4, pp. 6674-89.
- Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zaviyalov, A., Lam, E. & Bozhkov, P. 2011, 'Metacaspases', *Cell Death & Differentiation*, vol. 18, no. 8, pp. 1279-88.
- van Gurp, T.P., Wagemaker, N.C., Wouters, B., Vergeer, P., Ouborg, J.N. & Verhoeven, K.J. 2016, 'epiGBS: reference-free reduced representation bisulfite sequencing', *Nature methods*, vol. 13, no. 4, pp. 322-9.
- Vanyushin, B.F. & Ashapkin, V.V. 2011, 'DNA methylation in higher plants: past, present and future', *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, vol. 1809, no. 8, pp. 360-8.
- Vasileiadis, S., Puglisi, E., Arena, M., Cappa, F., Cocconcelli, P.S. & Trevisan, M. 2012, 'Soil bacterial diversity screening using single 16S rRNA gene V regions coupled with multi-million read generating sequencing technologies', *PloS one*, vol. 7, no. 8.
- Viggiano, L. & de Pinto, M.C. 2017, 'Dynamic DNA methylation patterns in stress response', *Plant epigenetics*, Springer, pp. 281-302.
- Vinocur, B. & Altman, A. 2005, 'Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations', *Current opinion in biotechnology*, vol. 16, no. 2, pp. 123-32.
- Wagemaker, N.C. 2018, 'Re: EpiGBS troubleshooting', personal communication, 15 February 2018.
- Wang, H., Beyene, G., Zhai, J., Feng, S., Fahlgren, N., Taylor, N.J., Bart, R., Carrington, J.C., Jacobsen, S.E. & Ausin, I. 2015, 'CG gene body DNA methylation changes and evolution of duplicated genes in cassava', *Proceedings of the National Academy of Sciences*, vol. 112, no. 44, pp. 13729-34.
- Wang, J., Mei, J. & Ren, G. 2019, 'Plant microRNAs: biogenesis, homeostasis and degradation', *Frontiers in Plant Science*, vol. 10, p. 360.
- Wang, L., Sun, J., Wu, H., Liu, S., Wang, J., Wu, B., Huang, S., Li, N., Wang, J. & Zhang, X. 2012, 'Systematic assessment of reduced representation bisulfite sequencing to human blood samples: A promising method for large-sample-scale epigenomic studies', *Journal of biotechnology*, vol. 157, no. 1, pp. 1-6.
- Wei, L., Xiao, M., An, Z., Ma, B., Mason, A.S., Qian, W., Li, J. & Fu, D. 2013, 'New insights into nested long terminal repeat retrotransposons in Brassica species', *Molecular plant*, vol. 6, no. 2, pp. 470-82.
- Weining, S. & Langridge, P. 1991, 'Identification and mapping of polymorphisms in cereals based on the polymerase chain reaction', *Theoretical and applied genetics*, vol. 82, no. 2, pp. 209-16.
- Weirather, J.L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D. & Au, K.F. 2017, 'Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis', *F1000Research*, vol. 6.

- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M. & Panaud, O. 2007, 'A unified classification system for eukaryotic transposable elements', *Nature Reviews Genetics*, vol. 8, no. 12, pp. 973-82.
- Wicker, T., Schulman, A.H., Tanskanen, J., Spannagl, M., Twardziok, S., Mascher, M., Springer, N.M., Li, Q., Waugh, R. & Li, C. 2017, 'The repetitive landscape of the 5100 Mbp barley genome', *Mobile DNA*, vol. 8, no. 1, p. 22.
- Wickham, H. 2016a, *ggplot2: elegant graphics for data analysis*, Springer.
- Wickham, H. 2016b, 'plyr: Tools for Splitting, Applying and Combining Data. R package version 1.8.4'.
- Williams, B.P. & Gehring, M. 2017, 'Stable transgenerational epigenetic inheritance requires a DNA methylation-sensing circuit', *Nature communications*, vol. 8, no. 1, pp. 1-8.
- Wu, R., Wang, X., Lin, Y., Ma, Y., Liu, G., Yu, X., Zhong, S. & Liu, B. 2013, 'Interspecies grafting caused extensive and heritable alterations of DNA methylation in Solanaceae plants', *PLOS one*, vol. 8, no. 4, p. e61995.
- Xia, Z., Zou, M., Zhang, S., Feng, B. & Wang, W. 2014, 'AFSM sequencing approach: a simple and rapid method for genome-wide SNP and methylation site discovery and genetic mapping', *Scientific reports*, vol. 4, no. 1, pp. 1-8.
- Xiao, Y., Hsiao, T.-H., Suresh, U., Chen, H.-I.H., Wu, X., Wolf, S.E. & Chen, Y. 2012, 'A novel significance score for gene selection and ranking', *Bioinformatics*, vol. 30, no. 6, pp. 801-7.
- Xie, M. & Yu, B. 2015, 'siRNA-directed DNA methylation in plants', *Current genomics*, vol. 16, no. 1, pp. 23-31.
- Xin, Z., Wang, Q., Yu, Z., Hu, L., Li, J., Xiang, C., Wang, B. & Lou, Y. 2014, 'Overexpression of a xylanase inhibitor gene, OsHI-XIP, enhances resistance in rice to herbivores', *Plant molecular biology reporter*, vol. 32, no. 2, pp. 465-75.
- Yaari, R., Katz, A., Domb, K., Harris, K.D., Zemach, A. & Ohad, N. 2019, 'RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs', *Nature communications*, vol. 10, no. 1, pp. 1-10.
- Yaish, M.W., Peng, M. & Rothstein, S.J. 2014, 'Global DNA methylation analysis using methyl-sensitive amplification polymorphism (MSAP)', *Arabidopsis Protocols*, Springer, pp. 285-98.
- Yang, J.-S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D. & Lai, E.C. 2010, 'Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis', *Proceedings of the National Academy of Sciences*, vol. 107, no. 34, pp. 15163-8.
- Yang, Y., Sebra, R., Pullman, B.S., Qiao, W., Peter, I., Desnick, R.J., Geyer, C.R., DeCoteau, J.F. & Scott, S.A. 2015, 'Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS)', *BMC genomics*, vol. 16, no. 1, p. 350.

- Yao, J., Shi, W. & Xu, W. 2008, 'Effects of salt stress on expression of nitrate transporter and assimilation-related genes in tomato roots', *Russian journal of plant physiology*, vol. 55, no. 2, pp. 232-40.
- You, W., Tyczewska, A., Spencer, M., Daxinger, L., Schmid, M.W., Grossniklaus, U., Simon, S.A., Meyers, B.C., Matzke, A.J. & Matzke, M. 2012, 'Atypical DNA methylation of genes encoding cysteine-rich peptides in *Arabidopsis thaliana*', *BMC plant biology*, vol. 12, no. 1, p. 51.
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. & Chen, X. 2005, 'Methylation as a crucial step in plant microRNA biogenesis', *Science*, vol. 307, no. 5711, pp. 932-5.
- Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z. & Forney, L.J. 2012, 'Evaluation of methods for the extraction and purification of DNA from the human microbiome', *PLoS one*, vol. 7, no. 3.
- Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L. & Zilberman, D. 2013, 'The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin', *Cell*, vol. 153, no. 1, pp. 193-205.
- Zhang, H., Lang, Z. & Zhu, J.-K. 2018, 'Dynamics and function of DNA methylation in plants', *Nature Reviews Molecular Cell Biology*, vol. 19, no. 8, pp. 489-506.
- Zhang, H. & Zhu, J.-K. 2011, 'RNA-directed DNA methylation', *Current opinion in plant biology*, vol. 14, no. 2, pp. 142-7.
- Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. 2014, 'PEAR: a fast and accurate Illumina Paired-End reAd mergeR', *Bioinformatics*, vol. 30, no. 5, pp. 614-20.
- Zhang, J., Mao, Z. & Chong, K. 2013, 'A global profiling of uncapped mRNAs under cold stress reveals specific decay patterns and endonucleolytic cleavages in *Brachypodium distachyon*', *Genome biology*, vol. 14, no. 8, p. R92.
- Zhang, J., Xu, Y., Huan, Q. & Chong, K. 2009, 'Deep sequencing of *Brachypodium* small RNAs at the global genome level identifies microRNAs involved in cold stress response', *BMC genomics*, vol. 10, no. 1, p. 449.
- Zhang, X., Xia, J., Lii, Y.E., Barrera-Figueroa, B.E., Zhou, X., Gao, S., Lu, L., Niu, D., Chen, Z. & Leung, C. 2012, 'Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function', *Genome biology*, vol. 13, no. 3, p. R20.
- Zhang, Y., Linghu, J., Wang, D., Liu, X., Yu, A., Li, F., Zhao, J. & Zhao, T. 2017, 'Foxtail millet CBL4 (SiCBL4) interacts with SiCIPK24, modulates plant salt stress tolerance', *Plant Molecular Biology Reporter*, vol. 35, no. 6, pp. 634-46.
- Zhao, W., He, X., Hoadley, K.A., Parker, J.S., Hayes, D.N. & Perou, C.M. 2014, 'Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling', *BMC genomics*, vol. 15, no. 1, p. 419.
- Zhao, X., Chai, Y. & Liu, B. 2007, 'Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids', *Plant science*, vol. 172, no. 5, pp. 930-8.

- Zhong, L., Xu, Y.-h. & Wang, J.-b. 2009, 'DNA-methylation changes induced by salt stress in wheat *Triticum aestivum*', *African Journal of Biotechnology*, vol. 8, no. 22.
- Ziemann, M., Kamboj, A., Hove, R.M., Loveridge, S., El-Osta, A. & Bhave, M. 2013, 'Analysis of the barley leaf transcriptome under salinity stress using mRNA-Seq', *Acta physiologiae plantarum*, vol. 35, no. 6, pp. 1915-24.
- Ziller, M.J., Hansen, K.D., Meissner, A. & Aryee, M.J. 2015, 'Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing', *Nature methods*, vol. 12, no. 3, p. 230.
- Zipor, G. & Oren-Shamir, M. 2013, 'Do vacuolar peroxidases act as plant caretakers?', *Plant science*, vol. 199, pp. 41-7.
- Zuker, M. 2003, 'Mfold web server for nucleic acid folding and hybridization prediction', *Nucleic acids research*, vol. 31, no. 13, pp. 3406-15.
- Zymo 2020, *EZ DNA Methylation-Lightning Kit*, viewed 3 February 2020, <<https://www.zymoresearch.com/collections/ez-dna-methylation-lightning-kits/products/ez-dna-methylation-lightning-kit>>.

9 Appendix A, Supplementary figures and tables

Table A.9.1. Differentially expressed genes in salt stressed barley leaf tissue. Genes found to be differentially expressed in 100 mM salt treated plants (third leaf tissue) relative to control plants using DESeq2 with adjusted p-value < 0.01 and $|\log_2(\text{fold change})| > \log_2(1.5)$. The adjusted p-value is shown as \log_{10} as the values are small.

Gene	$\log_2(\text{fold change})$	$\log_{10}(\text{adj-p})$	Gene	$\log_2(\text{fold change})$	$\log_{10}(\text{adj-p})$
HORVU1Hr1G093780	5.96	-102.3	HORVU3Hr1G022800	1.91	-7.5
HORVU7Hr1G090410	3.94	-32.4	HORVU7Hr1G086690	1.91	-12.1
HORVU1Hr1G093660	3.83	-30.4	HORVU5Hr1G125620	1.91	-15.4
HORVU3Hr1G069650	3.30	-53.3	HORVU6Hr1G073500	1.90	-7.5
HORVU2Hr1G043890	3.19	-48.9	HORVU2Hr1G030870	1.89	-6.6
HORVU6Hr1G084070	3.12	-20.2	HORVU2Hr1G074770	1.89	-9.7
HORVU7Hr1G008260	3.01	-23.7	HORVU7Hr1G111010	1.88	-6.5
HORVU2Hr1G103150	2.96	-20.1	HORVU2Hr1G001160	1.86	-7.1
HORVU7Hr1G073640	2.92	-40.6	HORVU7Hr1G051560	1.85	-6.1
HORVU7Hr1G120060	2.90	-54.4	HORVU2Hr1G079180	1.85	-14.0
HORVU7Hr1G021660	2.87	-15.4	HORVU2Hr1G081920	1.83	-6.7
HORVU5Hr1G010860	2.85	-18.9	HORVU5Hr1G030830	1.83	-6.0
HORVU1Hr1G093570	2.79	-14.4	HORVU2Hr1G012360	1.81	-7.7
HORVU4Hr1G084390	2.74	-69.4	HORVU2Hr1G013420	1.80	-7.5
HORVU2Hr1G042240	2.72	-14.0	HORVU0Hr1G005360	1.80	-10.0
HORVU4Hr1G076050	2.70	-23.0	HORVU5Hr1G089190	1.79	-7.4
HORVU4Hr1G074840	2.63	-30.0	HORVU7Hr1G050160	1.79	-11.9
HORVU1Hr1G088840	2.57	-13.0	HORVU4Hr1G087250	1.77	-5.7
HORVU5Hr1G103460	2.56	-14.5	HORVU5Hr1G001180	1.77	-17.5
HORVU7Hr1G098260	2.55	-28.4	HORVU5Hr1G022500	1.77	-5.8
HORVU6Hr1G090280	2.54	-28.3	HORVU4Hr1G010160	1.76	-8.6
HORVU3Hr1G085760	2.51	-11.8	HORVU5Hr1G025760	1.76	-10.0
HORVU6Hr1G079190	2.49	-12.0	HORVU5Hr1G095080	1.76	-7.5
HORVU5Hr1G077880	2.44	-11.6	HORVU1Hr1G088270	1.76	-19.9
HORVU1Hr1G002090	2.42	-15.4	HORVU7Hr1G001040	1.75	-17.3
HORVU5Hr1G073010	2.41	-15.8	HORVU1Hr1G070690	1.75	-12.1
HORVU5Hr1G077810	2.39	-12.4	HORVU5Hr1G068110	1.75	-6.6
HORVU3Hr1G009490	2.38	-14.0	HORVU1Hr1G085050	1.75	-6.1
HORVU4Hr1G076000	2.38	-10.4	HORVU3Hr1G002840	1.75	-6.2
HORVU1Hr1G087570	2.36	-19.2	HORVU3Hr1G015850	1.74	-23.5
HORVU7Hr1G122690	2.34	-10.1	HORVU5Hr1G106090	1.74	-10.8
HORVU3Hr1G108670	2.32	-9.8	HORVU3Hr1G116470	1.73	-6.1
HORVU5Hr1G077140	2.30	-65.8	HORVU1Hr1G056090	1.73	-6.5
HORVU3Hr1G004140	2.28	-9.8	HORVU3Hr1G110320	1.73	-5.3
HORVU6Hr1G088130	2.27	-9.3	HORVU4Hr1G072960	1.71	-5.4
HORVU7Hr1G028920	2.27	-16.4	HORVU4Hr1G063430	1.69	-5.4
HORVU7Hr1G098280	2.24	-27.0	HORVU2Hr1G071480	1.69	-5.0
HORVU7Hr1G098110	2.24	-10.0	HORVU5Hr1G125000	1.69	-5.2
HORVU5Hr1G019110	2.20	-18.1	HORVU3Hr1G087420	1.68	-5.1
HORVU1Hr1G054240	2.19	-16.9	HORVU3Hr1G110330	1.67	-12.4
HORVU4Hr1G052450	2.17	-13.0	HORVU1Hr1G087530	1.67	-6.8
HORVU2Hr1G073680	2.17	-8.6	HORVU3Hr1G019140	1.67	-5.8
HORVU5Hr1G063940	2.12	-65.7	HORVU0Hr1G031210	1.66	-5.3
HORVU2Hr1G042370	2.09	-27.8	HORVU7Hr1G098170	1.66	-11.2
HORVU1Hr1G022020	2.07	-10.3	HORVU6Hr1G059500	1.66	-4.9
HORVU1Hr1G016980	2.07	-9.3	HORVU7Hr1G038980	1.66	-7.7
HORVU6Hr1G034990	2.06	-12.7	HORVU3Hr1G081570	1.65	-4.8
HORVU2Hr1G000430	2.03	-7.3	HORVU4Hr1G066900	1.65	-6.6
HORVU3Hr1G077930	2.03	-11.7	HORVU5Hr1G082160	1.64	-5.9
HORVU5Hr1G021230	2.02	-11.3	HORVU4Hr1G076040	1.64	-4.7
HORVU1Hr1G005920	2.01	-7.3	HORVU3Hr1G078940	1.63	-60.9
HORVU1Hr1G001710	2.01	-7.2	HORVU6Hr1G084860	1.62	-4.7
HORVU6Hr1G000830	2.00	-7.1	HORVU5Hr1G098770	1.61	-10.3
HORVU1Hr1G039820	2.00	-21.1	HORVU7Hr1G083090	1.61	-4.8
HORVU3Hr1G009560	1.99	-32.4	HORVU7Hr1G038200	1.61	-5.8
HORVU3Hr1G087470	1.99	-16.7	HORVU4Hr1G066230	1.61	-9.5
HORVU1Hr1G020140	1.98	-9.5	HORVU6Hr1G069340	1.60	-6.1
HORVU1Hr1G079130	1.96	-11.0	HORVU5Hr1G065370	1.60	-6.4
HORVU7Hr1G011810	1.95	-6.9	HORVU2Hr1G000090	1.60	-4.6
HORVU4Hr1G004820	1.93	-6.6	HORVU1Hr1G072250	1.60	-6.5
HORVU5Hr1G094460	1.92	-7.2	HORVU3Hr1G058610	1.59	-12.5
HORVU2Hr1G115960	1.92	-16.1	HORVU3Hr1G058810	1.59	-4.9

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU2Hr1G108290	1.59	-19.3	HORVU3Hr1G067380	1.38	-8.4
HORVU3Hr1G053210	1.59	-5.0	HORVU7Hr1G028910	1.38	-19.4
HORVU3Hr1G059170	1.59	-4.7	HORVU6Hr1G067670	1.38	-8.0
HORVU7Hr1G000260	1.58	-4.6	HORVU3Hr1G074350	1.38	-4.6
HORVU2Hr1G085570	1.58	-12.0	HORVU1Hr1G087320	1.38	-6.5
HORVU1Hr1G039250	1.58	-6.4	HORVU2Hr1G035020	1.38	-3.3
HORVU0Hr1G025800	1.58	-4.4	HORVU1Hr1G051740	1.38	-3.4
HORVU6Hr1G005260	1.58	-5.2	HORVU3Hr1G108680	1.37	-3.3
HORVU5Hr1G082460	1.57	-11.7	HORVU4Hr1G022950	1.37	-9.9
HORVU1Hr1G094010	1.57	-7.0	HORVU2Hr1G099470	1.37	-3.6
HORVU6Hr1G062820	1.57	-5.1	HORVU4Hr1G011160	1.36	-7.4
HORVU4Hr1G085640	1.57	-5.2	HORVU7Hr1G010230	1.36	-6.4
HORVU7Hr1G012300	1.56	-4.7	HORVU7Hr1G049860	1.36	-3.4
HORVU3Hr1G061130	1.56	-6.1	HORVU3Hr1G037170	1.36	-4.1
HORVU3Hr1G028710	1.55	-4.5	HORVU2Hr1G003400	1.36	-3.3
HORVU4Hr1G054790	1.55	-6.4	HORVU2Hr1G083430	1.36	-4.2
HORVU3Hr1G002000	1.55	-4.3	HORVU1Hr1G065810	1.35	-18.1
HORVU3Hr1G092420	1.54	-5.3	HORVU7Hr1G104660	1.35	-3.5
HORVU7Hr1G040680	1.54	-8.6	HORVU3Hr1G104290	1.35	-4.2
HORVU1Hr1G016660	1.53	-5.8	HORVU1Hr1G037110	1.34	-3.6
HORVU5Hr1G070040	1.53	-4.2	HORVU7Hr1G122470	1.34	-3.3
HORVU2Hr1G038040	1.53	-4.7	HORVU3Hr1G018810	1.34	-3.2
HORVU4Hr1G062830	1.52	-4.5	HORVU4Hr1G073580	1.34	-18.1
HORVU7Hr1G056700	1.52	-7.2	HORVU3Hr1G069590	1.34	-3.8
HORVU1Hr1G025870	1.52	-5.7	HORVU7Hr1G107070	1.34	-3.1
HORVU3Hr1G009360	1.52	-12.6	HORVU3Hr1G034570	1.33	-5.7
HORVU4Hr1G087230	1.52	-7.4	HORVU5Hr1G114330	1.33	-4.6
HORVU2Hr1G043900	1.51	-15.2	HORVU1Hr1G061380	1.33	-4.0
HORVU7Hr1G088730	1.50	-7.4	HORVU4Hr1G063790	1.33	-3.6
HORVU7Hr1G000250	1.50	-18.3	HORVU5Hr1G047630	1.33	-5.2
HORVU2Hr1G025150	1.50	-3.9	HORVU3Hr1G065260	1.33	-3.1
HORVU4Hr1G002530	1.50	-5.9	HORVU2Hr1G115640	1.33	-3.2
HORVU2Hr1G120530	1.49	-7.6	HORVU2Hr1G012460	1.32	-3.1
HORVU6Hr1G061390	1.48	-3.9	HORVU6Hr1G066840	1.32	-3.5
HORVU7Hr1G102030	1.48	-5.2	HORVU3Hr1G003220	1.32	-12.6
HORVU3Hr1G106140	1.48	-8.9	HORVU3Hr1G054240	1.32	-3.4
HORVU7Hr1G037080	1.47	-12.3	HORVU1Hr1G080680	1.32	-3.7
HORVU0Hr1G003900	1.47	-5.9	HORVU3Hr1G084220	1.31	-13.7
HORVU6Hr1G029750	1.47	-5.0	HORVU5Hr1G096390	1.31	-9.7
HORVU5Hr1G101820	1.47	-4.0	HORVU4Hr1G081040	1.31	-5.5
HORVU0Hr1G010080	1.47	-7.4	HORVU4Hr1G087760	1.31	-3.3
HORVU3Hr1G015640	1.46	-10.6	HORVU2Hr1G025090	1.30	-3.1
HORVU5Hr1G082040	1.46	-3.9	HORVU2Hr1G115830	1.30	-3.7
HORVU6Hr1G067380	1.46	-4.0	HORVU4Hr1G023510	1.29	-3.1
HORVU2Hr1G083410	1.45	-5.6	HORVU2Hr1G095080	1.29	-5.5
HORVU7Hr1G050980	1.45	-3.7	HORVU3Hr1G000720	1.29	-2.9
HORVU7Hr1G085120	1.45	-5.0	HORVU2Hr1G117140	1.29	-2.9
HORVU7Hr1G001020	1.45	-3.7	HORVU2Hr1G023770	1.29	-4.2
HORVU5Hr1G087730	1.45	-3.8	HORVU1Hr1G010380	1.28	-7.3
HORVU5Hr1G109610	1.44	-4.3	HORVU4Hr1G079030	1.28	-4.2
HORVU7Hr1G010690	1.44	-5.9	HORVU1Hr1G066450	1.28	-4.7
HORVU2Hr1G001150	1.44	-7.0	HORVU2Hr1G122330	1.28	-2.8
HORVU6Hr1G002300	1.44	-6.4	HORVU1Hr1G046940	1.27	-7.5
HORVU5Hr1G077390	1.44	-7.5	HORVU3Hr1G023750	1.27	-3.3
HORVU7Hr1G098330	1.44	-12.5	HORVU3Hr1G090450	1.27	-3.5
HORVU4Hr1G063010	1.43	-7.7	HORVU1Hr1G075160	1.27	-3.9
HORVU5Hr1G103850	1.43	-7.1	HORVU4Hr1G061770	1.26	-5.5
HORVU7Hr1G070010	1.43	-3.8	HORVU5Hr1G052440	1.26	-6.9
HORVU7Hr1G026000	1.42	-4.8	HORVU3Hr1G019510	1.26	-4.5
HORVU7Hr1G002820	1.42	-8.1	HORVU5Hr1G062940	1.26	-16.4
HORVU5Hr1G064230	1.42	-6.2	HORVU5Hr1G073760	1.25	-6.2
HORVU2Hr1G094630	1.41	-3.7	HORVU4Hr1G010250	1.25	-3.2
HORVU1Hr1G002710	1.41	-3.7	HORVU3Hr1G081260	1.25	-2.8
HORVU4Hr1G056770	1.40	-4.5	HORVU7Hr1G048690	1.25	-6.0
HORVU4Hr1G020030	1.40	-12.4	HORVU7Hr1G100570	1.25	-2.7
HORVU7Hr1G113270	1.40	-12.2	HORVU7Hr1G092330	1.24	-3.5
HORVU1Hr1G047010	1.40	-4.6	HORVU2Hr1G002600	1.24	-16.0
HORVU3Hr1G085680	1.40	-4.6	HORVU3Hr1G071770	1.24	-4.8
HORVU2Hr1G088660	1.40	-3.4	HORVU4Hr1G067630	1.24	-6.9
HORVU3Hr1G096120	1.39	-3.9	HORVU6Hr1G062850	1.24	-4.2
HORVU7Hr1G075760	1.39	-3.4	HORVU7Hr1G046430	1.24	-4.2
HORVU5Hr1G026360	1.39	-9.8	HORVU7Hr1G022250	1.24	-6.5
HORVU7Hr1G090560	1.39	-10.6	HORVU0Hr1G015390	1.23	-3.2
HORVU2Hr1G114680	1.39	-3.5	HORVU4Hr1G087580	1.23	-3.5
HORVU7Hr1G082330	1.38	-6.1	HORVU4Hr1G001450	1.23	-2.8

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU6Hr1G067060	1.23	-3.7	HORVU4Hr1G077610	1.12	-3.1
HORVU3Hr1G066600	1.22	-2.8	HORVU7Hr1G076480	1.12	-3.7
HORVU7Hr1G061070	1.22	-4.6	HORVU7Hr1G028290	1.12	-2.4
HORVU3Hr1G003330	1.22	-3.2	HORVU7Hr1G007220	1.12	-2.1
HORVU3Hr1G093830	1.22	-7.7	HORVU4Hr1G089690	1.11	-2.3
HORVU3Hr1G022340	1.22	-2.7	HORVU7Hr1G020300	1.11	-3.1
HORVU4Hr1G085050	1.22	-4.1	HORVU2Hr1G109370	1.11	-3.1
HORVU7Hr1G029260	1.22	-7.0	HORVU4Hr1G025070	1.11	-2.6
HORVU2Hr1G093950	1.22	-2.9	HORVU1Hr1G009110	1.10	-2.4
HORVU4Hr1G052090	1.22	-3.6	HORVU7Hr1G062090	1.10	-4.3
HORVU4Hr1G068760	1.22	-9.8	HORVU3Hr1G026070	1.10	-4.5
HORVU5Hr1G093410	1.22	-2.6	HORVU2Hr1G090670	1.10	-2.8
HORVU4Hr1G066750	1.22	-2.7	HORVU3Hr1G001680	1.10	-3.5
HORVU7Hr1G036720	1.21	-2.6	HORVU1Hr1G029460	1.10	-3.8
HORVU1Hr1G070310	1.21	-2.7	HORVU0Hr1G000850	1.10	-2.1
HORVU4Hr1G062440	1.21	-21.9	HORVU2Hr1G079840	1.09	-11.3
HORVU1Hr1G037760	1.21	-8.3	HORVU3Hr1G115580	1.09	-10.0
HORVU3Hr1G076550	1.21	-3.0	HORVU5Hr1G037430	1.09	-2.0
HORVU6Hr1G076760	1.21	-3.6	HORVU3Hr1G036960	1.09	-2.7
HORVU3Hr1G009520	1.21	-8.9	HORVU6Hr1G080890	1.09	-2.2
HORVU2Hr1G118790	1.20	-3.9	HORVU7Hr1G008390	1.08	-7.4
HORVU1Hr1G018350	1.20	-2.7	HORVU7Hr1G048880	1.08	-6.5
HORVU7Hr1G017950	1.20	-8.7	HORVU4Hr1G084680	1.08	-2.0
HORVU1Hr1G046320	1.20	-3.3	HORVU3Hr1G002040	1.08	-2.3
HORVU1Hr1G063420	1.20	-8.5	HORVU7Hr1G030160	1.08	-2.0
HORVU6Hr1G073510	1.20	-6.9	HORVU1Hr1G051800	1.08	-2.8
HORVU3Hr1G030770	1.20	-3.8	HORVU6Hr1G004000	1.07	-2.0
HORVU4Hr1G088080	1.20	-5.5	HORVU3Hr1G075040	1.07	-2.4
HORVU7Hr1G045680	1.19	-10.4	HORVU4Hr1G063190	1.07	-3.5
HORVU2Hr1G108520	1.19	-2.8	HORVU3Hr1G085520	1.07	-4.2
HORVU4Hr1G080460	1.19	-4.2	HORVU3Hr1G096280	1.07	-3.3
HORVU7Hr1G111130	1.19	-5.9	HORVU1Hr1G015770	1.07	-2.5
HORVU4Hr1G068530	1.18	-5.9	HORVU4Hr1G088860	1.06	-3.3
HORVU5Hr1G059070	1.18	-3.6	HORVU2Hr1G103590	1.06	-2.3
HORVU6Hr1G090020	1.18	-2.5	HORVU6Hr1G030310	1.06	-3.7
HORVU1Hr1G030060	1.18	-3.3	HORVU7Hr1G120030	1.06	-9.9
HORVU2Hr1G007840	1.18	-5.6	HORVU5Hr1G055060	1.06	-2.7
HORVU4Hr1G008120	1.18	-5.6	HORVU2Hr1G075280	1.06	-4.5
HORVU4Hr1G082970	1.17	-2.6	HORVU4Hr1G054370	1.06	-2.9
HORVU2Hr1G063820	1.17	-4.7	HORVU0Hr1G005320	1.06	-3.4
HORVU3Hr1G000280	1.17	-5.9	HORVU3Hr1G079800	1.06	-6.4
HORVU2Hr1G013450	1.17	-2.9	HORVU4Hr1G066270	1.05	-3.3
HORVU6Hr1G076020	1.17	-2.8	HORVU1Hr1G002300	1.05	-2.1
HORVU4Hr1G027740	1.17	-16.6	HORVU6Hr1G057550	1.05	-3.3
HORVU7Hr1G006320	1.16	-2.8	HORVU3Hr1G027970	1.05	-4.6
HORVU4Hr1G076750	1.16	-2.8	HORVU5Hr1G019820	1.04	-2.5
HORVU5Hr1G103060	1.16	-2.5	HORVU3Hr1G007580	1.04	-2.4
HORVU5Hr1G095910	1.16	-6.1	HORVU1Hr1G083550	1.04	-8.2
HORVU4Hr1G089200	1.15	-3.5	HORVU4Hr1G083980	1.04	-4.7
HORVU2Hr1G108790	1.15	-4.3	HORVU4Hr1G023490	1.04	-3.3
HORVU6Hr1G087000	1.15	-2.7	HORVU1Hr1G054060	1.04	-3.3
HORVU4Hr1G057790	1.15	-2.3	HORVU5Hr1G120920	1.04	-4.2
HORVU6Hr1G000510	1.15	-2.4	HORVU7Hr1G114400	1.03	-2.5
HORVU4Hr1G085800	1.15	-2.3	HORVU5Hr1G046900	1.03	-2.6
HORVU6Hr1G027620	1.15	-4.2	HORVU2Hr1G003840	1.03	-5.0
HORVU1Hr1G014570	1.15	-2.4	HORVU5Hr1G081770	1.03	-2.6
HORVU6Hr1G023070	1.15	-6.4	HORVU4Hr1G054910	1.03	-5.9
HORVU2Hr1G028590	1.14	-4.3	HORVU1Hr1G073540	1.03	-3.6
HORVU2Hr1G127640	1.14	-4.4	HORVU6Hr1G033600	1.02	-17.6
HORVU5Hr1G000980	1.14	-4.0	HORVU0Hr1G009140	1.02	-3.4
HORVU6Hr1G079630	1.14	-2.4	HORVU3Hr1G077580	1.02	-4.9
HORVU4Hr1G024530	1.14	-4.4	HORVU5Hr1G099180	1.02	-2.4
HORVU3Hr1G095750	1.13	-3.9	HORVU3Hr1G090970	1.02	-3.1
HORVU2Hr1G108820	1.13	-4.5	HORVU5Hr1G042400	1.02	-3.0
HORVU3Hr1G003140	1.13	-5.5	HORVU4Hr1G064010	1.02	-2.9
HORVU3Hr1G078360	1.13	-4.8	HORVU0Hr1G012140	1.02	-2.8
HORVU5Hr1G111780	1.13	-2.4	HORVU3Hr1G029770	1.02	-2.7
HORVU3Hr1G112020	1.13	-2.2	HORVU3Hr1G056560	1.01	-2.7
HORVU5Hr1G080580	1.12	-4.8	HORVU3Hr1G059470	1.01	-3.4
HORVU5Hr1G089230	1.12	-3.6	HORVU7Hr1G000150	1.01	-3.8
HORVU3Hr1G070900	1.12	-3.1	HORVU3Hr1G037600	1.01	-15.2
HORVU5Hr1G077110	1.12	-6.3	HORVU2Hr1G081790	1.01	-4.2
HORVU7Hr1G074490	1.12	-7.0	HORVU2Hr1G093960	1.00	-2.3
HORVU7Hr1G085660	1.12	-2.3	HORVU1Hr1G060920	1.00	-2.3
HORVU2Hr1G107630	1.12	-4.2	HORVU4Hr1G056500	1.00	-2.3

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU7Hr1G030370	1.00	-3.0	HORVU4Hr1G084920	0.89	-6.9
HORVU5Hr1G032370	1.00	-2.8	HORVU3Hr1G003040	0.89	-3.8
HORVU1Hr1G058790	1.00	-3.6	HORVU2Hr1G020340	0.89	-3.0
HORVU1Hr1G050720	1.00	-2.6	HORVU4Hr1G007610	0.89	-2.2
HORVU2Hr1G011070	1.00	-2.4	HORVU0Hr1G016430	0.89	-2.1
HORVU1Hr1G074060	0.99	-5.9	HORVU5Hr1G012780	0.89	-2.6
HORVU7Hr1G108790	0.99	-3.1	HORVU2Hr1G121440	0.89	-4.6
HORVU2Hr1G035680	0.99	-2.6	HORVU4Hr1G012560	0.89	-2.3
HORVU1Hr1G089760	0.99	-4.1	HORVU2Hr1G028430	0.89	-2.8
HORVU7Hr1G048170	0.99	-2.4	HORVU0Hr1G020730	0.89	-4.2
HORVU0Hr1G020750	0.99	-2.8	HORVU3Hr1G116490	0.89	-2.9
HORVU2Hr1G096700	0.99	-2.7	HORVU1Hr1G064870	0.88	-2.9
HORVU4Hr1G068830	0.99	-5.6	HORVU4Hr1G043680	0.88	-6.8
HORVU3Hr1G074010	0.99	-2.3	HORVU2Hr1G059800	0.88	-2.7
HORVU7Hr1G043250	0.98	-2.4	HORVU6Hr1G069400	0.88	-2.2
HORVU7Hr1G046920	0.98	-2.5	HORVU1Hr1G016200	0.88	-3.1
HORVU5Hr1G112300	0.98	-2.9	HORVU7Hr1G021430	0.88	-6.5
HORVU5Hr1G022140	0.98	-2.4	HORVU2Hr1G080140	0.88	-3.4
HORVU3Hr1G036020	0.98	-2.7	HORVU4Hr1G048440	0.88	-6.4
HORVU2Hr1G080350	0.98	-2.1	HORVU2Hr1G123590	0.88	-2.8
HORVU2Hr1G038030	0.98	-2.3	HORVU4Hr1G077420	0.88	-11.0
HORVU1Hr1G029180	0.97	-2.5	HORVU1Hr1G080080	0.87	-4.3
HORVU7Hr1G095890	0.97	-5.6	HORVU6Hr1G041610	0.87	-2.4
HORVU2Hr1G077460	0.97	-3.5	HORVU3Hr1G026110	0.87	-2.1
HORVU6Hr1G014500	0.97	-6.9	HORVU2Hr1G010440	0.87	-8.5
HORVU5Hr1G074660	0.96	-8.5	HORVU1Hr1G047000	0.87	-2.1
HORVU7Hr1G041710	0.96	-11.3	HORVU4Hr1G061070	0.87	-2.4
HORVU3Hr1G059810	0.96	-4.1	HORVU4Hr1G089540	0.86	-2.1
HORVU4Hr1G024400	0.95	-2.9	HORVU4Hr1G063420	0.86	-3.0
HORVU1Hr1G026320	0.95	-2.6	HORVU0Hr1G001760	0.86	-2.1
HORVU4Hr1G050250	0.95	-2.6	HORVU5Hr1G064010	0.86	-2.5
HORVU3Hr1G028780	0.95	-4.8	HORVU4Hr1G065180	0.86	-3.6
HORVU1Hr1G083160	0.95	-3.3	HORVU6Hr1G062140	0.86	-2.0
HORVU3Hr1G019480	0.95	-3.9	HORVU7Hr1G098320	0.86	-2.0
HORVU4Hr1G083170	0.95	-3.1	HORVU1Hr1G051470	0.86	-2.8
HORVU7Hr1G057140	0.94	-2.7	HORVU1Hr1G038620	0.86	-5.5
HORVU7Hr1G088920	0.94	-2.3	HORVU2Hr1G028940	0.86	-2.3
HORVU6Hr1G025280	0.94	-2.1	HORVU3Hr1G002280	0.86	-14.0
HORVU7Hr1G041810	0.94	-2.1	HORVU4Hr1G002350	0.85	-4.0
HORVU1Hr1G051330	0.94	-5.3	HORVU7Hr1G096680	0.85	-3.1
HORVU5Hr1G124030	0.94	-3.1	HORVU2Hr1G045730	0.85	-2.2
HORVU5Hr1G074340	0.94	-6.7	HORVU1Hr1G037650	0.85	-2.5
HORVU2Hr1G003470	0.94	-2.7	HORVU2Hr1G100720	0.85	-3.1
HORVU4Hr1G078510	0.94	-5.3	HORVU2Hr1G104930	0.85	-5.5
HORVU3Hr1G116360	0.93	-2.8	HORVU4Hr1G060970	0.85	-3.3
HORVU6Hr1G076820	0.93	-3.1	HORVU1Hr1G089730	0.85	-2.3
HORVU7Hr1G043230	0.93	-2.4	HORVU1Hr1G083270	0.84	-2.2
HORVU2Hr1G072670	0.93	-2.9	HORVU4Hr1G021960	0.84	-3.8
HORVU7Hr1G092300	0.92	-5.2	HORVU5Hr1G121570	0.84	-4.1
HORVU6Hr1G034620	0.92	-5.9	HORVU1Hr1G088510	0.84	-2.2
HORVU4Hr1G049550	0.92	-3.2	HORVU3Hr1G078960	0.84	-2.8
HORVU3Hr1G116200	0.92	-10.3	HORVU5Hr1G046240	0.84	-3.2
HORVU7Hr1G006490	0.92	-6.9	HORVU7Hr1G035340	0.84	-9.1
HORVU2Hr1G010990	0.92	-2.3	HORVU5Hr1G047730	0.84	-3.7
HORVU4Hr1G002330	0.92	-3.1	HORVU2Hr1G077970	0.84	-2.5
HORVU6Hr1G087420	0.92	-3.7	HORVU4Hr1G084830	0.84	-3.7
HORVU6Hr1G070290	0.91	-2.9	HORVU6Hr1G026340	0.84	-3.2
HORVU3Hr1G094820	0.91	-4.1	HORVU1Hr1G058330	0.83	-4.3
HORVU2Hr1G113270	0.91	-2.9	HORVU5Hr1G053480	0.83	-5.3
HORVU0Hr1G009470	0.91	-2.8	HORVU2Hr1G019180	0.83	-2.7
HORVU3Hr1G018630	0.90	-2.6	HORVU1Hr1G053310	0.83	-4.2
HORVU1Hr1G092150	0.90	-2.3	HORVU5Hr1G124320	0.83	-2.4
HORVU3Hr1G091200	0.90	-7.2	HORVU2Hr1G071170	0.83	-3.1
HORVU7Hr1G020590	0.90	-3.8	HORVU2Hr1G009900	0.83	-2.6
HORVU2Hr1G014240	0.90	-4.1	HORVU2Hr1G005530	0.83	-2.3
HORVU6Hr1G075950	0.90	-4.5	HORVU4Hr1G063270	0.83	-3.6
HORVU5Hr1G037440	0.90	-3.5	HORVU6Hr1G013670	0.82	-5.7
HORVU2Hr1G066920	0.90	-8.2	HORVU6Hr1G050810	0.82	-2.3
HORVU1Hr1G017700	0.90	-2.5	HORVU6Hr1G050750	0.82	-7.1
HORVU7Hr1G054130	0.90	-2.4	HORVU1Hr1G001920	0.82	-2.3
HORVU4Hr1G022930	0.90	-3.5	HORVU5Hr1G107440	0.82	-2.2
HORVU7Hr1G035050	0.90	-3.1	HORVU3Hr1G088970	0.82	-7.5
HORVU3Hr1G089510	0.90	-4.0	HORVU2Hr1G022410	0.82	-4.5
HORVU7Hr1G012850	0.90	-2.3	HORVU4Hr1G056470	0.81	-3.6
HORVU6Hr1G029540	0.90	-2.3	HORVU5Hr1G115340	0.81	-3.2

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU7Hr1G037420	0.81	-16.4	HORVU2Hr1G090980	0.71	-7.9
HORVU6Hr1G016120	0.81	-2.8	HORVU4Hr1G084340	0.71	-2.7
HORVU5Hr1G080500	0.81	-2.1	HORVU3Hr1G038290	0.71	-2.5
HORVU5Hr1G092740	0.81	-4.8	HORVU2Hr1G106880	0.71	-3.9
HORVU3Hr1G070690	0.81	-10.0	HORVU1Hr1G062500	0.71	-3.4
HORVU1Hr1G090100	0.80	-10.1	HORVU6Hr1G087220	0.71	-4.2
HORVU1Hr1G050970	0.80	-4.5	HORVU1Hr1G093800	0.71	-3.7
HORVU5Hr1G053230	0.80	-6.4	HORVU5Hr1G011100	0.71	-3.5
HORVU1Hr1G009920	0.80	-4.0	HORVU3Hr1G087210	0.71	-2.7
HORVU1Hr1G045520	0.80	-4.8	HORVU5Hr1G106960	0.70	-3.8
HORVU7Hr1G033360	0.80	-3.8	HORVU5Hr1G104230	0.70	-3.0
HORVU3Hr1G024970	0.80	-4.2	HORVU1Hr1G048360	0.70	-3.4
HORVU2Hr1G102620	0.80	-3.7	HORVU4Hr1G019610	0.70	-2.1
HORVU4Hr1G000280	0.80	-7.0	HORVU3Hr1G085480	0.70	-3.1
HORVU0Hr1G000810	0.80	-2.8	HORVU7Hr1G058940	0.70	-4.3
HORVU7Hr1G045140	0.80	-2.7	HORVU4Hr1G049410	0.70	-6.1
HORVU3Hr1G070610	0.79	-2.6	HORVU3Hr1G019340	0.70	-5.4
HORVU5Hr1G064020	0.79	-4.6	HORVU5Hr1G106790	0.70	-4.0
HORVU0Hr1G021630	0.79	-4.2	HORVU4Hr1G038570	0.70	-4.1
HORVU1Hr1G091010	0.79	-2.3	HORVU1Hr1G030040	0.69	-3.5
HORVU5Hr1G016810	0.79	-8.0	HORVU4Hr1G021190	0.69	-2.3
HORVU3Hr1G116450	0.79	-3.1	HORVU2Hr1G099330	0.69	-4.7
HORVU4Hr1G068770	0.79	-2.5	HORVU1Hr1G008120	0.68	-4.8
HORVU7Hr1G078670	0.79	-3.7	HORVU0Hr1G038320	0.68	-2.2
HORVU3Hr1G065240	0.78	-2.4	HORVU7Hr1G096560	0.68	-2.7
HORVU7Hr1G092380	0.78	-2.0	HORVU2Hr1G123100	0.68	-2.1
HORVU6Hr1G002550	0.78	-3.2	HORVU1Hr1G029870	0.68	-9.6
HORVU1Hr1G000040	0.78	-3.7	HORVU1Hr1G073510	0.68	-2.1
HORVU1Hr1G013210	0.78	-4.6	HORVU5Hr1G003910	0.68	-5.9
HORVU7Hr1G002010	0.78	-2.4	HORVU3Hr1G073780	0.67	-6.6
HORVU2Hr1G072500	0.78	-6.9	HORVU3Hr1G097950	0.67	-6.0
HORVU3Hr1G114220	0.78	-3.4	HORVU1Hr1G043040	0.67	-3.1
HORVU5Hr1G007770	0.77	-6.1	HORVU3Hr1G089910	0.67	-2.2
HORVU6Hr1G047300	0.77	-6.3	HORVU3Hr1G025670	0.67	-2.1
HORVU4Hr1G058840	0.77	-2.1	HORVU2Hr1G005510	0.67	-2.5
HORVU6Hr1G015480	0.77	-2.8	HORVU6Hr1G036570	0.67	-4.5
HORVU1Hr1G046400	0.77	-4.6	HORVU0Hr1G040540	0.66	-3.2
HORVU1Hr1G010810	0.77	-2.0	HORVU2Hr1G059380	0.66	-5.4
HORVU3Hr1G105360	0.77	-2.5	HORVU5Hr1G111640	0.66	-3.0
HORVU2Hr1G031310	0.77	-12.8	HORVU5Hr1G018480	0.66	-2.2
HORVU6Hr1G075250	0.77	-2.8	HORVU2Hr1G085800	0.66	-2.4
HORVU5Hr1G118270	0.76	-4.6	HORVU7Hr1G095960	0.66	-2.4
HORVU6Hr1G082060	0.76	-3.1	HORVU7Hr1G081510	0.66	-3.2
HORVU4Hr1G000770	0.76	-4.7	HORVU5Hr1G081090	0.66	-2.6
HORVU2Hr1G109850	0.76	-3.0	HORVU4Hr1G084900	0.65	-3.0
HORVU3Hr1G056270	0.75	-2.5	HORVU1Hr1G057180	0.65	-4.6
HORVU1Hr1G010670	0.75	-3.6	HORVU1Hr1G036670	0.65	-2.1
HORVU5Hr1G020170	0.75	-2.5	HORVU1Hr1G051700	0.65	-2.5
HORVU3Hr1G005560	0.75	-4.0	HORVU4Hr1G015740	0.65	-4.0
HORVU3Hr1G030020	0.75	-5.8	HORVU5Hr1G000740	0.65	-5.3
HORVU3Hr1G113820	0.75	-3.3	HORVU4Hr1G000050	0.65	-3.3
HORVU6Hr1G081000	0.75	-2.4	HORVU5Hr1G000780	0.65	-4.8
HORVU5Hr1G001920	0.74	-2.4	HORVU4Hr1G069050	0.65	-2.5
HORVU3Hr1G078200	0.74	-4.0	HORVU3Hr1G074040	0.65	-2.5
HORVU3Hr1G114920	0.74	-4.8	HORVU3Hr1G081590	0.64	-3.1
HORVU1Hr1G070460	0.74	-2.8	HORVU7Hr1G033900	0.64	-4.9
HORVU2Hr1G056440	0.74	-3.2	HORVU1Hr1G091490	0.64	-3.1
HORVU2Hr1G096850	0.74	-5.9	HORVU3Hr1G078780	0.64	-2.7
HORVU3Hr1G079260	0.74	-2.0	HORVU3Hr1G094210	0.64	-3.2
HORVU7Hr1G082380	0.73	-8.4	HORVU1Hr1G030200	0.64	-6.3
HORVU4Hr1G053760	0.73	-2.4	HORVU2Hr1G089070	0.64	-2.4
HORVU4Hr1G086020	0.73	-4.7	HORVU4Hr1G082700	0.64	-2.3
HORVU1Hr1G056180	0.73	-5.4	HORVU3Hr1G059840	0.64	-4.8
HORVU5Hr1G024470	0.73	-4.4	HORVU2Hr1G034330	0.64	-2.1
HORVU7Hr1G120660	0.73	-2.7	HORVU6Hr1G031480	0.64	-2.5
HORVU7Hr1G040280	0.73	-2.1	HORVU4Hr1G004800	0.64	-2.7
HORVU7Hr1G084940	0.73	-2.0	HORVU7Hr1G047700	0.63	-5.7
HORVU7Hr1G093020	0.72	-3.2	HORVU5Hr1G024250	0.63	-2.2
HORVU6Hr1G078290	0.72	-2.5	HORVU2Hr1G048870	0.63	-2.7
HORVU4Hr1G090840	0.72	-3.8	HORVU1Hr1G030930	0.63	-2.2
HORVU5Hr1G111620	0.72	-9.3	HORVU6Hr1G020380	0.63	-5.0
HORVU4Hr1G024380	0.72	-2.1	HORVU2Hr1G091990	0.63	-2.1
HORVU5Hr1G103990	0.71	-3.3	HORVU2Hr1G115000	0.63	-2.3
HORVU3Hr1G082130	0.71	-2.2	HORVU6Hr1G075340	0.63	-3.9
HORVU7Hr1G002300	0.71	-3.3	HORVU4Hr1G031530	0.63	-4.0

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU7Hr1G108230	0.63	-3.1	HORVU7Hr1G030810	-0.64	-5.8
HORVU3Hr1G071530	0.62	-2.3	HORVU6Hr1G041590	-0.64	-5.6
HORVU3Hr1G033700	0.62	-2.4	HORVU5Hr1G119230	-0.64	-2.6
HORVU4Hr1G048970	0.62	-3.1	HORVU3Hr1G030580	-0.64	-5.7
HORVU1Hr1G057490	0.62	-2.2	HORVU4Hr1G013310	-0.64	-21.4
HORVU2Hr1G123200	0.62	-2.4	HORVU7Hr1G068230	-0.64	-2.6
HORVU3Hr1G056370	0.62	-2.2	HORVU2Hr1G005320	-0.64	-5.8
HORVU2Hr1G010560	0.62	-2.0	HORVU3Hr1G001940	-0.64	-3.4
HORVU2Hr1G110110	0.61	-5.2	HORVU0Hr1G019750	-0.64	-3.4
HORVU4Hr1G007890	0.61	-4.2	HORVU5Hr1G016340	-0.65	-2.5
HORVU2Hr1G094320	0.61	-3.6	HORVU1Hr1G001510	-0.65	-2.1
HORVU1Hr1G023510	0.61	-3.0	HORVU6Hr1G009090	-0.65	-3.1
HORVU2Hr1G018820	0.61	-3.1	HORVU7Hr1G100520	-0.65	-2.1
HORVU7Hr1G114170	0.61	-2.4	HORVU7Hr1G089480	-0.65	-2.5
HORVU4Hr1G003580	0.61	-3.4	HORVU1Hr1G061160	-0.65	-9.5
HORVU4Hr1G013170	0.61	-2.0	HORVU2Hr1G115620	-0.65	-8.4
HORVU3Hr1G062490	0.60	-2.8	HORVU1Hr1G009590	-0.65	-3.8
HORVU3Hr1G050340	0.60	-3.2	HORVU7Hr1G088620	-0.65	-3.6
HORVU4Hr1G084410	0.60	-2.8	HORVU7Hr1G098580	-0.65	-2.0
HORVU6Hr1G073010	0.60	-4.6	HORVU4Hr1G083650	-0.65	-4.1
HORVU3Hr1G003110	0.60	-2.9	HORVU7Hr1G041850	-0.66	-3.3
HORVU7Hr1G033820	0.60	-2.6	HORVU5Hr1G016320	-0.66	-2.3
HORVU6Hr1G066780	0.60	-3.0	HORVU6Hr1G050370	-0.66	-3.0
HORVU3Hr1G019840	0.60	-2.4	HORVU7Hr1G025910	-0.66	-12.2
HORVU3Hr1G055700	0.60	-3.8	HORVU6Hr1G093060	-0.66	-2.1
HORVU5Hr1G019590	0.60	-4.5	HORVU6Hr1G067840	-0.66	-4.4
HORVU4Hr1G000930	0.60	-3.2	HORVU6Hr1G031340	-0.66	-12.9
HORVU2Hr1G065960	0.60	-2.1	HORVU1Hr1G029370	-0.66	-4.2
HORVU4Hr1G063150	0.60	-3.1	HORVU1Hr1G046630	-0.66	-4.2
HORVU1Hr1G083710	0.60	-2.0	HORVU7Hr1G043620	-0.66	-5.6
HORVU4Hr1G084840	0.60	-2.2	HORVU2Hr1G049700	-0.66	-12.2
HORVU5Hr1G113630	0.60	-2.1	HORVU2Hr1G115010	-0.67	-2.8
HORVU5Hr1G062490	0.59	-3.0	HORVU7Hr1G097550	-0.67	-3.8
HORVU6Hr1G051070	0.59	-3.1	HORVU6Hr1G022950	-0.67	-2.2
HORVU4Hr1G069100	0.59	-2.2	HORVU3Hr1G035590	-0.68	-6.5
HORVU3Hr1G096500	0.59	-2.6	HORVU3Hr1G074250	-0.68	-5.1
HORVU2Hr1G095070	0.59	-3.5	HORVU3Hr1G059290	-0.68	-6.7
HORVU2Hr1G060730	0.59	-5.9	HORVU2Hr1G099510	-0.68	-3.5
HORVU1Hr1G000710	0.59	-2.4	HORVU5Hr1G106120	-0.68	-2.2
HORVU2Hr1G027640	0.59	-3.1	HORVU1Hr1G017380	-0.69	-4.5
HORVU7Hr1G051550	0.59	-3.3	HORVU2Hr1G095720	-0.69	-4.2
HORVU3Hr1G073560	-0.59	-5.1	HORVU3Hr1G055260	-0.69	-4.4
HORVU7Hr1G038300	-0.59	-2.2	HORVU6Hr1G094950	-0.69	-3.1
HORVU6Hr1G013290	-0.59	-4.0	HORVU7Hr1G113890	-0.69	-7.4
HORVU6Hr1G061580	-0.59	-2.6	HORVU5Hr1G093960	-0.70	-4.2
HORVU1Hr1G092820	-0.59	-4.6	HORVU6Hr1G054050	-0.70	-3.1
HORVU5Hr1G068520	-0.59	-3.8	HORVU3Hr1G115940	-0.70	-3.4
HORVU4Hr1G079230	-0.60	-3.9	HORVU3Hr1G014590	-0.70	-3.6
HORVU1Hr1G022840	-0.60	-3.0	HORVU1Hr1G081240	-0.70	-2.9
HORVU2Hr1G113300	-0.60	-5.1	HORVU2Hr1G070720	-0.70	-2.6
HORVU3Hr1G096780	-0.60	-2.8	HORVU4Hr1G001040	-0.71	-2.1
HORVU5Hr1G081250	-0.60	-2.7	HORVU2Hr1G080630	-0.71	-5.7
HORVU7Hr1G101590	-0.61	-4.8	HORVU3Hr1G085320	-0.71	-2.7
HORVU7Hr1G008800	-0.61	-2.9	HORVU6Hr1G054940	-0.71	-10.0
HORVU6Hr1G069150	-0.61	-4.4	HORVU1Hr1G019660	-0.71	-3.7
HORVU6Hr1G039740	-0.61	-4.9	HORVU3Hr1G012850	-0.71	-7.0
HORVU5Hr1G103180	-0.61	-2.3	HORVU5Hr1G099670	-0.72	-2.0
HORVU5Hr1G045150	-0.61	-6.3	HORVU5Hr1G082940	-0.72	-4.3
HORVU4Hr1G044460	-0.61	-3.9	HORVU7Hr1G101110	-0.72	-3.9
HORVU7Hr1G001160	-0.61	-2.5	HORVU7Hr1G049260	-0.72	-2.1
HORVU1Hr1G025390	-0.61	-3.9	HORVU1Hr1G086460	-0.72	-4.5
HORVU5Hr1G068060	-0.61	-6.9	HORVU6Hr1G087390	-0.72	-3.8
HORVU0Hr1G000200	-0.61	-11.8	HORVU3Hr1G066240	-0.73	-7.0
HORVU7Hr1G095730	-0.61	-3.7	HORVU6Hr1G059520	-0.73	-4.2
HORVU4Hr1G083870	-0.61	-2.6	HORVU5Hr1G062410	-0.73	-2.9
HORVU6Hr1G036760	-0.61	-2.2	HORVU3Hr1G013100	-0.73	-2.3
HORVU3Hr1G002820	-0.61	-4.2	HORVU3Hr1G057140	-0.73	-4.9
HORVU7Hr1G055090	-0.62	-6.9	HORVU0Hr1G016860	-0.73	-5.2
HORVU7Hr1G120130	-0.62	-3.4	HORVU4Hr1G022630	-0.74	-3.5
HORVU5Hr1G124630	-0.62	-4.5	HORVU1Hr1G049210	-0.74	-3.1
HORVU3Hr1G089290	-0.62	-6.7	HORVU6Hr1G063700	-0.74	-2.7
HORVU3Hr1G095700	-0.63	-6.4	HORVU4Hr1G028310	-0.74	-3.1
HORVU4Hr1G065350	-0.63	-5.4	HORVU2Hr1G107180	-0.74	-2.4
HORVU5Hr1G018840	-0.63	-23.7	HORVU1Hr1G027980	-0.74	-6.0
HORVU3Hr1G080500	-0.63	-9.8	HORVU7Hr1G040040	-0.75	-2.3

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU1Hr1G021620	-0.75	-2.8	HORVU3Hr1G012650	-0.85	-3.4
HORVU2Hr1G090310	-0.75	-2.9	HORVU3Hr1G021810	-0.85	-4.6
HORVU1Hr1G089310	-0.75	-3.0	HORVU1Hr1G008810	-0.85	-7.0
HORVU5Hr1G097440	-0.75	-2.8	HORVU3Hr1G013810	-0.85	-2.4
HORVU1Hr1G021760	-0.75	-3.8	HORVU7Hr1G078380	-0.86	-2.4
HORVU3Hr1G070340	-0.75	-7.6	HORVU7Hr1G072240	-0.86	-15.5
HORVU5Hr1G092990	-0.76	-2.2	HORVU7Hr1G008050	-0.86	-8.8
HORVU6Hr1G006310	-0.76	-6.4	HORVU1Hr1G066050	-0.86	-3.5
HORVU4Hr1G023260	-0.76	-4.5	HORVU7Hr1G040030	-0.87	-13.9
HORVU4Hr1G039410	-0.76	-10.5	HORVU4Hr1G054970	-0.87	-3.2
HORVU7Hr1G099560	-0.76	-9.9	HORVU3Hr1G029350	-0.87	-5.7
HORVU1Hr1G064300	-0.76	-2.0	HORVU7Hr1G002390	-0.87	-2.7
HORVU4Hr1G027660	-0.76	-12.4	HORVU4Hr1G063590	-0.88	-2.2
HORVU2Hr1G021090	-0.76	-9.4	HORVU7Hr1G028160	-0.88	-7.7
HORVU2Hr1G044680	-0.77	-2.4	HORVU3Hr1G035820	-0.88	-2.6
HORVU4Hr1G079600	-0.77	-4.8	HORVU6Hr1G013750	-0.88	-3.3
HORVU7Hr1G071300	-0.77	-3.0	HORVU2Hr1G005990	-0.88	-3.3
HORVU5Hr1G093310	-0.77	-4.6	HORVU6Hr1G092990	-0.88	-2.7
HORVU3Hr1G081600	-0.77	-3.2	HORVU2Hr1G065450	-0.88	-5.4
HORVU1Hr1G052930	-0.77	-3.0	HORVU2Hr1G106240	-0.89	-2.8
HORVU7Hr1G050030	-0.77	-3.1	HORVU1Hr1G063010	-0.89	-8.6
HORVU2Hr1G071880	-0.77	-4.6	HORVU3Hr1G083540	-0.89	-2.0
HORVU4Hr1G064440	-0.78	-4.2	HORVU6Hr1G058930	-0.90	-22.3
HORVU2Hr1G071850	-0.78	-4.0	HORVU2Hr1G027080	-0.90	-4.9
HORVU5Hr1G095490	-0.78	-2.1	HORVU7Hr1G098600	-0.90	-4.9
HORVU3Hr1G073740	-0.78	-8.4	HORVU4Hr1G026810	-0.90	-3.1
HORVU5Hr1G011650	-0.78	-2.4	HORVU5Hr1G093860	-0.90	-4.2
HORVU1Hr1G078120	-0.78	-2.5	HORVU7Hr1G006910	-0.91	-2.5
HORVU2Hr1G075010	-0.78	-8.5	HORVU7Hr1G001670	-0.91	-3.5
HORVU7Hr1G035430	-0.78	-5.6	HORVU7Hr1G089540	-0.92	-14.7
HORVU7Hr1G118010	-0.79	-3.3	HORVU1Hr1G004650	-0.92	-4.3
HORVU1Hr1G052860	-0.79	-4.2	HORVU6Hr1G068990	-0.92	-9.0
HORVU5Hr1G099410	-0.79	-4.4	HORVU0Hr1G022610	-0.92	-4.2
HORVU2Hr1G015980	-0.79	-4.2	HORVU5Hr1G120790	-0.93	-3.8
HORVU7Hr1G011890	-0.79	-4.8	HORVU6Hr1G022760	-0.93	-2.1
HORVU4Hr1G067620	-0.79	-14.9	HORVU6Hr1G004380	-0.93	-2.6
HORVU3Hr1G055910	-0.79	-5.1	HORVU4Hr1G080890	-0.93	-2.2
HORVU3Hr1G027700	-0.79	-3.8	HORVU6Hr1G016750	-0.94	-3.1
HORVU6Hr1G002460	-0.79	-2.0	HORVU6Hr1G017860	-0.94	-2.1
HORVU2Hr1G077630	-0.80	-5.7	HORVU7Hr1G010740	-0.94	-2.5
HORVU4Hr1G076690	-0.80	-3.8	HORVU3Hr1G092580	-0.94	-6.5
HORVU7Hr1G047210	-0.80	-2.3	HORVU6Hr1G085710	-0.94	-6.4
HORVU7Hr1G039800	-0.80	-2.8	HORVU5Hr1G062450	-0.94	-11.1
HORVU7Hr1G121100	-0.80	-2.0	HORVU6Hr1G079730	-0.94	-3.6
HORVU2Hr1G060350	-0.80	-6.0	HORVU3Hr1G030340	-0.95	-2.3
HORVU5Hr1G029060	-0.80	-5.1	HORVU5Hr1G059410	-0.95	-5.8
HORVU5Hr1G020320	-0.80	-5.2	HORVU1Hr1G093950	-0.95	-2.9
HORVU5Hr1G012240	-0.81	-2.1	HORVU3Hr1G013180	-0.95	-2.4
HORVU5Hr1G045850	-0.81	-3.4	HORVU5Hr1G023480	-0.96	-3.2
HORVU2Hr1G028540	-0.81	-2.3	HORVU4Hr1G078230	-0.96	-6.0
HORVU7Hr1G118090	-0.81	-3.3	HORVU0Hr1G008640	-0.96	-2.7
HORVU3Hr1G070540	-0.81	-5.6	HORVU3Hr1G050590	-0.96	-3.2
HORVU1Hr1G071430	-0.82	-2.0	HORVU4Hr1G055220	-0.97	-2.7
HORVU1Hr1G079070	-0.82	-4.2	HORVU6Hr1G031360	-0.97	-3.3
HORVU6Hr1G090560	-0.82	-6.9	HORVU1Hr1G057680	-0.97	-6.0
HORVU4Hr1G017860	-0.82	-3.3	HORVU4Hr1G013380	-0.97	-2.9
HORVU2Hr1G117490	-0.82	-2.4	HORVU3Hr1G002210	-0.97	-3.0
HORVU4Hr1G066000	-0.83	-8.4	HORVU3Hr1G070560	-0.98	-2.0
HORVU1Hr1G091540	-0.83	-3.7	HORVU1Hr1G020420	-0.98	-2.4
HORVU6Hr1G066740	-0.83	-4.9	HORVU5Hr1G014170	-0.98	-2.4
HORVU3Hr1G066220	-0.83	-2.8	HORVU3Hr1G010830	-0.98	-2.6
HORVU1Hr1G043550	-0.83	-3.4	HORVU1Hr1G017770	-0.99	-6.2
HORVU3Hr1G013390	-0.83	-7.0	HORVU3Hr1G059060	-0.99	-9.7
HORVU1Hr1G057860	-0.83	-2.0	HORVU6Hr1G004640	-1.00	-2.3
HORVU1Hr1G075580	-0.83	-3.7	HORVU4Hr1G002600	-1.00	-2.0
HORVU5Hr1G006350	-0.83	-6.8	HORVU2Hr1G025400	-1.00	-4.0
HORVU3Hr1G002270	-0.84	-4.2	HORVU0Hr1G040210	-1.00	-3.1
HORVU7Hr1G089450	-0.84	-4.7	HORVU1Hr1G044770	-1.01	-6.5
HORVU3Hr1G075790	-0.84	-8.9	HORVU3Hr1G004090	-1.01	-2.0
HORVU1Hr1G050220	-0.84	-4.3	HORVU3Hr1G098940	-1.01	-3.1
HORVU2Hr1G118110	-0.84	-3.1	HORVU5Hr1G098640	-1.01	-2.3
HORVU0Hr1G000250	-0.84	-3.2	HORVU3Hr1G068380	-1.01	-2.4
HORVU6Hr1G077770	-0.84	-2.5	HORVU4Hr1G050990	-1.02	-2.1
HORVU5Hr1G103430	-0.84	-9.2	HORVU1Hr1G026780	-1.02	-4.4
HORVU5Hr1G048810	-0.85	-3.3	HORVU6Hr1G011790	-1.02	-3.4

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU1Hr1G075570	-1.02	-2.4	HORVU2Hr1G112120	-1.27	-2.9
HORVU1Hr1G070640	-1.03	-2.7	HORVU3Hr1G063620	-1.27	-2.9
HORVU2Hr1G124530	-1.03	-3.2	HORVU2Hr1G125250	-1.27	-4.5
HORVU7Hr1G057460	-1.03	-5.5	HORVU2Hr1G030610	-1.27	-7.0
HORVU5Hr1G063480	-1.03	-3.5	HORVU1Hr1G089840	-1.28	-16.1
HORVU1Hr1G080480	-1.03	-4.8	HORVU5Hr1G004700	-1.29	-6.2
HORVU2Hr1G004230	-1.03	-2.6	HORVU3Hr1G002220	-1.29	-3.3
HORVU2Hr1G107250	-1.04	-6.7	HORVU5Hr1G122950	-1.31	-7.2
HORVU3Hr1G090000	-1.04	-4.6	HORVU6Hr1G090040	-1.31	-4.4
HORVU5Hr1G025690	-1.05	-3.1	HORVU3Hr1G002570	-1.32	-12.0
HORVU3Hr1G087180	-1.05	-3.1	HORVU5Hr1G092770	-1.32	-6.9
HORVU4Hr1G072310	-1.05	-5.6	HORVU5Hr1G124710	-1.32	-5.2
HORVU0Hr1G005140	-1.05	-8.6	HORVU3Hr1G006740	-1.32	-5.7
HORVU6Hr1G058780	-1.05	-9.8	HORVU2Hr1G109040	-1.33	-7.1
HORVU6Hr1G003050	-1.05	-4.5	HORVU3Hr1G116060	-1.34	-28.3
HORVU5Hr1G052310	-1.06	-2.2	HORVU3Hr1G085690	-1.35	-4.8
HORVU7Hr1G001570	-1.06	-2.7	HORVU7Hr1G079600	-1.35	-3.2
HORVU5Hr1G093660	-1.06	-4.6	HORVU5Hr1G112610	-1.35	-5.2
HORVU3Hr1G095360	-1.06	-3.2	HORVU5Hr1G025680	-1.35	-4.1
HORVU4Hr1G002390	-1.07	-2.2	HORVU3Hr1G002370	-1.36	-3.6
HORVU6Hr1G083680	-1.07	-2.5	HORVU7Hr1G109580	-1.37	-11.3
HORVU1Hr1G066530	-1.07	-6.3	HORVU6Hr1G004320	-1.37	-5.2
HORVU3Hr1G007720	-1.07	-2.3	HORVU1Hr1G016080	-1.38	-8.4
HORVU5Hr1G114700	-1.07	-2.3	HORVU2Hr1G014930	-1.38	-5.6
HORVU4Hr1G060230	-1.08	-4.5	HORVU3Hr1G117510	-1.39	-3.8
HORVU4Hr1G005500	-1.08	-6.5	HORVU3Hr1G062430	-1.39	-5.6
HORVU0Hr1G005210	-1.08	-2.4	HORVU4Hr1G077310	-1.39	-3.7
HORVU2Hr1G094870	-1.09	-2.1	HORVU2Hr1G029490	-1.40	-4.9
HORVU2Hr1G062890	-1.09	-4.4	HORVU2Hr1G038940	-1.40	-3.4
HORVU7Hr1G072670	-1.09	-2.1	HORVU2Hr1G079810	-1.40	-5.2
HORVU3Hr1G022960	-1.09	-3.3	HORVU6Hr1G004340	-1.42	-5.3
HORVU1Hr1G072890	-1.10	-3.4	HORVU3Hr1G064710	-1.42	-4.8
HORVU5Hr1G010870	-1.10	-2.1	HORVU7Hr1G074030	-1.44	-5.4
HORVU0Hr1G020630	-1.11	-3.7	HORVU5Hr1G001800	-1.46	-5.3
HORVU5Hr1G122390	-1.11	-2.6	HORVU7Hr1G099520	-1.46	-4.7
HORVU2Hr1G116670	-1.11	-9.3	HORVU5Hr1G077450	-1.47	-10.9
HORVU2Hr1G083100	-1.11	-13.9	HORVU6Hr1G090600	-1.47	-6.3
HORVU4Hr1G001060	-1.12	-8.2	HORVU2Hr1G099540	-1.48	-19.3
HORVU3Hr1G117590	-1.12	-15.7	HORVU1Hr1G001670	-1.48	-8.8
ENSRNA049482489	-1.12	-2.2	HORVU6Hr1G001350	-1.49	-5.7
HORVU5Hr1G095580	-1.12	-6.6	HORVU6Hr1G008550	-1.49	-6.0
HORVU2Hr1G014900	-1.12	-7.0	HORVU3Hr1G097810	-1.51	-8.9
HORVU2Hr1G010270	-1.13	-3.1	HORVU6Hr1G001420	-1.52	-8.5
HORVU4Hr1G081210	-1.13	-2.6	HORVU1Hr1G080520	-1.53	-7.2
HORVU1Hr1G015210	-1.13	-2.3	HORVU3Hr1G002520	-1.55	-5.6
HORVU1Hr1G088680	-1.13	-4.3	HORVU1Hr1G001660	-1.57	-5.4
HORVU4Hr1G071040	-1.13	-11.1	HORVU2Hr1G018570	-1.59	-6.5
HORVU3Hr1G027080	-1.14	-3.3	HORVU5Hr1G010960	-1.60	-5.7
HORVU3Hr1G035470	-1.14	-5.2	HORVU0Hr1G007620	-1.61	-4.7
HORVU2Hr1G063620	-1.14	-2.4	HORVU5Hr1G106370	-1.61	-5.0
HORVU7Hr1G096690	-1.15	-2.3	HORVU6Hr1G078260	-1.66	-5.3
HORVU5Hr1G097900	-1.15	-4.0	HORVU2Hr1G008140	-1.69	-7.6
HORVU3Hr1G013380	-1.15	-3.9	HORVU2Hr1G110080	-1.73	-6.5
HORVU1Hr1G005110	-1.16	-3.9	HORVU0Hr1G021970	-1.74	-13.9
HORVU3Hr1G056960	-1.17	-7.1	HORVU2Hr1G010250	-1.75	-6.9
HORVU7Hr1G028730	-1.17	-3.7	HORVU2Hr1G127440	-1.87	-17.1
HORVU1Hr1G076460	-1.19	-3.5	HORVU3Hr1G018390	-1.97	-18.8
HORVU2Hr1G123850	-1.19	-3.5	HORVU3Hr1G114800	-1.99	-29.7
HORVU2Hr1G121050	-1.20	-6.9	HORVU1Hr1G092290	-1.99	-11.2
HORVU1Hr1G095020	-1.21	-5.1	HORVU2Hr1G108460	-1.99	-8.9
HORVU2Hr1G034280	-1.22	-2.6	HORVU3Hr1G116010	-2.01	-8.4
HORVU2Hr1G060040	-1.22	-3.1	HORVU1Hr1G000920	-2.03	-7.6
HORVU3Hr1G095240	-1.22	-2.7	HORVU7Hr1G036960	-2.16	-9.9
HORVU4Hr1G024580	-1.23	-3.8	HORVU4Hr1G000030	-2.19	-26.8
HORVU7Hr1G103870	-1.23	-14.6	HORVU0Hr1G000910	-2.23	-13.7
HORVU4Hr1G076970	-1.23	-2.9	HORVU6Hr1G004440	-2.32	-12.1
HORVU7Hr1G018590	-1.24	-2.9	HORVU4Hr1G000040	-2.35	-12.9
HORVU5Hr1G041660	-1.24	-2.8	HORVU2Hr1G028780	-2.37	-13.1
HORVU2Hr1G010520	-1.25	-7.7	HORVU7Hr1G047910	-2.44	-32.4
HORVU2Hr1G007990	-1.25	-4.9	HORVU5Hr1G057090	-2.73	-15.9
HORVU3Hr1G067330	-1.25	-2.9	HORVU4Hr1G090860	-3.00	-40.9
HORVU5Hr1G120390	-1.25	-73.5			

Table A.9.2. Differentially expressed genes in salt stressed barley root tissue. Genes found to be differentially expressed in 100 mM salt treated plants relative to control plants using DESeq2 with adjusted p-value < 0.01 and $|\log_2(\text{fold change})| > \log_2(1.5)$. The adjusted p-value is shown as \log_{10} as the values are small.

Gene	$\log_2(\text{fold change})$	$\log_{10}(\text{adj-p})$	Gene	$\log_2(\text{fold change})$	$\log_{10}(\text{adj-p})$
HORVU3Hr1G065420	3.10	-48.9	HORVU2Hr1G085500	1.40	-4.7
HORVU7Hr1G106900	2.49	-34.7	HORVU7Hr1G122800	1.40	-3.7
HORVU2Hr1G030870	2.40	-37.6	HORVU7Hr1G117850	1.40	-4.3
HORVU3Hr1G009360	2.20	-14.3	HORVU5Hr1G080790	1.39	-6.8
HORVU4Hr1G087870	2.19	-16.0	HORVU2Hr1G004540	1.39	-6.5
HORVU3Hr1G093170	2.10	-12.8	HORVU0Hr1G000850	1.38	-8.2
HORVU2Hr1G073210	2.07	-10.6	HORVU1Hr1G069830	1.38	-7.0
HORVU1Hr1G092310	1.98	-9.0	HORVU3Hr1G082870	1.38	-13.9
HORVU2Hr1G011550	1.97	-8.0	HORVU6Hr1G012800	1.38	-19.8
HORVU4Hr1G071300	1.90	-8.3	HORVU5Hr1G006910	1.37	-3.6
HORVU3Hr1G082370	1.86	-10.8	HORVU7Hr1G116310	1.37	-10.3
HORVU2Hr1G085270	1.84	-11.6	HORVU2Hr1G004550	1.36	-6.9
HORVU2Hr1G079890	1.84	-12.6	HORVU4Hr1G087430	1.36	-8.4
HORVU3Hr1G085760	1.84	-15.3	HORVU7Hr1G030690	1.36	-8.5
HORVU3Hr1G013290	1.83	-12.8	HORVU3Hr1G053210	1.36	-4.2
HORVU3Hr1G093160	1.83	-6.9	HORVU7Hr1G019380	1.35	-6.4
HORVU4Hr1G018150	1.83	-21.4	HORVU1Hr1G000340	1.35	-4.3
HORVU2Hr1G012280	1.83	-14.0	HORVU2Hr1G117540	1.35	-8.2
HORVU2Hr1G012220	1.79	-10.6	HORVU4Hr1G075470	1.34	-5.8
HORVU1Hr1G059020	1.76	-11.9	HORVU1Hr1G005230	1.34	-7.8
HORVU5Hr1G094080	1.73	-8.6	HORVU5Hr1G075050	1.33	-6.4
HORVU3Hr1G076640	1.72	-11.7	HORVU3Hr1G074360	1.33	-3.2
HORVU6Hr1G000480	1.72	-10.7	HORVU2Hr1G116730	1.32	-4.5
HORVU2Hr1G118010	1.71	-7.0	HORVU0Hr1G002720	1.32	-5.3
HORVU7Hr1G036720	1.69	-5.3	HORVU5Hr1G000420	1.32	-3.9
HORVU1Hr1G092660	1.67	-5.6	HORVU2Hr1G073680	1.32	-15.5
HORVU2Hr1G004610	1.66	-11.4	HORVU2Hr1G015720	1.31	-3.8
HORVU2Hr1G002720	1.64	-5.4	HORVU6Hr1G091490	1.31	-4.9
HORVU6Hr1G008640	1.63	-12.9	HORVU2Hr1G124320	1.31	-5.8
HORVU2Hr1G097480	1.63	-9.1	HORVU4Hr1G022280	1.31	-3.7
HORVU4Hr1G073790	1.62	-5.1	HORVU7Hr1G020830	1.30	-3.5
HORVU2Hr1G012340	1.62	-10.2	HORVU2Hr1G114450	1.29	-3.9
HORVU3Hr1G076630	1.61	-11.7	HORVU7Hr1G008830	1.29	-5.0
HORVU2Hr1G029290	1.61	-8.6	HORVU5Hr1G007660	1.29	-6.1
HORVU5Hr1G057090	1.60	-4.9	HORVU2Hr1G073670	1.29	-6.6
HORVU3Hr1G100190	1.58	-7.2	HORVU3Hr1G092520	1.28	-3.3
HORVU1Hr1G074230	1.57	-5.3	HORVU2Hr1G120190	1.28	-14.5
HORVU7Hr1G086890	1.57	-12.9	HORVU7Hr1G001150	1.28	-4.6
HORVU1Hr1G005510	1.55	-11.7	HORVU5Hr1G111210	1.28	-5.6
HORVU5Hr1G067760	1.54	-7.5	HORVU2Hr1G111760	1.28	-9.0
HORVU3Hr1G076650	1.54	-8.2	HORVU6Hr1G034250	1.27	-3.6
HORVU6Hr1G025830	1.53	-13.2	HORVU2Hr1G117490	1.27	-8.0
HORVU5Hr1G047920	1.53	-6.1	HORVU4Hr1G013840	1.27	-5.4
HORVU7Hr1G020770	1.53	-6.2	HORVU3Hr1G111600	1.27	-3.8
HORVU0Hr1G021480	1.52	-5.3	HORVU7Hr1G021000	1.27	-3.0
HORVU6Hr1G090500	1.52	-13.0	HORVU3Hr1G065320	1.26	-3.4
HORVU3Hr1G006230	1.49	-5.9	HORVU5Hr1G012290	1.25	-4.4
HORVU2Hr1G004620	1.48	-8.0	HORVU2Hr1G044460	1.24	-10.4
HORVU2Hr1G004530	1.47	-8.3	HORVU3Hr1G109430	1.24	-3.3
HORVU3Hr1G115700	1.47	-9.8	HORVU2Hr1G040540	1.24	-4.9
HORVU1Hr1G008300	1.47	-13.2	HORVU5Hr1G059300	1.23	-10.4
HORVU2Hr1G114390	1.47	-5.3	HORVU4Hr1G022270	1.23	-4.0
HORVU3Hr1G099760	1.46	-16.2	HORVU1Hr1G080850	1.23	-10.1
HORVU2Hr1G003400	1.46	-4.5	HORVU1Hr1G007830	1.23	-2.7
HORVU2Hr1G004510	1.46	-11.7	HORVU5Hr1G125800	1.23	-3.4
HORVU1Hr1G084900	1.46	-9.2	HORVU3Hr1G076660	1.22	-3.4
HORVU7Hr1G008260	1.46	-12.1	HORVU4Hr1G018180	1.22	-2.9
HORVU3Hr1G098910	1.45	-6.1	HORVU2Hr1G038140	1.22	-8.6
HORVU2Hr1G004600	1.44	-11.0	HORVU0Hr1G015930	1.22	-4.8
HORVU7Hr1G083230	1.44	-5.0	HORVU7Hr1G106660	1.22	-19.8
HORVU2Hr1G004480	1.43	-11.4	HORVU5Hr1G016810	1.21	-5.0
HORVU1Hr1G058940	1.42	-8.5	HORVU2Hr1G001960	1.21	-5.1
HORVU4Hr1G087400	1.42	-13.2	HORVU2Hr1G106140	1.21	-3.1
HORVU4Hr1G081330	1.42	-8.6	HORVU3Hr1G006130	1.21	-7.2
HORVU2Hr1G033090	1.41	-7.1	HORVU5Hr1G084980	1.20	-7.6
HORVU2Hr1G126150	1.41	-6.6	HORVU4Hr1G011760	1.20	-4.4
HORVU5Hr1G065620	1.40	-9.1	HORVU7Hr1G019390	1.20	-5.0

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU1Hr1G092360	1.20	-4.0	HORVU3Hr1G064990	1.09	-3.2
HORVU3Hr1G113620	1.19	-3.1	HORVU5Hr1G024550	1.09	-4.1
HORVU5Hr1G103820	1.19	-4.9	HORVU6Hr1G054940	1.09	-7.4
HORVU3Hr1G093100	1.19	-2.7	HORVU2Hr1G101710	1.09	-3.7
HORVU4Hr1G075460	1.19	-2.9	HORVU3Hr1G116310	1.09	-6.5
HORVU4Hr1G059930	1.18	-2.5	HORVU7Hr1G119840	1.09	-2.8
HORVU2Hr1G070310	1.18	-5.8	HORVU1Hr1G083210	1.09	-2.8
HORVU1Hr1G005690	1.18	-3.4	HORVU5Hr1G051970	1.09	-2.6
HORVU2Hr1G096890	1.18	-4.7	HORVU3Hr1G026540	1.08	-3.5
HORVU4Hr1G066000	1.18	-7.3	HORVU2Hr1G109640	1.08	-5.5
HORVU6Hr1G013710	1.18	-4.2	HORVU2Hr1G044450	1.08	-2.8
HORVU5Hr1G047610	1.18	-6.2	HORVU5Hr1G014130	1.08	-2.5
HORVU0Hr1G020630	1.17	-3.5	HORVU3Hr1G059170	1.08	-2.5
HORVU0Hr1G039010	1.17	-4.4	HORVU2Hr1G114440	1.08	-2.7
HORVU4Hr1G043960	1.17	-4.3	HORVU7Hr1G084140	1.08	-5.8
HORVU2Hr1G116800	1.17	-4.1	HORVU5Hr1G118800	1.08	-3.7
HORVU1Hr1G000470	1.17	-3.7	HORVU6Hr1G015480	1.08	-2.3
HORVU3Hr1G087690	1.17	-6.1	HORVU3Hr1G074950	1.07	-2.3
HORVU4Hr1G053190	1.17	-3.6	HORVU4Hr1G060370	1.07	-2.8
HORVU6Hr1G061420	1.17	-5.8	HORVU2Hr1G117680	1.07	-3.3
HORVU7Hr1G091860	1.17	-3.2	HORVU2Hr1G035160	1.07	-4.4
HORVU3Hr1G081970	1.16	-2.6	HORVU2Hr1G010060	1.07	-2.7
HORVU1Hr1G064680	1.16	-2.7	HORVU4Hr1G088930	1.07	-2.6
HORVU3Hr1G074960	1.15	-5.1	HORVU5Hr1G023730	1.07	-4.0
HORVU7Hr1G089310	1.15	-3.1	HORVU1Hr1G000350	1.06	-3.2
HORVU5Hr1G105900	1.15	-2.6	HORVU6Hr1G020540	1.06	-3.2
HORVU7Hr1G116080	1.15	-2.4	HORVU5Hr1G094700	1.06	-2.7
HORVU7Hr1G048820	1.15	-2.4	HORVU7Hr1G002370	1.06	-4.1
HORVU7Hr1G001190	1.15	-5.4	HORVU7Hr1G009140	1.06	-2.5
HORVU6Hr1G084080	1.15	-5.2	HORVU4Hr1G071020	1.06	-3.7
HORVU6Hr1G014600	1.14	-2.7	HORVU4Hr1G072500	1.06	-3.7
HORVU5Hr1G056300	1.14	-2.4	HORVU3Hr1G076840	1.05	-2.6
HORVU5Hr1G016780	1.14	-7.3	HORVU5Hr1G094710	1.05	-3.3
HORVU6Hr1G091840	1.14	-4.2	HORVU2Hr1G003070	1.05	-2.4
HORVU1Hr1G008540	1.13	-7.3	HORVU7Hr1G078670	1.05	-4.2
HORVU3Hr1G006240	1.13	-2.6	HORVU4Hr1G001250	1.05	-3.3
HORVU3Hr1G004840	1.13	-2.9	HORVU5Hr1G110220	1.04	-3.0
HORVU2Hr1G076320	1.13	-3.0	HORVU7Hr1G078960	1.04	-3.8
HORVU7Hr1G087250	1.13	-3.4	HORVU4Hr1G080820	1.04	-2.4
HORVU3Hr1G107160	1.13	-6.0	HORVU5Hr1G123310	1.04	-2.2
HORVU5Hr1G070400	1.13	-5.6	HORVU7Hr1G120270	1.04	-2.5
HORVU2Hr1G095210	1.13	-3.6	HORVU4Hr1G090190	1.04	-3.4
HORVU1Hr1G088300	1.13	-4.2	HORVU3Hr1G081580	1.04	-3.8
HORVU5Hr1G052150	1.13	-3.2	HORVU3Hr1G076670	1.04	-9.5
HORVU2Hr1G124430	1.13	-5.3	HORVU2Hr1G018480	1.04	-2.1
HORVU2Hr1G000090	1.13	-2.7	HORVU2Hr1G022820	1.04	-3.4
HORVU2Hr1G124310	1.12	-4.4	HORVU2Hr1G043890	1.04	-5.2
HORVU2Hr1G021530	1.12	-2.3	HORVU1Hr1G091220	1.04	-5.8
HORVU3Hr1G071470	1.12	-2.4	HORVU6Hr1G084030	1.04	-3.8
HORVU5Hr1G110180	1.12	-4.2	HORVU1Hr1G082910	1.04	-2.4
HORVU7Hr1G028870	1.12	-3.1	HORVU0Hr1G019800	1.04	-3.8
HORVU0Hr1G011720	1.12	-3.8	HORVU1Hr1G057220	1.03	-3.0
HORVU4Hr1G004820	1.12	-3.0	HORVU1Hr1G061380	1.03	-3.0
HORVU3Hr1G091680	1.11	-8.6	HORVU5Hr1G055570	1.03	-3.1
HORVU7Hr1G027440	1.11	-4.5	HORVU5Hr1G119270	1.03	-5.7
HORVU3Hr1G068760	1.11	-3.1	HORVU2Hr1G085280	1.03	-2.9
HORVU2Hr1G108530	1.11	-2.4	HORVU5Hr1G006880	1.03	-4.6
HORVU2Hr1G124300	1.11	-3.6	HORVU1Hr1G092640	1.02	-2.9
HORVU5Hr1G009220	1.11	-4.8	HORVU6Hr1G020690	1.02	-3.8
HORVU4Hr1G079620	1.11	-3.3	HORVU4Hr1G078220	1.02	-3.5
HORVU5Hr1G124810	1.10	-2.3	HORVU5Hr1G099470	1.02	-3.5
HORVU1Hr1G044000	1.10	-6.7	HORVU4Hr1G079600	1.02	-4.4
HORVU2Hr1G011720	1.10	-3.4	HORVU3Hr1G083670	1.02	-2.2
HORVU2Hr1G105560	1.10	-13.8	HORVU2Hr1G006830	1.02	-2.2
HORVU1Hr1G089520	1.10	-2.4	HORVU5Hr1G064280	1.02	-2.7
HORVU1Hr1G000450	1.10	-2.5	HORVU2Hr1G105510	1.01	-7.1
HORVU1Hr1G012680	1.10	-3.9	HORVU1Hr1G013450	1.01	-2.6
HORVU1Hr1G092240	1.10	-6.4	HORVU2Hr1G110420	1.01	-4.4
HORVU4Hr1G000050	1.10	-4.0	HORVU5Hr1G023660	1.01	-3.4
HORVU6Hr1G084370	1.10	-2.2	HORVU4Hr1G023680	1.01	-2.6
HORVU3Hr1G018980	1.10	-4.1	HORVU1Hr1G045020	1.01	-2.1
HORVU7Hr1G104180	1.10	-3.2	HORVU3Hr1G003860	1.01	-2.3
HORVU5Hr1G023940	1.09	-7.0	HORVU3Hr1G081180	1.01	-2.6
HORVU2Hr1G108260	1.09	-9.3	HORVU6Hr1G065080	1.01	-5.8
HORVU3Hr1G108100	1.09	-8.6	HORVU7Hr1G088790	1.01	-2.2

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU2Hr1G125050	1.01	-5.7	HORVU1Hr1G093890	0.91	-3.8
HORVU7Hr1G093630	1.00	-2.9	HORVU6Hr1G022360	0.90	-11.4
HORVU2Hr1G019400	1.00	-2.6	HORVU0Hr1G008830	0.90	-4.3
HORVU5Hr1G097320	1.00	-3.1	HORVU1Hr1G088580	0.90	-2.6
HORVU0Hr1G031210	1.00	-2.7	HORVU0Hr1G013950	0.90	-2.1
HORVU2Hr1G047260	1.00	-5.1	HORVU4Hr1G088580	0.90	-2.2
HORVU4Hr1G074480	1.00	-3.6	HORVU7Hr1G077590	0.90	-3.0
HORVU7Hr1G105170	0.99	-3.8	HORVU5Hr1G066360	0.90	-2.5
HORVU7Hr1G101580	0.99	-3.4	HORVU5Hr1G108610	0.90	-4.5
HORVU5Hr1G012320	0.99	-2.2	HORVU6Hr1G031640	0.90	-4.3
HORVU0Hr1G011730	0.99	-2.2	HORVU4Hr1G027180	0.89	-5.6
HORVU2Hr1G090250	0.99	-2.0	HORVU5Hr1G075670	0.89	-6.9
HORVU2Hr1G123050	0.99	-2.5	HORVU1Hr1G019500	0.89	-4.4
HORVU2Hr1G007720	0.99	-3.8	HORVU6Hr1G086410	0.89	-2.3
HORVU1Hr1G072690	0.99	-3.1	HORVU7Hr1G088920	0.89	-2.1
HORVU7Hr1G098280	0.99	-3.2	HORVU2Hr1G034740	0.89	-3.4
HORVU2Hr1G086050	0.99	-6.7	HORVU5Hr1G096000	0.89	-3.5
HORVU6Hr1G027930	0.98	-3.9	HORVU4Hr1G053150	0.88	-2.9
HORVU4Hr1G073580	0.98	-6.9	HORVU3Hr1G096280	0.88	-3.5
HORVU7Hr1G040460	0.98	-2.2	HORVU1Hr1G071260	0.88	-4.7
HORVU7Hr1G045150	0.98	-3.2	HORVU4Hr1G072560	0.88	-3.1
HORVU2Hr1G102100	0.98	-5.2	HORVU7Hr1G098600	0.88	-2.7
HORVU5Hr1G122410	0.98	-2.6	HORVU3Hr1G026330	0.88	-2.7
HORVU1Hr1G054230	0.98	-4.2	HORVU4Hr1G022620	0.88	-2.3
HORVU5Hr1G023720	0.98	-3.3	HORVU3Hr1G002080	0.88	-3.1
HORVU1Hr1G049100	0.98	-3.0	HORVU5Hr1G012120	0.87	-2.9
HORVU6Hr1G091180	0.98	-3.8	HORVU5Hr1G047150	0.87	-2.2
HORVU5Hr1G102900	0.98	-2.4	HORVU5Hr1G122510	0.87	-2.6
HORVU2Hr1G104220	0.97	-2.6	HORVU6Hr1G067740	0.87	-2.4
HORVU1Hr1G057560	0.97	-2.1	HORVU6Hr1G035370	0.87	-2.0
HORVU5Hr1G119050	0.97	-3.4	HORVU7Hr1G104350	0.87	-2.2
HORVU4Hr1G088080	0.97	-4.0	HORVU5Hr1G015140	0.87	-2.3
HORVU6Hr1G023210	0.96	-3.4	HORVU5Hr1G019030	0.87	-4.5
HORVU5Hr1G122060	0.96	-4.1	HORVU4Hr1G064790	0.86	-4.8
HORVU1Hr1G015190	0.96	-2.4	HORVU5Hr1G123280	0.86	-2.1
HORVU3Hr1G091320	0.96	-5.9	HORVU1Hr1G019380	0.86	-4.2
HORVU3Hr1G080550	0.96	-2.2	HORVU7Hr1G028290	0.86	-2.2
HORVU0Hr1G030160	0.96	-2.6	HORVU3Hr1G022780	0.86	-2.3
HORVU5Hr1G106850	0.96	-2.9	HORVU3Hr1G097860	0.86	-2.2
HORVU5Hr1G097940	0.96	-3.0	HORVU4Hr1G010470	0.86	-2.5
HORVU7Hr1G029460	0.95	-2.7	HORVU4Hr1G017390	0.86	-3.8
HORVU4Hr1G043990	0.95	-4.0	HORVU7Hr1G026070	0.86	-6.4
HORVU5Hr1G014090	0.95	-2.1	HORVU5Hr1G115880	0.86	-2.1
HORVU7Hr1G118770	0.95	-5.2	HORVU3Hr1G076620	0.86	-3.6
HORVU4Hr1G053790	0.95	-2.5	HORVU4Hr1G086520	0.86	-3.4
HORVU2Hr1G043900	0.94	-6.7	HORVU5Hr1G067390	0.86	-2.2
HORVU3Hr1G116400	0.94	-2.0	HORVU5Hr1G053790	0.86	-2.9
HORVU1Hr1G025440	0.94	-4.1	HORVU0Hr1G005300	0.85	-3.2
HORVU5Hr1G077670	0.94	-2.3	HORVU1Hr1G000230	0.85	-4.5
HORVU7Hr1G002740	0.94	-3.8	HORVU3Hr1G000390	0.85	-6.2
HORVU2Hr1G110230	0.94	-2.4	HORVU1Hr1G090860	0.85	-3.3
HORVU2Hr1G034860	0.93	-3.9	HORVU5Hr1G063320	0.85	-3.2
HORVU7Hr1G116830	0.93	-7.4	HORVU4Hr1G020970	0.85	-3.1
HORVU4Hr1G055620	0.93	-3.7	HORVU2Hr1G122610	0.84	-2.5
HORVU4Hr1G058160	0.93	-2.0	HORVU4Hr1G078010	0.84	-2.2
HORVU2Hr1G016290	0.93	-2.4	HORVU4Hr1G010740	0.84	-2.8
HORVU3Hr1G013470	0.93	-6.3	HORVU5Hr1G093290	0.84	-8.8
HORVU2Hr1G003230	0.93	-2.4	HORVU1Hr1G018010	0.84	-2.3
HORVU6Hr1G094070	0.93	-2.2	HORVU2Hr1G102110	0.84	-5.1
HORVU5Hr1G094750	0.93	-3.2	HORVU0Hr1G009470	0.84	-2.3
HORVU5Hr1G062300	0.93	-2.4	HORVU2Hr1G014890	0.84	-3.1
HORVU5Hr1G045990	0.93	-7.5	HORVU2Hr1G096230	0.84	-2.1
HORVU2Hr1G121340	0.92	-2.2	HORVU1Hr1G021170	0.84	-3.0
HORVU5Hr1G081950	0.92	-4.3	HORVU3Hr1G111150	0.84	-8.5
HORVU3Hr1G115610	0.92	-6.0	HORVU3Hr1G077950	0.83	-7.4
HORVU1Hr1G065100	0.92	-2.1	HORVU7Hr1G101590	0.83	-4.6
HORVU4Hr1G068530	0.92	-19.4	HORVU5Hr1G028030	0.83	-9.0
HORVU2Hr1G062910	0.92	-4.0	HORVU2Hr1G029840	0.83	-2.3
HORVU4Hr1G010720	0.91	-3.8	HORVU3Hr1G097160	0.83	-3.6
HORVU4Hr1G053980	0.91	-4.1	HORVU0Hr1G016410	0.83	-5.3
HORVU6Hr1G018510	0.91	-2.8	HORVU7Hr1G019400	0.83	-3.4
HORVU7Hr1G043880	0.91	-4.6	HORVU7Hr1G113510	0.82	-2.6
HORVU3Hr1G016970	0.91	-2.3	HORVU6Hr1G090910	0.82	-2.2
HORVU2Hr1G116670	0.91	-2.9	HORVU1Hr1G020450	0.82	-2.0
HORVU5Hr1G010880	0.91	-2.9	HORVU7Hr1G029750	0.82	-3.1

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU3Hr1G031000	0.82	-2.3	HORVU5Hr1G124650	0.73	-3.8
HORVU7Hr1G029870	0.82	-4.2	HORVU1Hr1G065150	0.72	-3.1
HORVU6Hr1G080050	0.82	-2.6	HORVU0Hr1G009400	0.72	-3.1
HORVU2Hr1G079840	0.81	-8.1	HORVU1Hr1G048450	0.72	-3.1
HORVU5Hr1G005310	0.81	-6.6	HORVU1Hr1G064970	0.72	-2.3
HORVU4Hr1G020320	0.81	-3.0	HORVU1Hr1G075760	0.72	-4.4
HORVU1Hr1G039670	0.81	-2.8	HORVU3Hr1G067380	0.72	-2.5
HORVU1Hr1G058520	0.81	-2.9	HORVU4Hr1G071140	0.72	-4.1
HORVU4Hr1G007190	0.81	-2.5	HORVU2Hr1G116960	0.72	-2.9
HORVU5Hr1G103430	0.80	-5.8	HORVU3Hr1G114990	0.72	-3.0
HORVU6Hr1G001200	0.80	-5.0	HORVU2Hr1G062700	0.71	-3.0
HORVU2Hr1G081670	0.80	-3.1	HORVU0Hr1G000760	0.71	-3.2
HORVU5Hr1G047250	0.80	-2.1	HORVU2Hr1G072180	0.71	-2.3
HORVU4Hr1G005440	0.80	-2.3	HORVU6Hr1G058840	0.71	-4.0
HORVU2Hr1G045200	0.79	-2.5	HORVU7Hr1G012560	0.71	-2.4
HORVU2Hr1G082160	0.79	-4.0	HORVU2Hr1G044590	0.71	-2.7
HORVU7Hr1G115370	0.79	-5.8	HORVU1Hr1G019410	0.71	-2.6
HORVU4Hr1G083650	0.78	-3.7	HORVU1Hr1G063420	0.71	-3.7
HORVU5Hr1G019820	0.78	-4.4	HORVU2Hr1G005550	0.71	-3.8
HORVU5Hr1G062940	0.78	-7.4	HORVU4Hr1G072880	0.71	-2.4
HORVU2Hr1G108330	0.78	-2.1	HORVU2Hr1G124980	0.71	-3.6
HORVU4Hr1G054120	0.78	-6.0	HORVU2Hr1G036800	0.70	-3.4
HORVU4Hr1G080350	0.78	-2.6	HORVU5Hr1G104580	0.70	-3.1
HORVU4Hr1G010410	0.78	-8.8	HORVU2Hr1G013440	0.70	-2.8
HORVU5Hr1G087440	0.78	-4.3	HORVU6Hr1G075950	0.70	-2.5
HORVU4Hr1G008550	0.78	-2.0	HORVU3Hr1G104360	0.70	-3.4
HORVU6Hr1G026810	0.77	-2.7	HORVU1Hr1G049230	0.70	-5.0
HORVU2Hr1G093950	0.77	-2.8	HORVU1Hr1G076460	0.70	-3.8
HORVU1Hr1G074380	0.77	-2.6	HORVU3Hr1G029520	0.69	-3.6
HORVU4Hr1G018800	0.77	-2.4	HORVU1Hr1G093480	0.69	-7.8
HORVU1Hr1G070460	0.77	-2.3	HORVU5Hr1G119830	0.69	-2.0
HORVU1Hr1G021160	0.77	-2.4	HORVU7Hr1G085270	0.69	-2.1
HORVU6Hr1G021320	0.77	-3.0	HORVU2Hr1G018710	0.69	-2.4
HORVU7Hr1G017640	0.77	-2.5	HORVU5Hr1G058000	0.69	-2.6
HORVU2Hr1G005470	0.76	-4.6	HORVU2Hr1G113180	0.69	-3.7
HORVU7Hr1G035190	0.76	-2.3	HORVU2Hr1G102050	0.69	-2.5
HORVU5Hr1G095430	0.76	-3.2	HORVU2Hr1G040570	0.69	-2.9
HORVU6Hr1G065960	0.76	-3.6	HORVU7Hr1G079600	0.69	-2.7
HORVU2Hr1G033100	0.76	-6.3	HORVU1Hr1G001390	0.69	-3.2
HORVU0Hr1G014630	0.76	-3.0	HORVU6Hr1G083720	0.69	-4.3
HORVU1Hr1G067370	0.76	-5.5	HORVU4Hr1G000620	0.68	-2.3
HORVU0Hr1G000780	0.76	-5.0	HORVU7Hr1G114660	0.68	-4.2
HORVU3Hr1G065810	0.76	-4.3	HORVU3Hr1G099570	0.68	-2.2
HORVU2Hr1G102710	0.76	-3.0	HORVU3Hr1G013970	0.68	-2.1
HORVU0Hr1G018190	0.76	-2.2	HORVU3Hr1G078840	0.68	-3.4
HORVU5Hr1G064000	0.76	-3.8	HORVU7Hr1G108150	0.67	-3.2
HORVU5Hr1G067880	0.76	-6.0	HORVU6Hr1G035040	0.67	-2.8
HORVU2Hr1G111780	0.75	-2.1	HORVU2Hr1G024740	0.67	-2.6
HORVU7Hr1G024930	0.75	-2.4	HORVU7Hr1G055410	0.67	-2.9
HORVU4Hr1G089750	0.75	-3.0	HORVU2Hr1G012010	0.67	-2.5
HORVU6Hr1G005400	0.75	-2.2	HORVU7Hr1G114610	0.67	-3.5
HORVU3Hr1G083250	0.75	-2.6	HORVU7Hr1G084500	0.67	-4.5
HORVU4Hr1G063790	0.74	-3.0	HORVU2Hr1G084210	0.67	-2.2
HORVU3Hr1G067840	0.74	-3.0	HORVU7Hr1G012610	0.67	-2.7
HORVU3Hr1G072810	0.74	-2.4	HORVU6Hr1G072990	0.66	-2.0
HORVU7Hr1G105150	0.74	-2.5	HORVU1Hr1G018540	0.66	-2.8
HORVU7Hr1G042260	0.74	-2.3	HORVU7Hr1G053260	0.66	-2.4
HORVU3Hr1G012860	0.74	-2.1	HORVU6Hr1G034900	0.66	-2.0
HORVU5Hr1G083620	0.74	-2.5	HORVU3Hr1G109740	0.66	-2.6
HORVU2Hr1G066480	0.74	-3.9	HORVU7Hr1G074690	0.66	-3.5
HORVU2Hr1G071270	0.74	-2.2	HORVU3Hr1G019840	0.66	-3.5
HORVU7Hr1G121520	0.74	-2.5	HORVU3Hr1G107280	0.65	-2.1
HORVU3Hr1G112520	0.74	-2.5	HORVU4Hr1G061120	0.65	-3.6
HORVU1Hr1G017080	0.74	-2.0	HORVU1Hr1G092200	0.65	-2.7
HORVU7Hr1G077640	0.73	-2.2	HORVU7Hr1G101310	0.65	-2.3
HORVU7Hr1G082450	0.73	-5.3	HORVU4Hr1G010540	0.65	-2.6
HORVU4Hr1G078210	0.73	-2.1	HORVU3Hr1G098340	0.65	-2.7
HORVU2Hr1G122370	0.73	-7.0	HORVU5Hr1G093390	0.65	-4.7
HORVU4Hr1G089870	0.73	-3.6	HORVU4Hr1G087340	0.65	-2.7
HORVU6Hr1G064740	0.73	-2.4	HORVU1Hr1G092290	0.65	-2.3
HORVU3Hr1G066310	0.73	-3.4	HORVU7Hr1G114880	0.65	-4.5
HORVU3Hr1G024210	0.73	-4.6	HORVU3Hr1G092280	0.65	-2.9
HORVU2Hr1G063690	0.73	-5.1	HORVU1Hr1G069620	0.65	-3.5
HORVU3Hr1G036970	0.73	-7.7	HORVU2Hr1G016440	0.64	-4.0
HORVU3Hr1G096570	0.73	-2.1	HORVU3Hr1G076790	0.64	-2.5

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU0Hr1G012460	0.64	-2.2	HORVU4Hr1G009520	-0.63	-4.0
HORVU3Hr1G098360	0.64	-3.9	HORVU2Hr1G098860	-0.63	-3.1
HORVU7Hr1G047700	0.64	-3.4	HORVU2Hr1G071100	-0.63	-2.4
HORVU5Hr1G018240	0.64	-3.8	HORVU3Hr1G099220	-0.63	-2.1
HORVU3Hr1G006830	0.64	-3.1	HORVU3Hr1G024220	-0.63	-2.4
HORVU4Hr1G085310	0.64	-2.2	HORVU7Hr1G096430	-0.63	-4.6
HORVU2Hr1G114580	0.63	-2.5	HORVU6Hr1G003300	-0.63	-2.1
HORVU4Hr1G017430	0.63	-4.5	HORVU4Hr1G016080	-0.63	-2.3
HORVU4Hr1G085050	0.63	-2.4	HORVU0Hr1G032410	-0.64	-2.0
HORVU2Hr1G104690	0.63	-2.5	HORVU5Hr1G086670	-0.64	-2.1
HORVU5Hr1G070160	0.63	-4.5	HORVU3Hr1G047980	-0.64	-2.8
HORVU7Hr1G119020	0.63	-2.5	HORVU6Hr1G071950	-0.64	-2.3
HORVU1Hr1G075550	0.63	-6.2	HORVU7Hr1G073860	-0.64	-2.3
HORVU2Hr1G101140	0.62	-3.3	HORVU7Hr1G109020	-0.64	-2.1
HORVU3Hr1G079620	0.62	-4.2	HORVU7Hr1G118240	-0.64	-5.0
HORVU7Hr1G074640	0.62	-6.4	HORVU5Hr1G045150	-0.64	-5.2
HORVU5Hr1G006630	0.62	-3.4	HORVU3Hr1G098450	-0.64	-2.6
HORVU7Hr1G048570	0.62	-2.1	HORVU7Hr1G073510	-0.64	-3.0
HORVU4Hr1G018430	0.62	-2.3	HORVU2Hr1G005880	-0.64	-2.1
HORVU2Hr1G044890	0.62	-2.1	HORVU3Hr1G060920	-0.64	-2.8
HORVU0Hr1G019840	0.62	-2.9	HORVU4Hr1G026340	-0.64	-3.8
HORVU1Hr1G078350	0.62	-2.6	HORVU5Hr1G119790	-0.64	-12.7
HORVU7Hr1G036420	0.62	-3.1	HORVU1Hr1G063610	-0.64	-2.7
HORVU1Hr1G050780	0.61	-5.1	HORVU5Hr1G048810	-0.65	-2.4
HORVU6Hr1G081000	0.61	-4.5	HORVU2Hr1G014710	-0.65	-3.7
HORVU3Hr1G089170	0.61	-2.8	HORVU2Hr1G060460	-0.65	-4.4
HORVU7Hr1G041610	0.61	-2.0	HORVU2Hr1G123890	-0.65	-3.6
HORVU4Hr1G087390	0.61	-2.1	HORVU7Hr1G103870	-0.65	-4.6
HORVU7Hr1G028850	0.61	-2.5	HORVU7Hr1G021860	-0.65	-5.7
HORVU2Hr1G014130	0.60	-12.4	HORVU6Hr1G064620	-0.65	-2.7
HORVU6Hr1G058000	0.60	-3.3	HORVU7Hr1G076120	-0.66	-6.2
HORVU4Hr1G024530	0.60	-4.4	HORVU3Hr1G058580	-0.66	-2.9
HORVU4Hr1G083210	0.60	-3.3	HORVU5Hr1G042290	-0.66	-2.6
HORVU7Hr1G012530	0.60	-2.2	HORVU6Hr1G037460	-0.66	-2.9
HORVU5Hr1G082050	0.60	-2.3	HORVU0Hr1G010080	-0.66	-2.2
HORVU1Hr1G072780	0.60	-3.5	HORVU2Hr1G096400	-0.66	-2.7
HORVU2Hr1G114190	0.59	-2.3	HORVU1Hr1G017970	-0.66	-3.4
HORVU2Hr1G103180	0.59	-3.0	HORVU7Hr1G120660	-0.67	-2.7
HORVU2Hr1G022780	0.59	-3.7	HORVU2Hr1G097010	-0.67	-4.8
HORVU2Hr1G027640	0.59	-3.7	HORVU6Hr1G032760	-0.67	-2.3
HORVU7Hr1G025390	0.59	-3.5	HORVU7Hr1G023980	-0.67	-3.4
HORVU2Hr1G084750	0.59	-3.1	HORVU7Hr1G115960	-0.67	-2.5
HORVU6Hr1G095330	-0.59	-2.1	HORVU6Hr1G071840	-0.67	-2.4
HORVU1Hr1G035160	-0.59	-2.1	HORVU4Hr1G067410	-0.67	-2.4
HORVU2Hr1G027040	-0.59	-3.1	HORVU2Hr1G060510	-0.67	-3.7
HORVU4Hr1G082240	-0.59	-4.3	HORVU3Hr1G041820	-0.67	-2.2
HORVU5Hr1G095530	-0.59	-4.2	HORVU2Hr1G005570	-0.67	-2.0
HORVU5Hr1G086650	-0.59	-2.0	HORVU5Hr1G084010	-0.67	-9.0
HORVU2Hr1G016760	-0.59	-2.1	HORVU4Hr1G063730	-0.68	-3.3
HORVU3Hr1G108500	-0.59	-2.4	HORVU1Hr1G002820	-0.68	-2.6
HORVU4Hr1G062460	-0.60	-2.2	HORVU4Hr1G081600	-0.68	-5.1
HORVU1Hr1G092730	-0.60	-3.1	HORVU7Hr1G100090	-0.68	-2.2
HORVU1Hr1G013210	-0.60	-3.5	HORVU1Hr1G066610	-0.68	-2.8
HORVU1Hr1G026590	-0.60	-5.3	HORVU4Hr1G056950	-0.68	-3.8
HORVU4Hr1G067620	-0.60	-2.2	HORVU3Hr1G090350	-0.68	-5.3
HORVU2Hr1G003990	-0.60	-2.4	HORVU5Hr1G103790	-0.69	-2.7
HORVU6Hr1G072620	-0.60	-8.8	HORVU7Hr1G009610	-0.69	-2.6
HORVU7Hr1G007900	-0.60	-3.4	HORVU7Hr1G101270	-0.69	-15.3
HORVU1Hr1G092440	-0.60	-3.8	HORVU1Hr1G091880	-0.69	-3.2
HORVU4Hr1G022990	-0.60	-2.8	HORVU5Hr1G059770	-0.69	-7.5
HORVU1Hr1G043280	-0.60	-2.8	HORVU7Hr1G098560	-0.69	-2.6
HORVU1Hr1G010940	-0.60	-3.4	HORVU1Hr1G060530	-0.69	-2.3
HORVU3Hr1G062320	-0.61	-3.7	HORVU7Hr1G021890	-0.69	-2.3
HORVU4Hr1G019990	-0.61	-2.7	HORVU4Hr1G003480	-0.70	-2.0
HORVU5Hr1G098980	-0.61	-3.0	HORVU6Hr1G090780	-0.70	-4.6
HORVU4Hr1G028310	-0.61	-2.0	HORVU3Hr1G009940	-0.70	-2.0
HORVU6Hr1G040770	-0.61	-3.1	HORVU1Hr1G085790	-0.70	-2.4
HORVU5Hr1G109570	-0.61	-2.5	HORVU7Hr1G119570	-0.71	-8.3
HORVU0Hr1G012050	-0.62	-2.4	HORVU5Hr1G114220	-0.71	-2.6
HORVU4Hr1G060440	-0.62	-2.7	HORVU2Hr1G096360	-0.71	-2.4
HORVU3Hr1G047030	-0.62	-2.8	HORVU2Hr1G037550	-0.71	-3.8
HORVU4Hr1G075360	-0.62	-2.8	HORVU7Hr1G048710	-0.71	-3.1
HORVU2Hr1G113830	-0.62	-3.2	HORVU6Hr1G072710	-0.71	-2.6
HORVU2Hr1G097780	-0.63	-2.1	HORVU2Hr1G046050	-0.71	-3.6
HORVU0Hr1G040310	-0.63	-2.1	HORVU7Hr1G074050	-0.71	-4.4

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU2Hr1G072040	-0.71	-2.8	HORVU0Hr1G017570	-0.79	-2.7
HORVU5Hr1G122080	-0.71	-2.3	HORVU2Hr1G105890	-0.79	-5.1
HORVU3Hr1G018110	-0.71	-3.8	HORVU2Hr1G113270	-0.79	-2.2
HORVU5Hr1G016730	-0.72	-2.4	HORVU2Hr1G087860	-0.79	-6.3
HORVU2Hr1G029070	-0.72	-4.9	HORVU7Hr1G108000	-0.80	-2.9
HORVU2Hr1G094760	-0.72	-2.8	HORVU1Hr1G008130	-0.80	-3.6
HORVU0Hr1G008590	-0.72	-3.5	HORVU6Hr1G005580	-0.80	-2.5
HORVU1Hr1G091970	-0.72	-3.2	HORVU2Hr1G112050	-0.80	-2.5
HORVU5Hr1G076730	-0.72	-3.1	HORVU5Hr1G005650	-0.80	-2.7
HORVU6Hr1G059970	-0.72	-2.9	HORVU6Hr1G064430	-0.80	-2.7
HORVU4Hr1G077070	-0.72	-3.9	HORVU2Hr1G009980	-0.80	-2.0
HORVU1Hr1G040160	-0.72	-3.0	HORVU6Hr1G089540	-0.80	-4.0
HORVU1Hr1G046010	-0.72	-4.8	HORVU6Hr1G046980	-0.80	-5.2
HORVU1Hr1G076660	-0.72	-2.7	HORVU7Hr1G091250	-0.80	-3.0
HORVU2Hr1G018970	-0.72	-2.8	HORVU2Hr1G012440	-0.80	-2.7
HORVU6Hr1G000090	-0.72	-2.4	HORVU1Hr1G059550	-0.80	-3.7
HORVU4Hr1G085090	-0.73	-3.5	HORVU5Hr1G065200	-0.81	-3.0
HORVU4Hr1G083030	-0.73	-3.2	HORVU6Hr1G089600	-0.81	-4.6
HORVU1Hr1G003230	-0.73	-4.6	HORVU6Hr1G054110	-0.81	-2.3
HORVU1Hr1G085840	-0.73	-2.5	HORVU6Hr1G029520	-0.82	-2.3
HORVU5Hr1G122290	-0.73	-2.6	HORVU1Hr1G001850	-0.82	-6.0
HORVU3Hr1G070540	-0.73	-6.5	HORVU2Hr1G113260	-0.82	-2.6
HORVU6Hr1G075650	-0.73	-2.3	HORVU6Hr1G010150	-0.82	-2.3
HORVU5Hr1G062030	-0.73	-4.2	HORVU1Hr1G012240	-0.82	-2.0
HORVU1Hr1G049760	-0.73	-2.6	HORVU7Hr1G091100	-0.82	-2.3
HORVU4Hr1G003220	-0.73	-2.8	HORVU4Hr1G013170	-0.82	-2.4
HORVU7Hr1G108960	-0.73	-4.9	HORVU2Hr1G007580	-0.82	-2.2
HORVU7Hr1G038330	-0.73	-3.7	HORVU4Hr1G053530	-0.82	-3.0
HORVU1Hr1G022710	-0.74	-3.2	HORVU7Hr1G089240	-0.83	-2.6
HORVU2Hr1G064640	-0.74	-3.1	HORVU6Hr1G095020	-0.83	-3.4
HORVU2Hr1G007530	-0.74	-2.0	HORVU6Hr1G092430	-0.83	-3.9
HORVU4Hr1G082710	-0.74	-3.2	HORVU2Hr1G105130	-0.83	-2.1
HORVU7Hr1G089850	-0.74	-2.4	HORVU3Hr1G112020	-0.83	-2.3
HORVU3Hr1G112010	-0.74	-2.5	HORVU4Hr1G047260	-0.83	-2.6
HORVU4Hr1G074530	-0.74	-2.5	HORVU5Hr1G114330	-0.83	-2.8
HORVU2Hr1G116880	-0.75	-3.6	HORVU3Hr1G079560	-0.83	-2.4
HORVU5Hr1G001030	-0.75	-3.5	HORVU5Hr1G012900	-0.84	-3.0
HORVU7Hr1G049400	-0.75	-3.6	HORVU3Hr1G079230	-0.84	-2.3
HORVU4Hr1G089540	-0.76	-3.5	HORVU3Hr1G081600	-0.84	-3.2
HORVU2Hr1G015360	-0.76	-2.1	HORVU4Hr1G000770	-0.84	-4.5
HORVU7Hr1G036180	-0.76	-3.1	HORVU7Hr1G011250	-0.84	-2.3
HORVU4Hr1G067930	-0.76	-9.1	HORVU4Hr1G076690	-0.84	-7.1
HORVU3Hr1G075770	-0.76	-2.7	HORVU1Hr1G049440	-0.84	-2.9
HORVU4Hr1G052470	-0.76	-3.3	HORVU7Hr1G027630	-0.85	-2.9
HORVU1Hr1G002410	-0.76	-2.5	HORVU6Hr1G067430	-0.85	-4.2
HORVU3Hr1G024920	-0.76	-3.8	HORVU4Hr1G076760	-0.85	-2.0
HORVU4Hr1G006790	-0.76	-2.2	HORVU5Hr1G086970	-0.85	-5.3
HORVU1Hr1G047090	-0.76	-2.2	HORVU3Hr1G057140	-0.85	-3.5
HORVU4Hr1G089030	-0.77	-2.8	HORVU1Hr1G073870	-0.85	-2.6
HORVU3Hr1G023910	-0.77	-4.3	HORVU7Hr1G047060	-0.85	-2.8
HORVU6Hr1G005570	-0.77	-2.6	HORVU7Hr1G083360	-0.86	-2.1
HORVU1Hr1G017240	-0.77	-3.6	HORVU2Hr1G096430	-0.86	-2.8
HORVU4Hr1G004540	-0.77	-2.4	HORVU5Hr1G072370	-0.86	-2.3
HORVU7Hr1G122160	-0.77	-2.5	HORVU2Hr1G027210	-0.86	-4.2
HORVU5Hr1G001080	-0.77	-9.3	HORVU4Hr1G009140	-0.86	-10.7
HORVU7Hr1G028160	-0.77	-15.0	HORVU0Hr1G020400	-0.86	-2.1
HORVU4Hr1G004590	-0.78	-8.1	HORVU3Hr1G018070	-0.86	-3.8
HORVU0Hr1G027670	-0.78	-2.2	HORVU3Hr1G067280	-0.86	-5.2
HORVU4Hr1G064440	-0.78	-2.2	HORVU2Hr1G112620	-0.86	-3.3
HORVU3Hr1G069410	-0.78	-3.4	HORVU6Hr1G036760	-0.87	-3.1
HORVU6Hr1G093820	-0.78	-4.4	HORVU1Hr1G002210	-0.87	-3.8
HORVU3Hr1G031850	-0.78	-2.2	HORVU6Hr1G087050	-0.87	-3.7
HORVU3Hr1G068280	-0.78	-2.3	HORVU5Hr1G080150	-0.87	-2.5
HORVU6Hr1G010140	-0.78	-2.2	HORVU3Hr1G022840	-0.87	-2.7
HORVU2Hr1G091680	-0.78	-5.7	HORVU7Hr1G036900	-0.87	-2.3
HORVU1Hr1G005600	-0.78	-7.1	HORVU3Hr1G056630	-0.87	-5.6
HORVU3Hr1G051650	-0.78	-2.9	HORVU1Hr1G070190	-0.87	-2.9
HORVU7Hr1G001090	-0.79	-4.5	HORVU2Hr1G101370	-0.87	-2.4
HORVU3Hr1G010290	-0.79	-2.2	HORVU1Hr1G073760	-0.87	-5.2
HORVU5Hr1G118510	-0.79	-3.3	HORVU3Hr1G026920	-0.87	-6.7
HORVU5Hr1G056330	-0.79	-2.2	HORVU3Hr1G030580	-0.88	-11.0
HORVU7Hr1G081770	-0.79	-2.5	HORVU7Hr1G009770	-0.88	-6.9
HORVU4Hr1G049960	-0.79	-2.2	HORVU7Hr1G085570	-0.88	-3.7
HORVU1Hr1G006940	-0.79	-3.0	HORVU1Hr1G023690	-0.88	-3.1
HORVU2Hr1G121080	-0.79	-6.2	HORVU6Hr1G009480	-0.88	-2.5

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU4Hr1G072840	-0.88	-5.0	HORVU7Hr1G000630	-0.99	-2.8
HORVU0Hr1G040190	-0.88	-3.5	HORVU4Hr1G003340	-0.99	-2.5
HORVU4Hr1G062650	-0.89	-6.1	HORVU5Hr1G049420	-0.99	-5.8
HORVU2Hr1G117110	-0.89	-5.9	HORVU5Hr1G048820	-1.00	-3.0
HORVU3Hr1G082030	-0.89	-2.8	HORVU2Hr1G091910	-1.00	-4.1
HORVU7Hr1G021490	-0.89	-2.9	HORVU2Hr1G079810	-1.00	-3.5
HORVU6Hr1G073090	-0.89	-4.6	HORVU2Hr1G126440	-1.00	-2.4
HORVU4Hr1G008200	-0.89	-2.5	HORVU0Hr1G017850	-1.00	-2.4
HORVU3Hr1G060390	-0.90	-6.9	HORVU6Hr1G023050	-1.00	-8.5
HORVU7Hr1G074030	-0.90	-2.0	HORVU2Hr1G045090	-1.00	-4.6
HORVU5Hr1G032980	-0.90	-2.2	HORVU4Hr1G080430	-1.00	-2.2
HORVU5Hr1G066250	-0.90	-3.4	HORVU6Hr1G032680	-1.00	-2.7
HORVU5Hr1G125420	-0.90	-3.6	HORVU3Hr1G084230	-1.01	-5.0
HORVU3Hr1G031380	-0.90	-2.9	HORVU6Hr1G035280	-1.01	-2.7
HORVU3Hr1G064790	-0.90	-4.8	HORVU7Hr1G077740	-1.01	-4.0
HORVU0Hr1G022030	-0.91	-2.6	HORVU2Hr1G003840	-1.01	-3.6
HORVU5Hr1G041680	-0.91	-2.1	HORVU1Hr1G078280	-1.01	-6.3
HORVU0Hr1G012190	-0.91	-4.3	HORVU5Hr1G115500	-1.01	-3.1
HORVU4Hr1G072660	-0.91	-2.3	HORVU1Hr1G094180	-1.02	-3.0
HORVU0Hr1G022280	-0.91	-5.1	HORVU2Hr1G002830	-1.02	-11.0
HORVU3Hr1G085690	-0.91	-3.7	HORVU0Hr1G000020	-1.02	-3.8
HORVU1Hr1G029920	-0.91	-3.8	HORVU2Hr1G107060	-1.02	-3.2
HORVU5Hr1G086710	-0.91	-8.8	HORVU5Hr1G067480	-1.02	-7.5
HORVU4Hr1G072320	-0.92	-5.0	HORVU1Hr1G004650	-1.02	-2.0
HORVU5Hr1G081460	-0.92	-4.4	HORVU3Hr1G100060	-1.03	-3.8
HORVU4Hr1G013540	-0.92	-3.1	HORVU7Hr1G119790	-1.03	-2.1
HORVU6Hr1G010110	-0.92	-5.5	HORVU1Hr1G006920	-1.03	-3.4
HORVU7Hr1G116220	-0.92	-3.4	HORVU5Hr1G113490	-1.03	-2.0
HORVU4Hr1G023500	-0.92	-5.4	HORVU1Hr1G077560	-1.03	-2.2
HORVU4Hr1G081440	-0.92	-2.2	HORVU4Hr1G009060	-1.03	-9.2
HORVU7Hr1G045400	-0.93	-3.3	HORVU1Hr1G056270	-1.04	-3.0
HORVU5Hr1G041660	-0.93	-6.9	HORVU1Hr1G076670	-1.04	-4.4
HORVU2Hr1G113160	-0.93	-4.0	HORVU1Hr1G082410	-1.04	-4.9
HORVU4Hr1G077300	-0.93	-2.5	HORVU2Hr1G124530	-1.05	-16.6
HORVU1Hr1G065250	-0.93	-2.6	HORVU3Hr1G035820	-1.05	-10.0
HORVU6Hr1G023100	-0.93	-6.9	HORVU2Hr1G077680	-1.05	-5.4
HORVU1Hr1G003140	-0.94	-2.2	HORVU2Hr1G092170	-1.06	-2.0
HORVU5Hr1G121280	-0.94	-2.1	HORVU4Hr1G010050	-1.06	-2.1
HORVU3Hr1G068750	-0.94	-2.4	HORVU4Hr1G072580	-1.06	-2.6
HORVU5Hr1G097010	-0.94	-12.5	HORVU5Hr1G058690	-1.07	-2.0
HORVU4Hr1G087760	-0.94	-2.1	HORVU7Hr1G111190	-1.07	-2.5
HORVU6Hr1G054010	-0.94	-7.7	HORVU2Hr1G057700	-1.07	-2.2
HORVU7Hr1G030920	-0.94	-2.1	HORVU2Hr1G098890	-1.07	-5.6
HORVU2Hr1G090280	-0.94	-3.7	HORVU4Hr1G066140	-1.07	-4.2
HORVU7Hr1G007890	-0.94	-2.7	HORVU6Hr1G015390	-1.07	-5.7
HORVU2Hr1G091990	-0.95	-2.7	HORVU1Hr1G082460	-1.08	-7.2
HORVU0Hr1G033250	-0.95	-5.6	ENSRNA049486349	-1.08	-2.1
HORVU6Hr1G092210	-0.95	-5.9	HORVU4Hr1G050790	-1.08	-2.7
HORVU3Hr1G085520	-0.95	-2.8	HORVU7Hr1G122290	-1.08	-3.4
HORVU5Hr1G043200	-0.95	-2.4	HORVU2Hr1G046370	-1.08	-4.0
HORVU7Hr1G000320	-0.95	-3.8	HORVU4Hr1G082600	-1.08	-4.8
HORVU3Hr1G081570	-0.95	-4.5	HORVU6Hr1G055960	-1.08	-6.7
HORVU3Hr1G064240	-0.95	-2.6	HORVU5Hr1G063810	-1.08	-5.3
HORVU2Hr1G018530	-0.96	-4.6	HORVU0Hr1G038420	-1.09	-4.5
HORVU1Hr1G001960	-0.96	-6.0	HORVU1Hr1G044260	-1.09	-6.2
HORVU3Hr1G062490	-0.96	-5.3	HORVU5Hr1G099170	-1.09	-3.4
HORVU7Hr1G122510	-0.96	-3.8	HORVU1Hr1G061170	-1.09	-5.9
HORVU3Hr1G087180	-0.96	-4.0	HORVU2Hr1G045340	-1.09	-2.4
HORVU3Hr1G098950	-0.96	-5.0	HORVU2Hr1G007590	-1.09	-4.7
HORVU5Hr1G106350	-0.96	-3.1	HORVU2Hr1G060210	-1.09	-3.1
HORVU2Hr1G007540	-0.97	-6.5	HORVU3Hr1G018820	-1.09	-2.8
HORVU7Hr1G012180	-0.97	-5.3	HORVU4Hr1G055220	-1.10	-5.5
HORVU2Hr1G093210	-0.97	-2.8	HORVU7Hr1G100570	-1.10	-5.6
HORVU2Hr1G109040	-0.97	-4.6	HORVU4Hr1G026150	-1.10	-2.3
HORVU1Hr1G076700	-0.97	-5.3	HORVU6Hr1G094410	-1.10	-5.0
HORVU6Hr1G087070	-0.97	-10.7	HORVU3Hr1G095070	-1.10	-2.4
HORVU4Hr1G077310	-0.97	-3.3	HORVU7Hr1G045580	-1.10	-3.1
HORVU3Hr1G099080	-0.98	-4.6	HORVU7Hr1G077750	-1.10	-2.7
HORVU7Hr1G044100	-0.98	-4.2	HORVU7Hr1G038130	-1.10	-3.8
HORVU4Hr1G012540	-0.98	-2.5	HORVU0Hr1G023150	-1.11	-3.4
HORVU7Hr1G095030	-0.98	-4.3	HORVU5Hr1G077450	-1.11	-19.7
HORVU5Hr1G037890	-0.98	-4.4	HORVU2Hr1G088570	-1.11	-9.6
HORVU7Hr1G072670	-0.98	-3.2	HORVU4Hr1G015050	-1.11	-3.4
HORVU2Hr1G074210	-0.99	-3.5	HORVU2Hr1G016070	-1.11	-4.4
HORVU6Hr1G054910	-0.99	-6.3	HORVU1Hr1G007630	-1.12	-2.5

Gene	log ₂ (fold change)	log ₁₀ (adj-p)	Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU1Hr1G086780	-1.12	-5.9	HORVU4Hr1G076010	-1.27	-9.9
HORVU5Hr1G014170	-1.12	-3.5	HORVU6Hr1G001160	-1.28	-3.1
HORVU5Hr1G000480	-1.12	-2.2	HORVU2Hr1G066780	-1.28	-3.7
HORVU3Hr1G081980	-1.12	-4.4	HORVU6Hr1G080820	-1.28	-3.0
HORVU7Hr1G087750	-1.12	-3.4	HORVU0Hr1G022690	-1.29	-3.0
HORVU1Hr1G010210	-1.12	-2.3	HORVU2Hr1G003580	-1.29	-11.8
HORVU2Hr1G002520	-1.12	-2.3	HORVU3Hr1G095340	-1.29	-11.3
HORVU1Hr1G001060	-1.13	-2.6	HORVU2Hr1G007070	-1.30	-5.9
HORVU2Hr1G093700	-1.13	-2.5	HORVU1Hr1G069200	-1.30	-4.8
HORVU5Hr1G107540	-1.13	-6.1	HORVU7Hr1G075960	-1.31	-9.8
HORVU1Hr1G087010	-1.13	-2.4	HORVU5Hr1G012740	-1.31	-4.5
HORVU1Hr1G026780	-1.13	-2.9	HORVU2Hr1G105140	-1.32	-9.3
HORVU5Hr1G000690	-1.13	-2.3	HORVU7Hr1G109800	-1.32	-5.4
HORVU7Hr1G088260	-1.13	-7.1	HORVU2Hr1G012850	-1.32	-7.7
HORVU3Hr1G082300	-1.13	-7.5	HORVU3Hr1G106880	-1.33	-5.5
HORVU3Hr1G092580	-1.14	-7.1	HORVU2Hr1G004230	-1.33	-7.0
HORVU2Hr1G072120	-1.14	-3.4	HORVU2Hr1G040790	-1.33	-4.1
HORVU1Hr1G082420	-1.14	-14.6	HORVU0Hr1G022650	-1.34	-3.8
HORVU1Hr1G076650	-1.14	-5.5	HORVU5Hr1G007770	-1.34	-7.5
HORVU7Hr1G114950	-1.15	-8.5	HORVU3Hr1G079720	-1.34	-6.4
HORVU7Hr1G046250	-1.15	-3.4	HORVU1Hr1G054020	-1.34	-12.0
HORVU7Hr1G047750	-1.15	-2.6	HORVU6Hr1G087120	-1.34	-12.9
HORVU3Hr1G013390	-1.15	-3.1	HORVU1Hr1G053990	-1.35	-11.6
HORVU2Hr1G072150	-1.16	-4.6	HORVU0Hr1G020750	-1.35	-4.2
HORVU5Hr1G103040	-1.16	-5.9	HORVU3Hr1G098580	-1.37	-3.8
HORVU2Hr1G113110	-1.16	-3.9	HORVU7Hr1G000760	-1.37	-8.2
HORVU7Hr1G020270	-1.16	-3.1	HORVU7Hr1G000040	-1.37	-5.9
HORVU4Hr1G018450	-1.16	-4.5	HORVU2Hr1G023560	-1.37	-13.1
HORVU3Hr1G003410	-1.16	-3.7	HORVU7Hr1G031790	-1.38	-6.1
HORVU1Hr1G018480	-1.17	-9.2	HORVU6Hr1G059520	-1.38	-24.9
HORVU7Hr1G037000	-1.17	-3.1	HORVU2Hr1G089620	-1.38	-7.6
HORVU2Hr1G019720	-1.17	-4.0	HORVU7Hr1G012310	-1.38	-7.2
HORVU1Hr1G054520	-1.18	-3.7	HORVU2Hr1G092200	-1.39	-3.8
HORVU3Hr1G071170	-1.18	-3.7	HORVU4Hr1G060150	-1.40	-6.1
HORVU3Hr1G052710	-1.18	-5.2	HORVU2Hr1G007490	-1.40	-7.5
HORVU4Hr1G011110	-1.18	-4.9	HORVU6Hr1G079030	-1.41	-6.2
HORVU1Hr1G011990	-1.18	-10.8	HORVU6Hr1G089450	-1.41	-5.4
HORVU1Hr1G061970	-1.19	-3.2	HORVU1Hr1G053440	-1.41	-4.8
HORVU2Hr1G036680	-1.19	-5.4	HORVU4Hr1G065270	-1.43	-4.7
HORVU5Hr1G040620	-1.19	-6.0	HORVU3Hr1G061450	-1.43	-5.5
HORVU1Hr1G067080	-1.19	-2.6	HORVU2Hr1G109200	-1.43	-5.9
HORVU7Hr1G055930	-1.20	-6.6	HORVU6Hr1G005590	-1.43	-5.0
HORVU2Hr1G003780	-1.20	-2.7	HORVU7Hr1G096250	-1.45	-7.3
HORVU5Hr1G063650	-1.20	-2.6	HORVU5Hr1G048100	-1.45	-5.9
HORVU5Hr1G052760	-1.20	-5.7	HORVU2Hr1G001130	-1.45	-4.2
HORVU1Hr1G025890	-1.21	-9.2	HORVU6Hr1G080270	-1.45	-9.0
HORVU1Hr1G074790	-1.21	-10.0	HORVU1Hr1G082370	-1.46	-11.8
HORVU1Hr1G004720	-1.21	-5.5	HORVU3Hr1G006270	-1.46	-3.9
HORVU2Hr1G040800	-1.21	-3.5	HORVU6Hr1G016750	-1.46	-10.9
HORVU7Hr1G091350	-1.21	-9.9	HORVU3Hr1G092240	-1.47	-8.7
HORVU2Hr1G034650	-1.21	-3.0	HORVU1Hr1G045080	-1.47	-5.1
HORVU5Hr1G087290	-1.21	-2.6	HORVU2Hr1G100420	-1.47	-4.0
HORVU5Hr1G065350	-1.22	-3.4	HORVU3Hr1G088080	-1.48	-7.4
HORVU7Hr1G001550	-1.22	-2.9	HORVU7Hr1G041470	-1.48	-11.1
HORVU0Hr1G003480	-1.22	-3.2	HORVU2Hr1G030610	-1.48	-10.3
HORVU4Hr1G066120	-1.22	-5.0	HORVU3Hr1G095240	-1.49	-6.1
HORVU5Hr1G020500	-1.22	-2.7	HORVU6Hr1G059940	-1.49	-15.1
HORVU6Hr1G060720	-1.22	-5.4	HORVU3Hr1G021750	-1.49	-6.5
HORVU2Hr1G082150	-1.23	-3.6	HORVU3Hr1G023800	-1.53	-4.4
HORVU0Hr1G027070	-1.23	-5.0	HORVU3Hr1G085460	-1.54	-5.2
HORVU1Hr1G076730	-1.23	-3.9	HORVU5Hr1G095580	-1.54	-4.3
HORVU5Hr1G099030	-1.23	-3.8	HORVU7Hr1G002820	-1.54	-26.9
HORVU4Hr1G090860	-1.24	-6.8	HORVU1Hr1G063340	-1.55	-7.2
HORVU1Hr1G083910	-1.24	-3.9	HORVU2Hr1G118740	-1.55	-38.6
HORVU3Hr1G010240	-1.25	-3.6	HORVU1Hr1G056120	-1.55	-4.5
HORVU0Hr1G000620	-1.25	-4.3	HORVU5Hr1G076690	-1.56	-16.0
HORVU6Hr1G080150	-1.25	-7.8	HORVU5Hr1G013640	-1.56	-25.6
HORVU4Hr1G015740	-1.25	-8.6	HORVU0Hr1G019630	-1.56	-6.5
HORVU2Hr1G007500	-1.25	-10.9	HORVU3Hr1G085230	-1.56	-7.5
HORVU2Hr1G127670	-1.25	-3.7	HORVU1Hr1G027500	-1.56	-4.9
HORVU6Hr1G013400	-1.26	-10.2	HORVU7Hr1G079870	-1.57	-7.7
HORVU5Hr1G001060	-1.27	-19.3	HORVU1Hr1G018200	-1.58	-11.2
HORVU1Hr1G000920	-1.27	-6.1	HORVU2Hr1G018430	-1.59	-9.7
HORVU5Hr1G071270	-1.27	-5.0	HORVU7Hr1G025240	-1.60	-5.6
HORVU2Hr1G109210	-1.27	-5.1	HORVU5Hr1G074820	-1.60	-6.4

Gene	log ₂ (fold change)	log ₁₀ (adj-p)
HORVU6Hr1G030080	-1.61	-8.8
HORVU2Hr1G098110	-1.63	-5.9
HORVU2Hr1G065000	-1.64	-6.7
HORVU2Hr1G089210	-1.68	-5.2
HORVU7Hr1G009350	-1.73	-5.8
HORVU4Hr1G080890	-1.74	-13.7
HORVU7Hr1G001070	-1.77	-10.9
HORVU7Hr1G114020	-1.78	-19.8
HORVU7Hr1G025340	-1.78	-6.6
HORVU7Hr1G002560	-1.79	-17.1
HORVU7Hr1G114000	-1.79	-14.7
HORVU3Hr1G028840	-1.81	-7.8
HORVU1Hr1G051450	-1.81	-14.0
HORVU1Hr1G090990	-1.81	-11.3
HORVU2Hr1G012980	-1.84	-6.7
HORVU3Hr1G095700	-1.85	-15.3
HORVU0Hr1G017490	-1.93	-10.9
HORVU6Hr1G005600	-1.96	-8.7
HORVU6Hr1G051860	-1.97	-20.7
HORVU5Hr1G042230	-2.03	-18.2
HORVU5Hr1G057840	-2.05	-16.7
HORVU7Hr1G119800	-2.07	-21.5
HORVU5Hr1G056480	-2.07	-8.2
HORVU2Hr1G127480	-2.09	-10.6
HORVU3Hr1G071800	-2.12	-12.7
HORVU5Hr1G064040	-2.14	-13.3
HORVU2Hr1G007510	-2.28	-11.6
HORVU5Hr1G109980	-2.51	-12.3
HORVU5Hr1G114000	-2.59	-29.6
HORVU2Hr1G117610	-2.69	-21.1
HORVU5Hr1G076740	-2.80	-26.8
HORVU1Hr1G080790	-2.84	-20.9
HORVU3Hr1G066090	-2.91	-18.2

Table A.9.3. Complete biological process gene ontology hierarchy for leaf and root tissue in salt stressed barley.

GO term	GO term function	Fold over-represented	
		Leaf	Root
0009834	plant-type secondary cell wall biogenesis	29.3	
0009832	plant-type cell wall biogenesis	12.2	
0042546	cell wall biogenesis	6.9	
0071554	cell wall organization or biogenesis	5.2	
0009987	cellular process	1.5	
0071840	cellular component organization or biogenesis	1.8	
0071669	plant-type cell wall organization or biogenesis	5.3	
0006782	protoporphyrinogen IX biosynthetic process	24.5	
0051186	cofactor metabolic process	2.6	
0044237	cellular metabolic process	1.4	
0008152	metabolic process	1.6	
0071704	organic substance metabolic process	1.6	
0044249	cellular biosynthetic process	1.9	
0009058	biosynthetic process	1.9	
1901576	organic substance biosynthetic process	1.9	
0046501	protoporphyrinogen IX metabolic process	24.5	
0031408	oxylipin biosynthetic process	22.6	
0031407	oxylipin metabolic process	21.0	
0006631	fatty acid metabolic process	4.3	
0032787	monocarboxylic acid metabolic process	3.2	
0019752	carboxylic acid metabolic process	2.7	
0043436	oxoacid metabolic process	2.6	
0006082	organic acid metabolic process	2.6	
0044281	small molecule metabolic process	2.5	
0006629	lipid metabolic process	2.4	
0044238	primary metabolic process	1.6	
0006633	fatty acid biosynthetic process	5.6	
0072330	monocarboxylic acid biosynthetic process	5.0	
0046394	carboxylic acid biosynthetic process	3.9	
0016053	organic acid biosynthetic process	3.9	
0030244	cellulose biosynthetic process	13.8	
0051274	beta-glucan biosynthetic process	10.0	
0009250	glucan biosynthetic process	9.0	
0033692	cellular polysaccharide biosynthetic process	8.1	
0009059	macromolecule biosynthetic process	1.7	
0044264	cellular polysaccharide metabolic process	7.3	
0005976	polysaccharide metabolic process	6.1	
0005975	carbohydrate metabolic process	4.1	
0044262	cellular carbohydrate metabolic process	6.1	
0000271	polysaccharide biosynthetic process	8.1	
0016051	carbohydrate biosynthetic process	5.8	
0034637	cellular carbohydrate biosynthetic process	6.0	
0006073	cellular glucan metabolic process	7.8	
0044042	glucan metabolic process	7.8	
0051273	beta-glucan metabolic process	9.4	
0030243	cellulose metabolic process	11.6	

GO term	GO term function	Fold over-represented	
		Leaf	Root
0005985	sucrose metabolic process	10.9	
0000302	response to reactive oxygen species	8.7	
1901700	response to oxygen-containing compound	3.5	
0042221	response to chemical	2.5	
0050896	response to stimulus	1.7	
0046677	response to antibiotic	8.2	
0044247	cellular polysaccharide catabolic process	8.2	
0009251	glucan catabolic process	8.2	
0071555	cell wall organization	5.1	
0045229	external encapsulating structure organization	5.1	
0016043	cellular component organization	1.8	
0019318	hexose metabolic process	5.1	
0005996	monosaccharide metabolic process	4.7	
0044036	cell wall macromolecule metabolic process	5.0	
1901607	alpha-amino acid biosynthetic process	4.4	4.4
1901605	alpha-amino acid metabolic process	3.6	5.1
0006520	cellular amino acid metabolic process	2.5	3.9
0008652	cellular amino acid biosynthetic process	3.6	
0010035	response to inorganic substance	4.2	
0007017	microtubule-based process	3.9	
0001101	response to acid chemical	3.6	
0098869	cellular oxidant detoxification	3.0	
1990748	cellular detoxification	3.0	
0098754	detoxification	3.0	
0009636	response to toxic substance	3.2	
0097237	cellular response to toxic substance	3.0	
0070887	cellular response to chemical stimulus	2.9	
0051716	cellular response to stimulus	1.8	
0055114	oxidation-reduction process	1.9	2.5
0009627	systemic acquired resistance		23.7
0009814	defense response, incompatible interaction		14.7
0045087	innate immune response		12.1
0098542	defense response to other organism		7.1
0006952	defense response		6.0
0006950	response to stress		2.1
0051707	response to other organism		6.1
0043207	response to external biotic stimulus		6.1
0009605	response to external stimulus		3.9
0009607	response to biotic stimulus		5.5
0051704	multi-organism process		4.8
0006955	immune response		12.1
0002376	immune system process		10.8
0006002	fructose 6-phosphate metabolic process		20.2
0006796	phosphate-containing compound metabolic process		1.9
0006793	phosphorus metabolic process		1.9
0008152	metabolic process		1.5
0071704	organic substance metabolic process		1.3
0006568	tryptophan metabolic process		14.4
1901564	organonitrogen compound metabolic process		1.5

GO term	GO term function	Fold over-represented	
		Leaf	Root
0019752	carboxylic acid metabolic process		3.5
0043436	oxoacid metabolic process		3.5
0006082	organic acid metabolic process		3.4
0044281	small molecule metabolic process		2.4
0009072	aromatic amino acid family metabolic process		7.3
0006586	indolalkylamine metabolic process		14.4
0042430	indole-containing compound metabolic process		12.6
0006576	cellular biogenic amine metabolic process		11.6
0044106	cellular amine metabolic process		11.6
0009308	amine metabolic process		7.8
0042401	cellular biogenic amine biosynthetic process		12.7
0009309	amine biosynthetic process		12.7
0009664	plant-type cell wall organization		8.1
0071555	cell wall organization		3.4
0071554	cell wall organization or biogenesis		3.3
0045229	external encapsulating structure organization		3.3
0071669	plant-type cell wall organization or biogenesis		7.4
0008299	isoprenoid biosynthetic process		6.0
0048544	recognition of pollen		5.0
0009875	pollen-pistil interaction		4.8
0008037	cell recognition		5.0
0032269	negative regulation of cellular protein metabolic process		4.6
0051248	negative regulation of protein metabolic process		4.6
0046394	carboxylic acid biosynthetic process		3.2
0016053	organic acid biosynthetic process		3.2
0044283	small molecule biosynthetic process		2.5
0042737	drug catabolic process		4.0
0044248	cellular catabolic process		2.2
0009056	catabolic process		2.2
0017144	drug metabolic process		3.3
0017001	antibiotic catabolic process		3.6
0016999	antibiotic metabolic process		3.5
0005975	carbohydrate metabolic process		2.3
0006468	protein phosphorylation		1.9
0016310	phosphorylation		1.9
0055085	transmembrane transport		1.9

Table A.9.4. Complete molecular function gene ontology hierarchy for leaf and root tissue in salt stressed barley.

GO term	GO term function	Fold over-represented	
		Leaf	Root
0004575	sucrose alpha-glucosidase activity	21.0	
0090599	alpha-glucosidase activity	21.0	
0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	4.0	
0016798	hydrolase activity, acting on glycosyl bonds	3.9	
0016787	hydrolase activity	2.0	
0003824	catalytic activity	1.7	
0004564	beta-fructofuranosidase activity	21.0	
0016157	sucrose synthase activity	16.3	
0035251	UDP-glucosyltransferase activity	9.8	
0008194	UDP-glycosyltransferase activity	5.5	
0016757	transferase activity, transferring glycosyl groups	3.0	
0016740	transferase activity	1.5	
0046527	glucosyltransferase activity	8.3	
0016758	transferase activity, transferring hexosyl groups	3.1	
0016760	cellulose synthase (UDP-forming) activity	12.6	
0016759	cellulose synthase activity	12.6	
0005200	structural constituent of cytoskeleton	12.2	
0005198	structural molecule activity	2.1	
0033897	ribonuclease T2 activity	12.2	
0016892	endoribonuclease activity, producing 3'-phosphomonoesters	10.9	
0051087	chaperone binding	10.9	
0005488	binding	1.4	
0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	5.8	
0016903	oxidoreductase activity, acting on the aldehyde or oxo group of donors	4.9	
0016491	oxidoreductase activity	2.0	
0008017	microtubule binding	4.0	
0016209	antioxidant activity	3.0	
0005506	iron ion binding	2.5	5.0
0046872	metal ion binding	1.5	2.1
0043169	cation binding	1.6	2.1
0043167	ion binding	1.4	1.9
0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	2.3	
0020037	heme binding	2.1	3.9
0046906	tetrapyrrole binding	2.0	3.7
0048037	cofactor binding	1.9	
0000234	phosphoethanolamine N-methyltransferase activity		20.5
0016740	transferase activity		1.7
0003824	catalytic activity		1.9
0004350	glutamate-5-semialdehyde dehydrogenase activity		20.5
0016491	oxidoreductase activity		2.7
0004349	glutamate 5-kinase activity		20.5
0016772	transferase activity, transferring phosphorus-containing groups		1.8
0016301	kinase activity		2.0
0005504	fatty acid binding		18.7
0008289	lipid binding		4.6

GO term	GO term function	Fold over-represented	
		Leaf	Root
0005488	binding		1.4
0033293	monocarboxylic acid binding		16.9
0031406	carboxylic acid binding		11.8
0043168	anion binding		1.6
0043177	organic acid binding		11.8
0036094	small molecule binding		1.7
0005315	inorganic phosphate transmembrane transporter activity		16.5
0022857	transmembrane transporter activity		2.0
0005215	transporter activity		2.1
0015291	secondary active transmembrane transporter activity		4.2
0022804	active transmembrane transporter activity		4.1
0003872	6-phosphofructokinase activity		15.1
0016773	phosphotransferase activity, alcohol group as acceptor		2.0
0030410	nicotianamine synthase activity		15.1
0030598	rRNA N-glycosylase activity		8.6
0030597	RNA glycosylase activity		8.6
0016798	hydrolase activity, acting on glycosyl bonds		2.6
0016831	carboxy-lyase activity		7.5
0016830	carbon-carbon lyase activity		6.5
0016829	lyase activity		4.3
0030170	pyridoxal phosphate binding		6.3
0050662	coenzyme binding		2.2
0048037	cofactor binding		2.9
0070279	vitamin B6 binding		6.3
0097159	organic cyclic compound binding		1.4
0019842	vitamin binding		5.3
0008144	drug binding		1.8
1901363	heterocyclic compound binding		1.4
0030145	manganese ion binding		5.3
0046914	transition metal ion binding		2.6
0004497	monooxygenase activity		4.8
0042626	ATPase-coupled transmembrane transporter activity		4.7
0015399	primary active transmembrane transporter activity		4.1
0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen		4.4
0030246	carbohydrate binding		3.3
0016758	transferase activity, transferring hexosyl groups		2.6
0016757	transferase activity, transferring glycosyl groups		2.2
0004553	hydrolase activity, hydrolyzing O-glycosyl compounds		2.4
0004672	protein kinase activity		1.9
0140096	catalytic activity, acting on a protein		1.5
0005524	ATP binding		1.6
0032559	adenyl ribonucleotide binding		1.6
0030554	adenyl nucleotide binding		1.6
0097367	carbohydrate derivative binding		1.5

Table A.9.5. Complete cellular component gene ontology hierarchy for leaf and root tissue in salt stressed barley.

GO term	GO term function	Fold over-represented	
		Leaf	Root
0042644	chloroplast nucleoid	18.3	
0009570	chloroplast stroma	5.7	
0009507	chloroplast	2.4	
0009536	plastid	2.3	
0110165	cellular anatomical entity	1.2	
0009532	plastid stroma	5.6	
0005874	microtubule	6.1	
0015630	microtubule cytoskeleton	5.2	
0099513	polymeric cytoskeletal fiber	6.0	
0099512	supramolecular fiber	6.0	
0099081	supramolecular polymer	6.0	
0099080	supramolecular complex	4.9	
0005618	cell wall	3.5	6.4
0030312	external encapsulating structure	3.5	
0071944	cell periphery	2.4	2.2
0005576	extracellular region	2.4	
0030312	external encapsulating structure		6.4
0110165	cellular anatomical entity		1.2
0048046	apoplast		5.4
0005576	extracellular region		4.5
0016021	integral component of membrane		1.5
0031224	intrinsic component of membrane		1.4
0016020	membrane		1.4

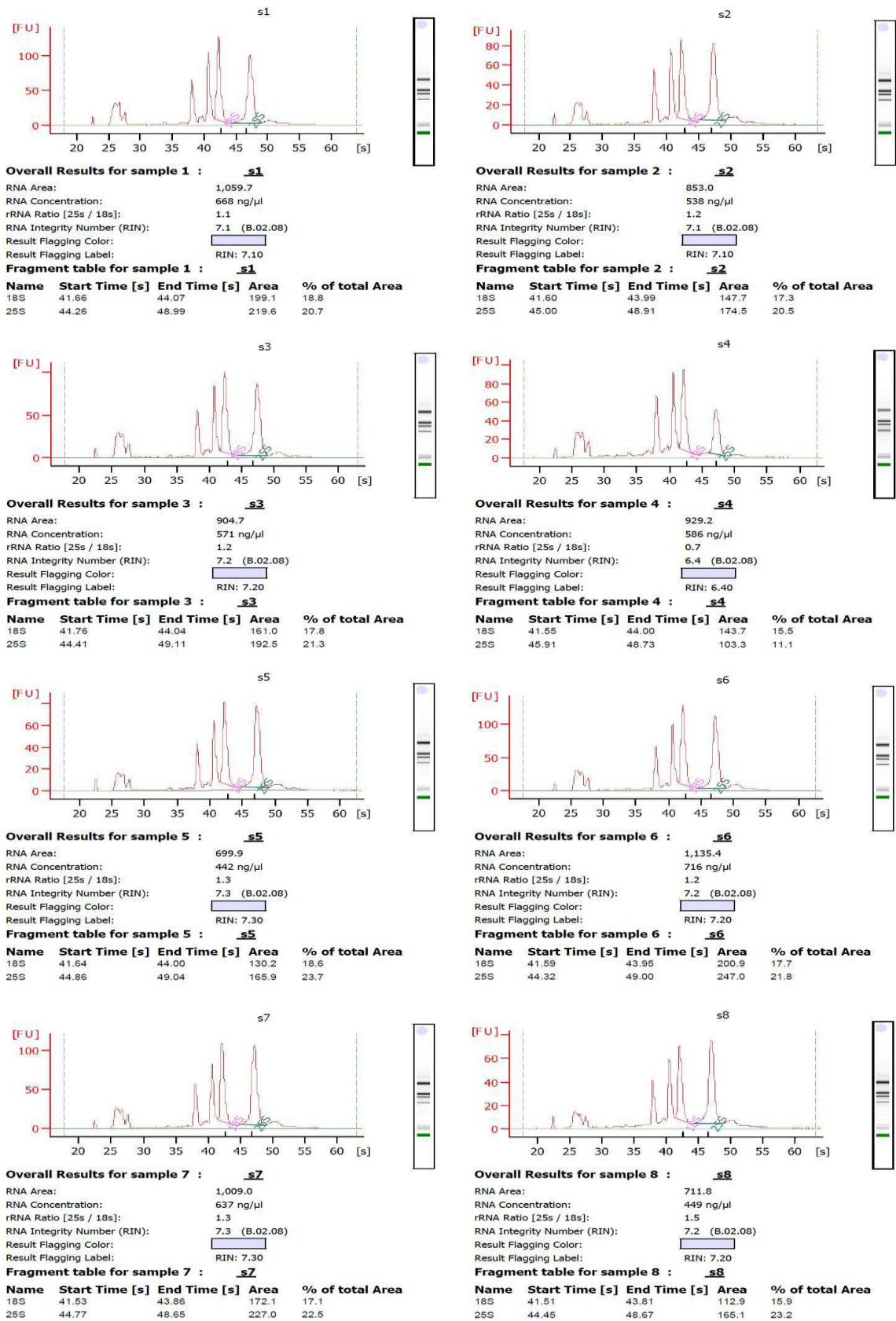


Figure A.9.1. Agilent Bioanalyser results for leaf RNA extractions. Samples designated s1-s8 for leaf tissue samples 1-8 where 1-4 were control plants and 5-8 were salt treated.

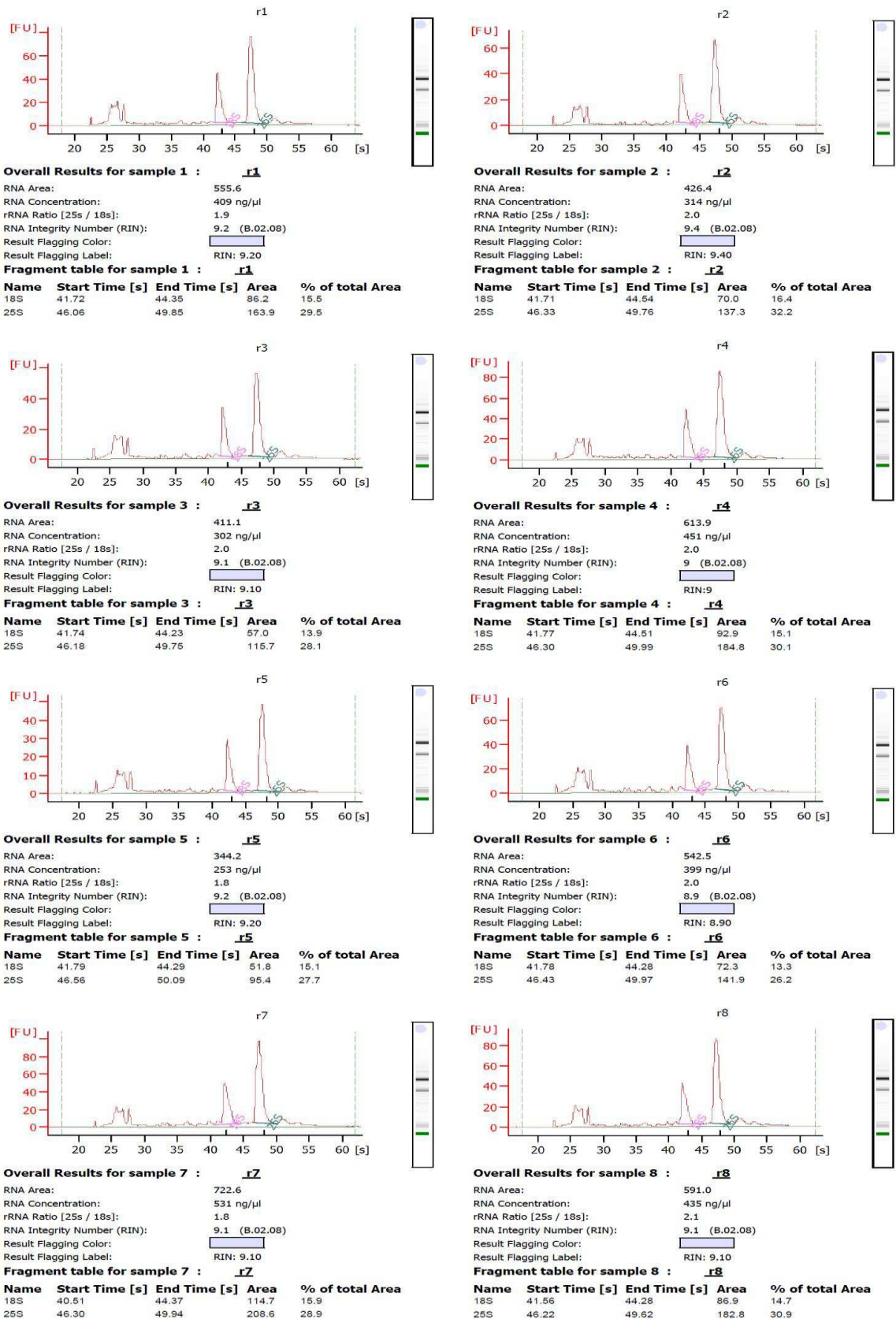


Figure A.9.2. Agilent Bioanalyser results for root RNA extractions. Samples designated r1-r8 for root tissue samples 1-8 where 1-4 were control plants and 5-8 were salt treated.

Table A.9.6. Illumina barcodes and corresponding samples for small RNA sequencing.

Sample Identifier	Tissue	Control or Salt treated	Barcode sequence
L1	Leaf	Control	GATCAG
L2	Leaf	Control	TAGCTT
L3	Leaf	Control	GGCTAC
L4	Leaf	Control	CTTGTA
L5	Leaf	Salt treated	AGTCAA
L6	Leaf	Salt treated	AGTTCC
L7	Leaf	Salt treated	ATGTCA
L8	Leaf	Salt treated	CCGTCC
R1	Root	Control	ATCACG
R2	Root	Control	CGATGT
R3	Root	Control	TTAGGC
R4	Root	Control	TGACCA
R5	Root	Salt treated	ACAGTG
R6	Root	Salt treated	GCCAAT
R7	Root	Salt treated	CAGATC
R8	Root	Salt treated	ACTTGA

Table A.9.7. Listing of previously undocumented mature miRNA in barley. These miRNA have not been previously categorised in miRBase. MicroRNA are named with hvu-new-miR for sequences that have not been found previously in barley. Where a similar miRNA has been found in other species, the numbering of that closest miRNA is used. When no similar sequences have been found in other species, the miRNA are named hvu-novel-miR900xx so that the numbers do not clash with any existing miRNA. The 5p or 3p at the end of the name is related to the mature miRNA read being closer to the 5' or 3' end of the hairpin-loop precursor.

```

>hvu-new-miR11115-3p
AGACTTCGCTGGGACATCAGC
>hvu-new-miR11115-5p
TGGTATGTCAGTTGTAATGCA
>hvu-new-miR11481a-3p
ATGCATCGATTTAAATGGACGAGC
>hvu-new-miR11481a-5p
TTGTTTGTGGATTTCGTGTGAAT
>hvu-new-miR1509a-3p
CCAGTTGTCAGACAACCGGGAC
>hvu-new-miR1509a-5p
CCCCGATTGTAACCTGCACAGGGC
>hvu-new-miR166a-3p
TCGGACCAGGATTCCATTATT
>hvu-new-miR166a-3p
TCGGACCAGGATTCCATTATTC
>hvu-new-miR3440-3p
CGGGTTGGACGGAGTACTAAAG
>hvu-new-miR3440-5p
TAAGCTACTGCATCTGATCCGAA
>hvu-new-miR3672-3p
CTGTCGCAAGGATTTCGTAATA
>hvu-new-miR3672-3p
CTGTCGCAAGGATTTCGTAATA
>hvu-new-miR396a-3p
GTTCAAGAAAGTCCCTGGAAAA
>hvu-new-miR396a-3p
GTTCAATAAAGCTGTGGAAA
>hvu-new-miR396a-5p
TCCACAGGCTTCTTGAAGCTG
>hvu-new-miR396a-5p
TTCCACAGCTTCTTGAAGCTG
>hvu-new-miR477-5p
TTCTCCCTCAAAGGCCTCCAACG
>hvu-new-miR528-3p
CCTGTGCCTGCCTCTCCATT
>hvu-new-miR528-5p
TGGAAAGTGGCATGCAGAGGAG
>hvu-new-miR5803-3p
CCCACTGTTGGATATTCACC
>hvu-new-miR5803-5p
ATTGTATCCGATATGGCCT
>hvu-new-miR8604-3p
AATGAAGGGGATAAATCCAGCCCT
>hvu-new-miR8604-5p
GCCATGAGTTATCCCCTTCATTAC
>hvu-new-miR867-3p
TTGAACATCCCAGAGCCACC
>hvu-new-miR867-3p
TTGAACATCCCAGAGCCACC
>hvu-new-miR867-5p
CGGCTCTGTGGTGTTCAGC
>hvu-new-miR867-5p
GCGGCTCTGTGGTGTTCAGC
>hvu-new-miR9555-3p
GTAGAAACCAGGGTTGTCCAGCAAG
>hvu-novel-miR90000-3p
ACACCAAATCACTTCCACATG
>hvu-novel-miR90000-5p
TGTAGAGGTGATTTGGTGATC
>hvu-novel-miR90001-3p
AAGGGGATAACTCATGGCGCC
>hvu-novel-miR90001-5p
CGCCATGAATTATCCCCTTCA
>hvu-novel-miR90002-3p
ATTTTCTTCCCTTGGTCCGTT
>hvu-novel-miR90002-5p
TGGGCCAAGGAAGGAAATAA
>hvu-novel-miR90003-3p
TTGGCTACACCTAGTCTCTT
>hvu-novel-miR90003-5p
GAGAACGGGATGCAGCCAAGG
>hvu-novel-miR90004-5p
TGGCTGCAAGGGCCTTATCTCTGA
>hvu-novel-miR90004a-3p
TGAGTGGCTCTTGACGCCCAT
>hvu-novel-miR90004b-3p
TGAGTGGCTCCTGCTGCCCAT
>hvu-novel-miR90004c-3p
TGAGTGGCTCCTGCGGCCCAT
>hvu-novel-miR90005-5p
TCGTGACCCGTACCATCATCTT
>hvu-novel-miR90006-5p
TCGTGACCCGTACCATTATCTT
>hvu-novel-miR90007-5p
TCGTGACCCGTACCATCATTTT
>hvu-novel-miR90008-5p
TCGTGACCCGTACAATGTTTC
>hvu-novel-miR90009-3p
TTTCTCCTTTGCTCGAGCAGAG
>hvu-novel-miR90009-5p
CTCGGAGCAACGGATGAATC
>hvu-novel-miR90010-3p
CCTTGCTCTTACCTCTGCTGT
>hvu-novel-miR90010-5p
GAAAGAGGCAGAGAGTGGAATG
>hvu-novel-miR90011-3p
CGTCAGTCGATGTGGTCAAGG
>hvu-novel-miR90011-5p
TTGGCCACATCGACTGACGGC
>hvu-novel-miR90012-3p
TGGCCAGAGCCTCGTCAGCTAG
>hvu-novel-miR90012-5p
AGCCGGCAGTCCAGGGCGATC
>hvu-novel-miR90013-3p
GTGCTATGGATAAATTTAACC
>hvu-novel-miR90013-5p
TTAAATTTCTCCATAGCATCA
>hvu-novel-miR90014-3p
ATATTTGACAATAACGATAGT
>hvu-novel-miR90014-5p
TATTGTTATCGTCTAAAACCT
>hvu-novel-miR90015-3p
CGCCGAATGTACATGCTCAGT
>hvu-novel-miR90015-5p
TGAGCATGATCAATCGGTGAT
>hvu-novel-miR90016-5p
TCGTGACCCGTACCATTATCT
>hvu-novel-miR90017-5p
TCGTGACCCGTACCATCATTT
>hvu-novel-miR90018-3p
ATAGACGCACGTCGCGACGCT
>hvu-novel-miR90019-3p
CGGGTTGGACGGAGTACTAAAGT

```

>hvu-novel-miR90019-5p
ATAAGCTACTGCATCTGATCCGAA
>hvu-novel-miR90020-3p
TTTTAGATTTGGATCGGCACCGTG
>hvu-novel-miR90020-5p
TTTGGTCGGTCGTTCTTTTTTG
>hvu-novel-miR90021-3p
CTCACGATTCTCGGCATTTT
>hvu-novel-miR90022-3p
CTATGTCCCATTTTTTCAGGCGA
>hvu-novel-miR90022-5p
TCTTGAACACCCATGGCTATAGTC
>hvu-novel-miR90023-3p
TTAAAACATCGAAACGTCCGACGA
>hvu-novel-miR90023-5p
GTGGGTTGTGTAGAGTGGTAGTG
>hvu-novel-miR90024-3p
TTCCGATAGGTGCAGTGGCATA
>hvu-novel-miR90024-5p
TGCACCTGCACCTACGGAGGA
>hvu-novel-miR90025-5p
TTTATGAACGCATTGAATTTTT
>hvu-novel-miR90026-3p
TCGATTCCCGCCTTCGGCACCT
>hvu-novel-miR90027-3p
TCGATTCCCCCCTCGGCACCC
>hvu-novel-miR90028-3p
TCGATTCCCCCCTCGGCACC
>hvu-novel-miR90029-3p
ATTGCTAATAGAATAATGTATTTC
>hvu-novel-miR90030-3p
TCGGCTCGTCGCATCGTGGTGT
>hvu-novel-miR90031-5p
GCCTTCGGTGTGTGGGATCT
>hvu-novel-miR90032-3p
TAGAGATCTGGTGGAACACC
>hvu-novel-miR90033-3p
GTTCCGAGTCCCCCGTCGGCGC
>hvu-novel-miR90034-3p
GTTCCGAGTCCCCCGTCGGCATC
>hvu-novel-miR90035-3p
TGAGATCTGCGCATCCCGGATGCA
>hvu-novel-miR90035-5p
GATCTGAGATCCGCGGTCTCATC
>hvu-novel-miR90036-5p
CCGGAGTCGGTGAGGGAACC
>hvu-novel-miR90037-5p
AACCGAAGCTGTTTCGATTGCG
>hvu-novel-miR90038-3p
CCTCGGGTGGGCACTGTAGC
>hvu-novel-miR90038-5p
GTGAGGATCGGCCGAGGCG
>hvu-novel-miR90039-5p
TCCGTGGATTCCGACTGGGG
>hvu-novel-miR90040-3p
TCGATTCCCCCCTCGGCACCT

Table A.9.8. Listing of previously undocumented hairpin-loop precursors for miRNAs in barley These miRNA have not been previously categorised in miRBase. MicroRNA are named with hvu-new-miR for sequences that have not been found previously in barley. Where a similar miRNA has been found in other species, the numbering of that closest miRNA is used. When no similar sequences have been found in other species, the miRNA are named hvu-novel-miR900xx so that the numbers do not clash with any existing miRNA.

```

>hvu-new-miR11115
CATACAAACTACTGGTATGTCTAGTTGTAATGCAAAGGGCAGATAAGTGATGACATGATATTACAAACACGCAAAGGACCTTCTACTCTGGTGGG
GCATCAGCTCGGTGGGGCATCATTAGACTTACAAACACGCAAAGGGCCTTCTACTTTTCGTGGGCATCATCAGACTTCGCTGGGACATCAGCA
TACAGTG
>hvu-new-miR11481a
CCATTTTCATCTGTCGTTGTTTGGATTTCGTGTGAATACGAAAGACGGTTTAATGCATCGATTTAAATGGACGAGCGTTTCGTTTTTTT
>hvu-new-miR1509a
CCTACATTAGCCCCGGATTGTAACCTGCACAGGGCAATACACATGAGGTCCTGAGTTCGATTCCCCCGAGTGCATTCGGTCGGAGGCATTATTT
TTTTCTAGCCTGGAAGACCTTTAGTCCCAGTTGTCAGACAACCGGACTAAAGGATCATG
>hvu-new-miR166a
AGACGCAACTATGAATTTTTCGAGTGCCGGTTCGGACCAGGATTCCATTATTCAGCAGCCAC
>hvu-new-miR3440
TATACAAATATAAGTACTGTCATCTGATCCGAATTTATTTGACGCATCCTCTACACAGTATAAAATGGGGTATTGTATAGAGGTTGCGTCAATTA
ATTCGGTTTGGACGAGTACTAAAGTTGAGGTAGGC
>hvu-new-miR3672
GCCGACAAATTTATGACTATCTTCCCTGGCACTGTGCGCAAGGATTCGTAACATTTTTTAT
>hvu-new-miR396a
CGGCCATGCTCTCCACAGGCTTTCTTGAACGTCAACTCGCGCGCCAGCCATCCATGGCCTGCTGCTCAATTCACCCGCTGCTGTCGCATA
TTGAGATCCGATCCATTGTTCAATCCATGGATCTTTGCCTTGCCCTTGCTTGTGATGGTTCAAGAAAGTCTTGAAAAACATGCC
>hvu-new-miR396b
TCTTTGCTGTCTTCCACAGCTTTCTTGAACGTCACTGCAATTTGGTGGATGATGGATTCTACCGCAAGATCTGCAGTTCAATAAAGCTGTGGG
AAATTGCAGAGAGA
>hvu-new-miR477
TTTCCTCTCCTTCTCCCTCAAAGGCCTCCAACGGAGGGCTTTGAGTGAGATGAAGAGGGCGCT
>hvu-new-miR528
CCGGACGAGCAGCGTGAAGGGGCATGCAGAGGAGCGGCATGCATGGAGCTTTGCTTTGCTTGCCCTCTCCTGCTCTGGGCTCTAGCTCTCT
CCTGTGCCTGCTCTTCCATTCCCTGCCCTAA
>hvu-new-miR5803
TCGTAGAATATGTAGTATCCGATATGGCCTTCCAGGTCCCCTGTTGGATATTCACCGGAGAGTGCT
>hvu-new-miR8604
TAGTCACAGGCGCCATGAGTTATCCCTTCACTAACAACAAAGTGAATGAAGGGGATAACTCATGCGCCTGTGACTATCTC
>hvu-new-miR867a
CGCGGAGCCCGCGGCTCTGTGGTGTCAAGCAGGAACCTCATGTACCGGCAGCATGCGGCGCTTGCTTGAACATCCCAGAGCCACCGGCGTG
CCAAA
>hvu-new-miR867b
CGCGGAGCCCGCGGCTCTGTGGTGTCAAGCAGGAACCTCATGTACCGGCAGCATGCGGCGCTTGCTTGAACATCCCAGAGCCACCGGCGTG
CCAAA
>hvu-new-miR9555
TTTCCTGTCCGATAAGTTGACAATGGTTTGTAGAAACCAGGGTTGTCCAGCAAGGGTGC
>hvu-novel-miR90000
AAATGAAAATTTGTAGAGGTGATTTGGTGTACCCGAATCTCTTGTATTCCGGTGCACACCAAATCACTTCCACATGCTTTTCGTTTTG
>hvu-novel-miR90001
TAGTCACAGGCGCCATGAATTTATCCCTTCACTACAGTTTTTGTAGTAATGAAGGAGATAACTCATAAGTTCCCTCGATGCGAAGGAGTTTTAG
GAGATAGTCACAGGCGCATGAGTTATCCCTTCACTACAGTTTTTGTAGTAATGAAGGGGATAACTCATGCGCCTGTGACTATCT
>hvu-novel-miR90002
GTCAAAGCTAATGGGCAAGGGAAGAAATAATGATGACATAGATCCAGCAGGGAACAGTGGGCTACATGCATGGCTCTTGAGTGAATAGGA
GTGTGGGAAATACATTTGACGAATTAGTTAATGTAAGAATCTTTTCACTACTATTTTCTTCCCTTGGTCCGTTAATTTGATAG
>hvu-novel-miR90003
CGCTAAGGGGCAGAGAACGGGATGCAGCCAAGGATGACTTGCCGGCTTCTGGTGTGGGAGTTTCGTAGAGCCTTAAGAATTAGCCGGCAAGCTG
TCCTTGGCTACACCTAGTTCTCTTCTTCTGGTGT
>hvu-novel-miR90004a
CAATCCTTACATGGCTGCAAGGGCCTTATCTCTGATAGCCAAGGATGACTTGCCCTGTGCTTGTCTCCTCCCTCAAGGAGCTTAATTGGCCTTG
GGGTGTGGTTTCATGGGCAGTCTCCTTGGCTAGCCTGAGTGGCTCTTGACGCCATGCTAGGCGCGCT
>hvu-novel-miR90004b
CAAGCCTTACATGGCTGCAAGGGCCTTATCTCTGATAGCCAAGGATGACTTGCCCTGTGCTTGTCTCCTCCCTCAAGGCTTAATTGGCCTTGGG
TGTCGTTTACATGGCAGTCTCCTTGGCTAGCCTGAGTGGCTCTGCTGCCATGCTAGGATGTCT
>hvu-novel-miR90004c
CAAGCCTTACATGGCTGCAAGGGCCTTATCTCTGATAGCCAAGGATGACTTGCCCTGTGCTTGTCTCCTCCCTCAAGGCTTAATTGGTGCCTTG
GGGTGTGGTTTCATGGGCAGTCTTCTTGGCTAGCCTGAGTGGCTCTTGCGGCCATGCTAGGCTGTCT
>hvu-novel-miR90005
GGCGTAGTCGTCGTGACCCTGACCATCATCTTGATTGTCGTCGTCACTATGCGCCTCGTTG
>hvu-novel-miR90006a
CGGGTAGTCATCGTGACCGTGACCATTATCTTGATTGTCGTCGTCACTATGCGCCTCATTG
>hvu-novel-miR90006b
GGCGTAGTCGTCGTGACCCTGACCATTATCTTGATTGTCGTCGTCACTATGCGCCTCGTTG

```

>hvu-novel-miR90007a
GGCGTAGTCGTCGTGACCGTGACCATCATTTTTGATTGTCATCGTCACTATGCGCCTCGTTGT
>hvu-novel-miR90007b
GGCGTAGTCGTCGTGACCGTGACCATCATTTTTGATTGTCGTCGTCACTATGCGCCTCGTTGT
>hvu-novel-miR90008
GGCAAGATGATCGTGACCGTGACAATGTTTCTAGCTTTGTTATGCCTAGTTGAGTCGTTT
>hvu-novel-miR90009
AGATTCTTTCTCTCGCGAGCAACGGATGAATCAGCCTCTCTGTTATCCAGAGAGATGGCCGATTTCTCCTTTGCTCGAGCAGAGAAAGAGCATC
G
>hvu-novel-miR90010
AGAGCCACGCAGAAAGAGGCAGAGAGTGGAAATGCAGCCAAGGATGACTTGCCGACATCACAGAACAAGGTGGTAGTAATCTTGCCGGCAAGTCC
GTCCTTGGCTACACCTTGCTCTCTACCTCTGCTGTGTGGCTTGGG
>hvu-novel-miR90011
AGCGTAGAGCCTTGCCACATCGACTGACGGCAACGACCGGTGCGGACCACGGGGGAGGGGGGTAGGGCGGAATCTGCTGCCGCGATCGATCGT
TGCCGTCAGTCGATGTGGTCAAGGCTCTACGCCAGGC
>hvu-novel-miR90012
GCTGAGCTAATAGCCGCGGAGTACGCGGCGATCGGGCGGCTGTCGATGTCGTTGAGGCGGTGACGGGCGTGCACGGCTGCCCGATGGCCAGAG
CCTCGTCAGCTAGCAGCTCAGCTC
>hvu-novel-miR90013
CAATCCTCTGGTTAAATTTCTCCATAGCATCAGCCATATTTCCATTAACCCGCTTTTTGTCACATGGGGATGTTGACTGGTGTATGGATAAA
TTTAACCCGATGGTTTCG
>hvu-novel-miR90014
TCCTTTAGTTATATGTTATCGTCTAAAAACCTGAAATGAAGATATTTGACAATAACGATAGTAGACAATTTT
>hvu-novel-miR90015
TATGGGATCACTGAGCATGATCAATCGGTGATCTCGACTTGAAGTCGGGGTCCCCACAGTTAGAAAATTTTGGTGGGTGCAGAAAATCTCATG
TGTTAAGGCTGAGTGCACACTCTGCTCGGCTATCGTCCATGCTACCTATGGTTTCTACAGTCAGGATGGCTCTAATACCAGCTCTGTGGGGA
CCTCAATTTACGAGTTGAGATCGCCGAATGTACATGCTCAGTGGTCCCAGAGATC
>hvu-novel-miR90016
CGGGTAGTCATCGTGACCGTGACCATTATCTTATTGTCGTCGTCCTATGCGCCTCAT
>hvu-novel-miR90018
TCATGCATGTGTGAAAGTGTGACATCGATAGACGCACGTCGCGACGCTCCGGTGTGTG
>hvu-novel-miR90019
CTTATACAAATATAAGCTACTGCATCTGATCCGAATATTTGACGCATCCTCTACACAGTATAAAAATGGGGTATTGTATAGAGGTTGCGTCAAT
TAATTCGGGTTGGACGGAGTACTAAAGTTGAGGTAGGC
>hvu-novel-miR90020
TTTTTGTGTTTGGTTCGGTCTTTTTTTGTCGTTTTTGGTTCGGCCGTCCTCCGACGTCGCGACTGTTTTAGATTTGGATCGGCACCGTG
TCCATGTTAAT
>hvu-novel-miR90021
TTCATCTTCGACGGTGGCATTGAGGAAAATCTCACGATTTCTCGGCATTTTTGAACACCTT
>hvu-novel-miR90022
GTCCTTCCCTCTTGAACACCCATGGCTATAGTCTTTTCTATGTCCATTTTTTCAGGCGACATGTTGTA
>hvu-novel-miR90023
TTCGTGGGCTGTGGGTTGTGTAGAGTGGTAGTGTGTTGAGATGAGAGGGAGTTTGTGGTCCGTGACACGCGTTTTAAACATCGAAACGTCGG
ACGACTAAACCAATG
>hvu-novel-miR90024
GGAGCTGCATTTGACCTGCACCTACGGAGGAAGATGGGCATGCATGTGTGCATGCATGCAAGAGAAAGAACCGCGCCGTCTTGGCCCTCTT
GCTCTCCGATAGGTGCGATGGCATATGCAACTCTAC
>hvu-novel-miR90025
TAGTTTTTCCCTTATGAACGCATTGAATTTTTTTGGCAAGTTGATAAACTCTTTGATATTAC
>hvu-novel-miR90026
GCGCGTGTGCTGATGAGAAGGGCGCGGTTTCGATTTCCGCTTCGGCACCTTATTTTTGA
>hvu-novel-miR90028
TGGTTTTGGTGTAGTTGGAGTCTGATTCGATTTCCCCCTTCGGCACCTTTTGCATT
>hvu-novel-miR90029
CATGGTTCACAAATAAGGCCATTACTCTAGATTGCTAATAGAATAATGTATTTCCATGCC
>hvu-novel-miR90030
CGCGGTGGCAGGTTTCATGAAGATTAGGCATCGGCTCGTCGATCGTGGTGTAGTGTGCTAG
>hvu-novel-miR90031
TCACCGAGCTGCCTTCGGTGTGTGGGATCTCCGCACAACCTCCGAATAGATCTACAGGGCG
>hvu-novel-miR90032
ATTTCACTAGAGTCATATTTTTACATCTCATAGAGATCTGGTGGAAACACCATTGGCAGTA
>hvu-novel-miR90033
CATATGCATGCGCCTGATGAGAGTCCAGGGTTCGAGTCCCCCGTCCGGCGCTTTTGTTC
>hvu-novel-miR90034
TGTGTGTTGTTGCCTGTGAGAGGGCGTGGGTTTCGAGTCCCCCGTCCGCATCCTTTTTACTT
>hvu-novel-miR90035
GCCATTGATGGATCTGAGATCCGCGGTCTCATCGCCTGAGATCTGCGCATCCCGGATGCAATCAACTACTA
>hvu-novel-miR90036
TGAGATCATAACCGAGTCGGTGAGGGAACCCATAGTAGAGTCGGCACCGGAGCCATGGAG
>hvu-novel-miR90037
GCTATCCTGCAACCGAAGCTGTTGATTCGATTCGATCGAACGGCTCGAATTGGTGAACCCCC
>hvu-novel-miR90038
ATCGAGGGGAGGTGAGAATCGGCCGAGGCTGGCCACCTTTTCTCGGGTGGGCACTGTAGCGTGCCTAGA
>hvu-novel-miR90039
GTCATTTTCTCCGTGGATTTCCGACTGGGGTGAAGAGGCTTAGGAGGTGAGGTGGTA

Table A.9.9. EpiGBS barcode combinations used with biological samples. Samples with L are for the third leaf tissue, R for root tissue, C for control and S for salt treated. The cytosine used in the 5' barcodes is 5-methylcytosine to ensure the barcode does not change during bisulphite treatment.

Sample	Ba Adapter		Co Adapter	
	Identifier	Barcode Sequence	Identifier	Barcode Sequence
L1C	1	AACT	2	CCAG
L2C	2	CCTA	2	CCAG
L3C	3	TTAC	3	TTGA
L4C	4	AGGC	3	TTGA
L5S	1	AACT	4	GGTC
L6S	2	CCTA	4	GGTC
L7S	3	TTAC	5	ACTA
L8S	4	AGGC	5	ACTA
R1C	1	AACT	6	CAGC
R2C	2	CCTA	6	CAGC
R3C	3	TTAC	7	TGAT
R4C	4	AGGC	7	TGAT
R5S	1	AACT	8	GTCG
R6S	2	CCTA	8	GTCG
R7S	3	TTAC	9	ATAC
R8S	4	AGGC	9	ATAC

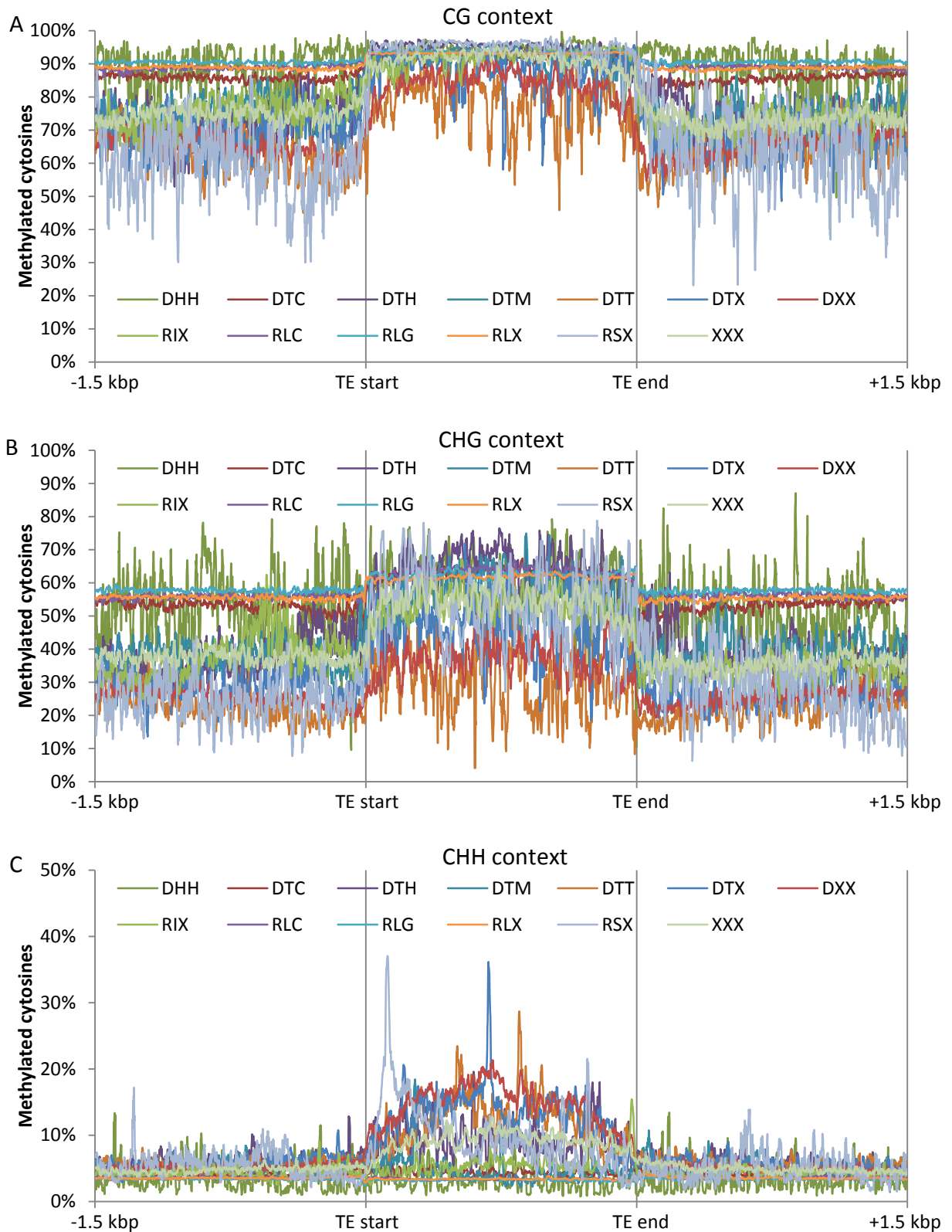


Figure A.9.3. Methylation patterns in barley around different TE classes. The methylation patterns for the three different cytosine contexts A: CG methylation, B: CHG methylation and C: CHH methylation. All cytosine reads for the relevant context that were within 1.5 kbp of a TE were included. The TE body from TE start to TE end was scaled for each TE to fit in the same width. TE classification based on the standard established by Wicker et al. (2007)

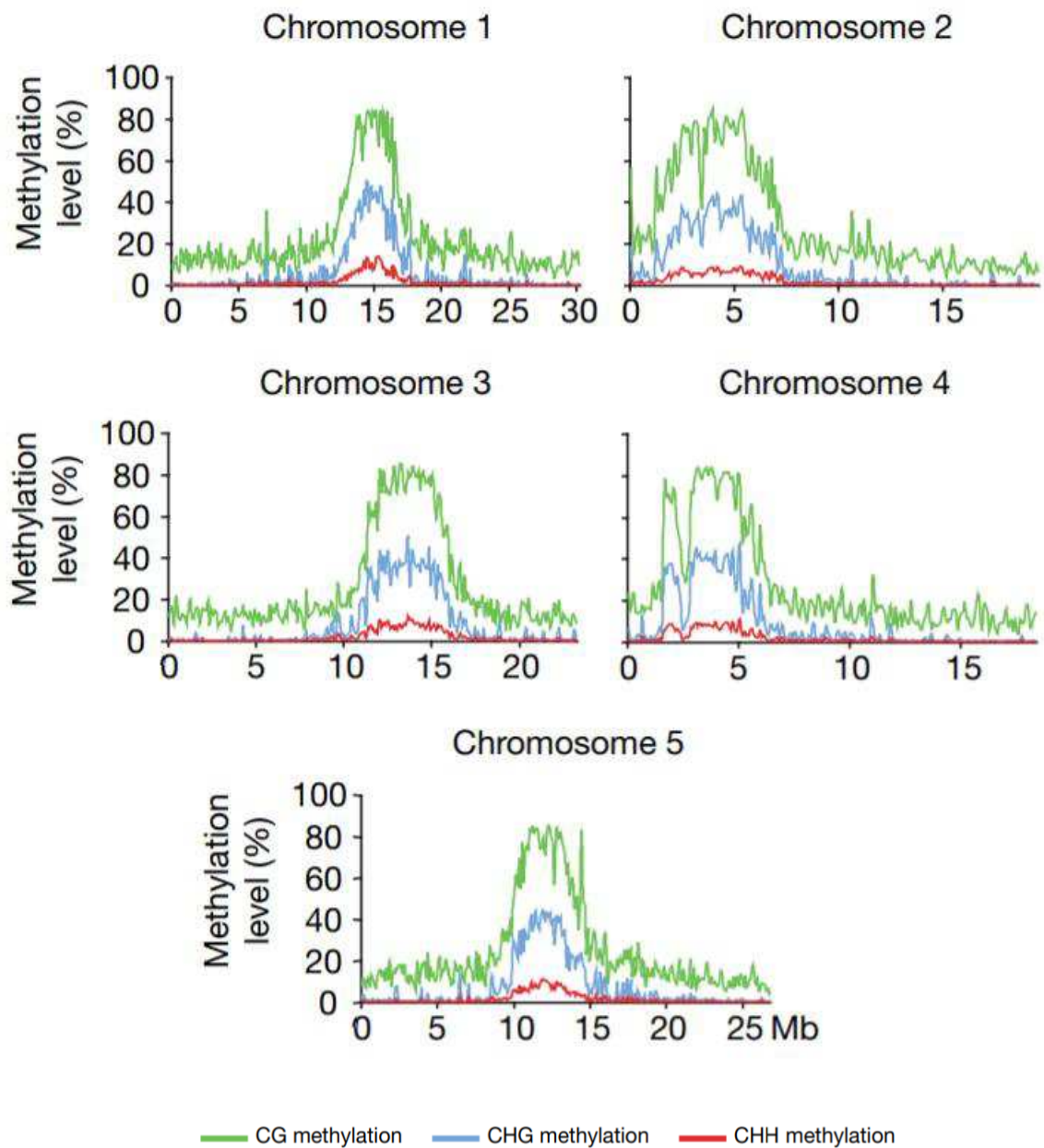


Figure A.9.4. *Arabidopsis thaliana* methylation pattern across each chromosome and cytosine context, reproduced from Cokus et al. (2008).

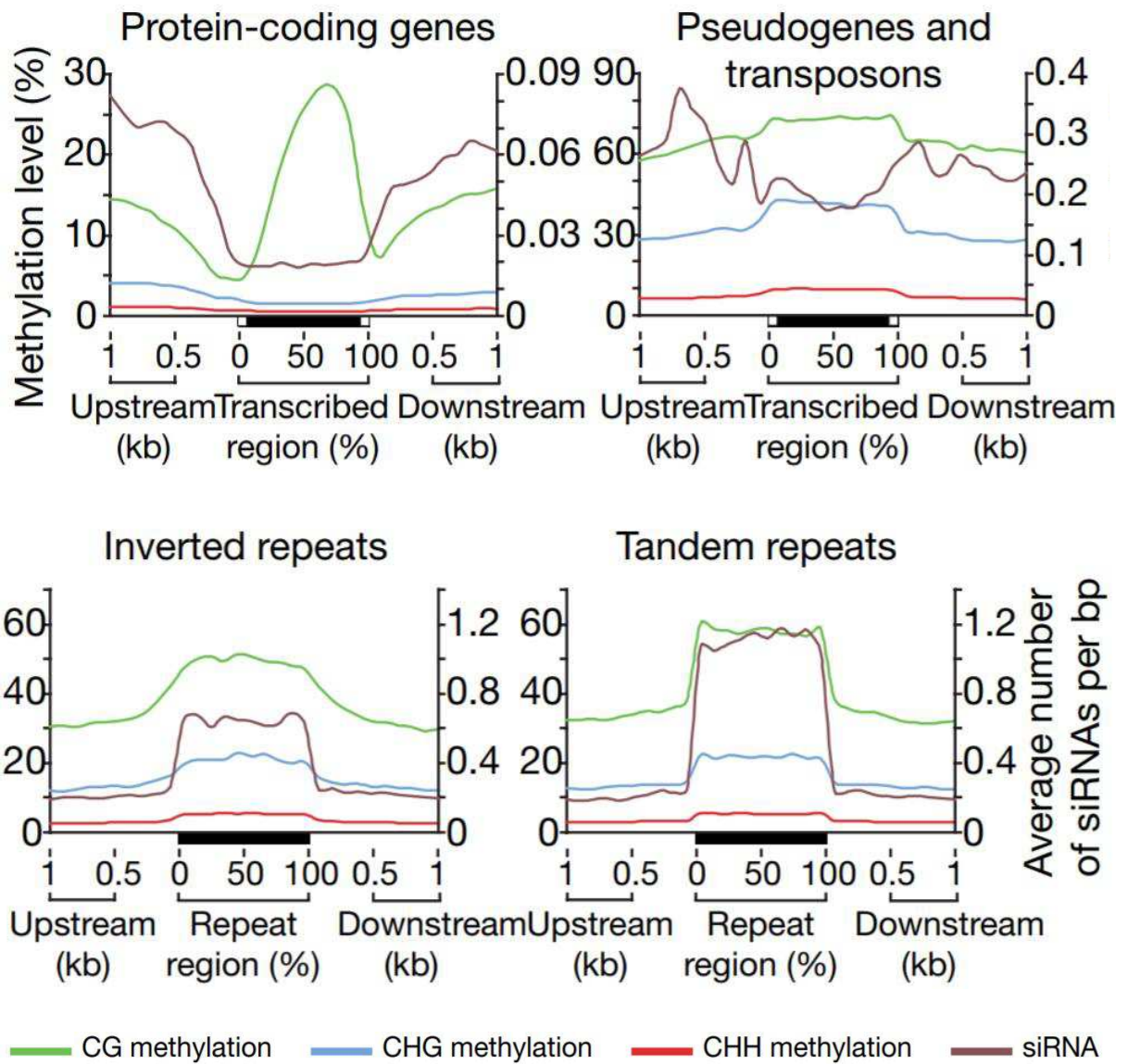


Figure A.9.5. Methylation patterns around different genomic features in *Arabidopsis thaliana*, reproduced from Cokus et al. (2008).

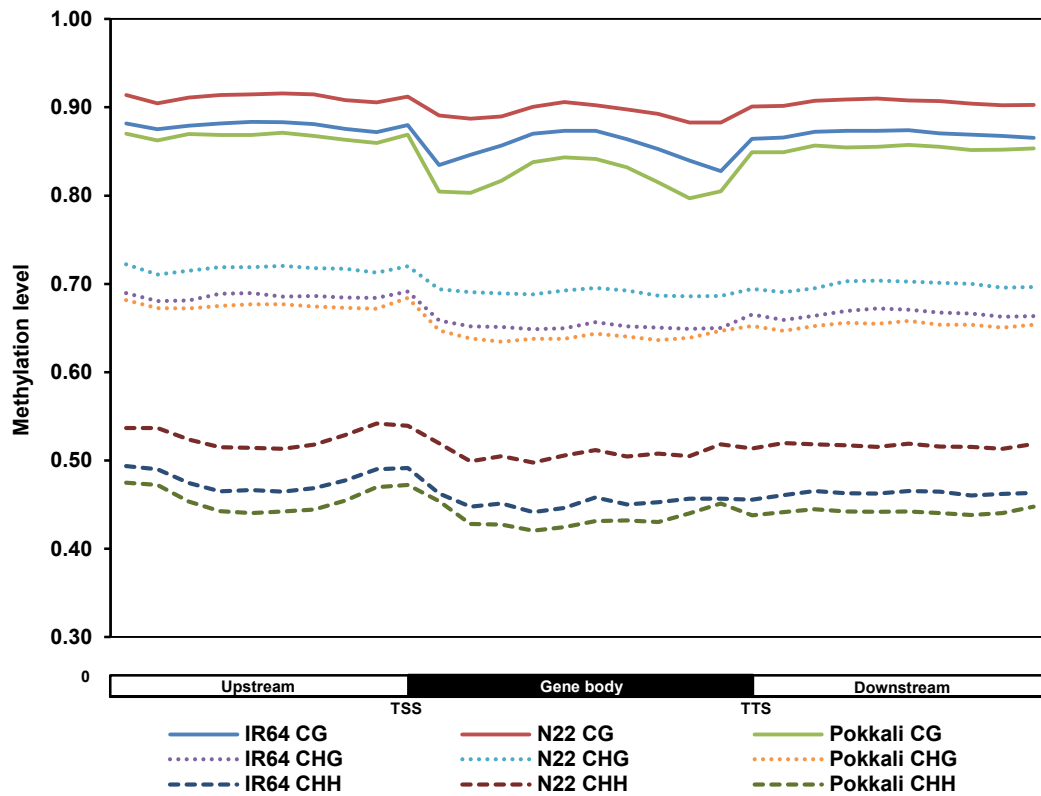


Figure A.9.6. DNA methylation patterns for all cytosine contexts around genes in three different rice varieties, reproduced from Garg et al. (2015).

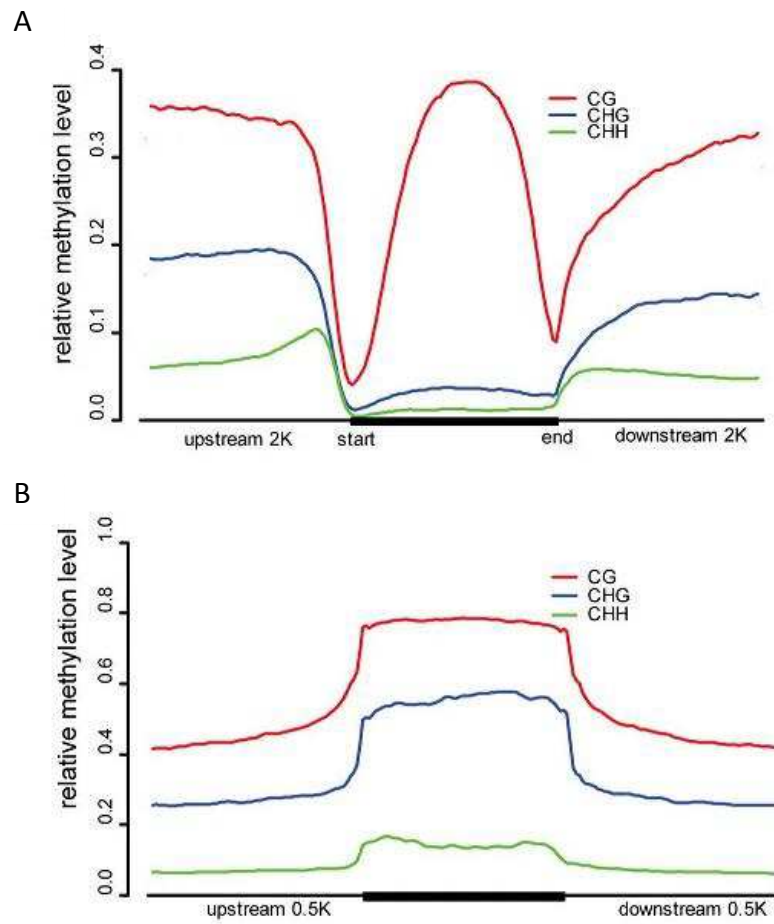


Figure A.9.7. DNA methylation patterns in wild rice for (A) genes and (B) transposable elements, reproduced from Li et al. (2012)

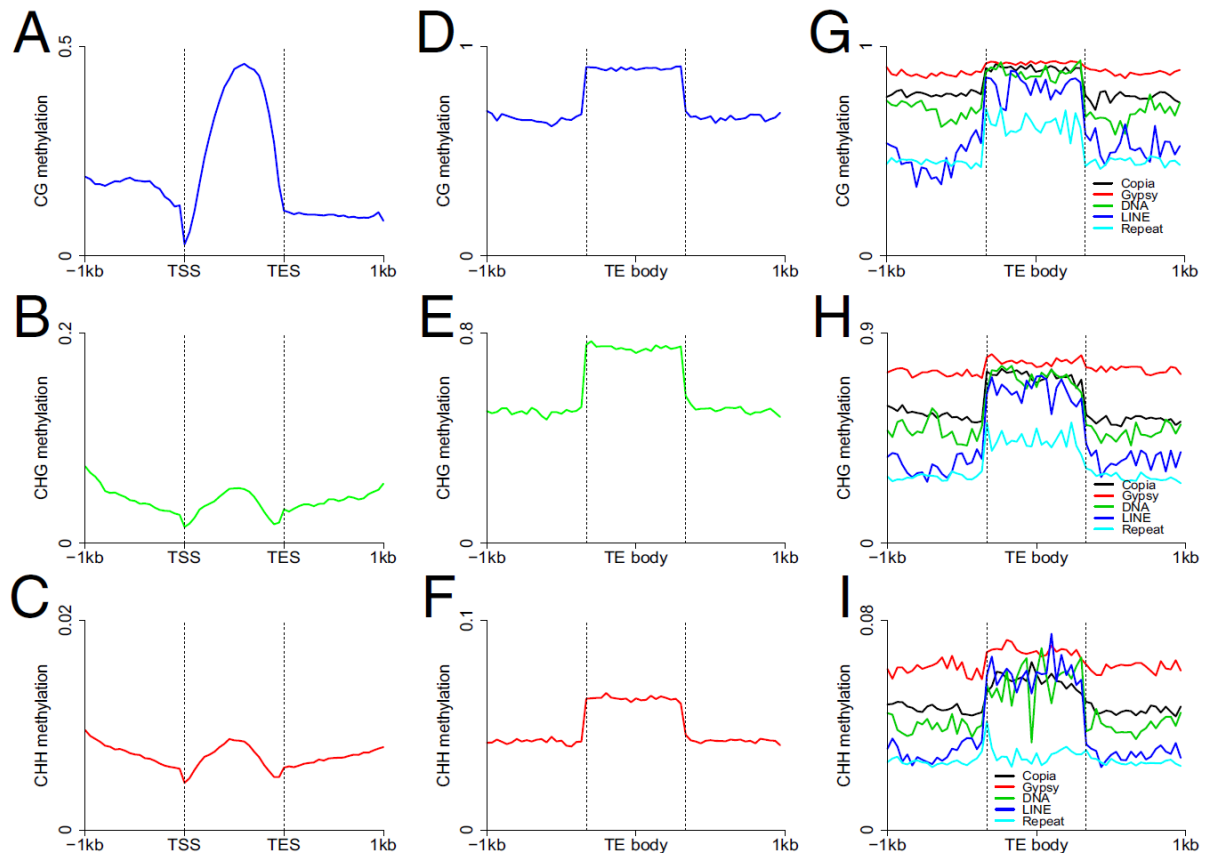


Figure A.9.8. DNA methylation patterns in cassava for all cytosine contexts for genes and Transposable Elements (TEs), reproduced from Wang et al. (2015). A-C: Methylation pattern for genes. D-F: Methylation pattery for TEs. G-I: Methylation pattern for certain selected classes of TEs.

Table A.9.10. Differentially methylated CG context genes in barley leaf tissue under salt stress. Genes within 1.5 kbp of differentially methylated sites in CG contexts in plant leaf tissue are listed. Difference in % methylation is the difference between methylation percentage in salt treated plants minus the methylation percentage in control plants. Some genes are listed more than once if more than one differentially methylated site fell in the region of the gene.

Gene	Difference in % methylation	log ₁₀ (adj-p)	Gene	Difference in % methylation	log ₁₀ (adj-p)
HORVU6Hr1G033380	100.0	-1.69	HORVU2Hr1G032280	13.5	-1.50
HORVU6Hr1G011260	92.3	-3.37	HORVU1Hr1G048900	12.8	-2.83
HORVU3Hr1G004480	88.2	-6.53	HORVU1Hr1G004930	12.5	-2.55
HORVU7Hr1G040550	78.6	-1.89	HORVU6Hr1G008870	12.3	-1.45
HORVU3Hr1G099160	74.1	-3.28	HORVU7Hr1G021000	11.3	-2.24
HORVU1Hr1G010390	73.3	-1.35	HORVU7Hr1G033530	-13.0	-2.47
HORVU3Hr1G113890	68.9	-1.80	HORVU1Hr1G024540	-15.0	-1.33
HORVU5Hr1G079970	66.5	-3.41	HORVU1Hr1G041030	-16.7	-2.11
HORVU4Hr1G070620	65.0	-1.69	HORVU4Hr1G001450	-20.0	-2.06
HORVU3Hr1G095430	58.0	-1.46	HORVU3Hr1G075970	-20.1	-4.38
HORVU7Hr1G120030	55.0	-2.71	HORVU7Hr1G028190	-27.5	-1.57
HORVU7Hr1G103020	54.2	-1.38	HORVU2Hr1G092080	-29.1	-2.61
HORVU2Hr1G122620	53.6	-1.42	HORVU5Hr1G110490	-29.2	-3.00
HORVU2Hr1G013370	52.7	-3.18	HORVU5Hr1G068590	-31.0	-1.57
HORVU6Hr1G001880	52.1	-1.72	HORVU3Hr1G003860	-31.7	-2.19
HORVU3Hr1G030560	51.7	-1.32	HORVU6Hr1G001960	-33.7	-1.73
HORVU1Hr1G091650	44.0	-1.31	HORVU7Hr1G117010	-35.7	-2.29
HORVU3Hr1G017270	42.5	-2.33	HORVU3Hr1G016230	-37.5	-2.09
HORVU2Hr1G122620	40.7	-1.58	HORVU1Hr1G085240	-38.2	-1.44
HORVU3Hr1G018250	39.8	-1.48	HORVU7Hr1G118180	-38.7	-2.61
HORVU7Hr1G002720	39.6	-1.89	HORVU7Hr1G117320	-38.8	-1.53
HORVU7Hr1G080200	37.9	-1.78	HORVU4Hr1G088760	-39.7	-1.48
HORVU3Hr1G107760	37.7	-2.19	HORVU2Hr1G070490	-40.0	-1.72
HORVU7Hr1G000040	37.5	-1.34	HORVU3Hr1G010140	-40.5	-1.38
HORVU1Hr1G061230	36.8	-1.76	HORVU7Hr1G081220	-42.2	-1.40
HORVU3Hr1G085760	35.5	-1.98	HORVU5Hr1G064830	-43.3	-1.40
HORVU3Hr1G021970	34.2	-3.28	HORVU5Hr1G075400	-44.0	-1.93
HORVU2Hr1G039030	33.3	-1.38	HORVU7Hr1G049190	-49.1	-1.85
HORVU2Hr1G101470	33.3	-2.14	HORVU5Hr1G064830	-50.0	-3.57
HORVU6Hr1G037950	28.9	-1.56	HORVU5Hr1G026160	-54.5	-1.54
HORVU7Hr1G120030	28.4	-2.19	HORVU1Hr1G075780	-61.8	-3.97
HORVU4Hr1G059270	28.2	-1.58	HORVU2Hr1G012540	-61.9	-2.43
HORVU3Hr1G090850	28.1	-2.49	HORVU5Hr1G097670	-62.9	-1.58
HORVU7Hr1G077810	27.9	-1.33	HORVU6Hr1G008870	-64.3	-2.09
HORVU5Hr1G064470	27.5	-1.32	HORVU7Hr1G068410	-64.3	-1.31
HORVU7Hr1G000040	25.8	-1.78	HORVU7Hr1G103870	-69.2	-1.41
HORVU3Hr1G047180	25.8	-3.46	HORVU1Hr1G062730	-70.0	-1.31
HORVU5Hr1G000940	24.6	-2.08	HORVU3Hr1G013650	-72.8	-4.56
HORVU6Hr1G020670	24.3	-1.43	HORVU7Hr1G000040	-73.0	-2.13
HORVU4Hr1G002120	19.4	-1.52	HORVU3Hr1G031320	-92.3	-1.79
HORVU7Hr1G085060	18.7	-1.31	HORVU4Hr1G080050	-92.9	-1.61
HORVU7Hr1G105580	17.5	-1.44			
HORVU1Hr1G005470	16.1	-2.06			

Table A.9.11. Differentially methylated CHG context genes in barley leaf tissue under salt stress. Genes within 1.5 kbp of differentially methylated sites in CHG contexts in plant leaf tissue are listed. Difference in % methylation is the difference between methylation percentage in salt treated plants minus the methylation percentage in control plants. Some genes are listed more than once if more than one differentially methylated site fell in the region of the gene.

Gene	Difference in % methylation	log ₁₀ (adj-p)	Gene	Difference in % methylation	log ₁₀ (adj-p)
HORVU4Hr1G044370	80.0	-2.13	HORVU2Hr1G091580	19.2	-1.59
HORVU2Hr1G060460	78.6	-2.36	HORVU7Hr1G042890	19.1	-1.42
HORVU4Hr1G081660	66.1	-1.39	HORVU5Hr1G122540	16.2	-1.42
HORVU3Hr1G081410	64.4	-1.50	HORVU4Hr1G039540	13.0	-1.39
HORVU5Hr1G123050	62.6	-1.49	HORVU3Hr1G035730	12.7	-1.52
HORVU2Hr1G059320	57.6	-1.82	HORVU2Hr1G025510	12.5	-1.39
HORVU4Hr1G085290	55.8	-1.38	HORVU5Hr1G051570	-11.2	-2.64
HORVU5Hr1G119490	54.5	-2.69	HORVU2Hr1G099160	-13.2	-1.83
HORVU1Hr1G013990	52.5	-1.34	HORVU2Hr1G066890	-16.0	-1.38
HORVU1Hr1G003140	46.2	-1.41	HORVU7Hr1G053070	-16.7	-1.44
HORVU7Hr1G111630	45.9	-1.98	HORVU1Hr1G092110	-18.2	-1.49
HORVU6Hr1G026400	45.8	-2.55	HORVU0Hr1G040200	-20.0	-1.70
HORVU2Hr1G008140	44.0	-2.00	HORVU3Hr1G110590	-20.0	-1.73
HORVU4Hr1G057520	42.9	-1.49	HORVU3Hr1G095940	-20.7	-1.54
HORVU5Hr1G031470	42.1	-1.64	HORVU5Hr1G068600	-22.2	-1.52
HORVU7Hr1G099800	41.5	-2.65	HORVU7Hr1G085570	-22.2	-1.75
HORVU6Hr1G018420	39.1	-1.69	HORVU2Hr1G056510	-27.3	-1.72
HORVU7Hr1G045660	38.7	-2.13	HORVU4Hr1G088760	-27.7	-1.48
HORVU5Hr1G056760	37.5	-1.34	HORVU3Hr1G099430	-28.6	-3.45
HORVU6Hr1G078520	37.5	-1.39	HORVU4Hr1G088760	-29.4	-2.55
HORVU6Hr1G083130	36.9	-2.69	HORVU2Hr1G013700	-33.3	-1.33
HORVU7Hr1G006070	36.9	-2.05	HORVU3Hr1G006760	-38.0	-1.31
HORVU5Hr1G017930	34.9	-1.43	HORVU3Hr1G090910	-41.9	-1.40
HORVU3Hr1G000590	31.8	-1.59	HORVU2Hr1G035080	-43.5	-1.47
HORVU3Hr1G098440	31.8	-2.42	HORVU1Hr1G032310	-45.3	-4.43
HORVU1Hr1G092000	30.0	-1.82	HORVU2Hr1G119310	-46.3	-2.00
HORVU2Hr1G103130	29.2	-1.47	HORVU2Hr1G108370	-48.7	-1.50
HORVU2Hr1G085850	28.6	-1.52	HORVU4Hr1G004270	-55.7	-1.52
HORVU5Hr1G037160	24.2	-1.69	HORVU3Hr1G057330	-58.3	-1.33
HORVU3Hr1G078300	24.1	-1.88	HORVU2Hr1G024680	-65.2	-1.41
HORVU2Hr1G103130	23.9	-1.64	HORVU2Hr1G034320	-66.0	-1.80
HORVU4Hr1G084910	22.1	-2.46	HORVU7Hr1G009770	-71.1	-2.53
HORVU5Hr1G048550	22.0	-2.48	HORVU3Hr1G050400	-73.7	-2.65
HORVU0Hr1G007540	21.7	-1.43	HORVU3Hr1G032660	-100.0	-2.55
HORVU7Hr1G092280	20.7	-2.00			

Table A.9.12. Differentially methylated CHH context genes in barley leaf tissue under salt stress. Genes within 1.5 kbp of differentially methylated sites in CHH contexts in plant leaf tissue are listed. Difference in % methylation is the difference between methylation percentage in salt treated plants minus the methylation percentage in control plants. A number of genes are listed more than once where more than one differentially methylated site fell in the region of the gene.

This table is too large to show here. Please follow the link below to view this table.
https://github.com/MagnificaScience/Salinity_stress_response_in_barley

Table A.9.13. Differentially methylated CG context genes in barley root tissue under salt stress. Genes within 1.5 kbp of differentially methylated sites in CG contexts in plant root tissue are listed. Difference in % methylation is the difference between methylation percentage in salt treated plants minus the methylation percentage in control plants. Some genes are listed more than once if more than one differentially methylated site fell in the region of the gene.

Gene	Difference in % methylation	log ₁₀ (adj-p)	Gene	Difference in % methylation	log ₁₀ (adj-p)
HORVU1Hr1G016600	100.0	-2.38	HORVU4Hr1G035180	36.7	-1.81
HORVU3Hr1G031320	100.0	-6.72	HORVU7Hr1G109020	36.1	-2.00
HORVU4Hr1G003360	100.0	-1.63	HORVU6Hr1G008550	35.2	-1.41
HORVU6Hr1G063130	100.0	-1.84	HORVU2Hr1G127440	35.0	-1.38
HORVU7Hr1G047720	100.0	-1.49	HORVU6Hr1G013290	34.8	-2.05
HORVU7Hr1G122200	100.0	-1.63	HORVU2Hr1G003990	34.5	-1.44
HORVU5Hr1G024800	91.7	-1.49	HORVU1Hr1G090670	33.3	-1.44
HORVU2Hr1G103020	90.9	-1.41	HORVU6Hr1G000440	32.5	-1.78
HORVU6Hr1G080620	90.0	-1.31	HORVU4Hr1G018610	31.7	-1.79
HORVU2Hr1G030040	85.5	-2.28	HORVU3Hr1G020490	31.3	-2.73
HORVU4Hr1G009870	83.3	-2.33	HORVU3Hr1G086610	30.9	-5.25
HORVU6Hr1G021260	83.3	-2.71	HORVU5Hr1G096560	30.3	-1.87
HORVU7Hr1G006800	80.0	-2.82	HORVU2Hr1G018430	29.5	-3.06
HORVU7Hr1G068450	80.0	-1.95	HORVU3Hr1G038610	29.4	-2.58
HORVU5Hr1G009040	76.9	-1.47	HORVU5Hr1G012770	29.3	-1.39
HORVU1Hr1G057910	75.0	-1.63	HORVU1Hr1G035960	28.6	-2.35
HORVU5Hr1G081840	72.7	-1.36	HORVU3Hr1G020900	28.6	-1.43
HORVU1Hr1G025280	72.2	-5.16	HORVU0Hr1G010390	28.2	-1.43
HORVU0Hr1G022520	68.8	-3.26	HORVU2Hr1G083430	28.0	-2.05
HORVU7Hr1G076810	68.4	-2.03	HORVU1Hr1G086110	27.4	-1.65
HORVU5Hr1G075920	67.6	-3.11	HORVU5Hr1G105780	26.7	-1.79
HORVU5Hr1G107990	67.6	-1.39	HORVU3Hr1G013150	25.9	-1.39
HORVU1Hr1G077910	66.7	-1.30	HORVU2Hr1G011070	25.7	-2.31
HORVU2Hr1G103490	66.7	-1.43	HORVU7Hr1G006980	25.7	-2.10
HORVU5Hr1G041660	63.5	-1.50	HORVU3Hr1G117400	24.6	-2.98
HORVU5Hr1G044220	62.5	-2.75	HORVU5Hr1G056760	24.6	-1.44
HORVU1Hr1G057460	58.3	-1.35	HORVU4Hr1G069790	24.3	-2.52
HORVU3Hr1G108500	58.3	-1.75	HORVU7Hr1G120030	23.8	-1.36
HORVU7Hr1G033530	58.3	-2.93	HORVU3Hr1G071480	23.6	-3.25
HORVU2Hr1G013370	57.6	-3.23	HORVU1Hr1G021000	22.6	-1.45
HORVU1Hr1G077080	56.4	-1.38	HORVU3Hr1G028020	22.2	-1.34
HORVU5Hr1G117220	54.5	-1.40	HORVU7Hr1G008140	22.1	-1.62
HORVU3Hr1G081180	53.9	-2.56	HORVU5Hr1G112930	21.9	-1.65
HORVU7Hr1G026510	53.6	-1.73	HORVU2Hr1G126180	21.7	-2.20
HORVU3Hr1G102840	52.9	-1.39	HORVU5Hr1G016670	21.6	-1.64
HORVU3Hr1G000230	52.4	-3.24	HORVU2Hr1G018430	21.5	-2.09
HORVU7Hr1G099320	52.3	-1.53	HORVU2Hr1G031120	21.4	-1.45
HORVU1Hr1G057910	52.2	-2.70	HORVU4Hr1G072840	21.1	-1.43
HORVU1Hr1G079410	50.0	-1.31	HORVU7Hr1G083580	20.6	-2.06
HORVU6Hr1G000810	50.0	-1.87	HORVU3Hr1G040500	20.2	-1.51
HORVU7Hr1G122750	50.0	-2.49	HORVU4Hr1G034660	20.0	-1.49
HORVU4Hr1G048210	49.0	-1.95	HORVU7Hr1G027680	19.6	-1.51
HORVU5Hr1G059030	48.5	-2.52	HORVU3Hr1G110540	19.2	-2.16
HORVU1Hr1G048870	47.9	-1.69	HORVU5Hr1G012120	19.1	-2.10
HORVU5Hr1G005910	45.8	-1.50	HORVU7Hr1G090440	19.0	-1.45
HORVU4Hr1G079800	45.8	-1.51	HORVU3Hr1G057990	19.0	-2.18
HORVU6Hr1G064370	45.0	-1.58	HORVU6Hr1G082310	18.9	-1.76
HORVU2Hr1G107600	44.4	-2.69	HORVU3Hr1G026970	18.6	-1.49
HORVU1Hr1G072160	44.4	-3.30	HORVU2Hr1G069470	18.1	-1.64
HORVU6Hr1G092270	43.5	-7.27	HORVU2Hr1G123270	18.1	-1.44
HORVU3Hr1G095940	43.5	-1.65	HORVU2Hr1G048070	17.9	-1.34
HORVU6Hr1G070170	43.4	-1.50	HORVU2Hr1G126180	17.9	-1.98
HORVU6Hr1G083960	43.3	-1.68	HORVU3Hr1G015800	17.8	-1.98
HORVU7Hr1G009420	42.9	-2.78	HORVU6Hr1G008870	17.8	-3.14
HORVU7Hr1G111260	41.8	-2.13	HORVU5Hr1G078500	16.7	-1.86
HORVU5Hr1G015220	41.2	-2.71	HORVU7Hr1G097070	16.7	-1.43
HORVU3Hr1G056830	40.0	-1.44	HORVU7Hr1G023100	16.0	-1.59
HORVU5Hr1G023720	40.0	-1.59	HORVU4Hr1G015530	15.7	-1.75
HORVU5Hr1G084410	39.3	-1.49	HORVU3Hr1G002130	15.3	-1.37
HORVU2Hr1G106480	38.7	-1.65	HORVU5Hr1G070390	15.0	-1.60
HORVU3Hr1G097400	38.5	-2.03	HORVU6Hr1G048850	15.0	-1.59
HORVU1Hr1G062740	37.5	-1.59	HORVU2Hr1G084170	14.6	-2.08
HORVU3Hr1G038050	37.1	-1.55	HORVU3Hr1G037030	14.6	-1.99

Gene	Difference in % methylation	log ₁₀ (adj-p)	Gene	Difference in % methylation	log ₁₀ (adj-p)
HORVU4Hr1G034660	14.3	-1.66	HORVU1Hr1G076670	-27.5	-2.82
HORVU2Hr1G023830	12.7	-1.44	HORVU1Hr1G087800	-27.6	-2.44
HORVU5Hr1G118500	12.2	-1.73	HORVU3Hr1G075840	-27.8	-1.88
HORVU3Hr1G112190	11.7	-1.64	HORVU1Hr1G055600	-27.9	-1.72
HORVU2Hr1G101470	11.4	-1.49	HORVU5Hr1G068350	-28.2	-2.04
HORVU2Hr1G125640	11.4	-1.58	HORVU1Hr1G016280	-28.6	-1.67
HORVU5Hr1G044640	11.1	-1.40	HORVU3Hr1G101540	-30.9	-4.11
HORVU2Hr1G036530	10.8	-1.41	HORVU7Hr1G083580	-31.0	-2.14
HORVU5Hr1G071530	10.7	-1.68	HORVU6Hr1G001960	-31.0	-1.34
HORVU1Hr1G083340	10.6	-1.70	HORVU3Hr1G079490	-31.3	-1.33
HORVU2Hr1G014240	10.3	-1.86	HORVU7Hr1G101050	-31.7	-2.35
HORVU5Hr1G067530	10.0	-1.82	HORVU3Hr1G031640	-32.6	-2.33
HORVU4Hr1G054940	-10.3	-1.30	HORVU1Hr1G052930	-33.3	-1.41
HORVU7Hr1G084580	-11.3	-1.44	HORVU4Hr1G073070	-33.3	-1.67
HORVU3Hr1G010920	-11.5	-1.54	HORVU2Hr1G092080	-33.9	-1.86
HORVU4Hr1G071690	-12.0	-1.41	HORVU4Hr1G052890	-34.3	-1.76
HORVU6Hr1G063790	-12.1	-1.37	HORVU6Hr1G079880	-36.8	-1.46
HORVU2Hr1G003340	-12.5	-1.86	HORVU0Hr1G011860	-37.6	-1.85
HORVU1Hr1G074260	-12.8	-1.43	HORVU0Hr1G011860	-38.3	-1.54
HORVU3Hr1G075970	-13.1	-1.86	HORVU5Hr1G023960	-41.4	-3.72
HORVU2Hr1G050270	-13.1	-1.30	HORVU7Hr1G095570	-42.4	-2.32
HORVU7Hr1G099870	-13.2	-1.34	HORVU3Hr1G018300	-44.1	-1.47
HORVU2Hr1G099160	-13.4	-1.65	HORVU7Hr1G118180	-44.5	-3.47
HORVU6Hr1G073170	-14.6	-1.92	HORVU5Hr1G079380	-45.8	-1.85
HORVU2Hr1G032540	-15.4	-1.31	HORVU5Hr1G042080	-46.1	-1.73
HORVU7Hr1G013300	-15.6	-1.89	HORVU3Hr1G039290	-46.7	-4.14
HORVU3Hr1G095800	-15.7	-1.37	HORVU5Hr1G124360	-46.7	-1.72
HORVU1Hr1G082670	-16.0	-1.36	HORVU7Hr1G117320	-49.2	-2.17
HORVU4Hr1G002240	-16.7	-1.50	HORVU3Hr1G039250	-49.2	-1.46
HORVU4Hr1G021270	-16.7	-1.32	HORVU2Hr1G101900	-50.0	-1.65
HORVU7Hr1G118740	-17.1	-1.42	HORVU1Hr1G039440	-50.4	-5.00
HORVU6Hr1G013110	-17.3	-1.66	HORVU1Hr1G059550	-54.4	-1.36
HORVU4Hr1G052810	-17.4	-1.34	HORVU0Hr1G038830	-54.5	-1.92
HORVU7Hr1G019770	-17.6	-1.51	HORVU4Hr1G072520	-55.0	-1.36
HORVU1Hr1G035760	-18.2	-1.43	HORVU7Hr1G117010	-55.2	-4.72
HORVU1Hr1G061160	-18.5	-1.52	HORVU1Hr1G085240	-57.7	-3.64
HORVU1Hr1G056340	-18.6	-2.22	HORVU5Hr1G078310	-58.8	-1.86
HORVU6Hr1G014360	-19.3	-2.78	HORVU6Hr1G058590	-59.7	-2.82
HORVU2Hr1G026540	-19.4	-2.71	HORVU7Hr1G036000	-60.4	-1.87
HORVU2Hr1G010690	-20.0	-1.69	HORVU1Hr1G076680	-61.1	-1.31
HORVU4Hr1G003660	-20.0	-1.49	HORVU7Hr1G106280	-62.2	-1.39
HORVU7Hr1G000320	-21.1	-1.50	HORVU7Hr1G122750	-62.5	-1.51
HORVU1Hr1G081570	-21.3	-2.53	HORVU2Hr1G126170	-65.2	-1.99
HORVU6Hr1G027460	-21.3	-2.80	HORVU2Hr1G050270	-70.0	-1.68
HORVU6Hr1G013110	-21.6	-1.31	HORVU7Hr1G008130	-71.4	-1.43
HORVU1Hr1G028250	-22.2	-1.72	HORVU1Hr1G094570	-75.0	-1.43
HORVU2Hr1G085910	-22.2	-1.45	HORVU5Hr1G079390	-75.0	-2.96
HORVU4Hr1G064660	-23.1	-1.36	HORVU5Hr1G006960	-76.9	-1.84
HORVU2Hr1G100660	-23.5	-1.36	HORVU3Hr1G113340	-78.3	-2.71
HORVU3Hr1G087640	-23.8	-1.37	HORVU2Hr1G039880	-80.0	-1.52
HORVU3Hr1G077170	-23.8	-1.49	HORVU3Hr1G082670	-80.0	-1.36
HORVU7Hr1G093790	-24.0	-1.34	HORVU3Hr1G093140	-80.0	-1.41
HORVU2Hr1G095130	-24.0	-1.38	HORVU4Hr1G067870	-80.0	-1.69
HORVU4Hr1G011850	-24.1	-3.17	HORVU2Hr1G056510	-85.7	-1.45
HORVU5Hr1G069620	-24.7	-1.40	HORVU5Hr1G006220	-90.0	-1.52
HORVU2Hr1G009580	-25.0	-1.36	HORVU1Hr1G001950	-100.0	-1.34
HORVU4Hr1G008370	-25.2	-1.71	HORVU1Hr1G070740	-100.0	-1.84
HORVU5Hr1G120770	-26.6	-1.36	HORVU3Hr1G033980	-100.0	-1.49
HORVU5Hr1G044150	-26.8	-2.17	HORVU3Hr1G060080	-100.0	-2.38
HORVU5Hr1G049810	-27.4	-1.85	HORVU7Hr1G089720	-100.0	-2.25
HORVU6Hr1G044840	-27.5	-1.91			

Table A.9.14. Differentially methylated CHG context genes in barley root tissue under salt stress. Genes within 1.5 kbp of differentially methylated sites in CHG contexts in plant root tissue are listed. Difference in % methylation is the difference between methylation percentage in salt treated plants minus the methylation percentage in control plants. Some genes are listed more than once if more than one differentially methylated site fell in the region of the gene.

Gene	Difference in % methylation	log ₁₀ (adj-p)	Gene	Difference in % methylation	log ₁₀ (adj-p)
HORVU6Hr1G025500	100.0	-2.23	HORVU2Hr1G056510	30.3	-1.59
HORVU5Hr1G092290	88.0	-2.05	HORVU1Hr1G092060	30.1	-1.33
HORVU4Hr1G045760	81.8	-2.26	HORVU0Hr1G016720	30.0	-1.31
HORVU5Hr1G079390	81.8	-1.37	HORVU4Hr1G087570	29.6	-1.63
HORVU1Hr1G060710	79.4	-2.92	HORVU7Hr1G091400	29.2	-1.51
HORVU5Hr1G079390	77.3	-2.90	HORVU4Hr1G018610	28.8	-1.45
HORVU2Hr1G088270	76.1	-1.57	HORVU5Hr1G095060	28.8	-1.36
HORVU6Hr1G039660	70.6	-5.63	HORVU2Hr1G088980	25.7	-1.34
HORVU7Hr1G068450	70.0	-1.34	HORVU7Hr1G034080	25.3	-1.44
HORVU5Hr1G079740	68.9	-1.58	HORVU7Hr1G023100	24.9	-1.73
HORVU5Hr1G021940	68.2	-2.05	HORVU4Hr1G071600	24.7	-1.83
HORVU2Hr1G041470	64.5	-2.06	HORVU0Hr1G010530	24.0	-1.40
HORVU4Hr1G062060	64.2	-1.52	HORVU2Hr1G015940	23.8	-1.56
HORVU1Hr1G048870	63.5	-5.98	HORVU0Hr1G018800	23.0	-2.31
HORVU2Hr1G050270	63.0	-2.04	HORVU2Hr1G007170	22.8	-1.41
HORVU3Hr1G043440	61.8	-1.37	HORVU2Hr1G031120	22.7	-2.92
HORVU3Hr1G056830	61.3	-2.48	HORVU6Hr1G028620	22.5	-1.63
HORVU6Hr1G002980	60.0	-1.41	HORVU7Hr1G006370	21.9	-1.34
HORVU6Hr1G029540	60.0	-1.47	HORVU1Hr1G034400	20.1	-1.62
HORVU2Hr1G107600	59.3	-2.48	HORVU2Hr1G091580	19.9	-1.73
HORVU5Hr1G111670	57.9	-3.83	HORVU5Hr1G019370	19.2	-1.95
HORVU5Hr1G092470	57.7	-2.61	HORVU2Hr1G031120	17.8	-1.84
HORVU1Hr1G092640	57.5	-2.04	HORVU5Hr1G028110	17.7	-2.34
HORVU3Hr1G083990	56.5	-1.68	HORVU4Hr1G022540	15.6	-1.51
HORVU7Hr1G090560	53.6	-1.83	HORVU5Hr1G108810	14.8	-2.31
HORVU4Hr1G050230	53.3	-1.55	HORVU2Hr1G073590	14.1	-1.35
HORVU5Hr1G095540	52.4	-2.04	HORVU7Hr1G122600	12.2	-2.18
HORVU2Hr1G001190	51.8	-1.35	HORVU6Hr1G039740	12.0	-1.63
HORVU3Hr1G107170	50.0	-1.94	HORVU3Hr1G113280	10.7	-1.65
HORVU4Hr1G025470	50.0	-1.74	HORVU5Hr1G039780	-10.6	-1.69
HORVU2Hr1G050270	48.4	-1.75	HORVU3Hr1G081720	-12.0	-2.05
HORVU7Hr1G000320	47.6	-1.34	HORVU5Hr1G086950	-12.3	-1.32
HORVU2Hr1G087820	47.3	-1.65	HORVU4Hr1G065620	-12.5	-1.58
HORVU7Hr1G077710	47.2	-1.40	HORVU3Hr1G117730	-13.3	-1.93
HORVU5Hr1G063070	46.7	-2.74	HORVU2Hr1G081360	-14.3	-1.50
HORVU1Hr1G082140	44.7	-2.25	HORVU2Hr1G012280	-14.6	-1.51
HORVU5Hr1G033480	43.7	-1.35	HORVU1Hr1G078450	-16.7	-1.44
HORVU6Hr1G087930	41.7	-1.44	HORVU3Hr1G077450	-16.7	-1.66
HORVU7Hr1G084240	41.6	-1.71	HORVU3Hr1G011330	-19.1	-1.35
HORVU7Hr1G002970	40.9	-1.33	HORVU3Hr1G039160	-19.6	-1.43
HORVU7Hr1G070080	39.6	-1.73	HORVU4Hr1G003660	-20.0	-2.39
HORVU6Hr1G031370	39.0	-1.34	HORVU7Hr1G117680	-22.9	-2.31
HORVU1Hr1G093250	37.7	-1.78	HORVU2Hr1G073590	-24.3	-1.31
HORVU2Hr1G067340	35.2	-1.78	HORVU1Hr1G038880	-25.0	-2.01
HORVU7Hr1G025850	35.1	-1.36	HORVU5Hr1G087670	-25.0	-1.69
HORVU3Hr1G042050	34.2	-1.40	HORVU5Hr1G009440	-28.6	-1.36
HORVU4Hr1G057520	33.3	-2.92	HORVU5Hr1G074660	-28.8	-1.72
HORVU4Hr1G057520	33.3	-1.33	HORVU4Hr1G075750	-31.6	-1.52
HORVU6Hr1G052490	33.0	-1.96	HORVU2Hr1G091720	-31.9	-1.96
HORVU6Hr1G022890	32.7	-1.97	HORVU6Hr1G074220	-33.2	-1.80
HORVU7Hr1G006250	32.4	-1.62	HORVU6Hr1G012570	-34.6	-1.35
HORVU3Hr1G027580	32.4	-2.85	HORVU4Hr1G022080	-35.3	-1.35
HORVU2Hr1G039100	32.3	-1.63	HORVU2Hr1G073590	-36.8	-3.71
HORVU6Hr1G085080	32.3	-1.51	HORVU4Hr1G067680	-40.6	-1.50
HORVU3Hr1G027580	31.8	-1.30	HORVU5Hr1G006960	-41.7	-1.30
HORVU6Hr1G030590	31.7	-1.63	HORVU7Hr1G062190	-41.7	-1.34
HORVU4Hr1G078830	31.5	-1.71	HORVU6Hr1G008200	-42.5	-1.71
HORVU5Hr1G089880	31.4	-1.79	HORVU6Hr1G071710	-43.5	-1.58
HORVU5Hr1G009150	31.4	-1.55	HORVU4Hr1G063360	-44.0	-1.36
HORVU5Hr1G121230	31.3	-1.39	HORVU0Hr1G027170	-49.3	-2.38
HORVU1Hr1G035960	31.0	-1.42	HORVU7Hr1G058250	-50.0	-1.32
HORVU7Hr1G035420	30.7	-2.10	HORVU6Hr1G011550	-50.8	-1.35
HORVU7Hr1G047800	30.7	-1.56	HORVU1Hr1G055600	-52.4	-1.34

Gene	Difference in % methylation	log ₁₀ (adj-p)
HORVU2Hr1G031110	-52.4	-1.53
HORVU3Hr1G031460	-52.5	-1.36
HORVU7Hr1G070490	-54.5	-1.99
HORVU2Hr1G114610	-55.2	-1.31
HORVU2Hr1G100680	-55.6	-1.74
HORVU3Hr1G073780	-57.1	-1.52
HORVU5Hr1G021830	-57.3	-1.46
HORVU6Hr1G036640	-59.6	-1.33
HORVU5Hr1G031470	-61.7	-1.82
HORVU3Hr1G065770	-66.7	-1.68
HORVU6Hr1G031080	-75.0	-2.10
HORVU2Hr1G089440	-83.8	-1.96
HORVU5Hr1G006960	-92.7	-13.78
HORVU1Hr1G081310	-100.0	-1.33
HORVU3Hr1G033980	-100.0	-4.19
HORVU3Hr1G033980	-100.0	-3.26
HORVU5Hr1G043010	-100.0	-2.68

Table A.9.15. Differentially methylated CHH context genes in barley root tissue under salt stress. Genes within 1.5 kbp of differentially methylated sites in CHH contexts in plant root tissue are listed. Difference in % methylation is the difference between methylation percentage in salt treated plants minus the methylation percentage in control plants. Some genes are listed more than once if more than one differentially methylated site fell in the region of the gene.

This table is too large to show here. Please follow the link below to view this table.
https://github.com/MagnificaScience/Salinity_stress_response_in_barley

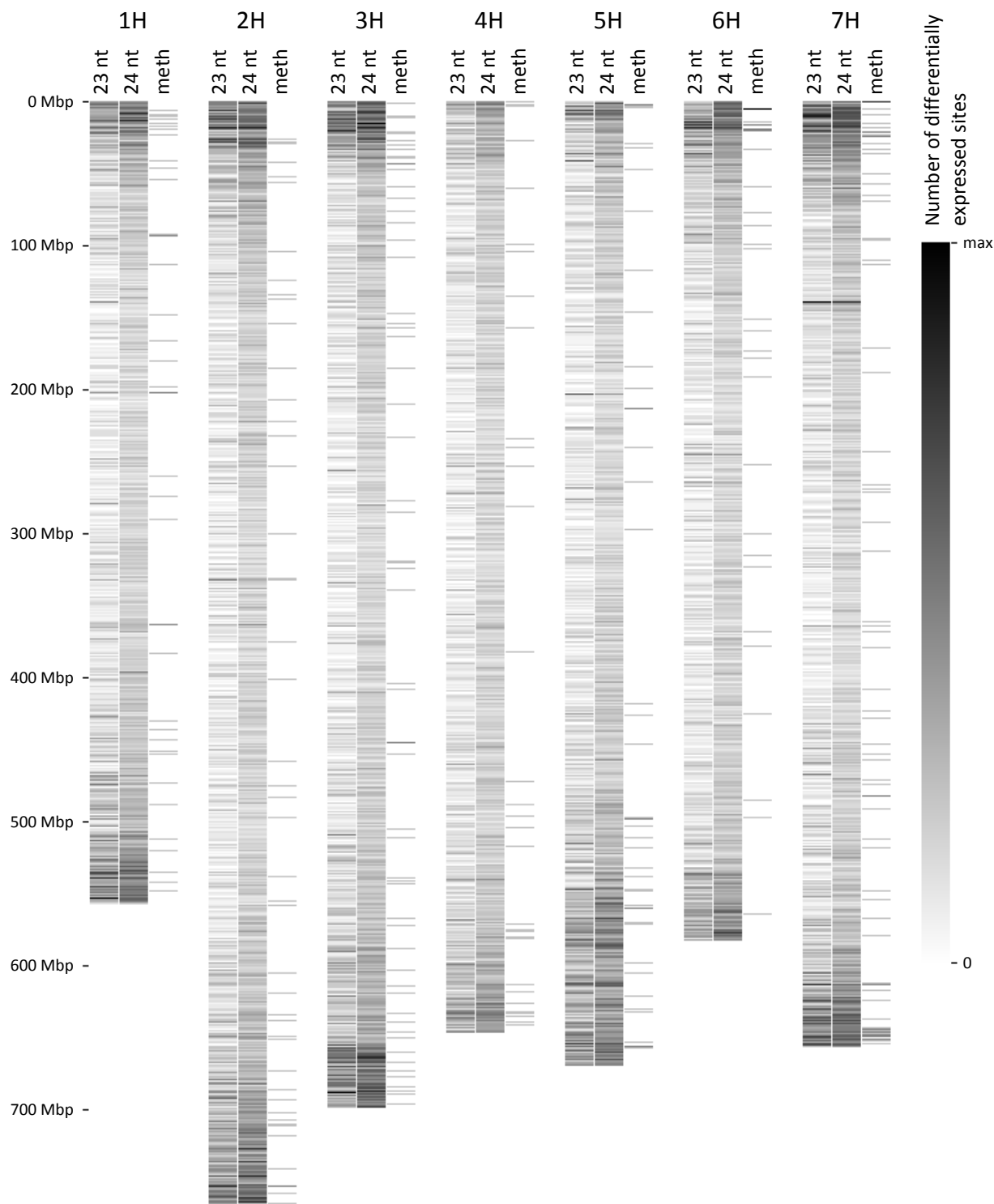


Figure A.9.9. Quantity of differentially expressed 23 nt miRNA, 24 nt siRNA, and CG context differential methylation mapped to 1 Mbp windows in barley under salt stress. Differential expression of 23 nt miRNA, 24 nt siRNA and CG context differential methylation was determined by mapping to 100 bp windows in the barley reference genome and finding differences between salt treated and control with adjusted p-values of less than 0.05. The number of differentially expressed 23/24 nt siRNA and CG context differentially methylated sites were counted for each 1 Mbp windows in the genome. Maximum count values for 23 nt miRNA is 24, for 24 nt siRNA is 90 and DNA methylation is 30 per 1 Mbp window.

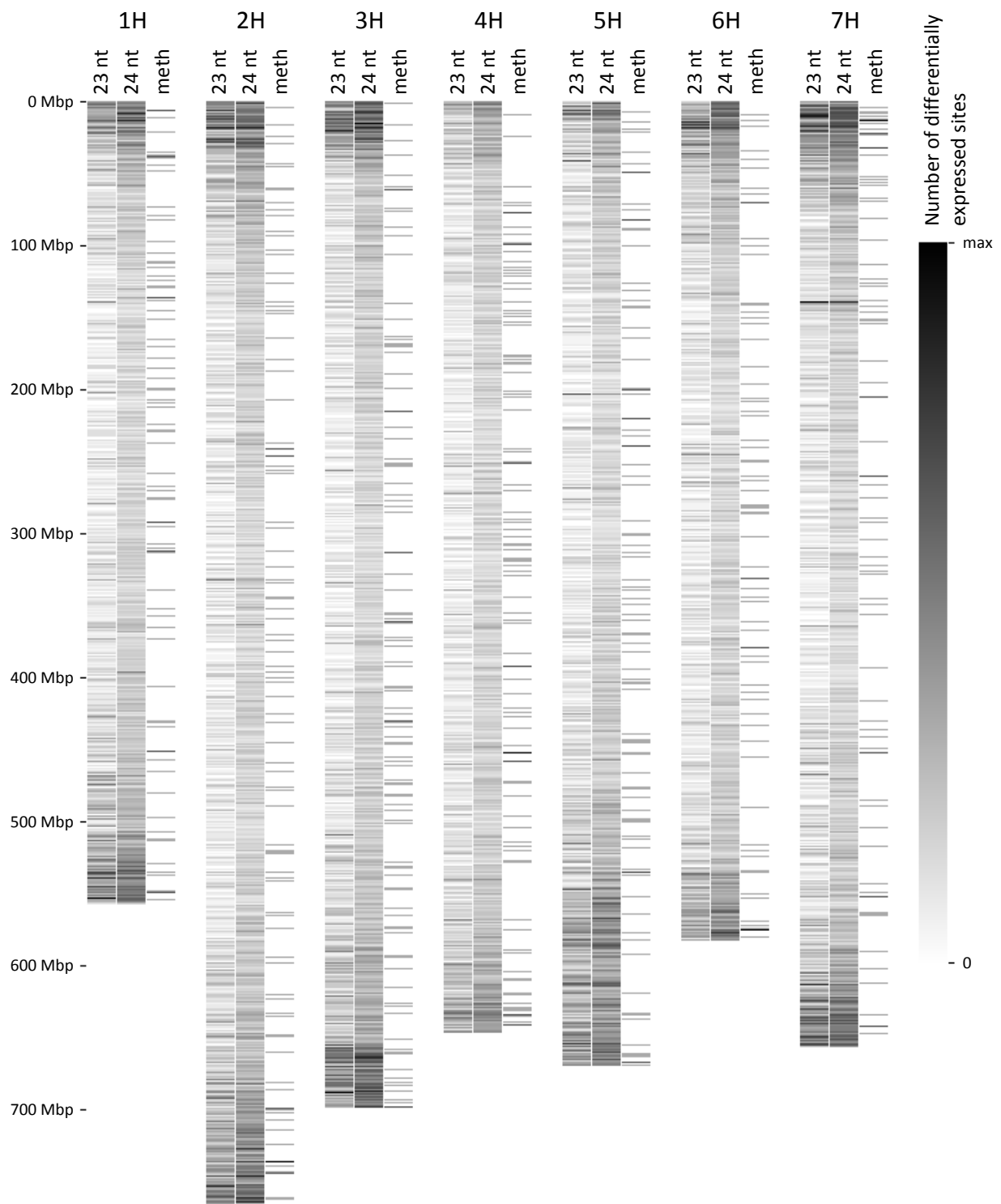


Figure A.9.10. Quantity of differentially expressed 23 nt miRNA, 24 nt siRNA, and CHG context differential methylation mapped to 1 Mbp windows in barley under salt stress. Differential expression of 23 nt miRNA, 24 nt siRNA and CHG context differential methylation was determined by mapping to 100 bp windows in the barley reference genome and finding differences between salt treated and control with adjusted p-values of less than 0.05. The number of differentially expressed 23/24 nt siRNA and CHG context differentially methylated sites were counted for each 1 Mbp windows in the genome. Maximum count values for 23 nt miRNA is 24, for 24 nt siRNA is 90 and DNA methylation is 30 per 1 Mbp window.

Table A.9.16 List of Protein Coding Genes (PCGs) targeted by differential 23/24 nt sRNA expression correlated with CHH context differential DNA methylation in leaf tissue in barley under salt stress. Correlations between differential sRNA expression and differential DNA methylation were within two 100 bp windows, and classified as targeting a PCG if the differential DNA methylation occurred between 1.5 kbp upstream of the transcription start site and 1.5 kbp downstream of the transcription end site.

HORVU0Hr1G000390	HORVU1Hr1G024860	HORVU1Hr1G066830	HORVU1Hr1G094880	HORVU2Hr1G032690
HORVU0Hr1G007860	HORVU1Hr1G025530	HORVU1Hr1G069130	HORVU2Hr1G000280	HORVU2Hr1G033370
HORVU0Hr1G008350	HORVU1Hr1G030390	HORVU1Hr1G069510	HORVU2Hr1G001160	HORVU2Hr1G033470
HORVU0Hr1G009460	HORVU1Hr1G031950	HORVU1Hr1G069650	HORVU2Hr1G001410	HORVU2Hr1G033820
HORVU0Hr1G012260	HORVU1Hr1G032470	HORVU1Hr1G070220	HORVU2Hr1G001690	HORVU2Hr1G034470
HORVU0Hr1G016350	HORVU1Hr1G032610	HORVU1Hr1G070390	HORVU2Hr1G002040	HORVU2Hr1G034740
HORVU0Hr1G017260	HORVU1Hr1G034400	HORVU1Hr1G071210	HORVU2Hr1G003150	HORVU2Hr1G035170
HORVU0Hr1G017690	HORVU1Hr1G034730	HORVU1Hr1G071380	HORVU2Hr1G003210	HORVU2Hr1G036750
HORVU0Hr1G018470	HORVU1Hr1G035200	HORVU1Hr1G071930	HORVU2Hr1G003360	HORVU2Hr1G039030
HORVU0Hr1G018510	HORVU1Hr1G035760	HORVU1Hr1G072140	HORVU2Hr1G003640	HORVU2Hr1G039040
HORVU0Hr1G020490	HORVU1Hr1G036500	HORVU1Hr1G072230	HORVU2Hr1G005150	HORVU2Hr1G039500
HORVU0Hr1G020750	HORVU1Hr1G037760	HORVU1Hr1G074030	HORVU2Hr1G005680	HORVU2Hr1G041180
HORVU0Hr1G021610	HORVU1Hr1G039400	HORVU1Hr1G074960	HORVU2Hr1G005960	HORVU2Hr1G041300
HORVU0Hr1G022520	HORVU1Hr1G039540	HORVU1Hr1G074970	HORVU2Hr1G006130	HORVU2Hr1G041530
HORVU0Hr1G022790	HORVU1Hr1G039670	HORVU1Hr1G075270	HORVU2Hr1G006250	HORVU2Hr1G042550
HORVU0Hr1G038730	HORVU1Hr1G039870	HORVU1Hr1G075580	HORVU2Hr1G006280	HORVU2Hr1G044280
HORVU0Hr1G040490	HORVU1Hr1G040450	HORVU1Hr1G076610	HORVU2Hr1G007170	HORVU2Hr1G044680
HORVU0Hr1G040540	HORVU1Hr1G041240	HORVU1Hr1G076690	HORVU2Hr1G007490	HORVU2Hr1G045020
HORVU1Hr1G000830	HORVU1Hr1G041790	HORVU1Hr1G077200	HORVU2Hr1G008890	HORVU2Hr1G045330
HORVU1Hr1G000990	HORVU1Hr1G043280	HORVU1Hr1G077910	HORVU2Hr1G009580	HORVU2Hr1G046160
HORVU1Hr1G001510	HORVU1Hr1G043620	HORVU1Hr1G078500	HORVU2Hr1G010690	HORVU2Hr1G046290
HORVU1Hr1G002970	HORVU1Hr1G043900	HORVU1Hr1G078650	HORVU2Hr1G010830	HORVU2Hr1G048870
HORVU1Hr1G003200	HORVU1Hr1G044440	HORVU1Hr1G078720	HORVU2Hr1G010960	HORVU2Hr1G050270
HORVU1Hr1G003340	HORVU1Hr1G046010	HORVU1Hr1G079410	HORVU2Hr1G011300	HORVU2Hr1G052220
HORVU1Hr1G004580	HORVU1Hr1G046610	HORVU1Hr1G080410	HORVU2Hr1G012280	HORVU2Hr1G052250
HORVU1Hr1G005120	HORVU1Hr1G047690	HORVU1Hr1G080950	HORVU2Hr1G012980	HORVU2Hr1G052430
HORVU1Hr1G005450	HORVU1Hr1G048010	HORVU1Hr1G081270	HORVU2Hr1G013080	HORVU2Hr1G052960
HORVU1Hr1G005460	HORVU1Hr1G051260	HORVU1Hr1G081890	HORVU2Hr1G013150	HORVU2Hr1G053480
HORVU1Hr1G005470	HORVU1Hr1G052010	HORVU1Hr1G081920	HORVU2Hr1G013700	HORVU2Hr1G054100
HORVU1Hr1G005520	HORVU1Hr1G052160	HORVU1Hr1G081990	HORVU2Hr1G014020	HORVU2Hr1G055230
HORVU1Hr1G006020	HORVU1Hr1G052280	HORVU1Hr1G082410	HORVU2Hr1G014150	HORVU2Hr1G055870
HORVU1Hr1G007290	HORVU1Hr1G053260	HORVU1Hr1G082820	HORVU2Hr1G016650	HORVU2Hr1G056440
HORVU1Hr1G007600	HORVU1Hr1G053280	HORVU1Hr1G083050	HORVU2Hr1G017380	HORVU2Hr1G056820
HORVU1Hr1G007830	HORVU1Hr1G053400	HORVU1Hr1G083160	HORVU2Hr1G017760	HORVU2Hr1G059170
HORVU1Hr1G008520	HORVU1Hr1G054300	HORVU1Hr1G084230	HORVU2Hr1G018430	HORVU2Hr1G059750
HORVU1Hr1G008690	HORVU1Hr1G055580	HORVU1Hr1G085140	HORVU2Hr1G019590	HORVU2Hr1G060550
HORVU1Hr1G009250	HORVU1Hr1G056000	HORVU1Hr1G085190	HORVU2Hr1G021280	HORVU2Hr1G063250
HORVU1Hr1G010130	HORVU1Hr1G056340	HORVU1Hr1G085240	HORVU2Hr1G022140	HORVU2Hr1G063840
HORVU1Hr1G013600	HORVU1Hr1G056560	HORVU1Hr1G085960	HORVU2Hr1G022180	HORVU2Hr1G064460
HORVU1Hr1G013720	HORVU1Hr1G058230	HORVU1Hr1G086200	HORVU2Hr1G023210	HORVU2Hr1G065000
HORVU1Hr1G014140	HORVU1Hr1G058810	HORVU1Hr1G088190	HORVU2Hr1G023660	HORVU2Hr1G065330
HORVU1Hr1G014570	HORVU1Hr1G058940	HORVU1Hr1G088510	HORVU2Hr1G024360	HORVU2Hr1G065430
HORVU1Hr1G014610	HORVU1Hr1G059010	HORVU1Hr1G088810	HORVU2Hr1G024680	HORVU2Hr1G066060
HORVU1Hr1G014950	HORVU1Hr1G059920	HORVU1Hr1G089380	HORVU2Hr1G024940	HORVU2Hr1G066100
HORVU1Hr1G017900	HORVU1Hr1G060090	HORVU1Hr1G090460	HORVU2Hr1G026550	HORVU2Hr1G067330
HORVU1Hr1G017920	HORVU1Hr1G060940	HORVU1Hr1G090860	HORVU2Hr1G026740	HORVU2Hr1G067790
HORVU1Hr1G018350	HORVU1Hr1G061090	HORVU1Hr1G091520	HORVU2Hr1G026750	HORVU2Hr1G068050
HORVU1Hr1G020240	HORVU1Hr1G061630	HORVU1Hr1G092060	HORVU2Hr1G027470	HORVU2Hr1G068120
HORVU1Hr1G020600	HORVU1Hr1G061830	HORVU1Hr1G092250	HORVU2Hr1G028210	HORVU2Hr1G070720
HORVU1Hr1G021100	HORVU1Hr1G062290	HORVU1Hr1G092340	HORVU2Hr1G028930	HORVU2Hr1G071860
HORVU1Hr1G021530	HORVU1Hr1G062470	HORVU1Hr1G092440	HORVU2Hr1G029150	HORVU2Hr1G073370
HORVU1Hr1G021990	HORVU1Hr1G064040	HORVU1Hr1G093250	HORVU2Hr1G029160	HORVU2Hr1G073740
HORVU1Hr1G022310	HORVU1Hr1G064150	HORVU1Hr1G093990	HORVU2Hr1G029290	HORVU2Hr1G074360
HORVU1Hr1G022470	HORVU1Hr1G064240	HORVU1Hr1G094320	HORVU2Hr1G029560	HORVU2Hr1G074530
HORVU1Hr1G023320	HORVU1Hr1G064400	HORVU1Hr1G094450	HORVU2Hr1G030140	HORVU2Hr1G075740
HORVU1Hr1G024190	HORVU1Hr1G066230	HORVU1Hr1G094610	HORVU2Hr1G030550	HORVU2Hr1G075970
HORVU1Hr1G024580	HORVU1Hr1G066760	HORVU1Hr1G094650	HORVU2Hr1G032280	HORVU2Hr1G076620

HORVU2Hr1G080480	HORVU2Hr1G115020	HORVU3Hr1G029430	HORVU3Hr1G085270	HORVU4Hr1G015060
HORVU2Hr1G080710	HORVU2Hr1G115200	HORVU3Hr1G030150	HORVU3Hr1G085760	HORVU4Hr1G015490
HORVU2Hr1G082170	HORVU2Hr1G115690	HORVU3Hr1G030270	HORVU3Hr1G085780	HORVU4Hr1G017200
HORVU2Hr1G082480	HORVU2Hr1G116680	HORVU3Hr1G031730	HORVU3Hr1G085900	HORVU4Hr1G017780
HORVU2Hr1G082840	HORVU2Hr1G118640	HORVU3Hr1G032090	HORVU3Hr1G086970	HORVU4Hr1G020000
HORVU2Hr1G085270	HORVU2Hr1G118680	HORVU3Hr1G033220	HORVU3Hr1G087430	HORVU4Hr1G021040
HORVU2Hr1G085850	HORVU2Hr1G118740	HORVU3Hr1G034020	HORVU3Hr1G087710	HORVU4Hr1G023100
HORVU2Hr1G085910	HORVU2Hr1G119180	HORVU3Hr1G034310	HORVU3Hr1G087800	HORVU4Hr1G025670
HORVU2Hr1G087030	HORVU2Hr1G119210	HORVU3Hr1G034950	HORVU3Hr1G088210	HORVU4Hr1G027080
HORVU2Hr1G087970	HORVU2Hr1G119430	HORVU3Hr1G035450	HORVU3Hr1G088310	HORVU4Hr1G031290
HORVU2Hr1G088390	HORVU2Hr1G119480	HORVU3Hr1G035810	HORVU3Hr1G088360	HORVU4Hr1G032120
HORVU2Hr1G088440	HORVU2Hr1G119520	HORVU3Hr1G035840	HORVU3Hr1G088910	HORVU4Hr1G033610
HORVU2Hr1G088840	HORVU2Hr1G120440	HORVU3Hr1G036180	HORVU3Hr1G090310	HORVU4Hr1G033650
HORVU2Hr1G089440	HORVU2Hr1G120780	HORVU3Hr1G036860	HORVU3Hr1G091360	HORVU4Hr1G033850
HORVU2Hr1G090280	HORVU2Hr1G121030	HORVU3Hr1G040230	HORVU3Hr1G091850	HORVU4Hr1G035180
HORVU2Hr1G092070	HORVU2Hr1G123080	HORVU3Hr1G041640	HORVU3Hr1G093140	HORVU4Hr1G036620
HORVU2Hr1G093610	HORVU2Hr1G123150	HORVU3Hr1G042500	HORVU3Hr1G094160	HORVU4Hr1G037750
HORVU2Hr1G093930	HORVU2Hr1G124030	HORVU3Hr1G045480	HORVU3Hr1G094210	HORVU4Hr1G038750
HORVU2Hr1G094620	HORVU2Hr1G124210	HORVU3Hr1G045900	HORVU3Hr1G095090	HORVU4Hr1G039640
HORVU2Hr1G096160	HORVU2Hr1G124720	HORVU3Hr1G049910	HORVU3Hr1G095580	HORVU4Hr1G040140
HORVU2Hr1G096350	HORVU2Hr1G124870	HORVU3Hr1G050430	HORVU3Hr1G096040	HORVU4Hr1G041810
HORVU2Hr1G096470	HORVU2Hr1G125910	HORVU3Hr1G050990	HORVU3Hr1G096600	HORVU4Hr1G045980
HORVU2Hr1G096550	HORVU2Hr1G126950	HORVU3Hr1G051540	HORVU3Hr1G097390	HORVU4Hr1G046590
HORVU2Hr1G096920	HORVU3Hr1G000030	HORVU3Hr1G053760	HORVU3Hr1G097570	HORVU4Hr1G046610
HORVU2Hr1G096960	HORVU3Hr1G000600	HORVU3Hr1G054120	HORVU3Hr1G098150	HORVU4Hr1G048620
HORVU2Hr1G097110	HORVU3Hr1G000880	HORVU3Hr1G054680	HORVU3Hr1G098940	HORVU4Hr1G050230
HORVU2Hr1G097360	HORVU3Hr1G001280	HORVU3Hr1G056270	HORVU3Hr1G098950	HORVU4Hr1G050680
HORVU2Hr1G097530	HORVU3Hr1G001400	HORVU3Hr1G056830	HORVU3Hr1G099120	HORVU4Hr1G051230
HORVU2Hr1G097780	HORVU3Hr1G001430	HORVU3Hr1G059230	HORVU3Hr1G099220	HORVU4Hr1G051590
HORVU2Hr1G098160	HORVU3Hr1G001910	HORVU3Hr1G059480	HORVU3Hr1G099740	HORVU4Hr1G052340
HORVU2Hr1G098550	HORVU3Hr1G002780	HORVU3Hr1G061210	HORVU3Hr1G100420	HORVU4Hr1G052430
HORVU2Hr1G099370	HORVU3Hr1G002840	HORVU3Hr1G061930	HORVU3Hr1G101540	HORVU4Hr1G052610
HORVU2Hr1G099550	HORVU3Hr1G002920	HORVU3Hr1G061980	HORVU3Hr1G104320	HORVU4Hr1G052890
HORVU2Hr1G099740	HORVU3Hr1G003140	HORVU3Hr1G062130	HORVU3Hr1G105880	HORVU4Hr1G052990
HORVU2Hr1G099820	HORVU3Hr1G003470	HORVU3Hr1G063470	HORVU3Hr1G107970	HORVU4Hr1G053150
HORVU2Hr1G100450	HORVU3Hr1G003960	HORVU3Hr1G064130	HORVU3Hr1G108370	HORVU4Hr1G054820
HORVU2Hr1G100600	HORVU3Hr1G004520	HORVU3Hr1G065000	HORVU3Hr1G108610	HORVU4Hr1G055070
HORVU2Hr1G100690	HORVU3Hr1G005430	HORVU3Hr1G065520	HORVU3Hr1G109380	HORVU4Hr1G055690
HORVU2Hr1G101600	HORVU3Hr1G005580	HORVU3Hr1G065770	HORVU3Hr1G109600	HORVU4Hr1G056320
HORVU2Hr1G102300	HORVU3Hr1G007680	HORVU3Hr1G066930	HORVU3Hr1G110580	HORVU4Hr1G056610
HORVU2Hr1G103000	HORVU3Hr1G009110	HORVU3Hr1G067020	HORVU3Hr1G110890	HORVU4Hr1G057510
HORVU2Hr1G103800	HORVU3Hr1G009360	HORVU3Hr1G068640	HORVU3Hr1G112450	HORVU4Hr1G057690
HORVU2Hr1G103930	HORVU3Hr1G009980	HORVU3Hr1G068660	HORVU3Hr1G112690	HORVU4Hr1G058350
HORVU2Hr1G104530	HORVU3Hr1G010030	HORVU3Hr1G070090	HORVU3Hr1G113670	HORVU4Hr1G058480
HORVU2Hr1G105700	HORVU3Hr1G010480	HORVU3Hr1G070150	HORVU3Hr1G113740	HORVU4Hr1G061440
HORVU2Hr1G105830	HORVU3Hr1G012260	HORVU3Hr1G070510	HORVU3Hr1G113790	HORVU4Hr1G062330
HORVU2Hr1G105960	HORVU3Hr1G012520	HORVU3Hr1G071320	HORVU3Hr1G113850	HORVU4Hr1G063350
HORVU2Hr1G106020	HORVU3Hr1G012650	HORVU3Hr1G071910	HORVU3Hr1G115610	HORVU4Hr1G063360
HORVU2Hr1G106480	HORVU3Hr1G012860	HORVU3Hr1G072070	HORVU3Hr1G115650	HORVU4Hr1G063980
HORVU2Hr1G106820	HORVU3Hr1G013110	HORVU3Hr1G072100	HORVU3Hr1G116580	HORVU4Hr1G064920
HORVU2Hr1G108210	HORVU3Hr1G013470	HORVU3Hr1G072340	HORVU4Hr1G000100	HORVU4Hr1G065310
HORVU2Hr1G108250	HORVU3Hr1G014580	HORVU3Hr1G073100	HORVU4Hr1G001450	HORVU4Hr1G066130
HORVU2Hr1G108370	HORVU3Hr1G016430	HORVU3Hr1G074160	HORVU4Hr1G002090	HORVU4Hr1G066150
HORVU2Hr1G108460	HORVU3Hr1G017020	HORVU3Hr1G074210	HORVU4Hr1G002360	HORVU4Hr1G066430
HORVU2Hr1G108520	HORVU3Hr1G017530	HORVU3Hr1G074920	HORVU4Hr1G003250	HORVU4Hr1G067140
HORVU2Hr1G109240	HORVU3Hr1G017930	HORVU3Hr1G075710	HORVU4Hr1G004150	HORVU4Hr1G067680
HORVU2Hr1G109590	HORVU3Hr1G019790	HORVU3Hr1G076220	HORVU4Hr1G004410	HORVU4Hr1G069100
HORVU2Hr1G109980	HORVU3Hr1G023230	HORVU3Hr1G078330	HORVU4Hr1G005450	HORVU4Hr1G069350
HORVU2Hr1G110080	HORVU3Hr1G025590	HORVU3Hr1G078470	HORVU4Hr1G010200	HORVU4Hr1G069700
HORVU2Hr1G110130	HORVU3Hr1G025820	HORVU3Hr1G078530	HORVU4Hr1G010410	HORVU4Hr1G070070
HORVU2Hr1G110800	HORVU3Hr1G026150	HORVU3Hr1G079050	HORVU4Hr1G011200	HORVU4Hr1G071040
HORVU2Hr1G110940	HORVU3Hr1G027220	HORVU3Hr1G080650	HORVU4Hr1G011500	HORVU4Hr1G071070
HORVU2Hr1G112790	HORVU3Hr1G027530	HORVU3Hr1G081170	HORVU4Hr1G012330	HORVU4Hr1G071690
HORVU2Hr1G113120	HORVU3Hr1G028550	HORVU3Hr1G081220	HORVU4Hr1G012780	HORVU4Hr1G072190
HORVU2Hr1G113380	HORVU3Hr1G028990	HORVU3Hr1G081290	HORVU4Hr1G013720	HORVU4Hr1G072660
HORVU2Hr1G114210	HORVU3Hr1G029010	HORVU3Hr1G082070	HORVU4Hr1G014300	HORVU4Hr1G072670
HORVU2Hr1G115010	HORVU3Hr1G029020	HORVU3Hr1G085040	HORVU4Hr1G014720	HORVU4Hr1G072700

HORVU4Hr1G072840	HORVU5Hr1G036160	HORVU5Hr1G085220	HORVU5Hr1G123520	HORVU6Hr1G057630
HORVU4Hr1G076010	HORVU5Hr1G036490	HORVU5Hr1G085550	HORVU5Hr1G123860	HORVU6Hr1G057700
HORVU4Hr1G076050	HORVU5Hr1G037920	HORVU5Hr1G085900	HORVU5Hr1G124210	HORVU6Hr1G057990
HORVU4Hr1G076260	HORVU5Hr1G038090	HORVU5Hr1G086120	HORVU5Hr1G124750	HORVU6Hr1G059800
HORVU4Hr1G076630	HORVU5Hr1G039730	HORVU5Hr1G086500	HORVU5Hr1G125320	HORVU6Hr1G059920
HORVU4Hr1G077300	HORVU5Hr1G039740	HORVU5Hr1G086630	HORVU5Hr1G125380	HORVU6Hr1G060720
HORVU4Hr1G077680	HORVU5Hr1G042400	HORVU5Hr1G087780	HORVU6Hr1G000060	HORVU6Hr1G061660
HORVU4Hr1G077720	HORVU5Hr1G044350	HORVU5Hr1G088260	HORVU6Hr1G000620	HORVU6Hr1G062320
HORVU4Hr1G078130	HORVU5Hr1G045630	HORVU5Hr1G088510	HORVU6Hr1G000680	HORVU6Hr1G063590
HORVU4Hr1G079300	HORVU5Hr1G046830	HORVU5Hr1G092000	HORVU6Hr1G000720	HORVU6Hr1G065120
HORVU4Hr1G079920	HORVU5Hr1G047250	HORVU5Hr1G092200	HORVU6Hr1G001200	HORVU6Hr1G065800
HORVU4Hr1G081130	HORVU5Hr1G048030	HORVU5Hr1G092490	HORVU6Hr1G001580	HORVU6Hr1G067660
HORVU4Hr1G081250	HORVU5Hr1G049050	HORVU5Hr1G093040	HORVU6Hr1G002980	HORVU6Hr1G068950
HORVU4Hr1G081290	HORVU5Hr1G049220	HORVU5Hr1G093410	HORVU6Hr1G003990	HORVU6Hr1G069260
HORVU4Hr1G081800	HORVU5Hr1G049440	HORVU5Hr1G093700	HORVU6Hr1G004320	HORVU6Hr1G069310
HORVU4Hr1G082040	HORVU5Hr1G050480	HORVU5Hr1G094080	HORVU6Hr1G005480	HORVU6Hr1G070010
HORVU4Hr1G082570	HORVU5Hr1G050720	HORVU5Hr1G094840	HORVU6Hr1G006480	HORVU6Hr1G070020
HORVU4Hr1G082630	HORVU5Hr1G051140	HORVU5Hr1G095990	HORVU6Hr1G006620	HORVU6Hr1G070230
HORVU4Hr1G083650	HORVU5Hr1G051370	HORVU5Hr1G096010	HORVU6Hr1G008670	HORVU6Hr1G070300
HORVU4Hr1G083680	HORVU5Hr1G051500	HORVU5Hr1G096760	HORVU6Hr1G009270	HORVU6Hr1G070330
HORVU4Hr1G083840	HORVU5Hr1G051530	HORVU5Hr1G097130	HORVU6Hr1G010650	HORVU6Hr1G070540
HORVU4Hr1G084140	HORVU5Hr1G052340	HORVU5Hr1G097290	HORVU6Hr1G010980	HORVU6Hr1G073180
HORVU4Hr1G084590	HORVU5Hr1G056350	HORVU5Hr1G098000	HORVU6Hr1G011120	HORVU6Hr1G074660
HORVU4Hr1G084730	HORVU5Hr1G056760	HORVU5Hr1G098340	HORVU6Hr1G011580	HORVU6Hr1G076020
HORVU4Hr1G084760	HORVU5Hr1G057140	HORVU5Hr1G098780	HORVU6Hr1G012290	HORVU6Hr1G076080
HORVU4Hr1G085050	HORVU5Hr1G059930	HORVU5Hr1G100300	HORVU6Hr1G012630	HORVU6Hr1G076600
HORVU4Hr1G085480	HORVU5Hr1G060090	HORVU5Hr1G101480	HORVU6Hr1G013530	HORVU6Hr1G077260
HORVU4Hr1G087500	HORVU5Hr1G060980	HORVU5Hr1G101820	HORVU6Hr1G013830	HORVU6Hr1G077320
HORVU4Hr1G087670	HORVU5Hr1G061150	HORVU5Hr1G102240	HORVU6Hr1G014600	HORVU6Hr1G077610
HORVU4Hr1G087780	HORVU5Hr1G061480	HORVU5Hr1G102240	HORVU6Hr1G014600	HORVU6Hr1G077790
HORVU4Hr1G088340	HORVU5Hr1G061530	HORVU5Hr1G103060	HORVU6Hr1G016530	HORVU6Hr1G077790
HORVU4Hr1G088350	HORVU5Hr1G062820	HORVU5Hr1G103280	HORVU6Hr1G016570	HORVU6Hr1G078650
HORVU4Hr1G088350	HORVU5Hr1G062820	HORVU5Hr1G103450	HORVU6Hr1G016810	HORVU6Hr1G079660
HORVU4Hr1G088570	HORVU5Hr1G062940	HORVU5Hr1G104380	HORVU6Hr1G017720	HORVU6Hr1G080270
HORVU4Hr1G089290	HORVU5Hr1G063390	HORVU5Hr1G104490	HORVU6Hr1G018350	HORVU6Hr1G081570
HORVU4Hr1G090490	HORVU5Hr1G065620	HORVU5Hr1G104690	HORVU6Hr1G018690	HORVU6Hr1G081790
HORVU5Hr1G000640	HORVU5Hr1G066020	HORVU5Hr1G104880	HORVU6Hr1G018790	HORVU6Hr1G081850
HORVU5Hr1G002970	HORVU5Hr1G067200	HORVU5Hr1G105000	HORVU6Hr1G018820	HORVU6Hr1G082310
HORVU5Hr1G003510	HORVU5Hr1G067530	HORVU5Hr1G105660	HORVU6Hr1G019450	HORVU6Hr1G082900
HORVU5Hr1G006950	HORVU5Hr1G068070	HORVU5Hr1G106030	HORVU6Hr1G020190	HORVU6Hr1G087120
HORVU5Hr1G007750	HORVU5Hr1G068590	HORVU5Hr1G106160	HORVU6Hr1G020310	HORVU6Hr1G087300
HORVU5Hr1G009230	HORVU5Hr1G068600	HORVU5Hr1G106200	HORVU6Hr1G025350	HORVU6Hr1G087800
HORVU5Hr1G009390	HORVU5Hr1G069350	HORVU5Hr1G107380	HORVU6Hr1G025500	HORVU6Hr1G088070
HORVU5Hr1G010440	HORVU5Hr1G069480	HORVU5Hr1G109380	HORVU6Hr1G029540	HORVU6Hr1G088420
HORVU5Hr1G010690	HORVU5Hr1G069510	HORVU5Hr1G110440	HORVU6Hr1G030350	HORVU6Hr1G088480
HORVU5Hr1G011230	HORVU5Hr1G069710	HORVU5Hr1G111000	HORVU6Hr1G031080	HORVU6Hr1G088580
HORVU5Hr1G011860	HORVU5Hr1G069960	HORVU5Hr1G111570	HORVU6Hr1G031190	HORVU6Hr1G088630
HORVU5Hr1G012160	HORVU5Hr1G069980	HORVU5Hr1G111620	HORVU6Hr1G031230	HORVU6Hr1G088680
HORVU5Hr1G012300	HORVU5Hr1G072170	HORVU5Hr1G113540	HORVU6Hr1G031470	HORVU6Hr1G088900
HORVU5Hr1G013160	HORVU5Hr1G072760	HORVU5Hr1G114260	HORVU6Hr1G031600	HORVU6Hr1G089670
HORVU5Hr1G013170	HORVU5Hr1G075400	HORVU5Hr1G115440	HORVU6Hr1G032050	HORVU6Hr1G090460
HORVU5Hr1G014920	HORVU5Hr1G075540	HORVU5Hr1G115470	HORVU6Hr1G032310	HORVU6Hr1G090990
HORVU5Hr1G016400	HORVU5Hr1G075670	HORVU5Hr1G116860	HORVU6Hr1G032610	HORVU6Hr1G091460
HORVU5Hr1G018010	HORVU5Hr1G076330	HORVU5Hr1G116920	HORVU6Hr1G032800	HORVU6Hr1G092010
HORVU5Hr1G019220	HORVU5Hr1G078000	HORVU5Hr1G117770	HORVU6Hr1G035210	HORVU6Hr1G092540
HORVU5Hr1G019310	HORVU5Hr1G078450	HORVU5Hr1G117790	HORVU6Hr1G036810	HORVU6Hr1G092910
HORVU5Hr1G020200	HORVU5Hr1G078850	HORVU5Hr1G117920	HORVU6Hr1G036840	HORVU6Hr1G094170
HORVU5Hr1G020990	HORVU5Hr1G079630	HORVU5Hr1G118010	HORVU6Hr1G037510	HORVU6Hr1G094480
HORVU5Hr1G022590	HORVU5Hr1G082210	HORVU5Hr1G118340	HORVU6Hr1G039140	HORVU6Hr1G094950
HORVU5Hr1G025710	HORVU5Hr1G082390	HORVU5Hr1G118410	HORVU6Hr1G044030	HORVU6Hr1G095190
HORVU5Hr1G027290	HORVU5Hr1G082480	HORVU5Hr1G119100	HORVU6Hr1G044080	HORVU7Hr1G000040
HORVU5Hr1G028000	HORVU5Hr1G082650	HORVU5Hr1G119850	HORVU6Hr1G044840	HORVU7Hr1G000320
HORVU5Hr1G028140	HORVU5Hr1G083330	HORVU5Hr1G121440	HORVU6Hr1G048280	HORVU7Hr1G001020
HORVU5Hr1G028870	HORVU5Hr1G083620	HORVU5Hr1G122220	HORVU6Hr1G050090	HORVU7Hr1G001160
HORVU5Hr1G030970	HORVU5Hr1G083890	HORVU5Hr1G122660	HORVU6Hr1G052250	HORVU7Hr1G001340
HORVU5Hr1G031720	HORVU5Hr1G084820	HORVU5Hr1G122840	HORVU6Hr1G055910	HORVU7Hr1G001570
HORVU5Hr1G033960	HORVU5Hr1G084980	HORVU5Hr1G122860	HORVU6Hr1G056230	HORVU7Hr1G002370
HORVU5Hr1G035300	HORVU5Hr1G085020	HORVU5Hr1G122960	HORVU6Hr1G056460	HORVU7Hr1G002600

HORVU7Hr1G002800	HORVU7Hr1G024990	HORVU7Hr1G049070	HORVU7Hr1G082590	HORVU7Hr1G111000
HORVU7Hr1G002880	HORVU7Hr1G025670	HORVU7Hr1G049190	HORVU7Hr1G082720	HORVU7Hr1G111120
HORVU7Hr1G005300	HORVU7Hr1G025800	HORVU7Hr1G049200	HORVU7Hr1G083440	HORVU7Hr1G111630
HORVU7Hr1G006840	HORVU7Hr1G026540	HORVU7Hr1G049290	HORVU7Hr1G083610	HORVU7Hr1G112470
HORVU7Hr1G006990	HORVU7Hr1G027600	HORVU7Hr1G050270	HORVU7Hr1G085060	HORVU7Hr1G112740
HORVU7Hr1G007590	HORVU7Hr1G028050	HORVU7Hr1G051860	HORVU7Hr1G085220	HORVU7Hr1G112980
HORVU7Hr1G007610	HORVU7Hr1G028730	HORVU7Hr1G052210	HORVU7Hr1G085240	HORVU7Hr1G113020
HORVU7Hr1G007900	HORVU7Hr1G029160	HORVU7Hr1G053780	HORVU7Hr1G086180	HORVU7Hr1G114130
HORVU7Hr1G009190	HORVU7Hr1G029420	HORVU7Hr1G053930	HORVU7Hr1G088300	HORVU7Hr1G114220
HORVU7Hr1G009420	HORVU7Hr1G030420	HORVU7Hr1G055880	HORVU7Hr1G089930	HORVU7Hr1G115120
HORVU7Hr1G009630	HORVU7Hr1G030860	HORVU7Hr1G056530	HORVU7Hr1G090350	HORVU7Hr1G115770
HORVU7Hr1G010430	HORVU7Hr1G031580	HORVU7Hr1G057640	HORVU7Hr1G090390	HORVU7Hr1G116310
HORVU7Hr1G010680	HORVU7Hr1G032100	HORVU7Hr1G057870	HORVU7Hr1G092360	HORVU7Hr1G116340
HORVU7Hr1G010690	HORVU7Hr1G033230	HORVU7Hr1G059870	HORVU7Hr1G093200	HORVU7Hr1G116940
HORVU7Hr1G010850	HORVU7Hr1G035430	HORVU7Hr1G062510	HORVU7Hr1G093380	HORVU7Hr1G117320
HORVU7Hr1G012820	HORVU7Hr1G035900	HORVU7Hr1G063030	HORVU7Hr1G093620	HORVU7Hr1G117450
HORVU7Hr1G012880	HORVU7Hr1G035970	HORVU7Hr1G065590	HORVU7Hr1G093790	HORVU7Hr1G117640
HORVU7Hr1G013890	HORVU7Hr1G036450	HORVU7Hr1G065740	HORVU7Hr1G094850	HORVU7Hr1G118470
HORVU7Hr1G017060	HORVU7Hr1G036760	HORVU7Hr1G065950	HORVU7Hr1G095410	HORVU7Hr1G118500
HORVU7Hr1G018000	HORVU7Hr1G037080	HORVU7Hr1G066510	HORVU7Hr1G096020	HORVU7Hr1G118740
HORVU7Hr1G018420	HORVU7Hr1G037140	HORVU7Hr1G067060	HORVU7Hr1G096460	HORVU7Hr1G119270
HORVU7Hr1G019670	HORVU7Hr1G037580	HORVU7Hr1G069520	HORVU7Hr1G096550	HORVU7Hr1G119680
HORVU7Hr1G019730	HORVU7Hr1G038040	HORVU7Hr1G069750	HORVU7Hr1G096930	HORVU7Hr1G120030
HORVU7Hr1G021840	HORVU7Hr1G038370	HORVU7Hr1G069840	HORVU7Hr1G097020	HORVU7Hr1G120960
HORVU7Hr1G022230	HORVU7Hr1G038420	HORVU7Hr1G070700	HORVU7Hr1G097780	HORVU7Hr1G121040
HORVU7Hr1G022250	HORVU7Hr1G038440	HORVU7Hr1G071060	HORVU7Hr1G098390	HORVU7Hr1G121170
HORVU7Hr1G022310	HORVU7Hr1G038710	HORVU7Hr1G072200	HORVU7Hr1G098550	HORVU7Hr1G121250
HORVU7Hr1G022530	HORVU7Hr1G039220	HORVU7Hr1G073020	HORVU7Hr1G099520	HORVU7Hr1G121600
HORVU7Hr1G023000	HORVU7Hr1G039550	HORVU7Hr1G074650	HORVU7Hr1G101740	HORVU7Hr1G121810
HORVU7Hr1G023760	HORVU7Hr1G040030	HORVU7Hr1G075120	HORVU7Hr1G105250	
HORVU7Hr1G024370	HORVU7Hr1G040550	HORVU7Hr1G079210	HORVU7Hr1G106480	
HORVU7Hr1G024550	HORVU7Hr1G045150	HORVU7Hr1G082120	HORVU7Hr1G108210	
HORVU7Hr1G024960	HORVU7Hr1G048730	HORVU7Hr1G082420	HORVU7Hr1G109770	

Table A.9.17 List of Protein Coding Genes (PCGs) targeted by differential 23/24 nt sRNA expression correlated with CHH context differential DNA methylation in root tissue in barley under salt stress. Correlations between differential sRNA expression and differential DNA methylation were within two 100 bp windows, and classified as targeting a PCG if the differential DNA methylation occurred between 1.5 kbp upstream of the transcription start site and 1.5 kbp downstream of the transcription end site.

HORVU0Hr1G000390	HORVU1Hr1G078650	HORVU2Hr1G059320	HORVU3Hr1G023230	HORVU3Hr1G113720
HORVU0Hr1G001570	HORVU1Hr1G079410	HORVU2Hr1G071570	HORVU3Hr1G024170	HORVU3Hr1G116930
HORVU0Hr1G010040	HORVU1Hr1G080640	HORVU2Hr1G073920	HORVU3Hr1G024200	HORVU4Hr1G001400
HORVU0Hr1G010760	HORVU1Hr1G081050	HORVU2Hr1G074360	HORVU3Hr1G027530	HORVU4Hr1G004160
HORVU0Hr1G012050	HORVU1Hr1G081990	HORVU2Hr1G074670	HORVU3Hr1G031000	HORVU4Hr1G005710
HORVU0Hr1G012230	HORVU1Hr1G082250	HORVU2Hr1G075690	HORVU3Hr1G032820	HORVU4Hr1G006870
HORVU0Hr1G016130	HORVU1Hr1G082670	HORVU2Hr1G0776910	HORVU3Hr1G033440	HORVU4Hr1G006930
HORVU0Hr1G017080	HORVU1Hr1G083160	HORVU2Hr1G080630	HORVU3Hr1G033660	HORVU4Hr1G007450
HORVU0Hr1G024100	HORVU1Hr1G085130	HORVU2Hr1G082090	HORVU3Hr1G034020	HORVU4Hr1G007990
HORVU0Hr1G040340	HORVU1Hr1G085680	HORVU2Hr1G084170	HORVU3Hr1G035450	HORVU4Hr1G008850
HORVU0Hr1G040490	HORVU1Hr1G085960	HORVU2Hr1G085560	HORVU3Hr1G035810	HORVU4Hr1G011040
HORVU1Hr1G000350	HORVU1Hr1G086330	HORVU2Hr1G087030	HORVU3Hr1G038950	HORVU4Hr1G012660
HORVU1Hr1G001400	HORVU1Hr1G087010	HORVU2Hr1G087380	HORVU3Hr1G041250	HORVU4Hr1G014120
HORVU1Hr1G001510	HORVU1Hr1G087900	HORVU2Hr1G089440	HORVU3Hr1G051050	HORVU4Hr1G014760
HORVU1Hr1G002970	HORVU1Hr1G089540	HORVU2Hr1G090280	HORVU3Hr1G055670	HORVU4Hr1G015890
HORVU1Hr1G003200	HORVU1Hr1G089620	HORVU2Hr1G091270	HORVU3Hr1G057530	HORVU4Hr1G017680
HORVU1Hr1G003340	HORVU1Hr1G089840	HORVU2Hr1G091530	HORVU3Hr1G061210	HORVU4Hr1G022680
HORVU1Hr1G003530	HORVU1Hr1G090850	HORVU2Hr1G091720	HORVU3Hr1G062730	HORVU4Hr1G024460
HORVU1Hr1G004580	HORVU1Hr1G091110	HORVU2Hr1G091800	HORVU3Hr1G063470	HORVU4Hr1G024590
HORVU1Hr1G005140	HORVU1Hr1G094320	HORVU2Hr1G093610	HORVU3Hr1G067020	HORVU4Hr1G026770
HORVU1Hr1G006020	HORVU2Hr1G000090	HORVU2Hr1G095050	HORVU3Hr1G068170	HORVU4Hr1G030380
HORVU1Hr1G010970	HORVU2Hr1G000280	HORVU2Hr1G095990	HORVU3Hr1G068830	HORVU4Hr1G032550
HORVU1Hr1G011400	HORVU2Hr1G000590	HORVU2Hr1G096550	HORVU3Hr1G069320	HORVU4Hr1G037750
HORVU1Hr1G016000	HORVU2Hr1G003720	HORVU2Hr1G097580	HORVU3Hr1G069560	HORVU4Hr1G038670
HORVU1Hr1G021800	HORVU2Hr1G005680	HORVU2Hr1G101690	HORVU3Hr1G069890	HORVU4Hr1G043620
HORVU1Hr1G024390	HORVU2Hr1G006180	HORVU2Hr1G101980	HORVU3Hr1G070220	HORVU4Hr1G045980
HORVU1Hr1G025530	HORVU2Hr1G006320	HORVU2Hr1G104170	HORVU3Hr1G071570	HORVU4Hr1G048400
HORVU1Hr1G028720	HORVU2Hr1G007980	HORVU2Hr1G107980	HORVU3Hr1G074210	HORVU4Hr1G049950
HORVU1Hr1G030200	HORVU2Hr1G010520	HORVU2Hr1G108460	HORVU3Hr1G075710	HORVU4Hr1G050900
HORVU1Hr1G030390	HORVU2Hr1G010830	HORVU2Hr1G109080	HORVU3Hr1G077210	HORVU4Hr1G051010
HORVU1Hr1G036720	HORVU2Hr1G011580	HORVU2Hr1G109980	HORVU3Hr1G078470	HORVU4Hr1G051710
HORVU1Hr1G037760	HORVU2Hr1G012010	HORVU2Hr1G110800	HORVU3Hr1G080040	HORVU4Hr1G052880
HORVU1Hr1G040050	HORVU2Hr1G012980	HORVU2Hr1G110900	HORVU3Hr1G081320	HORVU4Hr1G054350
HORVU1Hr1G040660	HORVU2Hr1G013370	HORVU2Hr1G112170	HORVU3Hr1G081710	HORVU4Hr1G058180
HORVU1Hr1G044500	HORVU2Hr1G014020	HORVU2Hr1G112590	HORVU3Hr1G083180	HORVU4Hr1G058210
HORVU1Hr1G046620	HORVU2Hr1G016300	HORVU2Hr1G112790	HORVU3Hr1G083350	HORVU4Hr1G060310
HORVU1Hr1G050780	HORVU2Hr1G017370	HORVU2Hr1G114950	HORVU3Hr1G086970	HORVU4Hr1G064440
HORVU1Hr1G051260	HORVU2Hr1G017640	HORVU2Hr1G116880	HORVU3Hr1G087400	HORVU4Hr1G065150
HORVU1Hr1G051450	HORVU2Hr1G019950	HORVU2Hr1G121110	HORVU3Hr1G087550	HORVU4Hr1G069810
HORVU1Hr1G053280	HORVU2Hr1G020100	HORVU2Hr1G123650	HORVU3Hr1G087720	HORVU4Hr1G070650
HORVU1Hr1G053420	HORVU2Hr1G021700	HORVU2Hr1G124850	HORVU3Hr1G090730	HORVU4Hr1G071510
HORVU1Hr1G056560	HORVU2Hr1G024360	HORVU2Hr1G125200	HORVU3Hr1G091170	HORVU4Hr1G071570
HORVU1Hr1G057910	HORVU2Hr1G024790	HORVU2Hr1G125430	HORVU3Hr1G091460	HORVU4Hr1G072670
HORVU1Hr1G058680	HORVU2Hr1G025160	HORVU2Hr1G127540	HORVU3Hr1G092600	HORVU4Hr1G073070
HORVU1Hr1G059130	HORVU2Hr1G027290	HORVU3Hr1G001020	HORVU3Hr1G094160	HORVU4Hr1G073290
HORVU1Hr1G059550	HORVU2Hr1G032280	HORVU3Hr1G001430	HORVU3Hr1G095360	HORVU4Hr1G073610
HORVU1Hr1G061630	HORVU2Hr1G034860	HORVU3Hr1G002710	HORVU3Hr1G095580	HORVU4Hr1G075740
HORVU1Hr1G062210	HORVU2Hr1G035170	HORVU3Hr1G004140	HORVU3Hr1G096290	HORVU4Hr1G075950
HORVU1Hr1G064830	HORVU2Hr1G036590	HORVU3Hr1G005810	HORVU3Hr1G098940	HORVU4Hr1G076150
HORVU1Hr1G066760	HORVU2Hr1G037260	HORVU3Hr1G006600	HORVU3Hr1G099120	HORVU4Hr1G078890
HORVU1Hr1G069310	HORVU2Hr1G038720	HORVU3Hr1G006770	HORVU3Hr1G099810	HORVU4Hr1G080610
HORVU1Hr1G069620	HORVU2Hr1G041530	HORVU3Hr1G015510	HORVU3Hr1G108000	HORVU4Hr1G081130
HORVU1Hr1G070710	HORVU2Hr1G042550	HORVU3Hr1G017020	HORVU3Hr1G108170	HORVU4Hr1G081550
HORVU1Hr1G072690	HORVU2Hr1G042560	HORVU3Hr1G018680	HORVU3Hr1G109010	HORVU4Hr1G082910
HORVU1Hr1G072850	HORVU2Hr1G045440	HORVU3Hr1G022030	HORVU3Hr1G110370	HORVU4Hr1G083340
HORVU1Hr1G073500	HORVU2Hr1G045730	HORVU3Hr1G022500	HORVU3Hr1G112190	HORVU4Hr1G084210
HORVU1Hr1G077910	HORVU2Hr1G055830	HORVU3Hr1G022770	HORVU3Hr1G112920	HORVU4Hr1G085030

HORVU4Hr1G086320	HORVU5Hr1G076400	HORVU6Hr1G019450	HORVU6Hr1G090890	HORVU7Hr1G060260
HORVU4Hr1G087780	HORVU5Hr1G076780	HORVU6Hr1G020570	HORVU6Hr1G091320	HORVU7Hr1G060310
HORVU5Hr1G001090	HORVU5Hr1G079390	HORVU6Hr1G020840	HORVU6Hr1G091700	HORVU7Hr1G065240
HORVU5Hr1G001260	HORVU5Hr1G079470	HORVU6Hr1G025850	HORVU6Hr1G092730	HORVU7Hr1G065740
HORVU5Hr1G006400	HORVU5Hr1G080150	HORVU6Hr1G031080	HORVU6Hr1G093640	HORVU7Hr1G070420
HORVU5Hr1G008270	HORVU5Hr1G080530	HORVU6Hr1G034660	HORVU6Hr1G093820	HORVU7Hr1G070490
HORVU5Hr1G009040	HORVU5Hr1G080540	HORVU6Hr1G035470	HORVU6Hr1G095190	HORVU7Hr1G071060
HORVU5Hr1G010120	HORVU5Hr1G080610	HORVU6Hr1G036840	HORVU7Hr1G000040	HORVU7Hr1G074970
HORVU5Hr1G010150	HORVU5Hr1G080700	HORVU6Hr1G040780	HORVU7Hr1G000320	HORVU7Hr1G075960
HORVU5Hr1G010420	HORVU5Hr1G082480	HORVU6Hr1G042410	HORVU7Hr1G000980	HORVU7Hr1G076810
HORVU5Hr1G010560	HORVU5Hr1G084230	HORVU6Hr1G044030	HORVU7Hr1G001020	HORVU7Hr1G076850
HORVU5Hr1G012020	HORVU5Hr1G084820	HORVU6Hr1G044080	HORVU7Hr1G006900	HORVU7Hr1G080820
HORVU5Hr1G012040	HORVU5Hr1G086960	HORVU6Hr1G052600	HORVU7Hr1G009700	HORVU7Hr1G081590
HORVU5Hr1G012300	HORVU5Hr1G089140	HORVU6Hr1G053490	HORVU7Hr1G010320	HORVU7Hr1G081770
HORVU5Hr1G013160	HORVU5Hr1G089290	HORVU6Hr1G053910	HORVU7Hr1G012910	HORVU7Hr1G084240
HORVU5Hr1G017890	HORVU5Hr1G093180	HORVU6Hr1G055910	HORVU7Hr1G013890	HORVU7Hr1G090350
HORVU5Hr1G018010	HORVU5Hr1G101340	HORVU6Hr1G056260	HORVU7Hr1G017660	HORVU7Hr1G090380
HORVU5Hr1G018160	HORVU5Hr1G104580	HORVU6Hr1G057570	HORVU7Hr1G018000	HORVU7Hr1G092680
HORVU5Hr1G018210	HORVU5Hr1G106140	HORVU6Hr1G058690	HORVU7Hr1G020300	HORVU7Hr1G094730
HORVU5Hr1G027290	HORVU5Hr1G111570	HORVU6Hr1G062050	HORVU7Hr1G022230	HORVU7Hr1G095180
HORVU5Hr1G028290	HORVU5Hr1G111870	HORVU6Hr1G062320	HORVU7Hr1G024290	HORVU7Hr1G095810
HORVU5Hr1G038090	HORVU5Hr1G112870	HORVU6Hr1G067360	HORVU7Hr1G024990	HORVU7Hr1G096020
HORVU5Hr1G039100	HORVU5Hr1G113040	HORVU6Hr1G069390	HORVU7Hr1G025400	HORVU7Hr1G096460
HORVU5Hr1G046050	HORVU5Hr1G114700	HORVU6Hr1G069570	HORVU7Hr1G026540	HORVU7Hr1G097020
HORVU5Hr1G048010	HORVU5Hr1G115670	HORVU6Hr1G072260	HORVU7Hr1G026650	HORVU7Hr1G097790
HORVU5Hr1G048030	HORVU5Hr1G116920	HORVU6Hr1G073090	HORVU7Hr1G028550	HORVU7Hr1G103170
HORVU5Hr1G052340	HORVU5Hr1G121700	HORVU6Hr1G074220	HORVU7Hr1G030540	HORVU7Hr1G104490
HORVU5Hr1G055260	HORVU5Hr1G122360	HORVU6Hr1G074440	HORVU7Hr1G033230	HORVU7Hr1G108810
HORVU5Hr1G055740	HORVU5Hr1G122860	HORVU6Hr1G076660	HORVU7Hr1G034290	HORVU7Hr1G114920
HORVU5Hr1G056350	HORVU5Hr1G125320	HORVU6Hr1G078130	HORVU7Hr1G035440	HORVU7Hr1G115040
HORVU5Hr1G060340	HORVU6Hr1G001200	HORVU6Hr1G080690	HORVU7Hr1G035770	HORVU7Hr1G115920
HORVU5Hr1G067230	HORVU6Hr1G002500	HORVU6Hr1G081160	HORVU7Hr1G038500	HORVU7Hr1G116680
HORVU5Hr1G068970	HORVU6Hr1G004270	HORVU6Hr1G082310	HORVU7Hr1G040250	HORVU7Hr1G116940
HORVU5Hr1G069350	HORVU6Hr1G005800	HORVU6Hr1G082630	HORVU7Hr1G041260	HORVU7Hr1G117320
HORVU5Hr1G071480	HORVU6Hr1G005900	HORVU6Hr1G083600	HORVU7Hr1G042940	HORVU7Hr1G118220
HORVU5Hr1G072020	HORVU6Hr1G008870	HORVU6Hr1G085030	HORVU7Hr1G045150	HORVU7Hr1G118740
HORVU5Hr1G072980	HORVU6Hr1G011120	HORVU6Hr1G087330	HORVU7Hr1G047800	HORVU7Hr1G119570
HORVU5Hr1G074260	HORVU6Hr1G012020	HORVU6Hr1G088570	HORVU7Hr1G049200	HORVU7Hr1G119680
HORVU5Hr1G074720	HORVU6Hr1G014360	HORVU6Hr1G089670	HORVU7Hr1G049240	HORVU7Hr1G121090
HORVU5Hr1G075300	HORVU6Hr1G017720	HORVU6Hr1G090040	HORVU7Hr1G055290	HORVU7Hr1G121250
HORVU5Hr1G075920	HORVU6Hr1G019300	HORVU6Hr1G090470	HORVU7Hr1G059240	

Table A.9.18. Log₂ fold difference between Protein Coding Genes (PCGs) and Transposable Element (TE) superfamilies targeted by differential methylation correlated with sRNA within two 100 bp windows when compared with random sampling. Methylation is classified as targeting a TE if the differential methylation is inside the TE, and targeting a PCG if it is between 1.5 kbp upstream of the transcription start site and 1.5 kbp downstream of the transcription end site. Classification of TEs is done according to the standard established by Wicker et al. (2007). Only TE classifications where random sampling was > 0.01% shown.

Differential methylation target of siRNA	log ₂ (fold difference relative to random sampling)					
	Leaf			Root		
	23 nt	24 nt	23/24 nt	23 nt	24 nt	23/24 nt
PCG without TE	1.5	1.0	1.0	2.0	1.7	1.8
PCG with TE	3.0	2.2	2.1	3.3	3.0	3.0
PCG with or without TE	2.3	1.6	1.6	2.7	2.4	2.4
DTC without PCG	0.3	1.0	1.0	-0.1	0.3	0.3
DTH without PCG	3.1	3.0	3.0	4.0	3.8	3.5
DTM without PCG	3.2	3.3	3.2	1.1	3.2	3.2
DTT without PCG	4.8	3.4	3.5	3.8	3.9	4.0
DTX without PCG	5.6	4.3	4.2	6.2	5.6	5.4
DXX without PCG	4.7	3.7	3.7	4.7	4.6	4.6
RIX without PCG	3.8	3.3	3.2	1.9	3.6	3.6
RLC without PCG	-1.5	-0.8	-0.8	-2.0	-1.7	-1.7
RLG without PCG	-2.5	-1.2	-1.2	-2.7	-2.4	-2.2
RLX without PCG	-1.3	-0.6	-0.6	-1.1	-1.1	-1.1
RSX without PCG	5.1	4.3	4.1	4.1	4.8	4.6
XXX without PCG	4.0	3.0	2.9	3.3	3.9	3.9
DTC with PCG	2.2	2.2	2.1	3.2	2.7	2.3
DTH with PCG	4.9	3.4	3.5	6.5	5.0	5.0
DTM with PCG	2.2	2.7	2.6	3.2	4.6	4.4
DTT with PCG	5.3	5.5	5.3	4.9	5.6	5.4
DTX with PCG	5.9	5.0	5.0	4.4	5.6	5.4
DXX with PCG	5.5	4.3	4.1	5.4	5.4	5.2
RIX with PCG	4.4	3.4	3.1	5.4	4.9	4.9
RLC with PCG	1.6	1.3	1.2	-0.1	1.2	1.7
RLG with PCG	1.8	0.9	0.9	2.3	1.5	1.9
RLX with PCG	3.0	2.0	2.0	3.4	2.7	2.7
RSX with PCG	5.5	4.2	4.2	5.2	5.9	6.0
XXX with PCG	4.5	3.4	3.3	4.6	4.5	4.5
DTC with or without PCG	0.6	1.1	1.1	0.5	0.7	0.6
DTH with or without PCG	3.7	3.1	3.1	5.0	4.2	4.0
DTM with or without PCG	3.1	3.2	3.1	1.8	3.6	3.6
DTT with or without PCG	5.0	4.4	4.3	4.2	4.7	4.6
DTX with or without PCG	5.7	4.5	4.4	5.9	5.6	5.4
DXX with or without PCG	4.9	3.9	3.8	5.0	4.9	4.8
RIX with or without PCG	3.9	3.3	3.2	3.7	4.1	4.0
RLC with or without PCG	-1.1	-0.6	-0.7	-1.9	-1.3	-1.2
RLG with or without PCG	-2.0	-1.2	-1.1	-2.0	-2.1	-1.8
RLX with or without PCG	-0.7	-0.4	-0.4	-0.4	-0.6	-0.7
RSX with or without PCG	5.2	4.3	4.1	4.5	5.2	5.2
XXX with or without PCG	4.1	3.1	3.0	3.7	4.1	4.1
none	0.3	0.3	0.3	-0.3	0.1	0.0

10 Appendix B, Custom Bioinformatics Code

This appendix records the code that was produced and used to analyse the sequence data set.

10.1 mRNA_STAR_PCA.R

```
library( Rsamtools )
library( compiler )
library( ggplot2 )
library( scales )
library( plotly )
library( ggfortify )
enableJIT(3)

ptm = proc.time()

L1_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\L1.ReadsPerGene.out.tab"
L2_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\L2.ReadsPerGene.out.tab"
L3_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\L3.ReadsPerGene.out.tab"
L4_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\L4.ReadsPerGene.out.tab"
L5_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\L5.ReadsPerGene.out.tab"
L6_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\L6.ReadsPerGene.out.tab"
L7_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\L7.ReadsPerGene.out.tab"
L8_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\L8.ReadsPerGene.out.tab"
R1_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\R1.ReadsPerGene.out.tab"
R2_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\R2.ReadsPerGene.out.tab"
R3_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\R3.ReadsPerGene.out.tab"
R4_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\R4.ReadsPerGene.out.tab"
R5_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\R5.ReadsPerGene.out.tab"
R6_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\R6.ReadsPerGene.out.tab"
R7_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\R7.ReadsPerGene.out.tab"
R8_in = "C:\\BaseSpace\\mRNA_04_STAR_align_count\\R8.ReadsPerGene.out.tab"

L1 = read.csv( L1_in, header = FALSE, sep = "\t" )
L2 = read.csv( L2_in, header = FALSE, sep = "\t" )
L3 = read.csv( L3_in, header = FALSE, sep = "\t" )
L4 = read.csv( L4_in, header = FALSE, sep = "\t" )
L5 = read.csv( L5_in, header = FALSE, sep = "\t" )
L6 = read.csv( L6_in, header = FALSE, sep = "\t" )
L7 = read.csv( L7_in, header = FALSE, sep = "\t" )
L8 = read.csv( L8_in, header = FALSE, sep = "\t" )
R1 = read.csv( R1_in, header = FALSE, sep = "\t" )
R2 = read.csv( R2_in, header = FALSE, sep = "\t" )
R3 = read.csv( R3_in, header = FALSE, sep = "\t" )
R4 = read.csv( R4_in, header = FALSE, sep = "\t" )
R5 = read.csv( R5_in, header = FALSE, sep = "\t" )
R6 = read.csv( R6_in, header = FALSE, sep = "\t" )
R7 = read.csv( R7_in, header = FALSE, sep = "\t" )
R8 = read.csv( R8_in, header = FALSE, sep = "\t" )

L1summary = L1[ 1:4, ]
L2summary = L2[ 1:4, ]
L3summary = L3[ 1:4, ]
L4summary = L4[ 1:4, ]
L5summary = L5[ 1:4, ]
L6summary = L6[ 1:4, ]
L7summary = L7[ 1:4, ]
L8summary = L8[ 1:4, ]
R1summary = R1[ 1:4, ]
R2summary = R2[ 1:4, ]
R3summary = R3[ 1:4, ]
R4summary = R4[ 1:4, ]
R5summary = R5[ 1:4, ]
R6summary = R6[ 1:4, ]
R7summary = R7[ 1:4, ]
R8summary = R8[ 1:4, ]
```

```

L1 = L1[ 5:nrow( L1 ), ]
L2 = L2[ 5:nrow( L2 ), ]
L3 = L3[ 5:nrow( L3 ), ]
L4 = L4[ 5:nrow( L4 ), ]
L5 = L5[ 5:nrow( L5 ), ]
L6 = L6[ 5:nrow( L6 ), ]
L7 = L7[ 5:nrow( L7 ), ]
L8 = L8[ 5:nrow( L8 ), ]
R1 = R1[ 5:nrow( R1 ), ]
R2 = R2[ 5:nrow( R2 ), ]
R3 = R3[ 5:nrow( R3 ), ]
R4 = R4[ 5:nrow( R4 ), ]
R5 = R5[ 5:nrow( R5 ), ]
R6 = R6[ 5:nrow( R6 ), ]
R7 = R7[ 5:nrow( R7 ), ]
R8 = R8[ 5:nrow( R8 ), ]

L1_un = data.frame( L1 = L1[,2], row.names = L1[,1] )
L2_un = data.frame( L2 = L2[,2], row.names = L2[,1] )
L3_un = data.frame( L3 = L3[,2], row.names = L3[,1] )
L4_un = data.frame( L4 = L4[,2], row.names = L4[,1] )
L5_un = data.frame( L5 = L5[,2], row.names = L5[,1] )
L6_un = data.frame( L6 = L6[,2], row.names = L6[,1] )
L7_un = data.frame( L7 = L7[,2], row.names = L7[,1] )
L8_un = data.frame( L8 = L8[,2], row.names = L8[,1] )
R1_un = data.frame( R1 = R1[,2], row.names = R1[,1] )
R2_un = data.frame( R2 = R2[,2], row.names = R2[,1] )
R3_un = data.frame( R3 = R3[,2], row.names = R3[,1] )
R4_un = data.frame( R4 = R4[,2], row.names = R4[,1] )
R5_un = data.frame( R5 = R5[,2], row.names = R5[,1] )
R6_un = data.frame( R6 = R6[,2], row.names = R6[,1] )
R7_un = data.frame( R7 = R7[,2], row.names = R7[,1] )
R8_un = data.frame( R8 = R8[,2], row.names = R8[,1] )

L1_1 = data.frame( L1 = L1[,3], row.names = L1[,1] )
L2_1 = data.frame( L2 = L2[,3], row.names = L2[,1] )
L3_1 = data.frame( L3 = L3[,3], row.names = L3[,1] )
L4_1 = data.frame( L4 = L4[,3], row.names = L4[,1] )
L5_1 = data.frame( L5 = L5[,3], row.names = L5[,1] )
L6_1 = data.frame( L6 = L6[,3], row.names = L6[,1] )
L7_1 = data.frame( L7 = L7[,3], row.names = L7[,1] )
L8_1 = data.frame( L8 = L8[,3], row.names = L8[,1] )
R1_1 = data.frame( R1 = R1[,3], row.names = R1[,1] )
R2_1 = data.frame( R2 = R2[,3], row.names = R2[,1] )
R3_1 = data.frame( R3 = R3[,3], row.names = R3[,1] )
R4_1 = data.frame( R4 = R4[,3], row.names = R4[,1] )
R5_1 = data.frame( R5 = R5[,3], row.names = R5[,1] )
R6_1 = data.frame( R6 = R6[,3], row.names = R6[,1] )
R7_1 = data.frame( R7 = R7[,3], row.names = R7[,1] )
R8_1 = data.frame( R8 = R8[,3], row.names = R8[,1] )

L1_2 = data.frame( L1 = L1[,4], row.names = L1[,1] )
L2_2 = data.frame( L2 = L2[,4], row.names = L2[,1] )
L3_2 = data.frame( L3 = L3[,4], row.names = L3[,1] )
L4_2 = data.frame( L4 = L4[,4], row.names = L4[,1] )
L5_2 = data.frame( L5 = L5[,4], row.names = L5[,1] )
L6_2 = data.frame( L6 = L6[,4], row.names = L6[,1] )
L7_2 = data.frame( L7 = L7[,4], row.names = L7[,1] )
L8_2 = data.frame( L8 = L8[,4], row.names = L8[,1] )
R1_2 = data.frame( R1 = R1[,4], row.names = R1[,1] )
R2_2 = data.frame( R2 = R2[,4], row.names = R2[,1] )
R3_2 = data.frame( R3 = R3[,4], row.names = R3[,1] )
R4_2 = data.frame( R4 = R4[,4], row.names = R4[,1] )
R5_2 = data.frame( R5 = R5[,4], row.names = R5[,1] )
R6_2 = data.frame( R6 = R6[,4], row.names = R6[,1] )
R7_2 = data.frame( R7 = R7[,4], row.names = R7[,1] )
R8_2 = data.frame( R8 = R8[,4], row.names = R8[,1] )

LeafCounts_un = data.frame( L1 = L1_un,
                             L2 = L2_un,
                             L3 = L3_un,
                             L4 = L4_un,
                             L5 = L5_un,
                             L6 = L6_un,
                             L7 = L7_un,
                             L8 = L8_un )

```

```

RootCounts_un = data.frame( R1 = R1_un,
                             R2 = R2_un,
                             R3 = R3_un,
                             R4 = R4_un,
                             R5 = R5_un,
                             R6 = R6_un,
                             R7 = R7_un,
                             R8 = R8_un )

LeafCounts_1 = data.frame( L1 = L1_1,
                             L2 = L2_1,
                             L3 = L3_1,
                             L4 = L4_1,
                             L5 = L5_1,
                             L6 = L6_1,
                             L7 = L7_1,
                             L8 = L8_1 )

RootCounts_1 = data.frame( R1 = R1_1,
                             R2 = R2_1,
                             R3 = R3_1,
                             R4 = R4_1,
                             R5 = R5_1,
                             R6 = R6_1,
                             R7 = R7_1,
                             R8 = R8_1 )

LeafCounts_2 = data.frame( L1 = L1_2,
                             L2 = L2_2,
                             L3 = L3_2,
                             L4 = L4_2,
                             L5 = L5_2,
                             L6 = L6_2,
                             L7 = L7_2,
                             L8 = L8_2 )

RootCounts_2 = data.frame( R1 = R1_2,
                             R2 = R2_2,
                             R3 = R3_2,
                             R4 = R4_2,
                             R5 = R5_2,
                             R6 = R6_2,
                             R7 = R7_2,
                             R8 = R8_2 )

fileoutLeaf_un = "C:\\BaseSpace\\mRNA_DE\\Leaf counts STAR unstranded.tsv"
fileoutLeaf_1 = "C:\\BaseSpace\\mRNA_DE\\Leaf counts STAR 1st read aligned.tsv"
fileoutLeaf_2 = "C:\\BaseSpace\\mRNA_DE\\Leaf counts STAR 2nd read aligned.tsv"
fileoutRoot_un = "C:\\BaseSpace\\mRNA_DE\\Root counts STAR unstranded.tsv"
fileoutRoot_1 = "C:\\BaseSpace\\mRNA_DE\\Root counts STAR 1st read aligned.tsv"
fileoutRoot_2 = "C:\\BaseSpace\\mRNA_DE\\Root counts STAR 2nd read aligned.tsv"

fileinLeaf = fileoutLeaf_un
fileinRoot = fileoutRoot_un
fileinLeaf = fileoutLeaf_1
fileinRoot = fileoutRoot_1
fileinLeaf = fileoutLeaf_2
fileinRoot = fileoutRoot_2

a = read.table( fileinLeaf, sep = "\t" )
b = a[ , 1:8 ]
c = t( b )
bsum = colSums( b )
bnorm = b / bsum
cnorm = t( bnorm )

library( edgeR )
UQ = calcNormFactors( b, method = "upperquartile" )
mCGt = t( sweep( b, 2, UQ, FUN = "*" ) )
cnorm = mCGt
#cnorm = log2( mCGt + 0.001 )

pc = prcomp( cnorm )
summary( pc )
plot( pc )

```



```

pcap = data.frame( PCA1 = pc$x[,1], PCA2 = pc$x[,2], PCA3 = pc$x[,3], group = c(
"LC","LC","LC","LC","LS","LS","LS","LS" ) )
p = plot_ly( pcap, x = ~PCA1, y = ~PCA2, z = ~PCA3, color = ~group, opacity = 0.7 ) %>%
  add_markers() %>%
  layout( scene = list( xaxis = list( title = "PCA1" ),
                        yaxis = list( title = "PCA2" ),
                        zaxis = list( title = "PCA3" )))

p

plot( pc$x[,1], pc$x[,2], main = sprintf( "PCA Leaf mRNA" ), xlab = "PCA1", ylab = "PCA2", pch
= 32 )
points( pc$x[1:4,1], pc$x[1:4,2], col = "blue", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,1], pc$x[5:8,2], col = "red", pch = c( 53,54,55,56 ) )

a = read.table( fileinRoot, sep = "\t" )
b = a[ , 1:8 ]
c = t( b )
bsum = colSums( b )
bnorm = b / bsum
cnorm = t( bnorm )

library( edgeR )
UQ = calcNormFactors( b, method = "upperquartile" )
mCGt = t( sweep( b, 2, UQ, FUN = "*" ) )
cnorm = mCGt

pc = prcomp( cnorm )
summary( pc )
plot( pc )

pcap = data.frame( PCA1 = pc$x[,1], PCA2 = pc$x[,2], PCA3 = pc$x[,3], group = c(
"LC","LC","LC","LC","LS","LS","LS","LS" ) )
p = plot_ly( pcap, x = ~PCA1, y = ~PCA2, z = ~PCA3, color = ~group, opacity = 0.7 ) %>%
  add_markers() %>%
  layout( scene = list( xaxis = list( title = "PCA1" ),
                        yaxis = list( title = "PCA2" ),
                        zaxis = list( title = "PCA3" )))

p

plot( pc$x[,1], pc$x[,2], main = sprintf( "PCA Root mRNA" ), xlab = "PCA1", ylab = "PCA2", pch
= 32 )
points( pc$x[1:4,1], pc$x[1:4,2], col = "darkgreen", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,1], pc$x[5:8,2], col = "darkorange", pch = c( 53,54,55,56 ) )

a1 = read.table( fileinLeaf, sep = "\t" )
a2 = read.table( fileinRoot, sep = "\t" )

a = cbind( a1, a2 )

b = a[ , 1:16 ]

mask = rowSums( b ) > 10
b = b[ mask, ]

c = t( b )
bsum = colSums( b )
bnorm = b / bsum
cnorm = t( bnorm )

library( edgeR )
UQ = calcNormFactors( b, method = "upperquartile" )
mCGt = t( sweep( b, 2, UQ, FUN = "*" ) )
cnorm = mCGt

pc = prcomp( cnorm )
summary( pc )
plot( pc )

pcap = data.frame( PCA1 = pc$x[,1], PCA2 = pc$x[,2], PCA3 = pc$x[,3], group = c(
"LC","LC","LC","LC","LS","LS","LS","LS","RC","RC","RC","RC","RS","RS","RS","RS" ) )
p = plot_ly( pcap, x = ~PCA1, y = ~PCA2, z = ~PCA3, color = ~group, opacity = 0.9 ) %>%
  add_markers() %>%
  layout( scene = list( xaxis = list( title = "PCA1" ),
                        yaxis = list( title = "PCA2" ),
                        zaxis = list( title = "PCA3" )))

p

```

```

plot( pc$x[,1], pc$x[,2], main = sprintf( "PCA" ), xlab = "PCA1", ylab = "PCA2", pch = 32 )
points( pc$x[1:4,1], pc$x[1:4,2], col = "blue", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,1], pc$x[5:8,2], col = "red", pch = c( 53,54,55,56 ) )
points( pc$x[9:12,1], pc$x[9:12,2], col = "darkgreen", pch = c( 49,50,51,52 ) )
points( pc$x[13:16,1], pc$x[13:16,2], col = "darkorange", pch = c( 53,54,55,56 ) )

plot( pc$x[,1], pc$x[,3], main = sprintf( "PCA" ), xlab = "PCA1", ylab = "PCA3", pch = 32 )
points( pc$x[1:4,1], pc$x[1:4,3], col = "blue", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,1], pc$x[5:8,3], col = "red", pch = c( 53,54,55,56 ) )
points( pc$x[9:12,1], pc$x[9:12,3], col = "darkgreen", pch = c( 49,50,51,52 ) )
points( pc$x[13:16,1], pc$x[13:16,3], col = "darkorange", pch = c( 53,54,55,56 ) )

```

10.2 mRNA_DE.R

```

library( DESeq2 )
library( pasilla )
library( ggplot2 )
library( reshape2 )
library( plyr )
library( scales )

Leaf = FALSE
#Leaf = TRUE

LeafCountsFile = "C:\\BaseSpace\\mRNA_DE\\Leaf counts STAR 2nd read aligned.tsv"
RootCountsFile = "C:\\BaseSpace\\mRNA_DE\\Root counts STAR 2nd read aligned.tsv"
LeafTreatFile = "C:/BaseSpace/mRNA_DE/Leaf treatment.csv"
RootTreatFile = "C:/BaseSpace/mRNA_DE/Root treatment.csv"

if( Leaf == TRUE )
{
  #cts = as.matrix( read.csv( LeafCountsFile, sep = "\t", row.names = "Geneid" ) )
  cts = as.matrix( read.table( LeafCountsFile, sep = "\t" ) )
  coldata = read.csv( LeafTreatFile, row.names = 1 )
  TotalReads = colSums( cts )
} else {
  #cts = as.matrix( read.csv( RootCountsFile, sep = "\t", row.names = "Geneid" ) )
  cts = as.matrix( read.table( RootCountsFile, sep = "\t" ) )
  coldata = read.csv( RootTreatFile, row.names = 1 )
  TotalReads = colSums( cts )
}
all( rownames( coldata ) == colnames( cts ) )

dds <- DESeqDataSetFromMatrix( countData = cts,
                              colData = coldata,
                              design = ~ condition )

dds

dds = dds[ rowMeans( counts( dds ) ) > 10, ]

dds$condition = factor( dds$condition, levels = c( "Control", "Salt" ) )

dds = DESeq( dds )
res = results( dds )
res

resOrdered = res[ order( res$padj ), ]
summary( res )
sum( res$padj < 0.01, na.rm = TRUE )

res0_01 = results( dds, alpha = 0.05 )
summary( res0_01 )

plotMA( res0_01, ylim = c( -3,3 ) )

plotCounts(dds, gene=which.min(res$padj), intgroup="condition")

restemp = res[ complete.cases( res$padj ), ]
resOnlySig = restemp[ restemp$padj < 0.05, ]
restemp$log2FoldChange > 1 ) ), ]
restemp$log2FoldChange > 0.585 ) ), ]

```

```

resOnlySig = resOnlySig[ order( resOnlySig$log2FoldChange ), ]

#Want to get the names from this list
SigGenes = row.names( resOnlySig )

SigGenesRoot0.001 = SigGenes
#SigGenesLeaf0.001 = SigGenes

rarray = NULL

counts1 = 0
counts2 = 0
counts3 = 0
counts4 = 0

for( GeneName in SigGenes )
{
  counts1 = counts1 + 1
  r = grep( GeneName, row.names( cts ) )
  if( !isEmpty(r) )
  {
    counts2 = counts2 + 1
    if( r > 0 )
    {
      counts3 = counts3 + 1
      rarray = append( rarray, r )
    }
  }
}

cts2 = counts(dds, normalized = TRUE )

SigGenesCounts = cts[ rarray, ]

SigGenesNormCol = SigGenesCounts

SigGenesNormCol.m = melt( SigGenesNormCol )
SigGenesNormCol.m = ddply( SigGenesNormCol.m, .(Var1), transform, rescale = rescale( value, to
= c(-1,1) ) )

(p = ggplot( SigGenesNormCol.m, aes( Var2, Var1, fill = rescale ) ) + geom_tile())# aes(fill =
rescale), colour = "white" ) + scale_fill_gradient(low = "white", high = "steelblue"))

nrow( SigGenesNormCol )

df1 = data.frame( SigGenesNormCol )

if( Leaf == TRUE )
{
  write.table( df1, file = "Significant mRNA DE Leaf.tsv", sep = "\t" )
  write.table( resOnlySig, file = "Significant mRNA DE Leaf resOnlySig.tsv", sep = "\t" )
} else {
  write.table( df1, file = "Significant mRNA DE Root.tsv", sep = "\t" )
  write.table( resOnlySig, file = "Significant mRNA DE Root resOnlySig.tsv", sep = "\t" )
}

ggsave( paste0( "Heatmap 0.001 Root test.png" ), width = 14, height = 20 )

nba <- read.csv("http://datasets.flowingdata.com/ppg2008.csv")
nba$Name <- with(nba, reorder(Name, PTS))
nba.m <- melt(nba)
nba.m <- ddply(nba.m, .(variable), transform, rescale = rescale(value))

(p <- ggplot(nba.m, aes(variable, Name)) + geom_tile(aes(fill = rescale), colour = "white") +
scale_fill_gradient(low = "white", high = "steelblue"))

(p <- ggplot(nba.m, aes(variable, Name)) + geom_tile(aes(fill = rescale)))

nba <- read.csv("http://datasets.flowingdata.com/ppg2008.csv")
nba$Name <- with(nba, reorder(Name, PTS))
nba.m <- melt(nba)
nba.m <- ddply(nba.m, .(variable), transform, rescale = rescale(value))

(p <- ggplot(nba.m, aes(variable, Name)) + geom_tile(aes(fill = rescale), colour = "white") +
scale_fill_gradient(low = "white", high = "steelblue"))

summary( SigGenesRoot )
summary( SigGenesLeaf )

```



```

library( reshape2 )
library( plyr )

enableJIT(3)

if( miRBase == TRUE )
{
  dir1 = 'C:\\BaseSpace\\Jason_SmallRNA_30_aligned_miRBase'
} else {
  dir1 = 'C:\\BaseSpace\\Jason_SmallRNA_34_aligned_sRNAbench'
}

FileListL = list.files( dir1, pattern = "^L[A-Za-z0-9_]*bam$" )
FileListR = list.files( dir1, pattern = "^R[A-Za-z0-9_]*bam$" )
FileCounts = paste0( dir1, "\\count_siRNA.tsv" )
FileList = c( FileListL, FileListR )

fc = 1
c_mapped = NULL
c_unmapped = NULL
c_na = NULL
A_df = NULL
CountsAll = NULL

for( fc in 1:length( FileList ) )
{
  cat("\nFile ", fc, " ", FileList[fc], "\n" )
  fl = paste0( dir1, "\\ ", FileList[fc] )

  pl = ScanBamParam( what = c( "rname", "strand", "pos", "qwidth", "flag", "qname" ) )
  bamFile = scanBam( fl, param = pl )
  pl = NULL

  rnames = levels( bamFile[[1]]$rname )
  counts_raw = array( 0, length( rnames ) )
  names( counts_raw ) = rnames

  ptm = proc.time()
  lbf = length( bamFile[[1]]$rname )

  for( i in 1:length( bamFile[[1]]$rname ) )
  {
    if( !is.na( bamFile[[1]]$rname[ i ] ) )
    {
      ind = which( names( counts_raw ) == bamFile[[1]]$rname[ i ] )
      counts_raw[ ind ] = counts_raw[ ind ] + 1
    }
    if( i%%1000 == 0 )
    {
      cat( sprintf( "\rChecked %i of %i", i, lbf ) )
    }
  }
  cat( "\n" )
  cat( proc.time() - ptm )
  CountsAll[[ fc ]] = data.frame( t( counts_raw ), row.names = sprintf( "%s", substring(
FileList[fc], 1, 2 ) ) )
}

a = rbind( CountsAll[[1]],
          CountsAll[[2]],
          CountsAll[[3]],
          CountsAll[[4]],
          CountsAll[[5]],
          CountsAll[[6]],
          CountsAll[[7]],
          CountsAll[[8]],
          CountsAll[[9]],
          CountsAll[[10]],
          CountsAll[[11]],
          CountsAll[[12]],
          CountsAll[[13]],
          CountsAll[[14]],
          CountsAll[[15]],
          CountsAll[[16]] )

if( miRBase == TRUE )
{

```

```

    save( a, file = paste0( "SmallRNA leaf and root rRNA mito plastid removed mapped to miRBase
optl.Rdata" ) )
} else {
    save( a, file = paste0( "SmallRNA leaf and root rRNA mito plastid removed mapped to
sRNAbase.Rdata" ) )
}

miRNAcountTable = read.csv( FileCounts, sep = "\t" )
cat( sprintf( "L1, %s", miRNAcountTable[17,1] ) )

mask = colSums( a ) > 10

b = a[ , mask ]

browsum = rowSums( b )

c = rbind( b[1,]/browsum[1],
           b[2,]/browsum[2],
           b[3,]/browsum[3],
           b[4,]/browsum[4],
           b[5,]/browsum[5],
           b[6,]/browsum[6],
           b[7,]/browsum[7],
           b[8,]/browsum[8],
           b[9,]/browsum[9],
           b[10,]/browsum[10],
           b[11,]/browsum[11],
           b[12,]/browsum[12],
           b[13,]/browsum[13],
           b[14,]/browsum[14],
           b[15,]/browsum[15],
           b[16,]/browsum[16] )

d = c * 1000000

# d is CPM

#edgeR is required for different normalisation methods
library( edgeR )

f = t( b )
fsum = colSums( f )
group = c( "LC", "LC", "LC", "LC", "LS", "LS", "LS", "LS", "RC", "RC", "RC", "RC", "RS", "RS", "RS", "RS" )

UQ = calcNormFactors( f, method = "upperquartile" )
UQn = t( sweep( f, 2, UQ, FUN = "*" ) )
UQn1 = log2( UQn + 0.1 )

pc = prcomp( d )

summary( pc )
plot( pc )

plot( pc$x[,1], pc$x[,2], main = sprintf( "PCA" ), xlab = "PCA1", ylab = "PCA2", pch = 32 )
points( pc$x[1:4,1], pc$x[1:4,2], col = "blue", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,1], pc$x[5:8,2], col = "red", pch = c( 53,54,55,56 ) )
points( pc$x[9:12,1], pc$x[9:12,2], col = "darkgreen", pch = c( 49,50,51,52 ) )
points( pc$x[13:16,1], pc$x[13:16,2], col = "darkorange", pch = c( 53,54,55,56 ) )

plot( pc$x[,3], pc$x[,4], main = sprintf( "PCA" ), xlab = "PCA3", ylab = "PCA4", pch = 32 )
points( pc$x[1:4,3], pc$x[1:4,4], col = "blue", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,3], pc$x[5:8,4], col = "red", pch = c( 53,54,55,56 ) )
points( pc$x[9:12,3], pc$x[9:12,4], col = "darkgreen", pch = c( 49,50,51,52 ) )
points( pc$x[13:16,3], pc$x[13:16,4], col = "darkorange", pch = c( 53,54,55,56 ) )

library( plotly )
#library( ggfortify )

pcap = data.frame( PCA1 = pc$x[,1], PCA2 = pc$x[,2], PCA3 = pc$x[,3], group = c(
"LC", "LC", "LC", "LC", "LS", "LS", "LS", "LS", "RC", "RC", "RC", "RC", "RS", "RS", "RS", "RS" ) )
p = plot_ly( pcap, x = ~PCA1, y = ~PCA2, z = ~PCA3, color = ~group, opacity = 0.5 ) %>%
add_markers() %>%
layout( scene = list( xaxis = list( title = "PCA1" ),
yaxis = list( title = "PCA2" ),
zaxis = list( title = "PCA3" ) ) )

p

```

```

d = UQn

pc = prcomp( d[1:8,] )
summary( pc )
plot( pc )

plot( pc$x[,1], pc$x[,2], main = sprintf( "PCA" ), xlab = "PCA1", ylab = "PCA2", pch = 32 )
points( pc$x[1:4,1], pc$x[1:4,2], col = "blue", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,1], pc$x[5:8,2], col = "red", pch = c( 53,54,55,56 ) )

plot( pc$x[,3], pc$x[,4], main = sprintf( "PCA" ), xlab = "PCA3", ylab = "PCA4", pch = 32 )
points( pc$x[1:4,3], pc$x[1:4,4], col = "blue", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,3], pc$x[5:8,4], col = "red", pch = c( 53,54,55,56 ) )

library( plotly )

pcap = data.frame( PCA1 = pc$x[,1], PCA2 = pc$x[,2], PCA3 = pc$x[,3], group = c(
"LC","LC","LC","LC","LS","LS","LS","LS" ) )
p = plot_ly( pcap, x = ~PCA1, y = ~PCA2, z = ~PCA3, color = ~group, opacity = 0.5 ) %>%
  add_markers() %>%
  layout( scene = list( xaxis = list( title = "PCA1" ),
                        yaxis = list( title = "PCA2" ),
                        zaxis = list( title = "PCA3" )))

p

pc = prcomp( b[9:16,] )

d = UQn1

mask = colSums( d[9:16,] ) > 10
e = d[9:16,]
e = e[ , mask ]

pc = prcomp( e )
summary( pc )
plot( pc )

plot( pc$x[,1], pc$x[,2], main = sprintf( "PCA" ), xlab = "PCA1", ylab = "PCA2", pch = 32 )
points( pc$x[1:4,1], pc$x[1:4,2], col = "darkgreen", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,1], pc$x[5:8,2], col = "darkorange", pch = c( 53,54,55,56 ) )

plot( pc$x[,3], pc$x[,4], main = sprintf( "PCA" ), xlab = "PCA3", ylab = "PCA4", pch = 32 )
points( pc$x[1:4,3], pc$x[1:4,4], col = "darkgreen", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,3], pc$x[5:8,4], col = "darkorange", pch = c( 53,54,55,56 ) )

library( plotly )

pcap = data.frame( PCA1 = pc$x[,1], PCA2 = pc$x[,2], PCA3 = pc$x[,3], group = c(
"LC","LC","LC","LC","LS","LS","LS","LS","RC","RC","RC","RC","RS","RS","RS","RS" ) )
p = plot_ly( pcap, x = ~PCA1, y = ~PCA2, z = ~PCA3, color = ~group, opacity = 0.5 ) %>%
  add_markers() %>%
  layout( scene = list( xaxis = list( title = "PCA1" ),
                        yaxis = list( title = "PCA2" ),
                        zaxis = list( title = "PCA3" )))

p
cat( "end\n" )

```

10.5 Barley_SmallRNA_DE.R

```

ptm = proc.time()

library( Rsamtools )
library( compiler )
library( ggplot2 )
library( scales )
library( reshape2 )
library( plyr )

enableJIT(3)

dir_miRBase = 'C:\\BaseSpace\\Jason_SmallRNA_31_aligned_miRBase_rm_mpr'
dir_sRNAbench = 'C:\\BaseSpace\\Jason_SmallRNA_34_aligned_sRNAbench'

```

```

#Need to load 2 files, :
# SmallRNA leaf and root rRNA mito plastid removed mapped to sRNAbase.Rdata
# SmallRNA leaf 2lnt with rRNA plastid mito removed mapped to miRBase.Rdata (or opt1)

miRNAcountTable = read.csv( FileCounts, sep = "\t" )
cat( sprintf( "L1, %s", miRNAcountTable[17,1] ) )

load( file = paste0( "SmallRNA leaf and root rRNA mito plastid removed mapped to
sRNAbase.Rdata" ) )
a_sRNAbase = a

load( file = paste0( "SmallRNA leaf and root rRNA mito plastid removed mapped to miRBase
opt1.Rdata" ) )
a_miRbase = a

a = cbind( a_sRNAbase, a_miRbase )

mask = colSums( a ) > 10

b = a[ , mask ]

# d is CPM

#edgeR is required for different normalisation methods
library( edgeR )

f = t( b )
fsum = colSums( f )
group = c( "LC", "LC", "LC", "LC", "LS", "LS", "LS", "LS", "RC", "RC", "RC", "RC", "RS", "RS", "RS", "RS" )

library( DESeq2 )
library( pasilla )
library( ggplot2 )
library( reshape2 )
library( plyr )
library( scales )

#Change this option to select only leaf or root tissue
#Leaf = TRUE
Leaf = FALSE

LeafTreatFile = "C:/BaseSpace/mRNA_DE/Leaf treatment.csv"
RootTreatFile = "C:/BaseSpace/mRNA_DE/Root treatment.csv"

if( Leaf == TRUE )
{
  cts = f[,1:8]
  coldata = read.csv( LeafTreatFile, row.names = 1 )
  TotalReads = colSums( cts )
} else {
  cts = f[,9:16]
  coldata = read.csv( RootTreatFile, row.names = 1 )
  TotalReads = colSums( cts )
}
all( rownames( coldata ) == colnames( cts ) )

dds <- DESeqDataSetFromMatrix( countData = cts,
                              colData = coldata,
                              design = ~ condition )

dds

dds$condition = factor( dds$condition, levels = c( "Control", "Salt" ) )

dds = DESeq( dds )
res = results( dds )
res

resOrdered = res[ order( res$padj ), ]
summary( res )
sum( res$padj < 0.05, na.rm = TRUE )

res0_01 = results( dds, alpha = 0.05 )
summary( res0_01 )

plotMA( res0_01, ylim = c( -3,3 ) )

plotCounts(dds, gene=which.min(res$padj), intgroup="condition")

```



```

restemp = res[ complete.cases( res$padj ), ]
resOnlySig = restemp[ restemp$padj < 0.05, ]
restemp$log2FoldChange > 1 ) ), ]
restemp$log2FoldChange > 0.585 ) ), ]
resOnlySig = resOnlySig[ order( resOnlySig$log2FoldChange ), ]

#Want to get the names from this list

SigGenes = row.names( resOnlySig )

SigGenesRoot0.001 = SigGenes

rarray = NULL

counts1 = 0
counts2 = 0
counts3 = 0
counts4 = 0

for( GeneName in SigGenes )
{
  counts1 = counts1 + 1
  r = grep( GeneName, row.names( cts ) )
  if( !isEmpty(r) )
  {
    counts2 = counts2 + 1
    if( r > 0 )
    {
      counts3 = counts3 + 1
      rarray = append( rarray, r )
    }
  }
}

cts2 = counts(dds, normalized = TRUE )

SigGenesCounts = cts[ rarray, ]
SigGenesNormCol = SigGenesCounts
SigGenesNormCol.m = melt( SigGenesNormCol )
SigGenesNormCol.m = ddply( SigGenesNormCol.m, .(Var1), transform, rescale = rescale( value, to
= c(-1,1) ) )

(p = ggplot( SigGenesNormCol.m, aes( Var2, Var1, fill = rescale ) ) + geom_tile())# aes(fill =
rescale), colour = "white" )+ scale_fill_gradient(low = "white", high = "steelblue"))

nrow( SigGenesNormCol )

df1 = data.frame( SigGenesNormCol )

if( Leaf == TRUE )
{
  write.table( df1, file = "Significant miRNA DE Leaf.tsv", sep = "\t" )
  write.table( resOnlySig, file = "Significant miRNA DE Leaf resOnlySig.tsv", sep = "\t" )
} else {
  write.table( df1, file = "Significant miRNA DE Root.tsv", sep = "\t" )
  write.table( resOnlySig, file = "Significant miRNA DE Root resOnlySig.tsv", sep = "\t" )
}

```

10.6 SmallRNA_size_grouping.py

This python program takes smallRNA that has had contaminants removed and sorts in to separate files based on the smallRNA sequence length.

```

import gzip
import os
import glob

filein0 = '/home/jason/Share/Jason_SmallRNA_25_remove_mito_plastid_rRNA_18-26nt/R'
#filein0 = '/home/jason/Share/Jason_SmallRNA_25_remove_mito_plastid_rRNA_18-26nt/L'

```

```

filein2 = '_remove_mito_plastid_rRNA.fastq.gz'
filesin = []
fileout0 = 'R'
#fileout0 = 'L'
fileout2 = '.fastq'
filesout = []
for x in range( 1, 9 ):
    filesin.append( filein0 + str(x) + filein2 )
    filesout.append( fileout0 + str(x) + fileout2 )

f20d = '/home/jason/Share/Jason_SmallRNA_27_Root_20nt/'
f21d = '/home/jason/Share/Jason_SmallRNA_27_Root_21nt/'
f22d = '/home/jason/Share/Jason_SmallRNA_27_Root_22nt/'
f23d = '/home/jason/Share/Jason_SmallRNA_27_Root_23nt/'
f24d = '/home/jason/Share/Jason_SmallRNA_27_Root_24nt/'
#f20d = '/home/jason/Share/Jason_SmallRNA_27_Leaf_20nt/'
#f21d = '/home/jason/Share/Jason_SmallRNA_27_Leaf_21nt/'
#f22d = '/home/jason/Share/Jason_SmallRNA_27_Leaf_22nt/'
#f23d = '/home/jason/Share/Jason_SmallRNA_27_Leaf_23nt/'
#f24d = '/home/jason/Share/Jason_SmallRNA_27_Leaf_24nt/'

count = 0

for x in range( 0, 8 ):
    print( filesin[x] )
    f20 = open( f20d + filesout[x], 'a' )
    f21 = open( f21d + filesout[x], 'a' )
    f22 = open( f22d + filesout[x], 'a' )
    f23 = open( f23d + filesout[x], 'a' )
    f24 = open( f24d + filesout[x], 'a' )
    with gzip.open( filesin[x], 'rb' ) as fin:
        for line in fin:
            if( count % 4000000 == 0 ):
                print( count/4 )
            count = count + 1
            if ( ( count%4 ) == 1 ):
                line1 = line
            if ( ( count%4 ) == 2 ):
                line2 = line
            if ( ( count%4 ) == 3 ):
                line3 = line
            if ( ( count%4 ) == 0 ):
                line4 = line
                if( len( line2 ) - 1 == 20 ):
                    f20.write( line1 + line2 + line3 + line4 )
                if( len( line2 ) - 1 == 21 ):
                    f21.write( line1 + line2 + line3 + line4 )
                if( len( line2 ) - 1 == 22 ):
                    f22.write( line1 + line2 + line3 + line4 )
                if( len( line2 ) - 1 == 23 ):
                    f23.write( line1 + line2 + line3 + line4 )
                if( len( line2 ) - 1 == 24 ):
                    f24.write( line1 + line2 + line3 + line4 )

    count = 0
    f20.close()
    f21.close()
    f22.close()
    f23.close()
    f24.close()

```

10.7 siRNA_windowing.py

This python program counts the number of siRNA reads in each 100 bp window of the genome and also the number of mapped and total smallRNA reads.

```

#!/usr/bin/python
#
# Batches siRNA reads in to windows
#

```

```

# Jason Smith
# Last modified 14/2/2020 (d/m/y)

import sys
import gzip
import os
import time

window_size = 100
step_size = 100

starttime = time.time()
if( sys.platform[0] == 'w' ):
    basedir = 'C:/BaseSpace/'
else:
    basedir = '/home/jason/Share/'

dir_in = basedir + 'Jason_SmallRNA_40_Root_aligned_23nt/'
dir_out = basedir + 'Jason_SmallRNA_40_Root_aligned_23nt/'

chromosomes = [ 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' ]
chr_sizes = [ 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000,
249774706 ]
s = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6',
'R7', 'R8' ]

sL = s[ 8:16 ]

for samp in sL:
    mapped_counts = [ None ] * 8
    for c1 in range( 8 ):
        mapped_counts[ c1 ] = [ 0.0 ] * ( chr_sizes[ c1 ] / step_size + 1 )
    fname = dir_in + samp + '.fastq.gz.bowtie.aligned.sam.gz'
    print( 'Reading %s' % fname )
    fh = gzip.open( fname, 'rb' )
    # Need to read and discard the header lines
    keep_going = True
    while( keep_going ):
        next_line = fh.readline().strip()
        if( next_line[0] != '@' ):
            keep_going = False
    keep_going = True
    loop_count = 0
    total_reads = 0
    mapped_reads = 0
    while( keep_going ):
        line_split = next_line.split()
        line_ID = line_split[0]
        # Need to find all instances with the same ID
        same_ID = True
        multilines = []
        multilines.append( line_split )
        total_reads += 1
        while( same_ID ):
            next_line = fh.readline().strip()
            if( next_line == '' ):
                keep_going = False
                same_ID = False
                break
            line_split = next_line.split()
            if( line_split[0] == line_ID ):
                # It is the same ID
                multilines.append( line_split )
            else:
                same_ID = False

        # Check that they are all the same match level
        if( ( int( multilines[ 0 ][ 1 ] ) & 4 ) == 4 ):
            pass
        else:
            mapped_reads += 1
            for c1 in range( len( multilines ) ):
                if( multilines[ c1 ][ 2 ] in chromosomes ):
                    ind1 = chromosomes.index( multilines[ c1 ][ 2 ] )
                    ind2 = int( ( int( multilines[ c1 ][ 3 ] ) + 11 ) / window_size )
                    mapped_counts[ ind1 ][ ind2 ] += 1.0 / len( multilines )

```

```

# Note that bowtie2 mapping location is relative to the genome and goes from that
# point to the right regardless of it being forward or reverse strand alignment
# BAM flags
# Bit      Description
#   1      0x1 template having multiple segments in sequencing
#   2      0x2 each segment properly aligned according to the aligner
#   4      0x4 segment unmapped
#   8      0x8 next segment in the template unmapped
#  16      0x10 SEQ being reverse complemented
#  32      0x20 SEQ of the next segment in the template being reverse complemented
#  64      0x40 the first segment in the template
# 128      0x80 the last segment in the template
# 256      0x100 secondary alignment
# 512      0x200 not passing filters, such as platform/vendor quality controls
#1024      0x400 PCR or optical duplicate
#2048      0x800 supplementary alignment

loop_count += 1
if( ( loop_count % 100000 ) == 0 ):
    print( loop_count )

fh.close()
fname_mc = dir_in + 'nl.mapped.counts.window.%d.step.%d.%s.totals.tsv' % ( window_size,
step_size, samp )
fo_mc = open( fname_mc, 'wb' )
fo_mc.write( 'Total reads = %d\n' % total_reads )
fo_mc.write( 'Mapped reads = %d\n' % mapped_reads )
fo_mc.close()
for chr_ind in range( len( chromosomes ) ):
    fname_mc = dir_in + 'nl.mapped.counts.window.%d.step.%d.%s.%s.tsv' % ( window_size,
step_size, samp, chromosomes[ chr_ind ] )
    fo_mc = open( fname_mc, 'wb' )
    for c2 in range( len( mapped_counts[ chr_ind ] ) ):
        fo_mc.write( '%.4g\n' % mapped_counts[ chr_ind ][ c2 ] )
    fo_mc.close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )

```

10.8 windowed_siRNA_ttests.R

This R script takes the output from `smallRNA_windowing.py` and performs t-tests for each genomic window with reads in at least three of the biological replicates. This output is not useful in itself as the t-tests need

```

# This R script takes the windowed small RNA data set from pypy smallRNA_windowing_04.py
# and determines statistical significance between groups

ptm = proc.time()

library( compiler )
library( data.table )

enableJIT( 3 )

lr = 'L'
#lr = 'R'
chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
s = c( 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6',
'R7', 'R8' )

if( lr == 'L' )
{
    ss = s[ 1:8 ]
    dirs = c( 'C:/BaseSpace/Jason_SmallRNA_40_Leaf_aligned_23nt/',
'C:/BaseSpace/Jason_SmallRNA_40_Leaf_aligned_24nt/' )

```

```

    dirout = 'C:/BaseSpace/Jason_SmallRNA_40_Leaf_aligned_23-24nt/'
} else {
    ss = s[ 9:16 ]
    dirs = c( 'C:/BaseSpace/Jason_SmallRNA_40_Root_aligned_23nt/',
'C:/BaseSpace/Jason_SmallRNA_40_Root_aligned_24nt/' )
    dirout = 'C:/BaseSpace/Jason_SmallRNA_40_Root_aligned_23-24nt/'
}
total_reads = numeric( length = length( chromosomes ) )
mapped_reads = numeric( length = length( chromosomes ) )

for( samp_ind in 1:length( ss ) )
{
    filename23 = paste0( dirs[ 1 ], 'mapped.counts.window.100.step.100.', ss[ samp_ind ],
'.totals.tsv' )
    totals23 = fread( filename23, sep = '=' )
    filename24 = paste0( dirs[ 2 ], 'mapped.counts.window.100.step.100.', ss[ samp_ind ],
'.totals.tsv' )
    totals24 = fread( filename24, sep = '=' )
    total_reads[ samp_ind ] = as.numeric( totals23[ 1, 2 ] ) + as.numeric( totals24[ 1, 2 ] )
    mapped_reads[ samp_ind ] = as.numeric( totals23[ 2, 2 ] ) + as.numeric( totals24[ 2, 2 ] )
}

chr_ind = 1
samp_ind = 1

for( chr_ind in 1:length( chromosomes ) )
{
    cpm = NULL
    for( samp_ind in 1:length( ss ) )
    {
        filename23 = paste0( dirs[ 1 ], 'n1.mapped.counts.window.100.step.100.', ss[ samp_ind ],
'.', chromosomes[ chr_ind ], '.tsv' )
        filename24 = paste0( dirs[ 2 ], 'n1.mapped.counts.window.100.step.100.', ss[ samp_ind ],
'.', chromosomes[ chr_ind ], '.tsv' )
        a1 = fread( filename23, sep = '\n' )
        a2 = fread( filename24, sep = '\n' )
        cpm[[ samp_ind ]] = ( a1 + a2 ) * 1000000 / mapped_reads[ samp_ind ]
    }
    rm( a1, a2 )

    cat( 'directory ', dirs[ 1 ], ', chromosome ', chromosomes[ chr_ind ], '\n', sep = '' )
    cat( 'directory ', dirs[ 2 ], ', chromosome ', chromosomes[ chr_ind ], '\n', sep = '' )
    # Need do lots of t-tests then work out the q-value (p-value adjusted for false discovery
rate)
    wins = NROW( cpm[[1]] )
    pvals = numeric( length = wins ) + 1
    mean_c = numeric( length = wins ) * NA
    mean_t = numeric( length = wins ) * NA
    indices = numeric( length = wins )
    counter = 0

    for( win_ind in 1:wins )
    {
        t1 = as.numeric( cpm[[ 1 ]][ win_ind ] )
        t2 = as.numeric( cpm[[ 2 ]][ win_ind ] )
        t3 = as.numeric( cpm[[ 3 ]][ win_ind ] )
        t4 = as.numeric( cpm[[ 4 ]][ win_ind ] )
        t5 = as.numeric( cpm[[ 5 ]][ win_ind ] )
        t6 = as.numeric( cpm[[ 6 ]][ win_ind ] )
        t7 = as.numeric( cpm[[ 7 ]][ win_ind ] )
        t8 = as.numeric( cpm[[ 8 ]][ win_ind ] )

        if( t1 + t2 + t3 + t4 + t5 + t6 + t7 + t8 > 0 )
        {
            counter = counter + 1
            x1 = c( t1, t2, t3, t4 )
            y1 = c( t5, t6, t7, t8 )
            if( ( sum( x1 != 0 ) >= 3 ) | ( sum( y1 != 0 ) >= 3 ) )
            {
                pt = t.test( x = x1, y = y1 )
                if( !is.nan( pt$p.value ) )
                {
                    pvals[ counter ] = pt$p.value
                    mean_c[ counter ] = pt$estimate[ 1 ]
                    mean_t[ counter ] = pt$estimate[ 2 ]
                    indices[ counter ] = win_ind
                } else {

```

```

        indices[ counter ] = win_ind
    }
    } else {
        indices[ counter ] = win_ind
    }
}
if( win_ind %% 100000 == 0 )
{
    temp_time = proc.time() - ptm
    cat( win_ind, ' of ', wins, ', user time = ', temp_time[ 1 ], 's, system time = ',
temp_time[ 2 ], 's, elapsed time = ', temp_time[ 3 ], 's\n', sep = '' )
}
}
temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's, elapsed time =
', temp_time[ 3 ], 's\n', sep = '' )
#shorten everything
pvals = pvals[ 1:counter ]
mean_c = mean_c[ 1:counter ]
mean_t = mean_t[ 1:counter ]
indices = indices[ 1:counter ]

stratr = data.table( index = indices, pvals = pvals, mean_c = mean_c, mean_t = mean_t )
filename = paste0( dirout, 'n1.mapped.counts.window.100.step.100.stats.', chromosomes[
chr_ind ], '.tsv' )
write.table( stratr, file = filename, col.names = TRUE, row.names = FALSE, sep = '\t' )
}

```

10.9 ttests_to_adjusted_p.R

This R script uses the output from `windowed_smallRNA_ttests.R` to determine adjusted p-values.

```

# This R script takes the p-values from windowed_smallRNA_ttests.R
# and determines statistical significance with q-value and false discovery rate adjustments

ptm = proc.time()

library( compiler )
library( data.table )
library( qvalue )

enableJIT( 3 )

dir1 = 'C:/BaseSpace/Jason_SmallRNA_40_Leaf_aligned_23nt/'
dir1 = '~/Share/Jason_SmallRNA_40_Leaf_aligned_23-24nt/'
lr = 'L'
#lr = 'R'
chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
chr_sizes = c( 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000,
249774706 )
s = c( 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6',
'R7', 'R8' )

for( lr in c( 'L', 'R' ) )
{
    if( lr == 'L' )
    {
        ss = s[ 1:8 ]
        dirs = c( '~/Share/Jason_SmallRNA_40_Leaf_aligned_23-24nt/' )
    } else {
        ss = s[ 9:16 ]
        dirs = c( '~/Share/Jason_SmallRNA_40_Root_aligned_23-24nt/' )
    }
    for( dir1 in dirs )
    {
        for( chr_ind in 1:NROW( chromosomes ) )
        {

```

```

    filenamein = paste0( dir1, 'n1.mapped.counts.window.100.step.100.stats.', chromosomes[
chr_ind ], '.tsv' )
    if( file.exists( filename = filenamein ) )
    {
        cat( filenamein, ' exists\n', sep = '' )
        filenameout = paste0( dir1, 'n1.mapped.counts.window.100.step.100.stats.qvals.',
chromosomes[ chr_ind ], '.tsv' )
        if( file.exists( filename = filenameout ) )
        {
            cat( filenameout, ' exists, so not overwriting\n', sep = '' )
        } else {
            cat( filenamein, ' does not exist, so determining q-values and fdr\n', sep = '' )
            # read the input file
            dtin = fread( filenamein, sep = '\t', showProgress = FALSE )
            cat( 'Number of rows = ', NROW( dtin ), sep = '' )
            # calculate the qvalues
            qvals = qvalue( dtin$pvals )
            summary( qvals )
            # save the new data.table
            statr = data.table( index = dtin$index, pvals = dtin$pvals, qvals = qvals$qvalues,
lfdr = qvals$lfdr, mean_c = dtin$mean_c, mean_t = dtin$mean_t )
            write.table( statr, file = filenameout, col.names = TRUE, row.names = FALSE, sep =
'\t' )
        }
    } else {
        cat( filenamein, ' does not exist yet, so cannot process that one yet\n', sep = '' )
    }
}
}
}
}

```

10.10 Annotate_siRNA.R

This custom R script takes the statistically significant differences in expression over 100 bp windows of heterochromatic siRNA in salt treated vs control plants from `ttests_to_adjusted_p.R` and finds genomic features nearby. Genes within 1.5 kbp are counted and other genomic features with direct overlap of the 100 bp window are counted.

```

library( 'methylKit' )
library( 'genomation' )
library( 'GenomicFeatures' )
library( 'data.table' )
library( 'compiler' )
enableJIT( 3 )

if( exists( 'alreadyread' ) == FALSE )
{
    gff = gffToGRanges( '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE )
    genes = gffToGRanges( '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE, filter = 'gene' )
    filter.types = c( 'gene', 'ncrna_gene', 'lnc_rna', 'trna', 'rrna', 'snrna', 'snorna',
'pre_miRNA' )
    gff.filtered = gff[ !is.na( match( gff$type, filter.types ) ) ]
    rm( gff )

    te.table = fread( file = '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Barley_TE_annotation_v2_18Aug16.tsv', sep = '\t', header = TRUE )
    t2 = data.table( chr = te.table$seq_id, start = te.table$start, end = te.table$end, TE_code
= te.table$TE_code )
}

```

```

te.chr1H = t2[ t2$chr == 'chr1H' ]
te.chr2H = t2[ t2$chr == 'chr2H' ]
te.chr3H = t2[ t2$chr == 'chr3H' ]
te.chr4H = t2[ t2$chr == 'chr4H' ]
te.chr5H = t2[ t2$chr == 'chr5H' ]
te.chr6H = t2[ t2$chr == 'chr6H' ]
te.chr7H = t2[ t2$chr == 'chr7H' ]
te.chrUn = t2[ t2$chr == 'chrUn' ]
rm( t2, te.table )

t1 = genes[ seqnames( genes ) == 'chrchr1H' ]
genes.chr1H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr2H' ]
genes.chr2H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr3H' ]
genes.chr3H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr4H' ]
genes.chr4H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr5H' ]
genes.chr5H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr6H' ]
genes.chr6H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr7H' ]
genes.chr7H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchrUn' ]
genes.chrUn = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
gff.chr1H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr1H' ]
gff.chr2H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr2H' ]
gff.chr3H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr3H' ]
gff.chr4H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr4H' ]
gff.chr5H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr5H' ]
gff.chr6H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr6H' ]
gff.chr7H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr7H' ]
gff.chrUn = gff.filtered[ seqnames( gff.filtered ) == 'chrchrUn' ]

}
alreadyread = 1

lr = 'R'
offset = 0
offset = 1000
offset = 1500
chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
chromosome = 'chrUn'
filebase = '/home/jason/Share/Jason_SmallRNA_40_Leaf_aligned_24nt/'

for( filebase in c( '/home/jason/Share/Jason_SmallRNA_40_Leaf_aligned_23nt/',
                  '/home/jason/Share/Jason_SmallRNA_40_Leaf_aligned_24nt/',
                  '/home/jason/Share/Jason_SmallRNA_40_Leaf_aligned_23-24nt/',
                  '/home/jason/Share/Jason_SmallRNA_40_Root_aligned_23nt/',
                  '/home/jason/Share/Jason_SmallRNA_40_Root_aligned_24nt/',
                  '/home/jason/Share/Jason_SmallRNA_40_Root_aligned_23-24nt/' ) )
{
  for( chromosome in chromosomes )
  {
    offset = 1500
    ptm = proc.time()

    file_in = paste0( filebase, 'nl.mapped.counts.window.100.step.100.stats.qvals.',
                    chromosome, '.tsv' )
    dat1 = fread( file = file_in, header = TRUE )
  }
}

```



```

dat2 = dat1[ dat1$pvals < 0.05 , ]
cat( paste0( file_in, '\n' ) )

len1 = NROW( dat2 )
prev_chr = 'chr1H'
gene_chr = gff.chr1H
te_chr = te.chr1H
annotation = NULL
for( ind1 in 1:len1 )
{
  seqname = chromosome
  flag_change = 0
  if( seqname != prev_chr ){
    if( seqname == 'chr2H' ){
      prev_chr = 'chr2H'
      gene_chr = gff.chr2H
      te_chr = te.chr2H
      flag_change = 1
    }
    if( seqname == 'chr3H' ){
      prev_chr = 'chr3H'
      gene_chr = gff.chr3H
      te_chr = te.chr3H
      flag_change = 1
    }
    if( seqname == 'chr4H' ){
      prev_chr = 'chr4H'
      gene_chr = gff.chr4H
      te_chr = te.chr4H
      flag_change = 1
    }
    if( seqname == 'chr5H' ){
      prev_chr = 'chr5H'
      gene_chr = gff.chr5H
      te_chr = te.chr5H
      flag_change = 1
    }
    if( seqname == 'chr6H' ){
      prev_chr = 'chr6H'
      gene_chr = gff.chr6H
      te_chr = te.chr6H
      flag_change = 1
    }
    if( seqname == 'chr7H' ){
      prev_chr = 'chr7H'
      gene_chr = gff.chr7H
      te_chr = te.chr7H
      flag_change = 1
    }
    if( seqname == 'chrUn' ){
      prev_chr = 'chrUn'
      gene_chr = gff.chrUn
      te_chr = te.chrUn
      flag_change = 1
    }
  }
}

# The location is from ( dat2$index - 1 ) * 100 + 1 to dat2$index * 100

temp1 = gene_chr[ ( ( dat2$index[ ind1 ]*100 < end( gene_chr ) + offset ) & (
dat2$index[ ind1 ]*100 > start( gene_chr ) - offset ) ) |
( ( ( dat2$index[ ind1 ] - 1 ) * 100 + 1 < end( gene_chr ) + offset )
& ( ( dat2$index[ ind1 ] - 1 ) * 100 + 1 > start( gene_chr ) - offset ) ) ]

mid = ( dat2$index[ ind1 ] * 100 + ( dat2$index[ ind1 ] - 1 ) * 100 + 1 ) / 2

temp2 = te_chr[ ( ( mid < te_chr$end ) & ( mid > te_chr$start ) ) |
( ( dat2$index[ ind1 ] * 100 < te_chr$end ) & ( dat2$index[ ind1 ] * 100 >
te_chr$start ) ) |
( ( ( dat2$index[ ind1 ] - 1 ) * 100 + 1 < te_chr$end ) & ( ( dat2$index[
ind1 ] - 1 ) * 100 + 1 > te_chr$start ) ) ]

temp3 = data.frame( chr = character(), start = integer(), end = integer(), TE_code =
character(), stringsAsFactors = FALSE )

```

```

# if everything is empty then the none flag is set
if( ( NROW( temp1 ) == 0 ) & ( NROW( temp2 ) == 0 ) )
{
  none = 1
} else {
  none = 0
}
#If it is a gene, then want to know the annotation
qv = dat2$pvals[ ind1 ]
mean_c = dat2$mean_c[ ind1 ]
mean_t = dat2$mean_t[ ind1 ]

if( as( sum( temp1$type == 'gene' ) > 0, 'integer' ) )
{
  temp4 = temp1[ temp1$type == 'gene' ]
  geneID = temp4$gene_id[ 1 ]
} else {
  geneID = ''
}
# if it is a gene at the same time as a TE then the TE label with 'g' at the start is
used rather than the TE label
if( ( as( sum( temp1$type == 'gene' ) > 0, 'integer' ) == 1 ) & ( NROW( temp2 ) != 0 ) )
{
  temp1 = temp1[ 0, ]
  temp3 = temp2
  temp2 = data.frame( chr = character(), start = integer(), end = integer(), TE_code =
character(), stringsAsFactors = FALSE )
}

df1 = data.frame( chr = seqname,
  start = ( dat2$index[ ind1 ] - 1 ) * 100 + 1,
  end = dat2$index[ ind1 ] * 100,
  gene = as( sum( temp1$type == 'gene' ) > 0, 'integer' ),
  ncRNA_gene = as( sum( temp1$type == 'ncRNA_gene' ) > 0, 'integer' ),
  lnc_RNA = as( sum( temp1$type == 'lnc_RNA' ) > 0, 'integer' ),
  tRNA = as( sum( temp1$type == 'tRNA' ) > 0, 'integer' ),
  rRNA = as( sum( temp1$type == 'rRNA' ) > 0, 'integer' ),
  snRNA = as( sum( temp1$type == 'snRNA' ) > 0, 'integer' ),
  snoRNA = as( sum( temp1$type == 'snoRNA' ) > 0, 'integer' ),
  pre_miRNA = as( sum( temp1$type == 'pre_miRNA' ) > 0, 'integer' ),
  DHH = as( sum( temp2$TE_code == 'DHH' ) > 0, 'integer' ),
  DTA = as( sum( temp2$TE_code == 'DTA' ) > 0, 'integer' ),
  DTC = as( sum( temp2$TE_code == 'DTC' ) > 0, 'integer' ),
  DTH = as( sum( temp2$TE_code == 'DTH' ) > 0, 'integer' ),
  DTM = as( sum( temp2$TE_code == 'DTM' ) > 0, 'integer' ),
  DTT = as( sum( temp2$TE_code == 'DTT' ) > 0, 'integer' ),
  DTX = as( sum( temp2$TE_code == 'DTX' ) > 0, 'integer' ),
  DXX = as( sum( temp2$TE_code == 'DXX' ) > 0, 'integer' ),
  RIX = as( sum( temp2$TE_code == 'RIX' ) > 0, 'integer' ),
  RLC = as( sum( temp2$TE_code == 'RLC' ) > 0, 'integer' ),
  RLG = as( sum( temp2$TE_code == 'RLG' ) > 0, 'integer' ),
  RLX = as( sum( temp2$TE_code == 'RLX' ) > 0, 'integer' ),
  RSX = as( sum( temp2$TE_code == 'RSX' ) > 0, 'integer' ),
  RXX = as( sum( temp2$TE_code == 'RXX' ) > 0, 'integer' ),
  XXX = as( sum( temp2$TE_code == 'XXX' ) > 0, 'integer' ),
  gDHH = as( sum( temp3$TE_code == 'DHH' ) > 0, 'integer' ),
  gDTA = as( sum( temp3$TE_code == 'DTA' ) > 0, 'integer' ),
  gDTC = as( sum( temp3$TE_code == 'DTC' ) > 0, 'integer' ),
  gDTH = as( sum( temp3$TE_code == 'DTH' ) > 0, 'integer' ),
  gDTM = as( sum( temp3$TE_code == 'DTM' ) > 0, 'integer' ),
  gDTT = as( sum( temp3$TE_code == 'DTT' ) > 0, 'integer' ),
  gDTX = as( sum( temp3$TE_code == 'DTX' ) > 0, 'integer' ),
  gDXX = as( sum( temp3$TE_code == 'DXX' ) > 0, 'integer' ),
  gRIX = as( sum( temp3$TE_code == 'RIX' ) > 0, 'integer' ),
  gRLC = as( sum( temp3$TE_code == 'RLC' ) > 0, 'integer' ),
  gRLG = as( sum( temp3$TE_code == 'RLG' ) > 0, 'integer' ),
  gRLX = as( sum( temp3$TE_code == 'RLX' ) > 0, 'integer' ),
  gRSX = as( sum( temp3$TE_code == 'RSX' ) > 0, 'integer' ),
  gRXX = as( sum( temp3$TE_code == 'RXX' ) > 0, 'integer' ),
  gXXX = as( sum( temp3$TE_code == 'XXX' ) > 0, 'integer' ),
  none = none,
  geneID = geneID,
  qvalue = qv,
  mean_c = mean_c,
  mean_t = mean_t )
annotation = rbind( annotation, df1 )

```

```

}

write.table( annotation, file = paste0( file_in, '.annotation.offset', offset, '.tsv' ),
sep = '\t', row.name = FALSE, col.name = TRUE, quote = FALSE )
annotationup = annotation[ annotation$mean_t > annotation$mean_c, ]
annotationdown = annotation[ annotation$mean_t < annotation$mean_c, ]
annotationsums = colSums( annotation[ , -c( 1, 2, 3, 43, 44, 45, 46 ) ] )
annotationupsums = colSums( annotationup[ , -c( 1, 2, 3, 43, 44, 45, 46 ) ] )
annotationdownsums = colSums( annotationdown[ , -c( 1, 2, 3, 43, 44, 45, 46 ) ] )
check = rowSums( annotation[ , -c( 1, 2, 3, 43, 44, 45, 46 ) ] )
write.table( annotationsums, file = paste0( file_in, '.annotationsums.offset', offset,
'.tsv' ), sep = '\t', col.name = FALSE, row.name = TRUE, quote = FALSE )
write.table( annotationupsums, file = paste0( file_in, '.annotationsums.up.offset',
offset, '.tsv' ), sep = '\t', col.name = FALSE, row.name = TRUE, quote = FALSE )
write.table( annotationdownsums, file = paste0( file_in, '.annotationsums.down.offset',
offset, '.tsv' ), sep = '\t', col.name = FALSE, row.name = TRUE, quote = FALSE )

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's, elapsed time
= ', temp_time[ 3 ], 's\n', sep = ' ' )
}
}

```

10.11 Annotate_siRNA_random_sampling.R

This custom R script is a modified version of Annotate_siRNA.R to determine the genomic features that would be targeted by random sampling of 100 bp windows of the genome.

```

library( 'methylKit' )
library( 'genomation' )
library( 'GenomicFeatures' )
library( 'data.table' )
library( 'compiler' )
enableJIT( 3 )

if( exists( 'alreadyread' ) == FALSE )
{
  gff = gffToGRanges( '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE )
  genes = gffToGRanges( '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE, filter = 'gene' )
  filter.types = c( 'gene', 'ncrRNA_gene', 'lnc_RNA', 'trRNA', 'rRNA', 'snRNA', 'snoRNA',
'pre_miRNA' )
  gff.filtered = gff[ !is.na( match( gff$type, filter.types ) ) ]
  rm( gff )

  te.table = fread( file = '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Barley_TE_annotation_v2_18Aug16.tsv', sep = '\t', header = TRUE )
  t2 = data.table( chr = te.table$seq_id, start = te.table$start, end = te.table$end, TE_code
= te.table$TE_code )
  te.chr1H = t2[ t2$chr == 'chr1H' ]
  te.chr2H = t2[ t2$chr == 'chr2H' ]
  te.chr3H = t2[ t2$chr == 'chr3H' ]
  te.chr4H = t2[ t2$chr == 'chr4H' ]
  te.chr5H = t2[ t2$chr == 'chr5H' ]
  te.chr6H = t2[ t2$chr == 'chr6H' ]
  te.chr7H = t2[ t2$chr == 'chr7H' ]
  te.chrUn = t2[ t2$chr == 'chrUn' ]
  rm( t2, te.table )

  t1 = genes[ seqnames( genes ) == 'chrchr1H' ]
  genes.chr1H = data.table( start = start( t1 ),
                          end = end( t1 ),
                          strand = as( strand( t1 ), 'vector' ) )
  t1 = genes[ seqnames( genes ) == 'chrchr2H' ]
  genes.chr2H = data.table( start = start( t1 ),

```

```

        end = end( t1 ),
        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr3H' ]
genes.chr3H = data.table( start = start( t1 ),
        end = end( t1 ),
        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr4H' ]
genes.chr4H = data.table( start = start( t1 ),
        end = end( t1 ),
        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr5H' ]
genes.chr5H = data.table( start = start( t1 ),
        end = end( t1 ),
        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr6H' ]
genes.chr6H = data.table( start = start( t1 ),
        end = end( t1 ),
        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr7H' ]
genes.chr7H = data.table( start = start( t1 ),
        end = end( t1 ),
        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchrUn' ]
genes.chrUn = data.table( start = start( t1 ),
        end = end( t1 ),
        strand = as( strand( t1 ), 'vector' ) )
gff.chr1H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr1H' ]
gff.chr2H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr2H' ]
gff.chr3H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr3H' ]
gff.chr4H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr4H' ]
gff.chr5H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr5H' ]
gff.chr6H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr6H' ]
gff.chr7H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr7H' ]
gff.chrUn = gff.filtered[ seqnames( gff.filtered ) == 'chrchrUn' ]
}
alreadyread = 1

lr = 'R'
offset = 0
offset = 1000
offset = 1500
chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
chromosome = 'chrUn'
filebase = '/home/jason/Share/Jason_SmallRNA_40_Leaf_aligned_24nt/'

chr_sizes = c( 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000,
249774706 )
nsamp = 10000

for( filebase in c( '/home/jason/Share/Jason_SmallRNA_40_Random/' ) )
{
  for( chromosome in chromosomes )
  {
    chr_sz = chr_sizes[ match( chromosome, chromosomes ) ]

    offset = 1500
    ptm = proc.time()

    dat2 = data.table( index = sort( sample( 1:floor( chr_sz/100 ), size = floor( nsamp *
chr_sz / sum( chr_sizes ) ) ) ) )

    file_in = paste0( filebase, 'random.', chromosome, '.tsv' )

    len1 = NROW( dat2 )
    prev_chr = 'chr1H'
    gene_chr = gff.chr1H
    te_chr = te.chr1H
    annotation = NULL
    for( ind1 in 1:len1 )
    {
      seqname = chromosome
      flag_change = 0
      if( seqname != prev_chr ){
        if( seqname == 'chr2H' ){
          prev_chr = 'chr2H'
          gene_chr = gff.chr2H

```

```

    te_chr = te.chr2H
    flag_change = 1
  }
  if( seqname == 'chr3H' ){
    prev_chr = 'chr3H'
    gene_chr = gff.chr3H
    te_chr = te.chr3H
    flag_change = 1
  }
  if( seqname == 'chr4H' ){
    prev_chr = 'chr4H'
    gene_chr = gff.chr4H
    te_chr = te.chr4H
    flag_change = 1
  }
  if( seqname == 'chr5H' ){
    prev_chr = 'chr5H'
    gene_chr = gff.chr5H
    te_chr = te.chr5H
    flag_change = 1
  }
  if( seqname == 'chr6H' ){
    prev_chr = 'chr6H'
    gene_chr = gff.chr6H
    te_chr = te.chr6H
    flag_change = 1
  }
  if( seqname == 'chr7H' ){
    prev_chr = 'chr7H'
    gene_chr = gff.chr7H
    te_chr = te.chr7H
    flag_change = 1
  }
  if( seqname == 'chrUn' ){
    prev_chr = 'chrUn'
    gene_chr = gff.chrUn
    te_chr = te.chrUn
    flag_change = 1
  }
}

# The location is from ( dat2$index - 1 ) * 100 + 1 to dat2$index * 100
temp1 = gene_chr[ ( ( dat2$index[ ind1 ]*100 < end( gene_chr ) + offset ) & (
dat2$index[ ind1 ]*100 > start( gene_chr ) - offset ) ) |
( ( ( dat2$index[ ind1 ] - 1 ) * 100 + 1 < end( gene_chr ) + offset )
& ( ( dat2$index[ ind1 ] - 1 ) * 100 + 1 > start( gene_chr ) - offset ) ) ]

mid = ( dat2$index[ ind1 ] * 100 + ( dat2$index[ ind1 ] - 1 ) * 100 + 1 ) / 2

temp2 = te_chr[ ( ( mid < te_chr$end ) & ( mid > te_chr$start ) ) |
( ( dat2$index[ ind1 ] * 100 < te_chr$end ) & ( dat2$index[ ind1 ] * 100 >
te_chr$start ) ) |
( ( ( dat2$index[ ind1 ] - 1 ) * 100 + 1 < te_chr$end ) & ( ( dat2$index[
ind1 ] - 1 ) * 100 + 1 > te_chr$start ) ) ]

temp3 = data.frame( chr = character(), start = integer(), end = integer(), TE_code =
character(), stringsAsFactors = FALSE )

# if everything is empty then the none flag is set
if( ( NROW( temp1 ) == 0 ) & ( NROW( temp2 ) == 0 ) )
{
  none = 1
} else {
  none = 0
}
#If it is a gene, then want to know the annotation
qv = 0.01
mean_c = 1
mean_t = 2

if( as( sum( temp1$type == 'gene' ) > 0, 'integer' ) )
{
  temp4 = temp1[ temp1$type == 'gene' ]
  geneID = temp4$gene_id[ 1 ]
}

```

```

    } else {
      geneID = ''
    }
    # if it is a gene at the same time as a TE then the TE label with 'g' at the start is
used rather than the TE label
    if( ( as( sum( temp1$type == 'gene' ) > 0, 'integer' ) == 1 ) & ( NROW( temp2 ) != 0 ) )
    {
      temp1 = temp1[ 0, ]
      temp3 = temp2
      temp2 = data.frame( chr = character(), start = integer(), end = integer(), TE_code =
character(), stringsAsFactors = FALSE )
    }

    df1 = data.frame( chr = seqname,
      start = ( dat2$index[ ind1 ] - 1 ) * 100 + 1,
      end = dat2$index[ ind1 ] * 100,
      gene = as( sum( temp1$type == 'gene' ) > 0, 'integer' ),
      ncRNA_gene = as( sum( temp1$type == 'ncRNA_gene' ) > 0, 'integer' ),
      lnc_RNA = as( sum( temp1$type == 'lnc_RNA' ) > 0, 'integer' ),
      tRNA = as( sum( temp1$type == 'tRNA' ) > 0, 'integer' ),
      rRNA = as( sum( temp1$type == 'rRNA' ) > 0, 'integer' ),
      snRNA = as( sum( temp1$type == 'snRNA' ) > 0, 'integer' ),
      snoRNA = as( sum( temp1$type == 'snoRNA' ) > 0, 'integer' ),
      pre_miRNA = as( sum( temp1$type == 'pre_miRNA' ) > 0, 'integer' ),
      DHH = as( sum( temp2$TE_code == 'DHH' ) > 0, 'integer' ),
      DTA = as( sum( temp2$TE_code == 'DTA' ) > 0, 'integer' ),
      DTC = as( sum( temp2$TE_code == 'DTC' ) > 0, 'integer' ),
      DTH = as( sum( temp2$TE_code == 'DTH' ) > 0, 'integer' ),
      DTM = as( sum( temp2$TE_code == 'DTM' ) > 0, 'integer' ),
      DTT = as( sum( temp2$TE_code == 'DTT' ) > 0, 'integer' ),
      DTX = as( sum( temp2$TE_code == 'DTX' ) > 0, 'integer' ),
      DXX = as( sum( temp2$TE_code == 'DXX' ) > 0, 'integer' ),
      RIX = as( sum( temp2$TE_code == 'RIX' ) > 0, 'integer' ),
      RLC = as( sum( temp2$TE_code == 'RLC' ) > 0, 'integer' ),
      RLG = as( sum( temp2$TE_code == 'RLG' ) > 0, 'integer' ),
      RLX = as( sum( temp2$TE_code == 'RLX' ) > 0, 'integer' ),
      RSX = as( sum( temp2$TE_code == 'RSX' ) > 0, 'integer' ),
      RXX = as( sum( temp2$TE_code == 'RXX' ) > 0, 'integer' ),
      XXX = as( sum( temp2$TE_code == 'XXX' ) > 0, 'integer' ),
      gDHH = as( sum( temp3$TE_code == 'DHH' ) > 0, 'integer' ),
      gDTA = as( sum( temp3$TE_code == 'DTA' ) > 0, 'integer' ),
      gDTC = as( sum( temp3$TE_code == 'DTC' ) > 0, 'integer' ),
      gDTH = as( sum( temp3$TE_code == 'DTH' ) > 0, 'integer' ),
      gDTM = as( sum( temp3$TE_code == 'DTM' ) > 0, 'integer' ),
      gDTT = as( sum( temp3$TE_code == 'DTT' ) > 0, 'integer' ),
      gDTX = as( sum( temp3$TE_code == 'DTX' ) > 0, 'integer' ),
      gDXX = as( sum( temp3$TE_code == 'DXX' ) > 0, 'integer' ),
      gRIX = as( sum( temp3$TE_code == 'RIX' ) > 0, 'integer' ),
      gRLC = as( sum( temp3$TE_code == 'RLC' ) > 0, 'integer' ),
      gRLG = as( sum( temp3$TE_code == 'RLG' ) > 0, 'integer' ),
      gRLX = as( sum( temp3$TE_code == 'RLX' ) > 0, 'integer' ),
      gRSX = as( sum( temp3$TE_code == 'RSX' ) > 0, 'integer' ),
      gRXX = as( sum( temp3$TE_code == 'RXX' ) > 0, 'integer' ),
      gXXX = as( sum( temp3$TE_code == 'XXX' ) > 0, 'integer' ),
      none = none,
      geneID = geneID,
      qvalue = qv,
      mean_c = mean_c,
      mean_t = mean_t )
    annotation = rbind( annotation, df1 )
  }

  write.table( annotation, file = paste0( file_in, '.annotation.offset', offset, '.tsv' ),
sep = '\t', row.name = FALSE, col.name = TRUE, quote = FALSE )
  check = rowSums( annotation[ , -c( 1, 2, 3, 43, 44, 45, 46 ) ] )
  write.table( annotationsums, file = paste0( file_in, '.annotationsums.offset', offset,
'.tsv' ), sep = '\t', col.name = FALSE, row.name = TRUE, quote = FALSE )
  temp_time = proc.time() - ptm
  cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's, elapsed time
= ', temp_time[ 3 ], 's\n', sep = ' ' )
}
}

```

10.12 DigestBarleyDouble.R

This is an R script which performs in-silico digestion of the barley genome. It generates a histogram of the number of fragments per genomic window of an inputted size selection and determines the coverage that should be generated by sequencing.

```
library( BSgenome )
library( Rsamtools )
library( ggplot2 )

BarleyGenomeFilenameHome = "~/Uni/2015 PhD/Not Synched/Barley
Genome/Barley_2017_genome/barley_morex_pseudomolecules.fasta"
# This takes a long time the first time, the result is saved as an .idx file (speeds up the
next read enormously)
idx = scanFaIndex( BarleyGenomeFilenameHome )

starts = NULL

EnzName = "AseI"
CutSeq = "ATTAAT" # Cutting seq AT|TAAT
Offset = 2 # shift by 2 bases to find the start of the cut
AseI = data.frame( EnzName, CutSeq, Offset )

EnzName = "NsiI" # Best in NEBuffer 3.1 at 37C
CutSeq = "ATGCAT" # Cutting seq ATGCA|T
Offset = 5 # shift by 5 bases to find the start of the cut
NsiI = data.frame( EnzName, CutSeq, Offset )

EnzName = "Csp6I" # NEB have CviQI, best in NEBuffer 3.1 at 25C, Csp6I 37C, thermo buffer B
CutSeq = "GTAC" # Cutting seq G|TAC
Offset = 1 # shift by 1 bases to find the start of the cut
Csp6I = data.frame( EnzName, CutSeq, Offset )

EnzName = "MseI"
CutSeq = "TTAA" # Cutting seq T|TAA
Offset = 1 # shift by 1 bases to find the start of the cut
MseI = data.frame( EnzName, CutSeq, Offset )

EnzName = "PstI" # POSSIBLE METHYLATION SENSITIVITY
CutSeq = "CTGCAG" # Cutting seq CTGCA|G
Offset = 5 # shift by 5 bases to find the start of the cut
PstI = data.frame( EnzName, CutSeq, Offset )

EnzName = ""
CutSeq = ""
Offset = 0
REnzNull = data.frame( EnzName, CutSeq, Offset )

#Change this if only using 1 restriction enzyme

SizeLimLower = 240
SizeLimUpper = 420
SeqWidth = 250-16

# Note they had problems with PstI not cutting near hypermethylation
# PstI is a 6 base cutter
#RestrictionEnzymesUsed = 1
#REnz1 = PstI
#REnz2 = REnzNull
#RestrictionEnzymesUsed = 2
#REnz1 = AseI
#REnz2 = NsiI
#RestrictionEnzymesUsed = 1
#REnz1 = MseI
```

```

#REnz2 = REnzNull
RestrictionEnzymesUsed = 2
REnz1 = Csp6I
REnz2 = NsiI
REnz3 = REnzNull
# Always assume that enzyme 2 and enzyme 3 have the same sticky end and are interchangeable (if
3 enzymes used)
#RestrictionEnzymesUsed = 1
#REnz1 = Csp6I
#REnz2 = REnzNull
#REnz3 = REnzNull

CutLoc_df = data.frame(NULL)
CutLocDiff_df = data.frame(NULL)
ChrSize = NULL

#for loop to go through each chromosome in barley
for( i in 1:length( idx ) )
{
  cat( "Current chromosome in use: ", seqlevelsInUse( idx[ i ] ), "\n", sep = "" )

  cat(">>> Reading chromosome: ", seqlevelsInUse( idx[ i ] ), "\n", sep = "" )
  Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ i ] )

  #Have to change the name of the label or it can't be run in a loop
  names(Chromosome) = c("chr")
  StartBase = 1
  EndBase = length(Chromosome$chr)

  cat(">>> Looking for Matches between: ", seqlevelsInUse( idx[ i ] ), " and ", levels(
REnz1$EnzName ), "\n", sep = "" )
  m1 = matchPattern( levels( REnz1$CutSeq ), Chromosome$chr )

  if( RestrictionEnzymesUsed == 1 )
  {
    m2 = NULL
    CutStarts = start( m1 ) + REnz1$Offset

    CutStarts = sort( CutStarts )
    CutEnds = c( CutStarts - 1, EndBase )
    CutStarts = c( 1, CutStarts )
    CutLengths = CutEnds - CutStarts
    ChrLabel = array( seqlevelsInUse( idx[ i ] ), length( CutStarts ) )

  }
  if( RestrictionEnzymesUsed == 2 )
  {
    cat(">>> Looking for Matches between: ", seqlevelsInUse( idx[ i ] ), " and ", levels(
REnz2$EnzName ), "\n", sep = "" )
    m2 = matchPattern( levels( REnz2$CutSeq ), Chromosome$chr )
    CutStarts = append( start( m1 ) + REnz1$Offset, start( m2 ) + REnz2$Offset )

    # I had to do this as a temporary fix to deal with the situation where there were 2 6-base
cutters
    CutStartsEnz = append( width( m1 ), width( m2 ) * 100 )

    # Am I able to keep the old code, get all the fragments then later identify the 'good'
fragments'

    indexOrder = order( CutStarts )
    CutStarts = sort( CutStarts )
    CutterLength = CutStartsEnz[indexOrder]
    CutEnds = c( CutStarts - 1, EndBase )
    CutStarts = c( 1, CutStarts )
    CutLengths = CutEnds - CutStarts
    CutDiffEnds = array( 0, length( CutStarts ) )
    for( ig in 2:( length( CutStarts ) - 2 ) )
    {
      if( CutterLength[ig-1] != CutterLength[ig] )
      {
        CutDiffEnds[ig] = 1
      }
    }
    ChrLabel = array( seqlevelsInUse( idx[ i ] ), length( CutStarts ) )
  }
}

```



```

}
if( RestrictionEnzymesUsed == 3 )
{
  cat(">>> Looking for Matches between: ", seqlevelsInUse( idx[ i ] ), " and ", levels(
REnz2$EnzName ), "\n", sep = "" )
  m2 = matchPattern( levels( REnz2$CutSeq ), Chromosome$chr )
  cat(">>> Looking for Matches between: ", seqlevelsInUse( idx[ i ] ), " and ", levels(
REnz3$EnzName ), "\n", sep = "" )
  m3 = matchPattern( levels( REnz3$CutSeq ), Chromosome$chr )
  CutStarts = append( start( m1 ) + RENz1$Offset, start( m2 ) + RENz2$Offset , start( m3 )
+ RENz3$Offset )
}

ChrSize[ i ] = EndBase
CutLoc_df = rbind( CutLoc_df, data.frame( ChrLabel, CutStarts, CutLengths, CutDiffEnds ) )
CutLocDiff_df = subset( CutLoc_df, CutDiffEnds == 1 )

#Set up a count window along each chromosome
WindowSize = 100000

#for( i in 1:8 ){
if( i == 1 )
  CountWChr1H = array( 0, length( 1:ceiling( ChrSize[ 1 ]/WindowSize ) ) )
if( i == 2 )
  CountWChr2H = array( 0, length( 1:ceiling( ChrSize[ 2 ]/WindowSize ) ) )
if( i == 3 )
  CountWChr3H = array( 0, length( 1:ceiling( ChrSize[ 3 ]/WindowSize ) ) )
if( i == 4 )
  CountWChr4H = array( 0, length( 1:ceiling( ChrSize[ 4 ]/WindowSize ) ) )
if( i == 5 )
  CountWChr5H = array( 0, length( 1:ceiling( ChrSize[ 5 ]/WindowSize ) ) )
if( i == 6 )
  CountWChr6H = array( 0, length( 1:ceiling( ChrSize[ 6 ]/WindowSize ) ) )
if( i == 7 )
  CountWChr7H = array( 0, length( 1:ceiling( ChrSize[ 7 ]/WindowSize ) ) )
if( i == 8 )
  CountWChrUn = array( 0, length( 1:ceiling( ChrSize[ 8 ]/WindowSize ) ) )
}

rm( m1 )
rm( m2 )
rm( Chromosome )

rm( ChrLabel )
rm( CutStarts )
rm( CutLengths )

gc() #Garbage collection - frees up memory

save( CutLoc_df, file = paste0( "CutLoc_df ", levels( RENz1$EnzName ), " ", levels(
REnz2$EnzName ), ".Rdata" ) )

SizeSelFrag = CutLoc_df[ ( CutLoc_df$CutLengths >= SizeLimLower ) & ( CutLoc_df$CutLengths <=
SizeLimUpper ) , ]

cat( ">>> Finding the frequency of cuts in a ", WindowSize, " bp window\n", sep = "" )

CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr1H"]
for( i3 in 1:length( CountWChr1H ) )
{
  CountWChr1H[ i3 ] = sum( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized <
WindowSize*( i3 ) ) )
}
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr2H"]
for( i3 in 1:length( CountWChr2H ) )
{
  CountWChr2H[ i3 ] = sum( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized <
WindowSize*( i3 ) ) )
}
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr3H"]
for( i3 in 1:length( CountWChr3H ) )
{
  CountWChr3H[ i3 ] = sum( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized <
WindowSize*( i3 ) ) )
}
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr4H"]

```

```

for( i3 in 1:length( CountWChr4H ) )
{
  CountWChr4H[ i3 ] = sum( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized <
WindowSize*( i3 ) ) )
}
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr5H"]
for( i3 in 1:length( CountWChr5H ) )
{
  CountWChr5H[ i3 ] = sum( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized <
WindowSize*( i3 ) ) )
}
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr6H"]
for( i3 in 1:length( CountWChr6H ) )
{
  CountWChr6H[ i3 ] = sum( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized <
WindowSize*( i3 ) ) )
}
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr7H"]
for( i3 in 1:length( CountWChr7H ) )
{
  CountWChr7H[ i3 ] = sum( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized <
WindowSize*( i3 ) ) )
}
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chrUn"]
for( i3 in 1:length( CountWChrUn ) )
{
  CountWChrUn[ i3 ] = sum( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized <
WindowSize*( i3 ) ) )
}

FragCounts = c( CountWChr1H, CountWChr2H, CountWChr3H, CountWChr4H, CountWChr5H, CountWChr6H,
CountWChr7H, CountWChrUn)
FragPos = c( ( 1:length( CountWChr1H ) ) * WindowSize,
( 1:length( CountWChr2H ) ) * WindowSize,
( 1:length( CountWChr3H ) ) * WindowSize,
( 1:length( CountWChr4H ) ) * WindowSize,
( 1:length( CountWChr5H ) ) * WindowSize,
( 1:length( CountWChr6H ) ) * WindowSize,
( 1:length( CountWChr7H ) ) * WindowSize,
( 1:length( CountWChrUn ) ) * WindowSize )
FragChr = c( array( seqlevelsInUse( idx[ 1 ] ), length( CountWChr1H ) ),
array( seqlevelsInUse( idx[ 2 ] ), length( CountWChr2H ) ),
array( seqlevelsInUse( idx[ 3 ] ), length( CountWChr3H ) ),
array( seqlevelsInUse( idx[ 4 ] ), length( CountWChr4H ) ),
array( seqlevelsInUse( idx[ 5 ] ), length( CountWChr5H ) ),
array( seqlevelsInUse( idx[ 6 ] ), length( CountWChr6H ) ),
array( seqlevelsInUse( idx[ 7 ] ), length( CountWChr7H ) ),
array( seqlevelsInUse( idx[ 8 ] ), length( CountWChrUn ) ) )

FragmentPos_df = data.frame( FragChr, FragPos, FragCounts )

save( FragmentPos_df, file = paste0( "FragmentPos_df ",
levels( REnz1$EnzName ),
" ",
levels( REnz2$EnzName ),
" ",
SizeLimLower,
"-",
SizeLimUpper,
".Rdata" ) )

ggplot( FragmentPos_df,
aes( x = FragPos, y = FragCounts ) ) +
geom_point( alpha = 1, size = .5 ) +
facet_grid( FragChr~. ) +
ggtitle( paste0( "Number of ",
SizeLimLower,
"-",
SizeLimUpper,
" bp size selected fragments in ",
WindowSize,
" bp windows in the barley genome using ",
levels( REnz1$EnzName ), " ", levels( REnz2$EnzName ) ) ) +
ylab( paste0( "Number of fragments per ", WindowSize, " bp window" ) ) +
xlab( "Genome position" )

```

```

ggsave( paste0( "plot_fragments ",
               levels( REnz1$EnzName ),
               " ",
               levels( REnz2$EnzName ),
               " ",
               SizeLimLower,
               "-",
               SizeLimUpper,
               ".png" ),
        width = 10,
        height = 7 )

cat( ">>> Total number of fragments sequenced = ", sum( FragCounts ), "\n", sep = "" )
cat( ">>> Total number of reads = ", sum( FragCounts )*2, "\n", sep = "" )
cat( ">>> Assuming 300 million reads on NextSeq, 20 samples, 15 million reads per sample,
coverage = ", 15e6/(sum( FragCounts )*2), "\n", sep = "" )
cat( ">>> Assuming 250 million reads on HiSeq lane, 20 samples, 12.5 million reads per sample,
coverage = ", 12.5e6/(sum( FragCounts )*2), "\n", sep = "" )
cat( ">>> Assuming 250 million reads on HiSeq lane, 12 samples, 20.8 million reads per sample,
coverage = ", 20.83e6/(sum( FragCounts )*2), "\n", sep = "" )
cat( ">>> Assuming 300 million reads on HiSeq4000 lane, 16 samples, 18.75 million reads per
sample, coverage = ", 18.75e6/(sum( FragCounts )*2), "\n", sep = "" )
cat( ">>> Total number of ", WindowSize, " bp window unrepresented in sequencing = ", sum(
FragCounts == 0 ), ", or ", 100 * sum( FragCounts == 0 )/length( FragCounts ), "% of
windows\n", sep = "" )

sink( paste0( "log for ",
             levels( REnz1$EnzName ),
             " ",
             levels( REnz2$EnzName ),
             " ",
             SizeLimLower,
             "-",
             SizeLimUpper,
             ".txt" ) )
cat( c( paste0( "***THESE ARE FRAGMENT SIZES BEFORE ADAPTORS ARE ADDED***\n",
              "Total number of fragments sequenced = ", sum( FragCounts ), "\n",
              "Total number of reads = ", sum( FragCounts )*2, "\n",
              "Assuming 300 million reads on NextSeq, 20 samples, 15 million reads per
sample, coverage = ", 15e6/(sum( FragCounts )*2 ), "\n",
              "Assuming 250 million reads on HiSeq lane, 20 samples, 12.5 million reads per
sample, coverage = ", 12.5e6/(sum( FragCounts )*2), "\n",
              "Assuming 250 million reads on HiSeq lane, 12 samples, 20.8 million reads per
sample, coverage = ", 20.83e6/(sum( FragCounts )*2), "\n",
              "Assuming 300 million reads on HiSeq4000 lane, 16 samples, 18.75 million reads
per sample, coverage = ", 18.75e6/(sum( FragCounts )*2), "\n",
              "Total number of ", WindowSize, " bp window unrepresented in sequencing = ",
sum( FragCounts == 0 ), ", or ", 100 * sum( FragCounts == 0 )/length( FragCounts ), "% of
windows\n" ) ) )
sink()

png( paste0( "histogram_fragments ",
           levels( REnz1$EnzName ),
           " ",
           levels( REnz2$EnzName ),
           " ",
           SizeLimLower,
           "-",
           SizeLimUpper,
           ".png" ),
    width = 800,
    height = 600 )
hist( FragCounts,
      breaks = 60,
      xlim = c(0, 40),
      main = paste0( "Frequency of fragments in ",
                    WindowSize,
                    " bp windows in the barley genome, size selection ",
                    SizeLimLower,
                    "-",
                    SizeLimUpper,
                    " bp using ",
                    levels( REnz1$EnzName ), " ", levels( REnz2$EnzName ) ),
      xlab = "Number of fragments per window")#
dev.off()

```

```

ptm = proc.time()

temprun = TRUE
if( temprun )
{
  cat("***** Note that this is only one direction, paired end reads are not included *****\n")
  ChrSize = end(ranges(idx))
  CountWChr1H = array( 0, length( 1:ceiling( ChrSize[ 1 ]/WindowSize ) ) )
  t = CountWChr1H
  CountWChr1CG = t
  CountWChr1CHG = t
  CountWChr1CHH = t
  CountWChr2H = array( 0, length( 1:ceiling( ChrSize[ 2 ]/WindowSize ) ) )
  t = CountWChr2H
  CountWChr2CG = t
  CountWChr2CHG = t
  CountWChr2CHH = t
  CountWChr3H = array( 0, length( 1:ceiling( ChrSize[ 3 ]/WindowSize ) ) )
  t = CountWChr3H
  CountWChr3CG = t
  CountWChr3CHG = t
  CountWChr3CHH = t
  CountWChr4H = array( 0, length( 1:ceiling( ChrSize[ 4 ]/WindowSize ) ) )
  t = CountWChr4H
  CountWChr4CG = t
  CountWChr4CHG = t
  CountWChr4CHH = t
  CountWChr5H = array( 0, length( 1:ceiling( ChrSize[ 5 ]/WindowSize ) ) )
  t = CountWChr5H
  CountWChr5CG = t
  CountWChr5CHG = t
  CountWChr5CHH = t
  CountWChr6H = array( 0, length( 1:ceiling( ChrSize[ 6 ]/WindowSize ) ) )
  t = CountWChr6H
  CountWChr6CG = t
  CountWChr6CHG = t
  CountWChr6CHH = t
  CountWChr7H = array( 0, length( 1:ceiling( ChrSize[ 7 ]/WindowSize ) ) )
  t = CountWChr7H
  CountWChr7CG = t
  CountWChr7CHG = t
  CountWChr7CHH = t
  CountWChrUn = array( 0, length( 1:ceiling( ChrSize[ 8 ]/WindowSize ) ) )
  t = CountWChrUn
  CountWChrUnCG = t
  CountWChrUnCHG = t
  CountWChrUnCHH = t
  #1
  Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 1 ] )
  names(Chromosome) = c("chr")
  CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr1H"]
  CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chr1H"] + CutLoc_sized
  cat( "Chromosome 1\n" )

  #ptm = proc.time()
  for( i3 in 1:length( CountWChr1H ) )
  #for( i3 in 1:20 )
  {
    if( i3%%10 == 0 )
    {
      cat( "Chr1, ", i3, "of", length( CountWChr1H ), "\n" )
    }
    t_ind = which( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized < WindowSize*(
i3 ) ) )
    CountWChr1CG[ i3 ] = 0
    CountWChr1CHG[ i3 ] = 0
    CountWChr1CHH[ i3 ] = 0
    for( i4 in t_ind )
    {
      t1 = narrow( Chromosome, start = CutLoc_sized[ i4 ], width = SeqWidth )
      t2 = maskMotif( t1$chr, "N" )
      CountWChr1CG[ i3 ] = CountWChr1CG[ i3 ] + length( matchPattern( DNASTring("CG"), t2 ) )
      CountWChr1CHG[ i3 ] = CountWChr1CHG[ i3 ] + length( matchPattern( DNASTring("CHG"), t2,
fixed = FALSE ) )
      CountWChr1CHH[ i3 ] = CountWChr1CHH[ i3 ] + length( matchPattern( DNASTring("CHH"), t2,
fixed = FALSE ) )
    }
  }
}

```

```

}
#2
cat( "Chromosome 2\n" )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 2 ] )
names(Chromosome) = c("chr")
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr2H"]
CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chr2H"] + CutLoc_sized
for( i3 in 1:length( CountWChr2H ) )
{
  if( i3%%10 == 0 )
  {
    cat( "Chr2, ", i3, "of", length( CountWChr2H ), "\n" )
  }
  t_ind = which( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized < WindowSize*(
i3 ) ) )
  CountWChr2CG[ i3 ] = 0
  CountWChr2CHG[ i3 ] = 0
  CountWChr2CHH[ i3 ] = 0
  for( i4 in t_ind )
  {
    t1 = narrow( Chromosome, start = CutLoc_sized[ i4 ], width = SeqWidth )
    t2 = maskMotif( t1$chr, "N" )
    CountWChr2CG[ i3 ] = CountWChr2CG[ i3 ] + length( matchPattern( DNASTring("CG"), t2 ) )
    CountWChr2CHG[ i3 ] = CountWChr2CHG[ i3 ] + length( matchPattern( DNASTring("CHG"), t2,
fixed = FALSE ) )
    CountWChr2CHH[ i3 ] = CountWChr2CHH[ i3 ] + length( matchPattern( DNASTring("CHH"), t2,
fixed = FALSE ) )
  }
}
#3
cat( "Chromosome 3\n" )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 3 ] )
names(Chromosome) = c("chr")
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr3H"]
CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chr3H"] + CutLoc_sized
for( i3 in 1:length( CountWChr3H ) )
{
  if( i3%%10 == 0 )
  {
    cat( "Chr3, ", i3, "of", length( CountWChr3H ), "\n" )
  }
  t_ind = which( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized < WindowSize*(
i3 ) ) )
  CountWChr3CG[ i3 ] = 0
  CountWChr3CHG[ i3 ] = 0
  CountWChr3CHH[ i3 ] = 0
  for( i4 in t_ind )
  {
    t1 = narrow( Chromosome, start = CutLoc_sized[ i4 ], width = SeqWidth )
    t2 = maskMotif( t1$chr, "N" )
    CountWChr3CG[ i3 ] = CountWChr3CG[ i3 ] + length( matchPattern( DNASTring("CG"), t2 ) )
    CountWChr3CHG[ i3 ] = CountWChr3CHG[ i3 ] + length( matchPattern( DNASTring("CHG"), t2,
fixed = FALSE ) )
    CountWChr3CHH[ i3 ] = CountWChr3CHH[ i3 ] + length( matchPattern( DNASTring("CHH"), t2,
fixed = FALSE ) )
  }
}
#4
cat( "Chromosome 4\n" )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 4 ] )
names(Chromosome) = c("chr")
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr4H"]
CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chr4H"] + CutLoc_sized
for( i3 in 1:length( CountWChr4H ) )
{
  if( i3%%10 == 0 )
  {
    cat( "Chr4, ", i3, "of", length( CountWChr4H ), "\n" )
  }
  t_ind = which( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized < WindowSize*(
i3 ) ) )
  CountWChr4CG[ i3 ] = 0
  CountWChr4CHG[ i3 ] = 0
  CountWChr4CHH[ i3 ] = 0
  for( i4 in t_ind )
  {
    t1 = narrow( Chromosome, start = CutLoc_sized[ i4 ], width = SeqWidth )

```

```

        t2 = maskMotif( t1$chr, "N" )
        CountWChr4CG[ i3 ] = CountWChr4CG[ i3 ] + length( matchPattern( DNASTring("CG"), t2 ) )
        CountWChr4CHG[ i3 ] = CountWChr4CHG[ i3 ] + length( matchPattern( DNASTring("CHG"), t2,
fixed = FALSE ) )
        CountWChr4CHH[ i3 ] = CountWChr4CHH[ i3 ] + length( matchPattern( DNASTring("CHH"), t2,
fixed = FALSE ) )
    }
}
#5
cat( "Chromosome 5\n" )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 5 ] )
names(Chromosome) = c("chr")
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr5H"]
CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chr5H"] + CutLoc_sized
for( i3 in 1:length( CountWChr5H ) )
{
    if( i3%%10 == 0 )
    {
        cat( "Chr5, ", i3, "of", length( CountWChr5H ), "\n" )
    }
    t_ind = which( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized < WindowSize*(
i3 ) ) )
    CountWChr5CG[ i3 ] = 0
    CountWChr5CHG[ i3 ] = 0
    CountWChr5CHH[ i3 ] = 0
    for( i4 in t_ind )
    {
        t1 = narrow( Chromosome, start = CutLoc_sized[ i4 ], width = SeqWidth )
        t2 = maskMotif( t1$chr, "N" )
        CountWChr5CG[ i3 ] = CountWChr5CG[ i3 ] + length( matchPattern( DNASTring("CG"), t2 ) )
        CountWChr5CHG[ i3 ] = CountWChr5CHG[ i3 ] + length( matchPattern( DNASTring("CHG"), t2,
fixed = FALSE ) )
        CountWChr5CHH[ i3 ] = CountWChr5CHH[ i3 ] + length( matchPattern( DNASTring("CHH"), t2,
fixed = FALSE ) )
    }
}
#6
cat( "Chromosome 6\n" )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 6 ] )
names(Chromosome) = c("chr")
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr6H"]
CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chr6H"] + CutLoc_sized
for( i3 in 1:length( CountWChr6H ) )
{
    if( i3%%10 == 0 )
    {
        cat( "Chr6, ", i3, "of", length( CountWChr6H ), "\n" )
    }
    t_ind = which( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized < WindowSize*(
i3 ) ) )
    CountWChr6CG[ i3 ] = 0
    CountWChr6CHG[ i3 ] = 0
    CountWChr6CHH[ i3 ] = 0
    for( i4 in t_ind )
    {
        t1 = narrow( Chromosome, start = CutLoc_sized[ i4 ], width = SeqWidth )
        t2 = maskMotif( t1$chr, "N" )
        CountWChr6CG[ i3 ] = CountWChr6CG[ i3 ] + length( matchPattern( DNASTring("CG"), t2 ) )
        CountWChr6CHG[ i3 ] = CountWChr6CHG[ i3 ] + length( matchPattern( DNASTring("CHG"), t2,
fixed = FALSE ) )
        CountWChr6CHH[ i3 ] = CountWChr6CHH[ i3 ] + length( matchPattern( DNASTring("CHH"), t2,
fixed = FALSE ) )
    }
}
#7
cat( "Chromosome 7\n" )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 7 ] )
names(Chromosome) = c("chr")
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr7H"]
CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chr7H"] + CutLoc_sized
for( i3 in 1:length( CountWChr7H ) )
{
    if( i3%%10 == 0 )
    {
        cat( "Chr7, ", i3, "of", length( CountWChr7H ), "\n" )
    }
}

```

```

t_ind = which( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized < WindowSize*(
i3 ) ) )
CountWChr7CG[ i3 ] = 0
CountWChr7CHG[ i3 ] = 0
CountWChr7CHH[ i3 ] = 0
for( i4 in t_ind )
{
t1 = narrow( Chromosome, start = CutLoc_sized[ i4 ], width = SeqWidth )
t2 = maskMotif( t1$chr, "N" )
CountWChr7CG[ i3 ] = CountWChr7CG[ i3 ] + length( matchPattern( DNASTring("CG"), t2 ) )
CountWChr7CHG[ i3 ] = CountWChr7CHG[ i3 ] + length( matchPattern( DNASTring("CHG"), t2,
fixed = FALSE ) )
CountWChr7CHH[ i3 ] = CountWChr7CHH[ i3 ] + length( matchPattern( DNASTring("CHH"), t2,
fixed = FALSE ) )
}
}
#Un
cat( "Chromosome Un\n" )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 8 ] )
names(Chromosome) = c("chr")
CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chrUn"]
CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chrUn"] + CutLoc_sized
for( i3 in 1:length( CountWChrUn ) )
{
if( i3%10 == 0 )
{
cat( "ChrUn, ", i3, "of", length( CountWChrUn ), "\n" )
}
t_ind = which( ( CutLoc_sized >= WindowSize*( i3 - 1 ) ) & ( CutLoc_sized < WindowSize*(
i3 ) ) )
CountWChrUnCG[ i3 ] = 0
CountWChrUnCHG[ i3 ] = 0
CountWChrUnCHH[ i3 ] = 0
for( i4 in t_ind )
{
t1 = narrow( Chromosome, start = CutLoc_sized[ i4 ], width = SeqWidth )
t2 = maskMotif( t1$chr, "N" )
CountWChrUnCG[ i3 ] = CountWChrUnCG[ i3 ] + length( matchPattern( DNASTring("CG"), t2 )
)
CountWChrUnCHG[ i3 ] = CountWChrUnCHG[ i3 ] + length( matchPattern( DNASTring("CHG"),
t2, fixed = FALSE ) )
CountWChrUnCHH[ i3 ] = CountWChrUnCHH[ i3 ] + length( matchPattern( DNASTring("CHH"),
t2, fixed = FALSE ) )
}
}

FragCounts = c( CountWChr1CG, CountWChr1CHG, CountWChr1CHH,
CountWChr2CG, CountWChr2CHG, CountWChr2CHH,
CountWChr3CG, CountWChr3CHG, CountWChr3CHH,
CountWChr4CG, CountWChr4CHG, CountWChr4CHH,
CountWChr5CG, CountWChr5CHG, CountWChr5CHH,
CountWChr6CG, CountWChr6CHG, CountWChr6CHH,
CountWChr7CG, CountWChr7CHG, CountWChr7CHH,
CountWChrUnCG, CountWChrUnCHG, CountWChrUnCHH)

FragPos = c( ( 1:length( CountWChr1H ) ) * WindowSize,
( 1:length( CountWChr1H ) ) * WindowSize,
( 1:length( CountWChr1H ) ) * WindowSize,
( 1:length( CountWChr2H ) ) * WindowSize,
( 1:length( CountWChr2H ) ) * WindowSize,
( 1:length( CountWChr2H ) ) * WindowSize,
( 1:length( CountWChr3H ) ) * WindowSize,
( 1:length( CountWChr3H ) ) * WindowSize,
( 1:length( CountWChr3H ) ) * WindowSize,
( 1:length( CountWChr4H ) ) * WindowSize,
( 1:length( CountWChr4H ) ) * WindowSize,
( 1:length( CountWChr4H ) ) * WindowSize,
( 1:length( CountWChr5H ) ) * WindowSize,
( 1:length( CountWChr5H ) ) * WindowSize,
( 1:length( CountWChr5H ) ) * WindowSize,
( 1:length( CountWChr6H ) ) * WindowSize,
( 1:length( CountWChr6H ) ) * WindowSize,
( 1:length( CountWChr6H ) ) * WindowSize,
( 1:length( CountWChr7H ) ) * WindowSize,
( 1:length( CountWChr7H ) ) * WindowSize,
( 1:length( CountWChr7H ) ) * WindowSize,
( 1:length( CountWChrUn ) ) * WindowSize,

```

```

      ( 1:length( CountWChrUn ) ) * WindowSize,
      ( 1:length( CountWChrUn ) ) * WindowSize )
FragPos = FragPos - WindowSize/2
FragChr = c( array( seqlevelsInUse( idx[ 1 ] ), length( CountWChr1H ) ),
            array( seqlevelsInUse( idx[ 1 ] ), length( CountWChr1H ) ),
            array( seqlevelsInUse( idx[ 1 ] ), length( CountWChr1H ) ),
            array( seqlevelsInUse( idx[ 2 ] ), length( CountWChr2H ) ),
            array( seqlevelsInUse( idx[ 2 ] ), length( CountWChr2H ) ),
            array( seqlevelsInUse( idx[ 2 ] ), length( CountWChr2H ) ),
            array( seqlevelsInUse( idx[ 3 ] ), length( CountWChr3H ) ),
            array( seqlevelsInUse( idx[ 3 ] ), length( CountWChr3H ) ),
            array( seqlevelsInUse( idx[ 3 ] ), length( CountWChr3H ) ),
            array( seqlevelsInUse( idx[ 4 ] ), length( CountWChr4H ) ),
            array( seqlevelsInUse( idx[ 4 ] ), length( CountWChr4H ) ),
            array( seqlevelsInUse( idx[ 4 ] ), length( CountWChr4H ) ),
            array( seqlevelsInUse( idx[ 5 ] ), length( CountWChr5H ) ),
            array( seqlevelsInUse( idx[ 5 ] ), length( CountWChr5H ) ),
            array( seqlevelsInUse( idx[ 5 ] ), length( CountWChr5H ) ),
            array( seqlevelsInUse( idx[ 6 ] ), length( CountWChr6H ) ),
            array( seqlevelsInUse( idx[ 6 ] ), length( CountWChr6H ) ),
            array( seqlevelsInUse( idx[ 6 ] ), length( CountWChr6H ) ),
            array( seqlevelsInUse( idx[ 7 ] ), length( CountWChr7H ) ),
            array( seqlevelsInUse( idx[ 7 ] ), length( CountWChr7H ) ),
            array( seqlevelsInUse( idx[ 7 ] ), length( CountWChr7H ) ),
            array( seqlevelsInUse( idx[ 8 ] ), length( CountWChrUn ) ),
            array( seqlevelsInUse( idx[ 8 ] ), length( CountWChrUn ) ),
            array( seqlevelsInUse( idx[ 8 ] ), length( CountWChrUn ) ) )
Motif = c( array( "CG", length( CountWChr1H ) ),
          array( "CHG", length( CountWChr1H ) ),
          array( "CHH", length( CountWChr1H ) ),
          array( "CG", length( CountWChr2H ) ),
          array( "CHG", length( CountWChr2H ) ),
          array( "CHH", length( CountWChr2H ) ),
          array( "CG", length( CountWChr3H ) ),
          array( "CHG", length( CountWChr3H ) ),
          array( "CHH", length( CountWChr3H ) ),
          array( "CG", length( CountWChr4H ) ),
          array( "CHG", length( CountWChr4H ) ),
          array( "CHH", length( CountWChr4H ) ),
          array( "CG", length( CountWChr5H ) ),
          array( "CHG", length( CountWChr5H ) ),
          array( "CHH", length( CountWChr5H ) ),
          array( "CG", length( CountWChr6H ) ),
          array( "CHG", length( CountWChr6H ) ),
          array( "CHH", length( CountWChr6H ) ),
          array( "CG", length( CountWChr7H ) ),
          array( "CHG", length( CountWChr7H ) ),
          array( "CHH", length( CountWChr7H ) ),
          array( "CG", length( CountWChrUn ) ),
          array( "CHG", length( CountWChrUn ) ),
          array( "CHH", length( CountWChrUn ) ) )

FragmentPos_df = data.frame( FragChr, FragPos, FragCounts, Motif )

ggplot( FragmentPos_df,
        aes( x = FragPos, y = FragCounts, col = Motif ) ) +
  geom_point( alpha = .2, size = .1 ) +
  facet_grid( FragChr~. ) +
  ggtitle( paste0( "Number of C's in ",
                  SizeLimLower,
                  "-",
                  SizeLimUpper,
                  " bp size selected fragments in ",
                  WindowSize,
                  " bp windows in the barley genome using ",
                  levels( REnz1$EnzName ), " ", levels( REnz2$EnzName ) ) ) +
  ylab( paste0( "Number of fragments per ", WindowSize, " bp window" ) ) +
  xlab( "Genome position" )

ggsave( paste0( "C frequencies ",
                levels( REnz1$EnzName ),
                " ",
                levels( REnz2$EnzName ),
                " ",
                SizeLimLower,
                "-",
                SizeLimUpper,

```



```

        ".png" ),
        width = 10,
        height = 7 )

    save( FragmentPos_df, file = paste0( "Fragment Cs df ", levels( REnz1$EnzName ), " ",
levels( REnz2$EnzName ), " ", SizeLimLower, "-", SizeLimUpper, ".Rdata" ) )

}

proc.time()-ptm
ptm2 = proc.time()

#Comparing fragments to the entire genome to see how unique fragments are
#for loop going through each fragment with 2 differences, mask all Ns

Mismatches = 5
TestStart = 100
TestEnd = 200

CutLoc_sized = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr1H"]
CutEnd_sized = SizeSelFrag$CutLengths[SizeSelFrag$ChrLabel == "chr1H"] + CutLoc_sized
cat( "Chromosome 1\n" )

#Mismatches
CountHits1 = array( 0, TestEnd )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 1 ] )
names(Chromosome) = c("chr")
Chromosome1 = Chromosome
ChromosomeMasked = maskMotif( Chromosome$chr, "N" )
rm( Chromosome )
cat( "***** Chromosome 1\n" )
for( i1 in TestStart:TestEnd )
{
    t1 = narrow( Chromosome1, start = CutLoc_sized[ i1 ], width = SeqWidth )

    tn = t1$chr
    C_loc = matchPattern( 'C', tn )
    G_loc = matchPattern( 'G', tn )
    tn = replaceLetterAt( tn, start( C_loc ), paste( replicate( length( start( C_loc ) ), "Y" ),
collapse = "" ) )
    tn = replaceLetterAt( tn, start( G_loc ), paste( replicate( length( start( G_loc ) ), "R" ),
collapse = "" ) )

    CountHits1[ i1 ] = length( matchPattern( tn, ChromosomeMasked, max.mismatch = Mismatches,
fixed = FALSE ) )
    cat( "Ch1, Frag =", i1, ", Hits = ", CountHits1[ i1 ], "\n" )
}

CountHits2 = array( 0, TestEnd )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 2 ] )
names(Chromosome) = c("chr")
ChromosomeMasked = maskMotif( Chromosome$chr, "N" )
rm( Chromosome )
cat( "***** Chromosome 2\n" )
for( i1 in TestStart:TestEnd )
{
    t1 = narrow( Chromosome1, start = CutLoc_sized[ i1 ], width = SeqWidth )

    tn = t1$chr
    C_loc = matchPattern( 'C', tn )
    G_loc = matchPattern( 'G', tn )
    tn = replaceLetterAt( tn, start( C_loc ), paste( replicate( length( start( C_loc ) ), "Y" ),
collapse = "" ) )
    tn = replaceLetterAt( tn, start( G_loc ), paste( replicate( length( start( G_loc ) ), "R" ),
collapse = "" ) )

    CountHits2[ i1 ] = length(matchPattern( tn, ChromosomeMasked, max.mismatch = Mismatches,
fixed = FALSE ) )
    cat( "Ch2, Frag =", i1, ", Hits = ", CountHits2[ i1 ], "\n" )
}

CountHits3 = array( 0, TestEnd )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 3 ] )
names(Chromosome) = c("chr")
ChromosomeMasked = maskMotif( Chromosome$chr, "N" )
rm( Chromosome )
cat( "***** Chromosome 3\n" )

```

```

for( i1 in TestStart:TestEnd )
{
  t1 = narrow( Chromosome1, start = CutLoc_sized[ i1 ], width = SeqWidth )

  tn = t1$chr
  C_loc = matchPattern( 'C', tn )
  G_loc = matchPattern( 'G', tn )
  tn = replaceLetterAt( tn, start( C_loc ), paste( replicate( length( start( C_loc ) ), "Y" ),
collapse = "" ) )
  tn = replaceLetterAt( tn, start( G_loc ), paste( replicate( length( start( G_loc ) ), "R" ),
collapse = "" ) )

  CountHits3[ i1 ] = length(matchPattern( tn, ChromosomeMasked, max.mismatch = Mismatches,
fixed = FALSE ))
  cat( "Ch3, Frag =", i1, ", Hits = ", CountHits3[ i1 ], "\n" )
}

CountHits4 = array( 0, TestEnd )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 4 ] )
names(Chromosome) = c("chr")
ChromosomeMasked = maskMotif( Chromosome$chr, "N" )
rm( Chromosome )
cat( "***** Chromosome 4\n" )
for( i1 in TestStart:TestEnd )
{
  t1 = narrow( Chromosome1, start = CutLoc_sized[ i1 ], width = SeqWidth )

  tn = t1$chr
  C_loc = matchPattern( 'C', tn )
  G_loc = matchPattern( 'G', tn )
  tn = replaceLetterAt( tn, start( C_loc ), paste( replicate( length( start( C_loc ) ), "Y" ),
collapse = "" ) )
  tn = replaceLetterAt( tn, start( G_loc ), paste( replicate( length( start( G_loc ) ), "R" ),
collapse = "" ) )

  CountHits4[ i1 ] = length(matchPattern( tn, ChromosomeMasked, max.mismatch = Mismatches,
fixed = FALSE ))
  cat( "Ch4, Frag =", i1, ", Hits = ", CountHits4[ i1 ], "\n" )
}

CountHits5 = array( 0, TestEnd )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 5 ] )
names(Chromosome) = c("chr")
ChromosomeMasked = maskMotif( Chromosome$chr, "N" )
rm( Chromosome )
cat( "***** Chromosome 5\n" )
for( i1 in TestStart:TestEnd )
{
  t1 = narrow( Chromosome1, start = CutLoc_sized[ i1 ], width = SeqWidth )

  tn = t1$chr
  C_loc = matchPattern( 'C', tn )
  G_loc = matchPattern( 'G', tn )
  tn = replaceLetterAt( tn, start( C_loc ), paste( replicate( length( start( C_loc ) ), "Y" ),
collapse = "" ) )
  tn = replaceLetterAt( tn, start( G_loc ), paste( replicate( length( start( G_loc ) ), "R" ),
collapse = "" ) )

  CountHits5[ i1 ] = length(matchPattern( tn, ChromosomeMasked, max.mismatch = Mismatches,
fixed = FALSE ))
  cat( "Ch5, Frag =", i1, ", Hits = ", CountHits5[ i1 ], "\n" )
}

CountHits6 = array( 0, TestEnd )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 6 ] )
names(Chromosome) = c("chr")
ChromosomeMasked = maskMotif( Chromosome$chr, "N" )
rm( Chromosome )
cat( "***** Chromosome 6\n" )
for( i1 in TestStart:TestEnd )
{
  t1 = narrow( Chromosome1, start = CutLoc_sized[ i1 ], width = SeqWidth )

  tn = t1$chr
  C_loc = matchPattern( 'C', tn )
  G_loc = matchPattern( 'G', tn )

```

```

    tn = replaceLetterAt( tn, start( C_loc ), paste( replicate( length( start( C_loc ) ), "Y" ),
collapse = "" ) )
    tn = replaceLetterAt( tn, start( G_loc ), paste( replicate( length( start( G_loc ) ), "R" ),
collapse = "" ) )

    CountHits6[ i1 ] = length(matchPattern( tn, ChromosomeMasked, max.mismatch = Mismatches,
fixed = FALSE ))
    cat( "Ch6, Frag =", i1, ", Hits = ", CountHits6[ i1 ], "\n" )
}

CountHits7 = array( 0, TestEnd )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 7 ] )
names(Chromosome) = c("chr")
ChromosomeMasked = maskMotif( Chromosome$chr, "N" )
rm( Chromosome )
cat( "***** Chromosome 7\n" )
for( i1 in TestStart:TestEnd )
{
    t1 = narrow( Chromosome1, start = CutLoc_sized[ i1 ], width = SeqWidth )

    tn = t1$chr
    C_loc = matchPattern( 'C', tn )
    G_loc = matchPattern( 'G', tn )
    tn = replaceLetterAt( tn, start( C_loc ), paste( replicate( length( start( C_loc ) ), "Y" ),
collapse = "" ) )
    tn = replaceLetterAt( tn, start( G_loc ), paste( replicate( length( start( G_loc ) ), "R" ),
collapse = "" ) )

    CountHits7[ i1 ] = length(matchPattern( tn, ChromosomeMasked, max.mismatch = Mismatches,
fixed = FALSE ))
    cat( "Ch7, Frag =", i1, ", Hits = ", CountHits7[ i1 ], "\n" )
}

CountHits8 = array( 0, TestEnd )
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 8 ] )
names(Chromosome) = c("chr")
ChromosomeMasked = maskMotif( Chromosome$chr, "N" )
rm( Chromosome )
cat( "***** Chromosome 8\n" )
for( i1 in TestStart:TestEnd )
{
    t1 = narrow( Chromosome1, start = CutLoc_sized[ i1 ], width = SeqWidth )

    tn = t1$chr
    C_loc = matchPattern( 'C', tn )
    G_loc = matchPattern( 'G', tn )
    tn = replaceLetterAt( tn, start( C_loc ), paste( replicate( length( start( C_loc ) ), "Y" ),
collapse = "" ) )
    tn = replaceLetterAt( tn, start( G_loc ), paste( replicate( length( start( G_loc ) ), "R" ),
collapse = "" ) )

    CountHits8[ i1 ] = length(matchPattern( tn, ChromosomeMasked, max.mismatch = Mismatches,
fixed = FALSE ))
    cat( "ChUn, Frag =", i1, ", Hits = ", CountHits8[ i1 ], "\n" )
}

FragCounts = c( CountHits1,
                CountHits2,
                CountHits3,
                CountHits4,
                CountHits5,
                CountHits6,
                CountHits7,
                CountHits8 )

FragPos = c( 1:TestEnd,
            1:TestEnd,
            1:TestEnd,
            1:TestEnd,
            1:TestEnd,
            1:TestEnd,
            1:TestEnd,
            1:TestEnd,
            1:TestEnd )

FragChr = c( array( seqlevelsInUse( idx[ 1 ] ), TestEnd ),
            array( seqlevelsInUse( idx[ 2 ] ), TestEnd ),

```

```

array( seqlevelsInUse( idx[ 3 ] ), TestEnd ),
array( seqlevelsInUse( idx[ 4 ] ), TestEnd ),
array( seqlevelsInUse( idx[ 5 ] ), TestEnd ),
array( seqlevelsInUse( idx[ 6 ] ), TestEnd ),
array( seqlevelsInUse( idx[ 7 ] ), TestEnd ),
array( seqlevelsInUse( idx[ 8 ] ), TestEnd ) )

Good = c( CountHits1 == 1,
          CountHits2 == 0,
          CountHits3 == 0,
          CountHits4 == 0,
          CountHits5 == 0,
          CountHits6 == 0,
          CountHits7 == 0,
          CountHits8 == 0 )

FragmentHits_df = data.frame( FragChr, FragPos, FragCounts, Good )

save( FragmentHits_df, file = paste0( "Fragment Hits using ",
                                     SizeLimLower,
                                     "-",
                                     SizeLimUpper,
                                     " bp size selected fragments with ",
                                     levels( REnz1$EnzName ),
                                     " ",
                                     levels( REnz2$EnzName ),
                                     " ",
                                     TestStart,
                                     " ",
                                     TestEnd,
                                     " ",
                                     Mismatches,
                                     " with Y and R replacements.Rdata" ) )

ggplot( FragmentHits_df,
        aes( x = FragPos, y = FragCounts, col = Good ) ) +
  geom_point( alpha = 1, size = .3 ) +
  facet_grid( FragChr~. ) +
  ggtitle( paste0( "Number of fragment hits in ",
                  SizeLimLower,
                  "-",
                  SizeLimUpper,
                  " bp size selected fragments using ",
                  levels( REnz1$EnzName ), " ", levels( REnz2$EnzName ) ) ) +
  ylab( paste0( "Number of hits" ) ) +
  xlab( "Fragment number" )

ggsave( paste0( "Fragment Hits using ",
                SizeLimLower,
                "-",
                SizeLimUpper,
                " bp size selected fragments with ",
                levels( REnz1$EnzName ),
                " ",
                levels( REnz2$EnzName ),
                " ",
                TestStart,
                " ",
                TestEnd,
                " ",
                Mismatches,
                " with Y and R replacements.png" ),
        width = 10,
        height = 7 )

Gooda = c( ( CountHits1 == 1 ) & ( CountHits2 == 0 ) & ( CountHits3 == 0 ) & ( CountHits4 == 0 )
          & ( CountHits5 == 0 ) & ( CountHits6 == 0 ) & ( CountHits7 == 0 ) & ( CountHits8 == 0 ) )
FragCounts_a = CountHits1 +
              CountHits2 +
              CountHits3 +
              CountHits4 +
              CountHits5 +
              CountHits6 +
              CountHits7 +
              CountHits8

FragPos_a = 1:TestEnd

```

```

FragmentHits_dfa = data.frame( FragPosa, FragCounts, Gooda )

ggplot( FragmentHits_dfa,
        aes( x = FragPosa, y = FragCounts, col = Gooda ) ) +
  geom_point( alpha = 1, size = .3 ) +
  ggtitle( paste0( "Number of fragment total hits in ",
                  SizeLimLower,
                  "-",
                  SizeLimUpper,
                  " bp size selected fragments using ",
                  levels( REnz1$EnzName ), " ", levels( REnz2$EnzName ) ) ) +
  ylab( paste0( "Number of hits" ) ) +
  xlab( "Fragment number" )

ggsave( paste0( "Fragment Hits Total using ",
                SizeLimLower,
                "-",
                SizeLimUpper,
                " bp size selected fragments with ",
                levels( REnz1$EnzName ),
                " ",
                levels( REnz2$EnzName ),
                " ",
                TestStart,
                " ",
                TestEnd,
                " ",
                Mismatches,
                " with Y and R replacements.png" ),
        width = 10,
        height = 7 )

gr = sum( Gooda[TestStart:TestEnd] == TRUE )
gt = length( Gooda[TestStart:TestEnd] )
cat( "Number of fragments with only 1 hit in the whole genome when using C->Y and G->R ", gr,
    ", out of ", gt, ", or ", 100*gr/gt, "% (assuming ", Mismatches, " bp error)\n", sep = " " )

proc.time()-ptm2

# This is checking to see if the different ends is working correctly by comparing ends

filename = ""

# If NsiI then Csp6I then fragment is T.....G
# If Csp6I then NsiI then fragment is TAC...ATGCA
# Check some of the sequences using
#
Chromosome = scanFa( BarleyGenomeFilenameHome, idx[ 3 ] )

CutLoc_start = SizeSelFrag$CutStarts[SizeSelFrag$ChrLabel == "chr3H"]
CutLoc_length = SizeSelFrag$CutLength[SizeSelFrag$ChrLabel == "chr3H"]
totalcounter = 0
diffcounter = 0

for( i3 in 101001:102000 )
{
  t1 = narrow( Chromosome, start = CutLoc_start[i3], width = CutLoc_length[i3]+1 )
  ta = start( vmatchPattern( "TAC", narrow( t1, start = 1, width = 3 ) ) )
  tb = start( vmatchPattern( "T", narrow( t1, start = 1, width = 1 ) ) )
  tc = start( vmatchPattern( "ATGCA", narrow( t1, start = CutLoc_length[i3]-3, width = 5 ) ) )
  td = start( vmatchPattern( "G", narrow( t1, start = CutLoc_length[i3]+1, width = 1 ) ) )
  E1 = 0
  E2 = 0
  if( length( ta[[1]] ) == 1 )
  {
    # The start restriction enzyme must have been Csp6I
    E1 = 1
  }
  else if( length( tb[[1]] ) == 1 )
  {
    # The start restriction enzyme must have been NsiI
    E1 = 2
  }
  if( length( td[[1]] ) == 1 )
  {
    # The start restriction enzyme must have been Csp6I
  }
}

```

```

    E2 = 1
}
else if( length( tc[[1]] ) == 1 )
{
    # The start restriction enzyme must have been NsiI
    E2 = 2
}
totalcounter = totalcounter + 1
if( E1 != E2 )
{
    diffcounter = diffcounter + 1
}
}
}

cat( "diffcounter ", diffcounter )
cat( "totalcounter ", totalcounter )
cat( "diffcounter/totalcounter ", diffcounter/totalcounter )

```

10.13 fastqfilter.py

This custom python program filters paired-read fastq files with arguments for the minimum acceptable phred value for any base and the minimum average phred value.

```

# fastq_filter
# Discards all reads with a single base quality less than filter1
# Discards all reads where the average quality is less than filter2
# Only works with paired end reads
#
# By Jason Smith, last modified 27/11/2019 (d/m/y)

import sys
import gzip
import os
import time
import argparse

parser = argparse.ArgumentParser( description = 'Filter fastq files' )
parser.add_argument( 'r1', help = 'Read 1 file in fastq format' )
parser.add_argument( 'r2', help = 'Read 2 file in fastq format' )
parser.add_argument( 'outdir', help = 'output directory (needs / at the end)' )
args = parser.parse_args()
filename1 = args.r1
filename2 = args.r2
outdir = args.outdir

f1 = open( filename1, 'r' )
f2 = open( filename2, 'r' )

countreads = 0
countkept = 0
countids = [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
countwc = 0

keptfile1 = 'kept.' + filename1[filename1.rindex('/')+1:]
keptfile2 = 'kept.' + filename2[filename2.rindex('/')+1:]
discardedfile1 = 'discarded.' + filename1[filename1.rindex('/')+1:]
discardedfile2 = 'discarded.' + filename2[filename2.rindex('/')+1:]

starttime = time.time()
of1 = open( outdir + keptfile1, "w" )
of2 = open( outdir + keptfile2, "w" )

#filter value needs 33 added for phred33 standard
filter1 = 5 + 33
filter2 = 35 + 33
t1 = ''

```

```

t2 = ''

for x in range( 1, 100000000 ):
    h1 = f1.readline()
    s1 = f1.readline()
    p1 = f1.readline()
    q1 = f1.readline()
    if q1 == "":
        print( "Breaking out of loop" )
        break
    h2 = f2.readline()
    s2 = f2.readline()
    p2 = f2.readline()
    q2 = f2.readline()
    countreads += 1
    #filter reads that have a single base in read 2 less than filter1
    if( all( [ ord( qt2 ) > filter1 for qt2 in q2.strip() ] ) ):
        #filter reads that have a single base in read 1 less than filter1
        if( all( [ ord( qt1 ) > filter1 for qt1 in q1.strip() ] ) ):
            #filter reads that have an average less than filter2
            if( sum( [ ord( c ) for c in ( q1.strip() + q2.strip() ) ] ) / 300 > filter2 ):
                of1.write( h1 + s1 + p1 + q1 )
                of2.write( h2 + s2 + p2 + q2 )
                countkept += 1

of1.close()
of2.close()
f1.close()
f2.close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )
print( "Total reads = %d" % countreads )
print( "Kept reads = %d" % countkept )

```

10.14 Barcodes.py

This custom python program does what demultiplex.py from the epiGBS bioinformatics pipeline should do. It determines the sample identity based on the barcode, works out if the read is Watson or Crick orientation, and reports bisulphite conversion rate based on the number of reads unmethylated at both ends of the Watson/Crick identifying cytosine. The generated files follow the file format required for the epiGBS pipeline and files need to be modified to be ready for Bismark.

```

# This program does what demultiplex.py in the epiGBS suite should do
# It determines the sample identity based on the barcode
# It determines if the read is a Watson or a Crick read
# It determines the conversion rate of unmethylated cytosine by counting instances where both
#   Cs in the adapter remain unchanged
#
# Jason Smith
# Last modified 27/11/2019 (d/m/y)

import sys
import gzip
import os
import time
import argparse

parser = argparse.ArgumentParser( description = 'Filter fastq files' )

```



```

if ( ( s1[ 7:8 ] == "C" ) & ( s2[ 7:8 ] == "T" ) ):
    countCT[ bcidn.get( bcsc ) ] = countCT[ bcidn.get( bcsc ) ] + 1
    #write to file output
    #The header consists of
    #1. The header from the fastq file up to the first space
    #2. BL:Z is left_bc (barcode, watson/crick identifier = Y, enzyme seq)
    #3. BR:Z is right_bc (barcode, watson/crick identifier = Y, enzyme seq)
    #4. RG:Z is the chip id, seq lane and sample ID, RG_id
    #5. ML:i is the left mismatch
    #6. MR:i is the right mismatch
    #7. ST:Z is the strand (watson or crick)
    #8. RN:Z is the wobble bases for de-duplication
    header = h1.split( ' ' )[0] + '\t' +\
        'BL:Z:' + s1[3:7] + 'YTAC' + '\t' +\
        'BR:Z:' + s2[3:7] + 'YTGCAT' + '\t' +\
        'RG:Z:' + chipid + id + '\t' +\
        'ML:i:0\tMR:i:0\t' +\
        'ST:Z:Crick\t' +\
        'RN:Z:' + s1[0:3] + '_' + s2[0:3]
    fo1.write( header + '\n' +\
        s1[ 8: ] +\
        p1 +\
        q1[ 8: ] )
    fo2.write( header + '\n' +\
        s2[ 8: ] +\
        p2 +\
        q2[ 8: ] )
elif ( ( s1[ 7:8 ] == "T" ) & ( s2[ 7:8 ] == "C" ) ):
    countTC[ bcidn.get( bcsc ) ] = countTC[ bcidn.get( bcsc ) ] + 1
    #write to file output
    header = h1.split( ' ' )[0] + '\t' +\
        'BL:Z:' + s1[3:7] + 'YTAC' + '\t' +\
        'BR:Z:' + s2[3:7] + 'YTGCAT' + '\t' +\
        'RG:Z:' + chipid + id + '\t' +\
        'ML:i:0\tMR:i:0\t' +\
        'ST:Z:Watson\t' +\
        'RN:Z:' + s1[0:3] + '_' + s2[0:3]
    fo1.write( header + '\n' +\
        s1[ 8: ] +\
        p1 +\
        q1[ 8: ] )
    fo2.write( header + '\n' +\
        s2[ 8: ] +\
        p2 +\
        q2[ 8: ] )
elif ( ( s1[ 7:8 ] == "C" ) & ( s2[ 7:8 ] == "C" ) ):
    countCC[ bcidn.get( bcsc ) ] = countCC[ bcidn.get( bcsc ) ] + 1
else:
    countNN[ bcidn.get( bcsc ) ] = countNN[ bcidn.get( bcsc ) ] + 1
countreads = countreads + 1
if( countreads % 100000 == 0):
    print( countreads )

fo1.close()
fo2.close()

outl = open( summaryfile, 'w' )
outl.write( 'Total reads processed\n' )
outl.write( str( countreads ).strip( '[' ] ) + '\n' )
outl.write( ' ,L1,L2,L3,L4,L5,L6,L7,L8,R1,R2,R3,R4,R5,R6,R7,R8\n' )
outl.write( 'countids,' + str( countids ).strip( '[' ] ) + '\n' )
outl.write( 'countCT,' + str( countCT ).strip( '[' ] ) + '\n' )
outl.write( 'countTC,' + str( countTC ).strip( '[' ] ) + '\n' )
outl.write( 'countCC,' + str( countCC ).strip( '[' ] ) + '\n' )
outl.write( 'countNN,' + str( countNN ).strip( '[' ] ) + '\n' )
outl.close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )
print( "Total reads = %d" % countreads )

f1.close()
f2.close()

```

10.15 WobbleIndexingToPickle.py

This custom python program takes the result from Barcodes.py and uses the pickle function to save the information sorted by barcode UMI (previously called the wobble) combination and sample combination. This is later used by WobblePickleToFiles.py and dedup.py to remove PCR duplicates.

```
import sys
import gzip
import os
import time
import pickle
import argparse

# This gets rid of PCR duplicates in 3 stages.
# WobbleIndexingToPickle__.py (This program) Finds the locations of reads and assigns it to
# the relevant sample and wobble combinatin
# WobblePickleToFiles__.py Saves the reads in the relevant files, this saves a lot of time
# opening and closing files
# dedup.py which uses dedupcy.pyx to compare reads and remove any duplicates
#
# Jason Smith
# Last modified 26/11/2019 (d/m/y)

parser = argparse.ArgumentParser( description = 'Finds the index of reads' )
parser.add_argument( 'r1in', help = 'Read 1 file in .fq format' )
parser.add_argument( 'r2in', help = 'Read 2 file in .fq format' )
parser.add_argument( 'outdir', help = 'Output directory (including start of filename if
multiple)' )
args = parser.parse_args()

filename1 = args.r1in
filename2 = args.r2in
outdir = args.outdir

print( "Starting" )
f1 = open( filename1, 'r' )
f2 = open( filename2, 'r' )

L1i = [ [] for i in range( 4096 ) ]
L2i = [ [] for i in range( 4096 ) ]
L3i = [ [] for i in range( 4096 ) ]
L4i = [ [] for i in range( 4096 ) ]
L5i = [ [] for i in range( 4096 ) ]
L6i = [ [] for i in range( 4096 ) ]
L7i = [ [] for i in range( 4096 ) ]
L8i = [ [] for i in range( 4096 ) ]
R1i = [ [] for i in range( 4096 ) ]
R2i = [ [] for i in range( 4096 ) ]
R3i = [ [] for i in range( 4096 ) ]
R4i = [ [] for i in range( 4096 ) ]
R5i = [ [] for i in range( 4096 ) ]
R6i = [ [] for i in range( 4096 ) ]
R7i = [ [] for i in range( 4096 ) ]
R8i = [ [] for i in range( 4096 ) ]

countsaves = 0

countreads = 0
approxsizecalc = 1

starttime = time.time()
testing = 0

currentindex1 = 0
currentindex2 = 0
```

```

bpdict = { 'A':0, 'C':1, 'G':2, 'T':3 }
kept = 0

while 1:
    h1 = f1.readline()
    s1 = f1.readline()
    p1 = f1.readline()
    q1 = f1.readline()
    h2 = f2.readline()
    s2 = f2.readline()
    p2 = f2.readline()
    q2 = f2.readline()
    if q1 == "":
        print( "Breaking out of loop" )
        break
    if( len( t[ 7 ] ) == 12 ):
        bc = t[ 3 ][ len( t[ 3 ] ) - 2 : len( t[ 3 ] ) ]
        wob = t[ 7 ][ len( t[ 7 ] ) - 7 : len( t[ 7 ] ) ]
        wob = wob.strip( ":" )
        wob = wob[0:3] + wob[4:7]
        wobv = 0
        mul = 0
        for bp in wob:
            wobv += ( bpdict.get( bp, 10000 ) )*( 4**mul )
            mul += 1
        if( wobv < 4096 ):
            kept += 1
            if( bc == "L1" ):
                L1i[ wobv ].append( currentindex1 )
                L1i[ wobv ].append( currentindex2 )
            elif( bc == "L2" ):
                L2i[ wobv ].append( currentindex1 )
                L2i[ wobv ].append( currentindex2 )
            elif( bc == "L3" ):
                L3i[ wobv ].append( currentindex1 )
                L3i[ wobv ].append( currentindex2 )
            elif( bc == "L4" ):
                L4i[ wobv ].append( currentindex1 )
                L4i[ wobv ].append( currentindex2 )
            elif( bc == "L5" ):
                L5i[ wobv ].append( currentindex1 )
                L5i[ wobv ].append( currentindex2 )
            elif( bc == "L6" ):
                L6i[ wobv ].append( currentindex1 )
                L6i[ wobv ].append( currentindex2 )
            elif( bc == "L7" ):
                L7i[ wobv ].append( currentindex1 )
                L7i[ wobv ].append( currentindex2 )
            elif( bc == "L8" ):
                L8i[ wobv ].append( currentindex1 )
                L8i[ wobv ].append( currentindex2 )
            elif( bc == "R1" ):
                R1i[ wobv ].append( currentindex1 )
                R1i[ wobv ].append( currentindex2 )
            elif( bc == "R2" ):
                R2i[ wobv ].append( currentindex1 )
                R2i[ wobv ].append( currentindex2 )
            elif( bc == "R3" ):
                R3i[ wobv ].append( currentindex1 )
                R3i[ wobv ].append( currentindex2 )
            elif( bc == "R4" ):
                R4i[ wobv ].append( currentindex1 )
                R4i[ wobv ].append( currentindex2 )
            elif( bc == "R5" ):
                R5i[ wobv ].append( currentindex1 )
                R5i[ wobv ].append( currentindex2 )
            elif( bc == "R6" ):
                R6i[ wobv ].append( currentindex1 )
                R6i[ wobv ].append( currentindex2 )
            elif( bc == "R7" ):
                R7i[ wobv ].append( currentindex1 )
                R7i[ wobv ].append( currentindex2 )
            elif( bc == "R8" ):
                R8i[ wobv ].append( currentindex1 )
                R8i[ wobv ].append( currentindex2 )
            else:

```

```

        fdiscard = open( outdir + 'discarded' + filename1[ filename1.rfind( '/' ) + 1: ],
'a' )
        fdiscard.write( h1 + s1 + p1 + q1 )
        fdiscard.close()
currentindex1 += len(h1) + len(s1) + len(p1) + len(q1)
currentindex2 += len(h2) + len(s2) + len(p2) + len(q2)
countreads = countreads + 1
if ( countreads % 1000000 ) == 0:
'a' )
    fprog = open( outdir + 'progress' + filename1[ filename1.rfind( '/' ) + 1: ] + '.txt',
) + '\n' )
    fprog.write( 'Reads ' + "{:,}".format( countreads ) + ' kept ' + "{:,}".format( kept
) + '\n' )
    fprog.close()
)
if ( countreads % 1000000000 ) == 0:
    print( 'Saving pickle files...' )
    fnp = open( outdir + 'L1i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( L1i, fnp )
    fnp.close()
    fnp = open( outdir + 'L2i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( L2i, fnp )
    fnp.close()
    fnp = open( outdir + 'L3i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( L3i, fnp )
    fnp.close()
    fnp = open( outdir + 'L4i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( L4i, fnp )
    fnp.close()
    fnp = open( outdir + 'L5i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( L5i, fnp )
    fnp.close()
    fnp = open( outdir + 'L6i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( L6i, fnp )
    fnp.close()
    fnp = open( outdir + 'L7i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( L7i, fnp )
    fnp.close()
    fnp = open( outdir + 'L8i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( L8i, fnp )
    fnp.close()
    fnp = open( outdir + 'R1i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( R1i, fnp )
    fnp.close()
    fnp = open( outdir + 'R2i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( R2i, fnp )
    fnp.close()
    fnp = open( outdir + 'R3i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( R3i, fnp )
    fnp.close()
    fnp = open( outdir + 'R4i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( R4i, fnp )
    fnp.close()
    fnp = open( outdir + 'R5i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( R5i, fnp )
    fnp.close()
    fnp = open( outdir + 'R6i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( R6i, fnp )
    fnp.close()
    fnp = open( outdir + 'R7i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( R7i, fnp )
    fnp.close()
    fnp = open( outdir + 'R8i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
    pickle.dump( R8i, fnp )
    fnp.close()
    countsaves += 1
    L1i = [ [] for i in range( 4096 ) ]
    L2i = [ [] for i in range( 4096 ) ]
    L3i = [ [] for i in range( 4096 ) ]
    L4i = [ [] for i in range( 4096 ) ]
    L5i = [ [] for i in range( 4096 ) ]
    L6i = [ [] for i in range( 4096 ) ]
    L7i = [ [] for i in range( 4096 ) ]
    L8i = [ [] for i in range( 4096 ) ]
    R1i = [ [] for i in range( 4096 ) ]
    R2i = [ [] for i in range( 4096 ) ]
    R3i = [ [] for i in range( 4096 ) ]
    R4i = [ [] for i in range( 4096 ) ]

```

```

R5i = [ [] for i in range( 4096 ) ]
R6i = [ [] for i in range( 4096 ) ]
R7i = [ [] for i in range( 4096 ) ]
R8i = [ [] for i in range( 4096 ) ]

print( 'Saving pickle files...' )
fnp = open( outdir + 'L1i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( L1i, fnp )
fnp.close()
fnp = open( outdir + 'L2i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( L2i, fnp )
fnp.close()
fnp = open( outdir + 'L3i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( L3i, fnp )
fnp.close()
fnp = open( outdir + 'L4i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( L4i, fnp )
fnp.close()
fnp = open( outdir + 'L5i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( L5i, fnp )
fnp.close()
fnp = open( outdir + 'L6i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( L6i, fnp )
fnp.close()
fnp = open( outdir + 'L7i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( L7i, fnp )
fnp.close()
fnp = open( outdir + 'L8i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( L8i, fnp )
fnp.close()
fnp = open( outdir + 'R1i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( R1i, fnp )
fnp.close()
fnp = open( outdir + 'R2i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( R2i, fnp )
fnp.close()
fnp = open( outdir + 'R3i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( R3i, fnp )
fnp.close()
fnp = open( outdir + 'R4i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( R4i, fnp )
fnp.close()
fnp = open( outdir + 'R5i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( R5i, fnp )
fnp.close()
fnp = open( outdir + 'R6i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( R6i, fnp )
fnp.close()
fnp = open( outdir + 'R7i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( R7i, fnp )
fnp.close()
fnp = open( outdir + 'R8i_' + str( countsaves ).zfill(6) + '.pickle', 'wb' )
pickle.dump( R8i, fnp )
fnp.close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )
print( "Total reads = %d" % countreads )

f1.close()
f2.close()

```

10.16 WobblePickleToFiles.py

This custom python program reads in the pickle files created by WobbleIndexingToPickel.py and then saves as a separate file for each UMI

(previously called the wobble) combination and each sample. The output of this program is used by dedup.py to remove PCR duplicates.

```

# This gets rid of PCR duplicates in 3 stages.
# WobbleIndexingToPickle__.py (This program) Finds the locations of reads and assigns it to
# the relevant sample and wobble combinatin
# WobblePickleToFiles__.py (This program) Saves the reads in the relevant files, this saves a
# lot of time opening and closing files
# dedup.py which uses dedupcy.pyx to compare reads and remove any duplicates
#
# Jason Smith
# Last modified 26/11/2019 (d/m/y)

import sys
import gzip
import os
import time
import pickle
import argparse

$HOME/Share/epiGBS_89_demux_new/test.2.fq          $HOME/Share/epiGBS_90_wobble/
$HOME/Share/epiGBS_90_wobble/test L1

parser = argparse.ArgumentParser( description = 'Uses read indexes to generate files for each
wobble combination' )
parser.add_argument( 'r1in', help = 'Read 1 file in .fq format' )
parser.add_argument( 'r2in', help = 'Read 2 file in .fq format' )
parser.add_argument( 'indir', help = 'Input directory for pickle files (including start of
filename if multiple)' )
parser.add_argument( 'outdir', help = 'Output directory (including start of filename if
multiple)')
parser.add_argument( 'samp', help = 'Sample name (L1, L2, ... R8 )' )

args      = parser.parse_args()
filename1 = args.r1in
filename2 = args.r2in
indir     = args.indir
outdir   = args.outdir
samp     = args.samp

f1 = open( filename1, 'r' )
f2 = open( filename2, 'r' )

starttime = time.time()

samples = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5',
'R6', 'R7', 'R8' ]

fp = open( indir + samp + 'i_000000.pickle', 'rb' )
ind = pickle.load( fp )
fp.close()
for c in range( 0, 4096 ):
    fo1 = open( outdir + samp + str( c ).zfill( 4 ) + '.1.fq', 'w' )
    fo2 = open( outdir + samp + str( c ).zfill( 4 ) + '.2.fq', 'w' )
    for i in range( 0, len( ind[ c ] ) ):
        if ( i % 2 ) == 0 ):
            t = f1.seek( ind[c][i] )
            fo1.write( f1.readline() )
            fo1.write( f1.readline() )
            fo1.write( f1.readline() )
            fo1.write( f1.readline() )
        else:
            t = f2.seek( ind[c][i] )
            fo2.write( f2.readline() )
            fo2.write( f2.readline() )
            fo2.write( f2.readline() )
            fo2.write( f2.readline() )
    fo1.close()
    fo2.close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )

```

```
f1.close()
f2.close()
```

10.17 dedup.py

This custom python program takes the output files from WobblePickleToFiles.py and removes any PCR duplicates, keeping only the read with the highest phred quality score. This program utilises 8 CPU cores and requires the cython compiled program dedupcy.pyx.

```
from multiprocessing import Pool, TimeoutError
import time
import os
import dedupcy

indir = "/scratch/user/smit1191/epiGBS_90_wobble/joined/"
outdir = "/scratch/user/smit1191/epiGBS_91_dedup/"
samples = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5',
'R6', 'R7', 'R8' ]
logh = open( outdir + 'log.txt', 'a' )
logh.write( 'Starting...\n' )
logh.close()

if __name__ == '__main__':
    # start 8 worker processes
    with Pool( processes = 8 ) as pool:
        counter = 0
        samp = samples[ 0 ]
        dir = os.listdir( indir + samp + '/' )
        dir1 = [ x for x in dir if ".1.fq" in x ]
        dir2 = [ samp + '/' + file for file in dir1 ]
        logh = open( outdir + samp + 'log.txt', 'a' )
        logh.write( 'Files found in ' + indir + samp + '/\n' + str( dir2 ) + '\n' )
        logh.close()
        dir2 = []
        for c in range( 0, 10 ):
            dir2.append( samp + '/' + samp + str( c ).zfill( 4 ) + '.1.fq.gz' + '\t' + indir +
'\t' + outdir )
            for i in pool.imap_unordered( dedupcy.f, dir2 ):
                logh = open( outdir + samp + 'log.txt', 'a' )
                logh.write( i + ' ' + str( counter ) + '\n' )
                logh.close()
                counter += 1
```

10.18 dedupcy.pyx

This custom program was written in cython for speed improvements when comparing reads to remove PCR duplicates. It is used by dedup.py to find and remove PCR duplicates.

```
import os
import array
from cpython cimport array
import cython
import gzip
```

```

@cython.boundscheck(False)
cpdef str f( str inputstrs ):
    inputs = inputstrs.split('\t')
    cdef str filein = inputs[0]
    cdef str indir = inputs[1]
    cdef str outdir = inputs[2]
    cdef int fq = 1
    if( filein[-2:] == 'gz' ):
        fq = 0

    if( fq ):
        filein1 = indir + filein
        filein2 = indir + filein[:-4] + '2.fq'
        fileout1 = outdir + filein
        fileout2 = outdir + filein[:-4] + '2.fq'
        logh = open( outdir + 'log.txt', 'a' )
        logh.write( 'Starting ' + filein1 + '\n' )
        logh.close()
        fh = open( filein1, 'r' )
        r1buf = fh.read().splitlines()
        fh.close()
        fh = open( filein2, 'r' )
        r2buf = fh.read().splitlines()
        fh.close()
    else:
        filein1 = indir + filein
        filein2 = indir + filein[:-7] + '2.fq.gz'
        fileout1 = outdir + filein
        fileout2 = outdir + filein[:-7] + '2.fq.gz'
        logh = open( outdir + 'log.txt', 'a' )
        logh.write( 'Starting ' + filein1 + '\n' )
        logh.close()
        fh = gzip.open( filein1, 'r' )
        r1buf = fh.read().decode( encoding = 'UTF-8' ).splitlines()
        fh.close()
        fh = gzip.open( filein2, 'r' )
        r2buf = fh.read().decode( encoding = 'UTF-8' ).splitlines()
        fh.close()

    cdef int cont = 1
    cdef ccount = 0
    cdef long r1len = len( r1buf )
    cdef list nrlwi = []
    cdef list nr2wi = []
    cdef list nrlci = []
    cdef list nr2ci = []
    while( ccount < r1len ):
        if( 'Watson' in r1buf[ ccount ] ):
            nrlwi.append( r1buf[ ccount ] )
            nrlwi.append( r1buf[ ccount + 1 ] )
            nrlwi.append( r1buf[ ccount + 2 ] )
            nrlwi.append( r1buf[ ccount + 3 ] )
            nr2wi.append( r2buf[ ccount ] )
            nr2wi.append( r2buf[ ccount + 1 ] )
            nr2wi.append( r2buf[ ccount + 2 ] )
            nr2wi.append( r2buf[ ccount + 3 ] )
        else:
            nrlci.append( r1buf[ ccount ] )
            nrlci.append( r1buf[ ccount + 1 ] )
            nrlci.append( r1buf[ ccount + 2 ] )
            nrlci.append( r1buf[ ccount + 3 ] )
            nr2ci.append( r2buf[ ccount ] )
            nr2ci.append( r2buf[ ccount + 1 ] )
            nr2ci.append( r2buf[ ccount + 2 ] )
            nr2ci.append( r2buf[ ccount + 3 ] )
        ccount += 4

    cdef list r1w = []
    cdef list r1c = []
    cdef list r2w = []
    cdef list r2c = []
    cdef list nrlw = []
    cdef list nr1c = []
    cdef list nr2w = []
    cdef list nr2c = []
    cdef int l

```



```

cdef list r1wh = []
cdef list r1wf = []
cdef array.array r1wq = array.array( 'i', [] )
cdef list r2wh = []
cdef list r2wf = []
cdef array.array r2wq = array.array( 'i', [] )
for l in range( 0, int( len( nrlwi )/4 ) ):
    if( min( [ ord(x)-33 for x in nrlwi[ 1*4+3 ] ] + [ ord(x)-33 for x in nr2wi[ 1*4+3 ] ]
) > 11 ):
        nrlw.append( nrlwi[ 1*4 ] )
        nrlw.append( nrlwi[ 1*4+1 ] )
        nrlw.append( nrlwi[ 1*4+2 ] )
        nrlw.append( nrlwi[ 1*4+3 ] )
        nr2w.append( nr2wi[ 1*4 ] )
        nr2w.append( nr2wi[ 1*4+1 ] )
        nr2w.append( nr2wi[ 1*4+2 ] )
        nr2w.append( nr2wi[ 1*4+3 ] )
    for l in range( 0, int( len( nrlci )/4 ) ):
        if( min( [ ord(x)-33 for x in nrlci[ 1*4+3 ] ] + [ ord(x)-33 for x in nr2ci[ 1*4+3 ] ]
) > 11 ):
            nrlc.append( nrlci[ 1*4 ] )
            nrlc.append( nrlci[ 1*4+1 ] )
            nrlc.append( nrlci[ 1*4+2 ] )
            nrlc.append( nrlci[ 1*4+3 ] )
            nr2c.append( nr2ci[ 1*4 ] )
            nr2c.append( nr2ci[ 1*4+1 ] )
            nr2c.append( nr2ci[ 1*4+2 ] )
            nr2c.append( nr2ci[ 1*4+3 ] )
r2wi[l].letter_annotations['phred_quality'] > 11 ):
r2ci[l].letter_annotations['phred_quality'] > 11 ):
cdef list r1wo = []
cdef list r1co = []
cdef list r2wo = []
cdef list r2co = []
cdef array.array r1ws = array.array( 'i', [] )
cdef array.array r2ws = array.array( 'i', [] )
cdef array.array r1cs = array.array( 'i', [] )
cdef array.array r2cs = array.array( 'i', [] )
cdef list r1wt = []
cdef list r2wt = []
cdef list r1ct = []
cdef list r2ct = []
cdef long i
cdef long j
cdef long k
cdef long matchc
cdef long keepind
cdef long currentmax
cdef long currentrefind = 0
cdef array.array indsw = array.array( 'l', range( 0, ( len( nrlw )/4 ) ) )
cdef int stopping = 0
while( stopping == 0 ):
    r1ws = array.array( 'i', [0] * ( len( nrlw )/4 ) )
    seqref = nrlw[currentrefind*4+1]
    for i in indsw:
        matchc = 0
        for j in range( 0, len( seqref ) ):
            if( seqref[j] == nrlw[i*4+1][j] ) :
                matchc += 1
        r1ws[ i ] = matchc
    r1wsind = [ ind for ind,v in enumerate( r1ws ) if v > 135 ]
    r1wsind.sort( reverse = True )
    r1wt = []
    r2wt = []
    for k in r1wsind:
        r1wt.append( nrlw[ k*4 ] )
        r1wt.append( nrlw[ k*4+1 ] )
        r1wt.append( nrlw[ k*4+2 ] )
        r1wt.append( nrlw[ k*4+3 ] )
        r2wt.append( nr2w[ k*4 ] )
        r2wt.append( nr2w[ k*4+1 ] )
        r2wt.append( nr2w[ k*4+2 ] )
        r2wt.append( nr2w[ k*4+3 ] )
        del indsw[ indsw.index( k ) ]
    if( not indsw ):
        stopping = 1
    elif( currentrefind == max( indsw ) ):

```

```

        stopping = 1
        currentrefind += 1 #Stop it from running next time
    else:
        currentrefind = min( indsw )
min( indsw ) ) + ', max( indsw ) = ' + str( max( indsw ) ) )
        keepind = 0
        currentmax = 0
        for l in range( 0, len( r1wt )/4 ):
            tmin = min( [ ord(x)-33 for x in r1wt[ l*4+3 ] ] + [ ord(x)-33 for x in r2wt[
l*4+3 ] ] )
            if currentmax < tmin:
                keepind = l
                currentmax = tmin
currentmax, keepind ) )
        r1wo.append( r1wt[ keepind*4 ] )
        r1wo.append( r1wt[ keepind*4+1 ] )
        r1wo.append( r1wt[ keepind*4+2 ] )
        r1wo.append( r1wt[ keepind*4+3 ] )
        r2wo.append( r2wt[ keepind*4 ] )
        r2wo.append( r2wt[ keepind*4+1 ] )
        r2wo.append( r2wt[ keepind*4+2 ] )
        r2wo.append( r2wt[ keepind*4+3 ] )
if( fq ):
    fout1 = open( fileout1, 'w' )
    for item in r1wo:
        fout1.write( item + '\n' )
    fout1.close()
    fout2 = open( fileout2, 'w' )
    for item in r2wo:
        fout2.write( item + '\n' )
    fout2.close()
else:
    buffer = ''
    for item in r1wo:
        buffer += item + '\n'
    fout1 = gzip.open( fileout1, 'w' )
    fout1.write( buffer.encode( encoding = 'UTF-8' ) )
    fout1.close()
    buffer = ''
    for item in r2wo:
        buffer += item + '\n'
    fout2 = gzip.open( fileout2, 'w' )
    fout2.write( buffer.encode( encoding = 'UTF-8' ) )
    fout2.close()

cdef array.array indsc = array.array( 'l', range( 0, ( len( nr1c )/4 ) ) )
stopping = 0
currentrefind = 0
while( stopping == 0 ):
    rlcs = array.array( 'i', [0] * ( len( nr1c )/4 ) )
    seqref = nr1c[currentrefind*4+1]
    for i in indsc:
        #print( "i %d" % i )
        matchc = 0
        for j in range( 0, len( seqref ) ):
            #print( "j %d" % j )
            if( seqref[j] == nr1c[i*4+1][j] ) :
                matchc += 1
            #print( "i %d, matchc %d" % ( i, matchc ) )
        rlcs[ i ] = matchc
    rlcsind = [ ind for ind,v in enumerate( rlcs ) if v > 135 ]
    rlcsind.sort( reverse = True )
    rlct = []
    r2ct = []
    for k in rlcsind:
        rlct.append( nr1c[ k*4 ] )
        rlct.append( nr1c[ k*4+1 ] )
        rlct.append( nr1c[ k*4+2 ] )
        rlct.append( nr1c[ k*4+3 ] )
        r2ct.append( nr2c[ k*4 ] )
        r2ct.append( nr2c[ k*4+1 ] )
        r2ct.append( nr2c[ k*4+2 ] )
        r2ct.append( nr2c[ k*4+3 ] )
        del indsc[ indsc.index( k ) ]
    if( not indsc ):
        stopping = 1
    elif( currentrefind == max( indsc ) ):

```

```

        stopping = 1
        currentrefind += 1 #Stop it from running next time
    else:
        currentrefind = min( indsc )
    keepind = 0
    currentmax = 0
    for l in range( 0, len( rlct )/4 ):
        tmin = min( [ ord(x)-33 for x in rlct[ l*4+3 ] ] + [ ord(x)-33 for x in r2ct[
l*4+3 ] ] )
        if currentmax < tmin:
            keepind = l
            currentmax = tmin
    rlco.append( rlct[ keepind*4 ] )
    rlco.append( rlct[ keepind*4+1 ] )
    rlco.append( rlct[ keepind*4+2 ] )
    rlco.append( rlct[ keepind*4+3 ] )
    r2co.append( r2ct[ keepind*4 ] )
    r2co.append( r2ct[ keepind*4+1 ] )
    r2co.append( r2ct[ keepind*4+2 ] )
    r2co.append( r2ct[ keepind*4+3 ] )
if( fq ):
    fout1 = open( fileout1, 'a' )
    for item in rlco:
        fout1.write( item + '\n' )
    fout1.close()
    fout2 = open( fileout2, 'a' )
    for item in r2co:
        fout2.write( item + '\n' )
    fout2.close()
else:
    buffer = ''
    for item in rlco:
        buffer += item + '\n'
    fout1 = gzip.open( fileout1, 'a' )
    fout1.write( buffer.encode( encoding = 'UTF-8' ) )
    fout1.close()
    buffer = ''
    for item in r2co:
        buffer += item + '\n'
    fout2 = gzip.open( fileout2, 'a' )
    fout2.write( buffer.encode( encoding = 'UTF-8' ) )
    fout2.close()
    retstr = rlwo[0][-7:] + ' ' + fileinl[-11:] + ', Input length = ' + str( len( rlbuf )/4 )
+ ', Output length = ' + str( len( rlwo )/4 + len( rlco )/4 )
    return retstr

```

10.19 epiGBS2Bismark.py

This custom python program takes the library which has been assembled by PEAR and converts it in to a format that can be used by Bismark.

```

# This converts a library assembled by PEAR in to something Bismark can use
# 1st argument is the unassembled input read 1 (forward) from pear
# 2nd argument is the unassembled input read 2 (reverse) from pear
# 3rd argument is the assembled input read from pear
# 4th argument is the output directory and start of filename
#
# IMPORTANT NOTE, pear reverse complements the second unassembled read
#
# Watson R1 reads get w1 added to the header and reads are unmodified and sent to .1.fq
# Watson R2 reads get w1 added to the header and reads are reverse complemented (because pear
# already reverse complemented it) and sent to .2.fq
# Crick R1 reads get c1 added to the header and reads are sent to .2.fq
# Crick R2 reads get c1 added to the header and reads reverse complemented (because pear
# already reverse complemented it) and sent to .1.fq
# Watson assembled reads get wa added to the header and are unmodified and sent to .a.fq (a
# for assembled)
# Crick assembled reads get ca added to the header and are reverse complemented and sent to
# .a.fq (a for assembled)
#

```

```

# Jason Smith
# Last modified 08/12/2019 (d/m/y)

import sys
import gzip
import os
import time
import argparse
import string

parser = argparse.ArgumentParser( description = 'Finds the index of reads' )
parser.add_argument( 'r1in', help = 'Unassembled read 1 file in .fq or .fq.gz format' )
parser.add_argument( 'r2in', help = 'Unassembled read 2 file in .fq or .fq.gz format' )
parser.add_argument( 'rain', help = 'Assembled file in .fq or .fq.gz format' )
parser.add_argument( 'outfile', help = 'Output directory and start of file name (.1.fq, .2.fq
or .a.fq will be added)' )
args = parser.parse_args()

filename1 = args.r1in
filename2 = args.r2in
filenameea = args.rain
filenameeo = args.outfile

if( filename1[-2:] == 'gz' ):
    f1 = gzip.open( filename1, 'rb' )
    f2 = gzip.open( filename2, 'rb' )
    fa = gzip.open( filenameea, 'rb' )
else:
    f1 = open( filename1, 'r' )
    f2 = open( filename2, 'r' )
    fa = open( filenameea, 'r' )

fo1 = open( filenameeo + '.1.fq', 'w' )
fo2 = open( filenameeo + '.2.fq', 'w' )

countreads = 0
starttime = time.time()
testing = 0

trans = string.maketrans( 'ATGC', 'TACG' )

#NOTE pear already reverse complemented read 2
while 1:
    h1 = f1.readline()
    s1 = f1.readline()
    p1 = f1.readline()
    q1 = f1.readline()
    h2 = f2.readline()
    s2 = f2.readline()
    p2 = f2.readline()
    q2 = f2.readline()
    if q1 == "":
        print( "Finished R1" )
        break
    if( h1.find( 'Watson' ) > 0 ):
        fo1.write( '@w1' + h1[1:] + s1 + p1 + q1 )
        fo2.write( '@w1' + h2[1:] + s2.strip().translate( trans )[::-1] + '\n' + p2 + q2[
::-1 ].strip() + '\n' )
    else:
        fo1.write( '@c1' + h2[1:] + s2.strip().translate( trans )[::-1] + '\n' + p2 + q2[
::-1 ].strip() + '\n' )
        fo2.write( '@c1' + h1[1:] + s1 + p1 + q1 )
    countreads += 1
fo1.close()
fo2.close()

fo = open( filenameeo + '.a.fq', 'w' )
while 1:
    h1 = fa.readline()
    s1 = fa.readline()
    p1 = fa.readline()
    q1 = fa.readline()
    if q1 == "":
        print( "Finished R assembled" )
        break
    if( h1.find( 'Watson' ) > 0 ):
        fo.write( '@wa' + h1[1:] + s1 + p1 + q1 )

```

```

else:
    fo.write( '@ca' + h1[1:] + s1.strip().translate( trans )[::-1] + '\n' + p1 + q1[ ::-1
1 ].strip() + '\n' )
    countreads += 1
fo.close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )

print( "Total reads = %d" % countreads )

f1.close()
f2.close()
fa.close()

```

10.20 splitcov.py

This is a custom pypy program that takes the .cov file output from Bismark and splits the file by chromosome.

```

# This splits .cov.gz files by chromosome to be handled later
# Input file has columns separated with \t
#   Chromosome, Position, Position, % methylation, Methylated count, Unmethylated count
# Output files are by chromosome and has columns separated with \t
#   Chromosome, Position, Methylated count, Unmethylated count
#
# Jason Smith
# Last modified 09/12/2019 (d/m/y)

import sys
import gzip
import os
import time
import glob

dir_in = 'C:/BaseSpace/epiGBS_98_bismark/important'
dir_out = 'C:/BaseSpace/epiGBS_98_bismark/split'
dir_in = '/home/jason/Share/epiGBS_98_bismark/important'
dir_out = '/home/jason/Share/epiGBS_98_bismark/split'

chromosomes = [ 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn', 'Mt',
'Pt' ]

files_in = glob.glob( dir_in + '/*.cov.gz' )

for file_in_temp in files_in:
    file_in = file_in_temp.replace( '\\', '/' )
    file_base = dir_out + '/' + file_in[ file_in.rfind( '/' ) + 1 : file_in.rfind( '.bismark'
) ]
    print( file_in )
    print( file_base )

    filenames = []
    outfiles = []

    sort_chr = [[]]
    for i in range( 0, len( chromosomes ) ):
        sort_chr.append( [] )
        outfiles.append( open( file_base + '.' + chromosomes[ i ] + '.tsv', 'w' ) )

    countlines = 0
    starttime = time.time()
    testing = 0
    unknown_chr = 0

    fi = gzip.open( file_in, 'rb' )

    while( True ):
        line = fi.readline()
        if( line == '' ):

```

```

        break
    ll = line.split( '\t' )
    try:
        outfile[ chromosomes.index( ll[ 0 ] ) ].write( ll[0] + '\t' + ll[1] + '\t' +
ll[4] + '\t' + ll[5] )
    except: # Exception:
        #There is an error here
        unknown_chr += 1
        countlines += 1
    if( countlines % 10000000 == 0 ):
        print( "Current line = %d" % countlines )
fi.close()

for i in range( 0, len( chromosomes ) ):
    outfile[ i ].close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )
print( "Total lines = %d" % countlines )
print( "Lines with unknown chromosome labels = %d" % countlines )

```

10.21 mergcov.py

This is a custom pypy program that takes the output from splitcov.py and merges by sample resulting in .tsv files that have all samples per chromosome with number of methylated reads and total reads for cytosines.

```

import sys
import gzip
import os
import time

# This merges .tsv from splitcov01.py files by chromosome to be used in R
#
# Jason Smith
# Last modified 10/12/2019 (d/m/y)

dir_in = 'C:/BaseSpace/epiGBS_98_bismark/split'
dir_out = 'C:/BaseSpace/Share/epiGBS_98_bismark/merged'
dir_in = '/home/jason/Share/epiGBS_98_bismark/split'
dir_out = '/home/jason/Share/epiGBS_98_bismark/merged'

chromosomes = [ 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn', 'Mt',
'Pt' ]
s = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6',
'R7', 'R8' ]
sample = []
for st in s:
    sample.append( st + '.assembled' )
    sample.append( st + '.unassembled_pe' )

for chromosome in chromosomes:
    print( chromosome )
    infiles = []
    outfile = dir_out + '/cov.' + chromosome + '.tsv'
    header = 'position'
    for samp in sample:
        infiles.append( open( dir_in + '/' + samp + '.' + chromosome + '.tsv', 'r' ) )
        header = header + '\t' + samp + '.methylated.count\t' + samp + '.unmethylated.count'

    countlines = 0
    starttime = time.time()
    testing = 0
    unknown_chr = 0
    latest_line = []
    latest_pos = []
    latest_meth = []

```

```

latest_unmeth = []
fo = open( outfile, 'w' )
fo.write( header + '\n' )
# Set up the different identifiers
current_pos = 0
for fi in infiles:
    latest_line.append( fi.readline() )
for i in range( 0, len( infiles ) ):
    line_split = latest_line[ i ].strip().split( '\t' )
    latest_pos.append( int( line_split[ 1 ] ) )
    latest_meth.append( line_split[ 2 ] )
    latest_unmeth.append( line_split[ 3 ] )

while( True ):
    current_pos = min( latest_pos )
    line_build = str( current_pos )
    for i in range( 0, len( latest_pos ) ):
        if( latest_pos[ i ] == current_pos ):
            line_build = line_build + '\t' + latest_meth[ i ] + '\t' + latest_unmeth[ i ]
            latest_line[ i ] = infiles[ i ].readline()
            if( latest_line[ i ] == '' ):
                latest_pos[ i ] = max( latest_pos )
                latest_meth[ i ] = '0'
                latest_unmeth[ i ] = '0'
            else:
                line_split = latest_line[ i ].strip().split( '\t' )
                latest_pos[ i ] = ( int( line_split[ 1 ] ) )
                latest_meth[ i ] = ( line_split[ 2 ] )
                latest_unmeth[ i ] = ( line_split[ 3 ] )
        else:
            line_build = line_build + '\t0\t0'
    fo.write( line_build + '\n' )
    if( latest_line == ['']*len( latest_line ) ):
        break
    #if all lines are empty then go to the next chromosome
fo.close()
endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )
print( "Total lines = %d" % countlines )
print( "Lines with unknown chromosome labels = %d" % countlines )

```

10.22 covcontext.py

This is a pypy (or python) program that takes the files generated in mergecov.py and references against the genome to find the context of each cytosine.

```

# This looks at the context of cytosines in the .tsv from mergecov01.py and saves to
# <dir_out>/context.<chromosome>.tsv
#
# Jason Smith
# Last modified 10/12/2019 (d/m/y)

import sys
import gzip
import os
import time

dir_in = '/home/jason/Share/epiGBS_98_bismark/merged'
dir_out = '/home/jason/Share/epiGBS_98_bismark/merged'
genome = '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-05/Hordeum_vulgare.IBSC_v2.dna.toplevel.fa.gz'

chromosomes = [ 'chr1H', 'Mt', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' ]
s = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6', 'R7', 'R8' ]

fg = gzip.open( genome, 'rb' )

for chromosome in chromosomes:
    print( 'Searching genome for ' + chromosome )

```

```

while( True ):
    gline = fg.readline()
    if( gline == '' ):
        print( 'Reached end of genome, chromosomes must have been out of order, sorry, try
reordering' )
        print( 'zcat <genome.fa.gz> | grep \>\>' )
        sys.exit()
    if( gline.find( '>' + chromosome ) == 0 ):
        print( 'Found ' + chromosome )
        gline = fg.readline().strip()
        chr_line_start = 1
        chr_line_end = chr_line_start + len( gline ) - 1
        break
infile = dir_in + '/cov.' + chromosome + '.tsv'
outfile = dir_out + '/context.' + chromosome + '.tsv'

countlines = 0
starttime = time.time()
testing = 0
unknown_chr = 0
fi = open( infile, 'r' )
fo = open( outfile, 'w' )
header
'position\tcontext_1.CGfwd_2.CGrev_3.CHGfwd_4.CHGrev_5.CHHfwd_6.CHHrev_7.N_8.CNfwd_9.CNrev_10.
CHNfwd_11.CHNrev_12.Error\n'
fo.write( header )

fi.readline() #get rid of the header line

while( True ):
    lt = fi.readline()
    if( lt == '' ):
        break
    line = lt.split( '\t' )
    position = int( line[0] )
    # Need to get 5 bases from the genome, 2 before, the base in question and 2 after
    wanted_start = position - 2
    wanted_end = position + 2
    while( wanted_end > chr_line_end ):
        # need to go forward until the desired section of the genome is in gline_p
        # and gline
        gline_p = gline
        chr_line_start_p = chr_line_start
        chr_line_end_p = chr_line_end
        gline = fg.readline().strip()
        chr_line_start = chr_line_end_p + 1
        chr_line_end = chr_line_start + len( gline ) - 1
    if( wanted_start >= chr_line_start ):
        seq = gline[ ( wanted_start - chr_line_start ) : ( wanted_end - chr_line_start + 1 )
]

else:
    seq = gline_p[ ( wanted_start - chr_line_start_p ) : ] + \
        gline[ 0 : ( wanted_end - chr_line_start + 1 ) ]
    seq = seq.upper()
    # Context notation
    # 1.CGfwd
    # 2.CGrev
    # 3.CHGfwd
    # 4.CHGrev
    # 5.CHHfwd
    # 6.CHHrev
    # 7.N
    # 8.CNfwd
    # 9.CNrev
    # 10.CHNfwd
    # 11.CHNrev
    # 12.Error
    if( seq[ 2 ] == 'C' ):
        #Narrowed it down to forward CG, CHG, CHH, CN, CHN
        if( seq[ 3 ] == 'G' ):
            #It is forward CG
            context = '1'
        elif( ( seq[ 3 ] == 'A' ) or ( seq[ 3 ] == 'C' ) or ( seq[ 3 ] == 'T' ) ):
            #Narrowed it down to forward CHG, CHH or CHN
            if( seq[ 4 ] == 'G' ):
                #It is forward CHG
                context = '3'

```



```

elif( ( seq[ 4 ] == 'A' ) or ( seq[ 4 ] == 'C' ) or ( seq[ 4 ] == 'T' ) ):
    #It is forward CHH
    context = '5'
elif( seq[ 4 ] == 'N' ):
    #It is forward CHN
    context = '10'
else:
    #Otherwise it is an error
    context = '12'
elif( seq[ 3 ] == 'N' ):
    #It is forward CN
    context = '8'
else:
    #Otherwise it is an error
    context = '12'
elif( seq[ 2 ] == 'G' ):
    #Narrowed it down to reverse CG, CHG, CHH, CN, CHN
    if( seq[ 1 ] == 'C' ):
        #It is reverse CG
        context = '2'
    elif( ( seq[ 1 ] == 'A' ) or ( seq[ 1 ] == 'G' ) or ( seq[ 1 ] == 'T' ) ):
        #Narrowed it down to reverse CHG, CHH or CHN
        if( seq[ 0 ] == 'C' ):
            #It is reverse CHG
            context = '4'
        elif( ( seq[ 0 ] == 'A' ) or ( seq[ 0 ] == 'G' ) or ( seq[ 0 ] == 'T' ) ):
            #It is reverse CHH
            context = '6'
        elif( seq[ 0 ] == 'N' ):
            #It is reverse CHN
            context = '11'
        else:
            #Otherwise it is an error
            context = '12'
    elif( seq[ 1 ] == 'N' ):
        #It is reverse CN
        context = '9'
    else:
        #Otherwise it is an error
        context = '12'
elif( seq[ 2 ] == 'N' ):
    context = '7'
else:
    #Otherwise it is an error
    context = '12'
fo.write( line[0] + '\t' + context + '\n' )

fo.close()
fi.close()
endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )
print( "Total lines = %d" % countlines )
print( "Lines with unknown chromosome labels = %d" % countlines )

```

10.23 cov2R.py

This pypy program takes the files generated by covcontext and separates by context.

Each input file generates 3 .tsv output files, one for each of the contexts CG, CHG, and CHH.

```

# This looks at the context of cytosines in the .tsv from mergecov01.py and saves to
# <dir_out>/contect.<chromosome>.tsv
#
# Jason Smith
# Last modified 10/12/2019 (d/m/y)

```

```

import sys
import gzip
import os
import time

dir_in = '/home/jason/Share/epiGBS_98_bismark/merged'
dir_out = '/home/jason/Share/epiGBS_98_bismark/merged'

chromosomes = [ 'chr1H', 'Mt', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' ]
s = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6', \
      'R7', 'R8' ]

for chromosome in chromosomes:
    infile_cov = dir_in + '/cov.' + chromosome + '.tsv'
    infile_context = dir_out + '/context.' + chromosome + '.tsv'
    outfile = dir_out + '/R.' + chromosome + '.tsv'
    outfile_CG = dir_out + '/R.' + chromosome + '.CG.tsv'
    outfile_CHG = dir_out + '/R.' + chromosome + '.CHG.tsv'
    outfile_CHH = dir_out + '/R.' + chromosome + '.CHH.tsv'

    countlines = 0
    starttime = time.time()
    testing = 0
    unknown_chr = 0
    fi_cov = open( infile_cov, 'r' )
    fi_context = open( infile_context, 'r' )
    fo = open( outfile, 'w' )
    fo_CG = open( outfile_CG, 'w' )
    fo_CHG = open( outfile_CHG, 'w' )
    fo_CHH = open( outfile_CHH, 'w' )
    header = 'position\tcontext'
    for samp in s:
        header += '\t' + samp + '_methylation_reads\t' + samp + '_total_reads'
    fo.write( header + '\n' )
    fo_CG.write( header + '\n' )
    fo_CHG.write( header + '\n' )
    fo_CHH.write( header + '\n' )

    fi_cov.readline() #get rid of the header line
    fi_context.readline() #get rid of the header line

    while( True ):
        lt_cov = fi_cov.readline()
        lt_context = fi_context.readline()
        if( lt_cov == '' ):
            break
        line_cov = lt_cov.split( '\t' )
        line_context = lt_context.split( '\t' )
        if( line_cov[ 0 ] != line_context[ 0 ] ):
            print( 'Error' )
            print( line_cov )
            print( line_context )
            sys.exit()
        meth_array = [ 0 ] * len( s )
        unmeth_array = [ 0 ] * len( s )
        reads_array = [ 0 ] * len( s )
        read_filter_pass = 0
        for i in range( len( s ) ):
            meth_array[ i ] = int( line_cov[ i*4 + 1 ] ) + int( line_cov[ i*4 + 3 ] )
            unmeth_array[ i ] = int( line_cov[ i*4 + 2 ] ) + int( line_cov[ i*4 + 4 ] )
            reads_array[ i ] = meth_array[ i ] + unmeth_array[ i ]
            if( reads_array[ i ] >= 2 ):
                read_filter_pass += 1
        if( read_filter_pass >= 6 ):
            write_line = line_context[ 0 ] + '\t' + line_context[ 1 ].strip()
            for i in range( len( s ) ):
                write_line += '\t%d' % meth_array[ i ] + '\t%d' % reads_array [ i ]
            fo.write( write_line + '\n' )
            lc1 = line_context[ 1 ].strip()
            if( ( lc1 == '1' ) or ( lc1 == '2' ) ):
                fo_CG.write( write_line + '\n' )
            elif( ( lc1 == '3' ) or ( lc1 == '4' ) ):
                fo_CHG.write( write_line + '\n' )
            elif( ( lc1 == '5' ) or ( lc1 == '6' ) ):
                fo_CHH.write( write_line + '\n' )

    fo.close()

```

```

fo_CG.close()
fo_CHG.close()
fo_CHH.close()
fi_cov.close()
fi_context.close()
endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )
print( "Total lines = %d" % countlines )

```

10.24 PCA_methylation.R

This R script generates PCA plots of methylation levels from the files generated by

cov2R.py.

```

library( data.table )
library( compiler )
enableJIT(3)

ptm = proc.time()

sysinfo = Sys.info()
if( sysinfo['sysname'] == "Windows" ){
  dir_base = 'C:/BaseSpace/'
} else {
  dir_base = '/scratch/user/smit1191/'
}

dir_in_b = paste0( dir_base, 'epiGBS_98_bismark/merged' )
dir_in_e = paste0( dir_base, 'epiGBS_93_varcall/merged' )

samples = c( 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5',
'R6', 'R7', 'R8' )
chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
context = c( 'CG', 'CHG', 'CHH' )

for( j in 1:3 ){
  s_threshold = 50
  i = 1
  samp_in = paste( dir_in_b, '/R.', chromosomes[ i ], '.', context[ j ], '.tsv', sep = '' )
  dts1 = fread( samp_in )
  dts1 = dts1[ ( ( L1_total_reads + L2_total_reads + L3_total_reads + L4_total_reads >
s_threshold ) & ( L5_total_reads + L6_total_reads + L7_total_reads + L8_total_reads >
s_threshold ) ) | ( ( R1_total_reads + R2_total_reads + R3_total_reads + R4_total_reads >
s_threshold ) & ( R5_total_reads + R6_total_reads + R7_total_reads + R8_total_reads >
s_threshold ) ) ]
  i = 2
  samp_in = paste( dir_in_b, '/R.', chromosomes[ i ], '.', context[ j ], '.tsv', sep = '' )
  dts2 = fread( samp_in )
  dts2 = dts2[ ( ( L1_total_reads + L2_total_reads + L3_total_reads + L4_total_reads >
s_threshold ) & ( L5_total_reads + L6_total_reads + L7_total_reads + L8_total_reads >
s_threshold ) ) | ( ( R1_total_reads + R2_total_reads + R3_total_reads + R4_total_reads >
s_threshold ) & ( R5_total_reads + R6_total_reads + R7_total_reads + R8_total_reads >
s_threshold ) ) ]
  i = 3
  samp_in = paste( dir_in_b, '/R.', chromosomes[ i ], '.', context[ j ], '.tsv', sep = '' )
  dts3 = fread( samp_in )
  dts3 = dts3[ ( ( L1_total_reads + L2_total_reads + L3_total_reads + L4_total_reads >
s_threshold ) & ( L5_total_reads + L6_total_reads + L7_total_reads + L8_total_reads >
s_threshold ) ) | ( ( R1_total_reads + R2_total_reads + R3_total_reads + R4_total_reads >
s_threshold ) & ( R5_total_reads + R6_total_reads + R7_total_reads + R8_total_reads >
s_threshold ) ) ]
  i = 4
  samp_in = paste( dir_in_b, '/R.', chromosomes[ i ], '.', context[ j ], '.tsv', sep = '' )
  dts4 = fread( samp_in )
  dts4 = dts4[ ( ( L1_total_reads + L2_total_reads + L3_total_reads + L4_total_reads >
s_threshold ) & ( L5_total_reads + L6_total_reads + L7_total_reads + L8_total_reads >
s_threshold ) ) | ( ( R1_total_reads + R2_total_reads + R3_total_reads + R4_total_reads >
s_threshold ) & ( R5_total_reads + R6_total_reads + R7_total_reads + R8_total_reads >
s_threshold ) ) ]
  i = 5

```



```

ylation_reads,R3_total_reads,R4_methylation_reads,R4_total_reads,R5_methylation_reads,R5_total
_reads,R6_methylation_reads,R6_total_reads,R7_methylation_reads,R7_total_reads,R8_methylation_
reads,R8_total_reads )],
    dts7[
        , .( chr = 'chr7H',
position,L1_methylation_reads,L1_total_reads,L2_methylation_reads,L2_total_reads,L3_methylatio
n_reads,L3_total_reads,L4_methylation_reads,L4_total_reads,L5_methylation_reads,L5_total_reads
,L6_methylation_reads,L6_total_reads,L7_methylation_reads,L7_total_reads,L8_methylation_reads,
L8_total_reads,R1_methylation_reads,R1_total_reads,R2_methylation_reads,R2_total_reads,R3_meth
ylation_reads,R3_total_reads,R4_methylation_reads,R4_total_reads,R5_methylation_reads,R5_total
_reads,R6_methylation_reads,R6_total_reads,R7_methylation_reads,R7_total_reads,R8_methylation_
reads,R8_total_reads )],
    dtsU[
        , .( chr = 'chrUn',
position,L1_methylation_reads,L1_total_reads,L2_methylation_reads,L2_total_reads,L3_methylatio
n_reads,L3_total_reads,L4_methylation_reads,L4_total_reads,L5_methylation_reads,L5_total_reads
,L6_methylation_reads,L6_total_reads,L7_methylation_reads,L7_total_reads,L8_methylation_reads,
L8_total_reads,R1_methylation_reads,R1_total_reads,R2_methylation_reads,R2_total_reads,R3_meth
ylation_reads,R3_total_reads,R4_methylation_reads,R4_total_reads,R5_methylation_reads,R5_total
_reads,R6_methylation_reads,R6_total_reads,R7_methylation_reads,R7_total_reads,R8_methylation_
reads,R8_total_reads ) ] )

rm( dts1, dts2, dts3, dts4, dts5, dts6, dts7, dtsU )

dtsf_tots_b = dtsf_b[
L1_total_reads,L2_total_reads,L3_total_reads,L4_total_reads,L5_total_reads,L6_total_reads,L7_t
otal_reads,L8_total_reads,R1_total_reads,R2_total_reads,R3_total_reads,R4_total_reads,R5_total
_reads,R6_total_reads,R7_total_reads,R8_total_reads ) ]
sum_b = sum( dtsf_tots_b )

dts_pca = dtsf_b[
L2_methylation_reads/L2_total_reads, L3pc = L3_methylation_reads/L3_total_reads, L4pc =
L4_methylation_reads/L4_total_reads, L5pc = L5_methylation_reads/L5_total_reads, L6pc =
L6_methylation_reads/L6_total_reads, L7pc = L7_methylation_reads/L7_total_reads, L8pc =
L8_methylation_reads/L8_total_reads, R1pc = R1_methylation_reads/R1_total_reads, R2pc =
R2_methylation_reads/R2_total_reads, R3pc = R3_methylation_reads/R3_total_reads, R4pc =
R4_methylation_reads/R4_total_reads, R5pc = R5_methylation_reads/R5_total_reads, R6pc =
R6_methylation_reads/R6_total_reads, R7pc = R7_methylation_reads/R7_total_reads, R8pc =
R8_methylation_reads/R8_total_reads ) ]
dts_pca[ is.na( dts_pca ) ] = 0
pc = prcomp( t( dts_pca ) )
summary( pc )
t1 = summary( pc )
t2 = t1$importance

plot( pc$x[1:16,1], pc$x[1:16,2], main = sprintf( "%s", context[ j ] ),
      xlab = paste0( "PCA1, (", sprintf( "%.1f", t2[ 2, 1 ] * 100 ), "% of variance)" ),
      ylab = paste0( "PCA2, (", sprintf( "%.1f", t2[ 2, 2 ] * 100 ), "% of variance)" ),
      pch = 32 )
points( pc$x[1:4,1], pc$x[1:4,2], col = "blue", pch = c( 49,50,51,52 ) )
points( pc$x[5:8,1], pc$x[5:8,2], col = "red", pch = c( 53,54,55,56 ) )
points( pc$x[9:12,1], pc$x[9:12,2], col = "darkgreen", pch = c( 49,50,51,52 ) )
points( pc$x[13:16,1], pc$x[13:16,2], col = "darkorange", pch = c( 53,54,55,56 ) )

t_b_rs = tryCatch( { tt = t.test( x = pc$x[ 1:8, 1 ], y = pc$x[ 9:16, 1 ] )
tt$p.value }, error = function( cond ) { return( 1 ) }, warning = function( cond ) { return(
1 ) } )
if( sysinfo['sysname'] == "Windows" ){
    if( t_b_rs != 1 ){
        png( paste0( dir_in_b, '/pca.nomt.b_rs.s_thr', s_threshold, '.', context[ j ],
'largetxt.png' ), width = 1000, height = 1000, pointsize = 24 )
        plot( pc$x[1:16,1], pc$x[1:16,2], main = sprintf( "%s", context[ j ] ),
              xlab = paste0( "PCA1, (", sprintf( "%.1f", t2[ 2, 1 ] * 100 ), "% of variance)" ),
              ylab = paste0( "PCA2, (", sprintf( "%.1f", t2[ 2, 2 ] * 100 ), "% of variance)" ),
              pch = 32 )
        points( pc$x[1:4,1], pc$x[1:4,2], col = "blue", pch = c( 49,50,51,52 ) )
        points( pc$x[5:8,1], pc$x[5:8,2], col = "red", pch = c( 53,54,55,56 ) )
        points( pc$x[9:12,1], pc$x[9:12,2], col = "darkgreen", pch = c( 49,50,51,52 ) )
        points( pc$x[13:16,1], pc$x[13:16,2], col = "darkorange", pch = c( 53,54,55,56 ) )
        dev.off()
    }
}

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[1], 's, system time = ', temp_time[2], 's, elapsed time = ',
temp_time[3], 's', sep = ' ' )

```

10.25 plot_chromosome_methylation.R

This custom R script plots the percentage methylation of cytosines in 1 million base pair windows across each chromosome of the genome. Each cytosine context (CG, CHG, and CHH) is plotted separately as these have different levels and different functions. The average methylation level for each context is also calculated.

```
library( GenomicRanges )
library( data.table )
library( compiler )
library( ggplot2 )
enableJIT(3)

ptm = proc.time()

sysinfo = Sys.info()
if( sysinfo['sysname'] == "Windows" ){
  dir_base = 'C:/BaseSpace/'
} else {
  dir_base = '/scratch/user/smit1191/'
}

dir_in_b = paste0( dir_base, 'epiGBS_98_bismark/merged' )
dir_in_e = paste0( dir_base, 'epiGBS_93_varcall/merged' )

samples = c( 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5',
             'R6', 'R7', 'R8' )
chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
contexts = c( 'CG', 'CHG', 'CHH' )

context = 'CG'

files_in = c( paste0( dir_in_b, '/R.for.win.4reps.windowed1000000.step1000000.CG.L.Rdata' ),
              paste0( dir_in_b, '/R.for.win.4reps.windowed1000000.CHG.L.Rdata' ),
              paste0( dir_in_b, '/R.for.win.4reps.windowed1000000.CHH.L.Rdata' ) )

for( ind1 in 1:3 )
{
  file_in = files_in[ ind1 ]
  load( file_in )
  reads.dt = data.table( r1 = meth.granges$coverage1,
                        r2 = meth.granges$coverage2,
                        r3 = meth.granges$coverage3,
                        r4 = meth.granges$coverage4,
                        r5 = meth.granges$coverage5,
                        r6 = meth.granges$coverage6,
                        r7 = meth.granges$coverage7,
                        r8 = meth.granges$coverage8 )
  meths.dt = data.table( r1 = meth.granges$numCs1,
                        r2 = meth.granges$numCs2,
                        r3 = meth.granges$numCs3,
                        r4 = meth.granges$numCs4,
                        r5 = meth.granges$numCs5,
                        r6 = meth.granges$numCs6,
                        r7 = meth.granges$numCs7,
                        r8 = meth.granges$numCs8 )
  chrs = as( seqnames( meth.granges ), 'factor' )
  starts = start( meth.granges )

  sumreads = rowSums( reads.dt, na.rm = TRUE )
  summeths = rowSums( meths.dt, na.rm = TRUE )
  cat( 'Average methylation level = ', sum( summeths )/sum( sumreads ), '\n' )

  min( sumreads )
}
```

```

pc.meth = summeths/sumreads

plot.df = data.frame( x = NULL, y = NULL, chrs = NULL )

png( filename = paste0( file_in, '.asterisk.new.png' ), width = 1000, height = 1200, units =
'px', pointsize = 12 )
par( mfrow = c( 7, 1 ), mar = c( 3, 4.5, 0.4, 0.1 ), xaxt = 's' )
if( ind1 == 1 )
{
  maxy = 100
} else if( ind1 == 2 ) {
  maxy = 80
} else {
  maxy = 10
}
maxx = 8e8
for( chromosome in chromosomes[1:7] )
{
  mask1 = which( chrs == chromosome )
  plot( x = c( starts[ mask1 ], maxx )/1000000, y = c( pc.meth[ mask1 ]*100, 100), pch =
'*', main = '', xlab = '', ylab = '', ylim = c( 0, maxy ), xlim = c( 0, maxx/1000000 ), axes =
FALSE )
  axis( 2, at = c( 0, maxy/2, maxy ), las = 1, cex.axis = 2.5 )
  plot.df = rbind( plot.df, data.frame( x = starts[ mask1 ], y = pc.meth[ mask1 ]*100, chrs
= chrs[ mask1 ] ) )
}
axis( 1, at = seq( 0, maxx/1000000, maxx/8000000 ), cex.axis = 2.5, padj = 0.5 )
dev.off()
}

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[1], 's, system time = ', temp_time[2], 's, elapsed time = ',
temp_time[3], 's', sep = ' ' )

```

10.26 methylation_levels_genes.py

This program determines the methylation pattern in and 1.5 kbp either side of genes.

```

# Gets methylation information around genes for plotting of methylation patterns
#
# Jason Smith
# Last modified 6/3/2020 (d/m/y)

import sys
import gzip
import os
import time
import csv

#lr = 'L'
lr = 'R'

starttime = time.time()
if( sys.platform[0] == 'w' ):
  basedir = 'C:/BaseSpace/'
else:
  basedir = '/home/jason/Share/'

dir_in_m = basedir + 'epiGBS_98_bismark/merged/'
dir_out = basedir + 'epiGBS_98_bismark/merged/'

genes_file = basedir + 'Reference_Genome/Hvu_dl_2019-12-05/Hordeum_vulgare.IBSC_v2.45.gff3.genes.tsv'

chromosomes = [ 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' ]
chr_sizes = [ 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000,
249774706 ]
s = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6',
'R7', 'R8' ]
contexts = [ 'CHH', 'CG', 'CHG' ]
context = 'CHH'

```

```

before_gene = 1500
in_gene = 1000
after_gene = 1500

total_count_f = [ 0 ] * ( before_gene + in_gene + after_gene )
meth_count_f = [ 0 ] * ( before_gene + in_gene + after_gene )
total_count_r = [ 0 ] * ( before_gene + in_gene + after_gene )
meth_count_r = [ 0 ] * ( before_gene + in_gene + after_gene )

f_g = open( genes_file, 'rb' )
genes_header = f_g.readline().strip().split()

gene_line = f_g.readline().strip()
gene_split = gene_line.split()
chr_count = 0

# skip the gene forward to chr1H
current_chr_meth = chromosomes[ 0 ]
current_chr_gene = 'chr' + current_chr_meth
while( gene_split[0] != 'chrchr1H' ):
    gene_line = f_g.readline().strip()
    gene_split = gene_line.split()

# open the relevant chromosome and context combination
meth_file = dir_in_m + 'R.' + current_chr_meth + '.' + context + '.tsv.gz'
f_m = gzip.open( meth_file, 'rb' )
meth_header = f_m.readline().strip().split()
meth_line = f_m.readline().strip()
meth_split = meth_line.split()
print( current_chr_meth )
cont = True
exitnext = False
nextChr = False

while( cont ):
    while( int( gene_split[ 2 ] ) + after_gene < int( meth_split[ 0 ] ) ):
        gene_line = f_g.readline().strip()
        gene_split = gene_line.split()
        if( gene_split[ 0 ] != current_chr_gene ):
            current_chr_gene = gene_split[ 0 ]
            current_chr_meth = current_chr_gene[ 3: ]
            print( current_chr_meth )
            f_m.close()
            meth_file = dir_in_m + 'R.' + current_chr_meth + '.' + context + '.tsv.gz'
            f_m = gzip.open( meth_file, 'rb' )
            meth_header = f_m.readline().strip().split()
            meth_line = f_m.readline().strip()
            meth_split = meth_line.split()
        if( gene_split[ 4 ] == '+' ):
            # first deal with upstream
            minval = int( gene_split[ 1 ] ) - before_gene
            maxval = int( gene_split[ 1 ] ) - 1
            if( meth_line == 'T' ):
                meth_split[ 0 ] = 1000000000
                nextChr = True
                break

            while( int( meth_split[ 0 ] ) <= maxval ):
                while( int( meth_split[ 0 ] ) < minval ):
                    meth_line = f_m.readline().strip()
                    if( meth_line == ' ' ):
                        meth_split[ 0 ] = 1000000000
                        nextChr = True
                        break
                    meth_split = meth_line.split()
                if( int( meth_split[ 0 ] ) <= maxval ):
                    # add it to the relevant arrays
                    ind1 = int( meth_split[ 0 ] ) - minval
                    meth_add = 0
                    count_add = 0
                    for c1 in range( 0, 16 ):
                        meth_add += int( meth_split[ 2 * c1 + 2 ] )
                        count_add += int( meth_split[ 2 * c1 + 3 ] )
                    meth_count_f[ ind1 ] += meth_add
                    total_count_f[ ind1 ] += count_add
                    meth_line = f_m.readline().strip()
                    if( meth_line == ' ' ):

```



```

        meth_split[ 0 ] = 1000000000
        nextChr = True
        break
    meth_split = meth_line.split()
# then deal with the gene body
minval = int( gene_split[ 1 ] )
maxval = int( gene_split[ 2 ] )
while( int( meth_split[ 0 ] ) <= maxval ):
    while( int( meth_split[ 0 ] ) < minval ):
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()
    if( int( meth_split[ 0 ] ) <= maxval ):
        # add it to the relevant arrays
        # This time the array position is based on the position relative
        # to the TSS and TES
        dl = int( meth_split[ 0 ] ) - minval
        dr = maxval - int( meth_split[ 0 ] )
        ind1 = int( round( before_gene + dl * 1000.0 / ( dl + dr ) ) )
        meth_add = 0
        count_add = 0
        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
        meth_count_f[ ind1 ] += meth_add
        total_count_f[ ind1 ] += count_add
        meth_line = f_m.readline().strip()
        if( meth_line == '' ):
            meth_split[ 0 ] = 1000000000
            nextChr = True
            break
        meth_split = meth_line.split()
# then deal with downstream
minval = int( gene_split[ 2 ] ) + 1
maxval = int( gene_split[ 2 ] ) + after_gene
while( int( meth_split[ 0 ] ) <= maxval ):
    while( int( meth_split[ 0 ] ) < minval ):
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()
    if( int( meth_split[ 0 ] ) <= maxval ):
        # add it to the relevant arrays
        ind1 = int( meth_split[ 0 ] ) - minval + before_gene + in_gene
        meth_add = 0
        count_add = 0
        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
        meth_count_f[ ind1 ] += meth_add
        total_count_f[ ind1 ] += count_add
        meth_line = f_m.readline().strip()
        if( meth_line == '' ):
            meth_split[ 0 ] = 1000000000
            nextChr = True
            break
        meth_split = meth_line.split()

if( gene_split[ 4 ] == '-' ):
# first deal with upstream
minval = int( gene_split[ 1 ] ) - before_gene
maxval = int( gene_split[ 1 ] ) - 1
if( meth_line == '' ):
    meth_split[ 0 ] = 1000000000
    nextChr = True
    break

while( int( meth_split[ 0 ] ) <= maxval ):
    while( int( meth_split[ 0 ] ) < minval ):
        meth_line = f_m.readline().strip()
        if( meth_line == '' ):
            meth_split[ 0 ] = 1000000000
            nextChr = True
            break
        meth_split = meth_line.split()
    if( int( meth_split[ 0 ] ) <= maxval ):
        # add it to the relevant arrays
        ind1 = int( meth_split[ 0 ] ) - minval
        meth_add = 0
        count_add = 0

```

```

        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
            meth_count_r[ ind1 ] += meth_add
            total_count_r[ ind1 ] += count_add
meth_line = f_m.readline().strip()
if( meth_line == '' ):
    meth_split[ 0 ] = 1000000000
    nextChr = True
    break
    meth_split = meth_line.split()
# then deal with the gene body
minval = int( gene_split[ 1 ] )
maxval = int( gene_split[ 2 ] )
while( int( meth_split[ 0 ] ) <= maxval ):
    while( int( meth_split[ 0 ] ) < minval ):
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()
    if( int( meth_split[ 0 ] ) <= maxval ):
        # add it to the relevant arrays
        # This time the array position is based on the position relative
        # to the TSS and TES
        dl = int( meth_split[ 0 ] ) - minval
        dr = maxval - int( meth_split[ 0 ] )
        ind1 = int( round( before_gene + dl * 1000.0 / ( dl + dr ) ) )
        meth_add = 0
        count_add = 0
        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
            meth_count_r[ ind1 ] += meth_add
            total_count_r[ ind1 ] += count_add
        meth_line = f_m.readline().strip()
        if( meth_line == '' ):
            meth_split[ 0 ] = 1000000000
            nextChr = True
            break
        meth_split = meth_line.split()
# then deal with downstream
minval = int( gene_split[ 2 ] ) + 1
maxval = int( gene_split[ 2 ] ) + after_gene
while( int( meth_split[ 0 ] ) <= maxval ):
    while( int( meth_split[ 0 ] ) < minval ):
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()
    if( int( meth_split[ 0 ] ) <= maxval ):
        # add it to the relevant arrays
        ind1 = int( meth_split[ 0 ] ) - minval + before_gene + in_gene
        meth_add = 0
        count_add = 0
        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
            meth_count_r[ ind1 ] += meth_add
            total_count_r[ ind1 ] += count_add
        meth_line = f_m.readline().strip()
        if( meth_line == '' ):
            meth_split[ 0 ] = 1000000000
            nextChr = True
            break
        meth_split = meth_line.split()

if( nextChr == True ):
    nextChr = False
    # Need to skip forward to the next chromosome
    #print( 'nextChr' )
    while( gene_split[ 0 ] == current_chr_gene ):
        if( current_chr_gene == 'chrchrUn' ):
            cont = False
            break
        gene_line = f_g.readline().strip()
        gene_split = gene_line.split()
        if( gene_split[ 0 ] != current_chr_gene ):
            #print( 'if gene_split' )
            current_chr_gene = gene_split[ 0 ]
            current_chr_meth = current_chr_gene[ 3: ]
            print( current_chr_meth )

```

```

        f_m.close()
        meth_file = dir_in_m + 'R.' + current_chr_meth + '.' + context + '.tsv.gz'
        f_m = gzip.open( meth_file, 'rb' )
        meth_header = f_m.readline().strip().split()
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()
        break

    else:
        gene_line = f_g.readline().strip()
        if( gene_line == '' ):
            nextChr = True
            break
        gene_split = gene_line.split()

    if( gene_split[ 0 ] != current_chr_gene ):
        #print( 'it ended up getting the next chromosome here' )
        current_chr_gene = gene_split[ 0 ]
        current_chr_meth = current_chr_gene[ 3: ]
        print( current_chr_meth )
        f_m.close()
        meth_file = dir_in_m + 'R.' + current_chr_meth + '.' + context + '.tsv.gz'
        f_m = gzip.open( meth_file, 'rb' )
        meth_header = f_m.readline().strip().split()
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()

f_o_f = open( dir_out + 'gene.meth.forward.' + context + '.tsv', 'wb' )
f_o_f.write( 'meth\ttotal\n' )
f_o_r = open( dir_out + 'gene.meth.reverse.' + context + '.tsv', 'wb' )
f_o_r.write( 'meth\ttotal\n' )

for c2 in range( 0, before_gene + in_gene + after_gene ):
    f_o_f.write( '%d\t%d\n' % ( meth_count_f[ c2 ], total_count_f[ c2 ] ) )
    f_o_r.write( '%d\t%d\n' % ( meth_count_r[ c2 ], total_count_r[ c2 ] ) )

f_o_f.close()
f_o_r.close()
f_m.close()
f_g.close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )

sys.exit()

```

10.27 methylation_levels_TEs.py

This program determines the methylation pattern in and 1.5 kbp either side of TEs for each TE class.

```

#!/usr/bin/python
#
# Gets methylation information around TEs (not differential meth)
#
# Jason Smith
# Last modified 7/3/2020 (d/m/y)

import sys
import gzip
import os
import time
import csv

#lr = 'L'
lr = 'R'

starttime = time.time()
if( sys.platform[0] == 'w' ):
    basedir = 'C:/BaseSpace/'

```

```

else:
    basedir = '/home/jason/Share/'

dir_in_m = basedir + 'epiGBS_98_bismark/merged/'
dir_out = basedir + 'epiGBS_98_bismark/merged/'

genes_file = basedir + 'Reference_Genome/Hvu_dl_2019-12-05/Hordeum_vulgare.IBSC_v2.45.gff3.genes.tsv'
TE_file_base = basedir + 'Reference_Genome/Hvu_dl_2019-12-05/Barley_TE_annotation_v2_18Aug16.'

chromosomes = [ 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' ]
chr_sizes = [ 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000, 249774706 ]
s = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6', 'R7', 'R8' ]
contexts = [ 'CHH', 'CG', 'CHG' ]
context = 'CHH'
TE_types = [ 'DHH', 'DTA', 'DTC', 'DTH', 'DTM', 'DTT', 'DTX', 'DXX', 'RIX', 'RLC', 'RLG', 'RLX', 'RSX', 'RXX', 'XXX' ]

for TE_type in TE_types:
    print( TE_type )
    before_gene = 1500
    in_gene = 1000
    after_gene = 1500

    total_count_f = [ 0 ] * ( before_gene + in_gene + after_gene )
    meth_count_f = [ 0 ] * ( before_gene + in_gene + after_gene )
    total_count_r = [ 0 ] * ( before_gene + in_gene + after_gene )
    meth_count_r = [ 0 ] * ( before_gene + in_gene + after_gene )

    TE_file = TE_file_base + TE_type + '.tsv'

    f_g = open( TE_file, 'rb' )
    genes_header = f_g.readline().strip().split()

    gene_line = f_g.readline().strip()
    gene_split = gene_line.split()
    chr_count = 0

    # This starts at chrUn
    current_chr_gene = 'chrUn'
    current_chr_meth = current_chr_gene

    # open the relevant chromosome and context combination
    meth_file = dir_in_m + 'R.' + current_chr_meth + '.' + context + '.tsv.gz'
    f_m = gzip.open( meth_file, 'rb' )
    meth_header = f_m.readline().strip().split()
    meth_line = f_m.readline().strip()
    meth_split = meth_line.split()
    print( current_chr_meth )
    cont = True
    exitnext = False
    nextChr = False

    while( cont ):
        while( int( gene_split[ 2 ] ) + after_gene < int( meth_split[ 0 ] ) ):
            gene_line = f_g.readline().strip()
            if( gene_line == '' ):
                cont = False
                break
            gene_split = gene_line.split()
            if( gene_split[ 0 ] != current_chr_gene ):
                current_chr_gene = gene_split[ 0 ]
                current_chr_meth = current_chr_gene
                print( current_chr_meth )
                f_m.close()
                meth_file = dir_in_m + 'R.' + current_chr_meth + '.' + context + '.tsv.gz'
                f_m = gzip.open( meth_file, 'rb' )
                meth_header = f_m.readline().strip().split()
                meth_line = f_m.readline().strip()
                meth_split = meth_line.split()
            if( gene_line == '' ):
                cont = False
                break
            if( gene_split[ 4 ] == 'F' ):
                # first deal with upstream

```

```

minval = int( gene_split[ 1 ] ) - before_gene
maxval = int( gene_split[ 1 ] ) - 1
if( meth_line == '' ):
    meth_split[ 0 ] = 1000000000
    nextChr = True
    break

while( int( meth_split[ 0 ] ) <= maxval ):
    while( int( meth_split[ 0 ] ) < minval ):
        meth_line = f_m.readline().strip()
        if( meth_line == '' ):
            meth_split[ 0 ] = 1000000000
            nextChr = True
            break
        meth_split = meth_line.split()
    if( int( meth_split[ 0 ] ) <= maxval ):
        # add it to the relevant arrays
        ind1 = int( meth_split[ 0 ] ) - minval
        meth_add = 0
        count_add = 0
        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
        meth_count_f[ ind1 ] += meth_add
        total_count_f[ ind1 ] += count_add
    meth_line = f_m.readline().strip()
    if( meth_line == '' ):
        meth_split[ 0 ] = 1000000000
        nextChr = True
        break
    meth_split = meth_line.split()
# then deal with the gene body
minval = int( gene_split[ 1 ] )
maxval = int( gene_split[ 2 ] )
while( int( meth_split[ 0 ] ) <= maxval ):
    while( int( meth_split[ 0 ] ) < minval ):
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()
    if( int( meth_split[ 0 ] ) <= maxval ):
        # add it to the relevant arrays
        # This time the array position is based on the position relative
        # to the TSS and TES
        dl = int( meth_split[ 0 ] ) - minval
        dr = maxval - int( meth_split[ 0 ] )
        ind1 = int( round( before_gene + dl * 1000.0 / ( dl + dr ) ) )
        meth_add = 0
        count_add = 0
        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
        meth_count_f[ ind1 ] += meth_add
        total_count_f[ ind1 ] += count_add
    meth_line = f_m.readline().strip()
    if( meth_line == '' ):
        meth_split[ 0 ] = 1000000000
        nextChr = True
        break
    meth_split = meth_line.split()
# then deal with downstream
minval = int( gene_split[ 2 ] ) + 1
maxval = int( gene_split[ 2 ] ) + after_gene
while( int( meth_split[ 0 ] ) <= maxval ):
    while( int( meth_split[ 0 ] ) < minval ):
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()
    if( int( meth_split[ 0 ] ) <= maxval ):
        # add it to the relevant arrays
        ind1 = int( meth_split[ 0 ] ) - minval + before_gene + in_gene
        meth_add = 0
        count_add = 0
        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
        meth_count_f[ ind1 ] += meth_add
        total_count_f[ ind1 ] += count_add
    meth_line = f_m.readline().strip()
    if( meth_line == '' ):

```

```

        meth_split[ 0 ] = 1000000000
        nextChr = True
        break
    meth_split = meth_line.split()

if( gene_split[ 4 ] == 'R' ):
    # first deal with upstream
    minval = int( gene_split[ 1 ] ) - before_gene
    maxval = int( gene_split[ 1 ] ) - 1
    if( meth_line == 'T' ):
        meth_split[ 0 ] = 1000000000
        nextChr = True
        break

    while( int( meth_split[ 0 ] ) <= maxval ):
        while( int( meth_split[ 0 ] ) < minval ):
            meth_line = f_m.readline().strip()
            if( meth_line == 'T' ):
                meth_split[ 0 ] = 1000000000
                nextChr = True
                break
            meth_split = meth_line.split()
        if( int( meth_split[ 0 ] ) <= maxval ):
            # add it to the relevant arrays
            ind1 = int( meth_split[ 0 ] ) - minval
            meth_add = 0
            count_add = 0
            for c1 in range( 0, 16 ):
                meth_add += int( meth_split[ 2 * c1 + 2 ] )
                count_add += int( meth_split[ 2 * c1 + 3 ] )
            meth_count_r[ ind1 ] += meth_add
            total_count_r[ ind1 ] += count_add
            meth_line = f_m.readline().strip()
            if( meth_line == 'T' ):
                meth_split[ 0 ] = 1000000000
                nextChr = True
                break
            meth_split = meth_line.split()
        # then deal with the gene body
        minval = int( gene_split[ 1 ] )
        maxval = int( gene_split[ 2 ] )
        while( int( meth_split[ 0 ] ) <= maxval ):
            while( int( meth_split[ 0 ] ) < minval ):
                meth_line = f_m.readline().strip()
                meth_split = meth_line.split()
            if( int( meth_split[ 0 ] ) <= maxval ):
                # add it to the relevant arrays
                # This time the array position is based on the position relative
                # to the TSS and TES
                dl = int( meth_split[ 0 ] ) - minval
                dr = maxval - int( meth_split[ 0 ] )
                ind1 = int( round( before_gene + dl * 1000.0 / ( dl + dr ) ) )
                meth_add = 0
                count_add = 0
                for c1 in range( 0, 16 ):
                    meth_add += int( meth_split[ 2 * c1 + 2 ] )
                    count_add += int( meth_split[ 2 * c1 + 3 ] )
                meth_count_r[ ind1 ] += meth_add
                total_count_r[ ind1 ] += count_add
                meth_line = f_m.readline().strip()
                if( meth_line == 'T' ):
                    meth_split[ 0 ] = 1000000000
                    nextChr = True
                    break
                meth_split = meth_line.split()
            # then deal with downstream
            minval = int( gene_split[ 2 ] ) + 1
            maxval = int( gene_split[ 2 ] ) + after_gene
            while( int( meth_split[ 0 ] ) <= maxval ):
                while( int( meth_split[ 0 ] ) < minval ):
                    meth_line = f_m.readline().strip()
                    meth_split = meth_line.split()
                if( int( meth_split[ 0 ] ) <= maxval ):
                    # add it to the relevant arrays
                    ind1 = int( meth_split[ 0 ] ) - minval + before_gene + in_gene
                    meth_add = 0
                    count_add = 0

```

```

        for c1 in range( 0, 16 ):
            meth_add += int( meth_split[ 2 * c1 + 2 ] )
            count_add += int( meth_split[ 2 * c1 + 3 ] )
            meth_count_r[ ind1 ] += meth_add
            total_count_r[ ind1 ] += count_add
meth_line = f_m.readline().strip()
if( meth_line == '' ):
    meth_split[ 0 ] = 1000000000
    nextChr = True
    break
meth_split = meth_line.split()

if( nextChr == True ):
    nextChr = False
    # Need to skip forward to the next chromosome
    while( gene_split[ 0 ] == current_chr_gene ):
        if( current_chr_gene == 'chr7H' ):
            cont = False
            break
        gene_line = f_g.readline().strip()
        gene_split = gene_line.split()
        if( gene_split[ 0 ] != current_chr_gene ):
            current_chr_gene = gene_split[ 0 ]
            current_chr_meth = current_chr_gene
            print( current_chr_meth )
            f_m.close()
            meth_file = dir_in_m + 'R.' + current_chr_meth + '.' + context + '.tsv.gz'
            f_m = gzip.open( meth_file, 'rb' )
            meth_header = f_m.readline().strip().split()
            meth_line = f_m.readline().strip()
            meth_split = meth_line.split()
            break

    else:
        gene_line = f_g.readline().strip()
        if( gene_line == '' ):
            nextChr = True
            break
        gene_split = gene_line.split()

    if( gene_split[ 0 ] != current_chr_gene ):
        #print( 'it ended up getting the next chromosome here' )
        current_chr_gene = gene_split[ 0 ]
        current_chr_meth = current_chr_gene
        print( current_chr_meth )
        f_m.close()
        meth_file = dir_in_m + 'R.' + current_chr_meth + '.' + context + '.tsv.gz'
        f_m = gzip.open( meth_file, 'rb' )
        meth_header = f_m.readline().strip().split()
        meth_line = f_m.readline().strip()
        meth_split = meth_line.split()

f_o = open( dir_out + 'TE.meth.' + context + '.' + TE_type + '.tsv', 'wb' )
f_o.write( 'meth\t\ttotal\n' )

for c2 in range( 0, before_gene + in_gene + after_gene ):
    c3 = before_gene + in_gene + after_gene - c2 - 1
    f_o.write( '%d\t%d\n' % ( meth_count_f[ c2 ] + meth_count_r[ c3 ], total_count_f[ c2 ]
+ total_count_r[ c3 ] ) )

f_o.close()
f_m.close()
f_g.close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )

sys.exit()

```

10.28 find_differentially_methylated.R

```

library( 'methylKit' )

setwd( '/home/jason/Share/epiGBS_98_bismark/merged/' )

```

```

for(     filebase     in     c(     '/home/jason/Share/epiGBS_98_bismark/merged/',
'/home/jason/Share/epiGBS_93_varcall/merged/' ) )
{
  for( context1 in c( 'CG', 'CHG', 'CHH' ) )
  {
    for( lr in c( 'L', 'R' ) )
    {
      file.list = list( paste0( filebase, 'R.methylKit.', lr, '1.', context1, '.tsv' ),
                        paste0( filebase, 'R.methylKit.', lr, '2.', context1, '.tsv' ),
                        paste0( filebase, 'R.methylKit.', lr, '3.', context1, '.tsv' ),
                        paste0( filebase, 'R.methylKit.', lr, '4.', context1, '.tsv' ),
                        paste0( filebase, 'R.methylKit.', lr, '5.', context1, '.tsv' ),
                        paste0( filebase, 'R.methylKit.', lr, '6.', context1, '.tsv' ),
                        paste0( filebase, 'R.methylKit.', lr, '7.', context1, '.tsv' ),
                        paste0( filebase, 'R.methylKit.', lr, '8.', context1, '.tsv' ) )

      myobj = methRead( file.list,
                       sample.id = list( 'c1', 'c2', 'c3', 'c4', 's1', 's2', 's3', 's4' ),
                       assembly = 'Hvu', treatment = c( 0, 0, 0, 0, 1, 1, 1, 1 ),
                       context = context1, mincov = 5 )

      tiles = tileMethylCounts( myobj, win.size = 100, step.size = 100 )
      meth = unite( tiles, destrand = FALSE, min.per.group = 1L, mc.cores = 3 )
      myDiff = calculateDiffMeth( meth, mc.cores = 1 )
      myDiff10p = getMethylDiff( myDiff, difference = 10, qvalue = 0.05 )
      cat( paste0( filebase, 'R.methylKit.diff.4reps.mincov5.windowed100.step100.',
                  context1, '.', lr, '.tsv\n' ) )
      cat( nrow( myDiff10p ), '\n' )

      #need to save the values as .tsv and .Rdata
      write.table( myDiff10p,
                  file = paste0( filebase,
                                'R.methylKit.diff.4reps.mincov5.windowed100.step100.',
                                context1, '.', lr, '.tsv' ),
                  sep = '\t', row.name = FALSE, col.name = TRUE, quote = FALSE )
      save( myDiff10p,
           file = paste0( filebase,
                         'R.methylKit.diff.4reps.mincov5.windowed100.step100.',
                         context1, '.', lr, '.Rdata' ) )
    }
  }
}

```

10.29 dist_to_TSS.R

This custom R script compares the distances between the gene transcription start site and the location of differentially methylated regions and generates histograms of these distances.

```

library( 'methylKit' )
library( 'genomation' )
library( 'GenomicFeatures' )
library( 'data.table' )
library( 'compiler' )
enableJIT( 3 )

if( exists( 'alreadyread' ) == FALSE )
{
  gff      =      gffToGRanges(      '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE )
  genes    =      gffToGRanges(      '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE, filter = 'gene' )

  t1 = genes[ seqnames( genes ) == 'chrchr1H' ]
  genes.chr1H = data.table( start = start( t1 ),
                           end = end( t1 ),
                           strand = as( strand( t1 ), 'vector' ) )
}

```



```

t1 = genes[ seqnames( genes ) == 'chrchr2H' ]
genes.chr2H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr3H' ]
genes.chr3H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr4H' ]
genes.chr4H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr5H' ]
genes.chr5H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr6H' ]
genes.chr6H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr7H' ]
genes.chr7H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchrUn' ]
genes.chrUn = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
}
alreadyread = 1

setwd( '/home/jason/Share/epiGBS_98_bismark/merged/' )

for( filebase in c( '/home/jason/Share/epiGBS_98_bismark/merged/',
                  '/home/jason/Share/epiGBS_93_varcall/merged/' ) )
{
  for( context1 in c( 'CG', 'CHG', 'CHH' ) )
  {
    for( lr in c( 'L', 'R' ) )
    {
      file_in = paste0( filebase, 'R.methylKit.diff.4reps.mincov5.windowed100.step100.',
                      context1, '.', lr, '.Rdata' )
      load( file = file_in )
      cat( paste0( file_in, '\n' ) )
      diffval = 10
      #diffval = 25
      #qval = 0.05
      qval = 0.01
      myDiff_sl = getMethylDiff( myDiff10p, difference = diffval, qvalue = qval )
      myDiff.GRanges = as( myDiff_sl, 'GRanges' )
      len1 = NROW( myDiff.GRanges )
      prev_chr = 'chr1H'
      gene_chr = genes.chr1H
      promoters = NULL
      for( ind2 in 1:NROW( gene_chr ) )
      {
        if( gene_chr[ ind2 ]$strand == '+' )
        {
          promoters[ ind2 ] = gene_chr[ ind2 ]$start
        } else {
          promoters[ ind2 ] = gene_chr[ ind2 ]$end
        }
      }
      promoter.dt = data.table( promoters, val = promoters )
      setattr( promoter.dt, 'sorted', 'promoters' )
      dt.chr = NULL
      dt.start = NULL
      dt.end = NULL
      dt.mid = NULL
      dt.qvalue = NULL
      dt.meth.diff = NULL
      dt.dist.to.gene.start = NULL
      for( ind1 in 1:len1 )
      {
        seqname = as( seqnames( myDiff.GRanges[ ind1 ] ), 'vector' )
        flag_change = 0
        if( seqname != prev_chr ){

```

```

if( seqname == 'chr2H' ){
  prev_chr = 'chr2H'
  gene_chr = genes.chr2H
  flag_change = 1
}
if( seqname == 'chr3H' ){
  prev_chr = 'chr3H'
  gene_chr = genes.chr3H
  flag_change = 1
}
if( seqname == 'chr4H' ){
  prev_chr = 'chr4H'
  gene_chr = genes.chr4H
  flag_change = 1
}
if( seqname == 'chr5H' ){
  prev_chr = 'chr5H'
  gene_chr = genes.chr5H
  flag_change = 1
}
if( seqname == 'chr6H' ){
  prev_chr = 'chr6H'
  gene_chr = genes.chr6H
  flag_change = 1
}
if( seqname == 'chr7H' ){
  prev_chr = 'chr7H'
  gene_chr = genes.chr7H
  flag_change = 1
}
if( seqname == 'chrUn' ){
  prev_chr = 'chrUn'
  gene_chr = genes.chrUn
  flag_change = 1
}
if( flag_change == 1 )
{
  promoters = NULL
  for( ind2 in 1:NROW( gene_chr ) )
  {
    if( gene_chr[ ind2 ]$strand == '+' )
    {
      promoters[ ind2 ] = gene_chr[ ind2 ]$start
    } else {
      promoters[ ind2 ] = gene_chr[ ind2 ]$end
    }
  }
  promoter.dt = data.table( promoters, val = promoters )
  setattr( promoter.dt, 'sorted', 'promoters' )
}
}
start = start( myDiff.GRanges[ ind1 ] )
end = end( myDiff.GRanges[ ind1 ] )
mid = ( start + end )/2

ind3 = promoter.dt[ J( as( mid, 'integer' ) ), roll = 'nearest', which = TRUE ]
closest_pr = promoter.dt[ ind3 ]$promoters
if( gene_chr[ ind3 ]$strand == '+' )
{
  dist = - ( closest_pr - mid )
} else {
  dist = closest_pr - mid
}
dt.chr[ ind1 ] = seqname
dt.start[ ind1 ] = start
dt.end[ ind1 ] = end
dt.mid[ ind1 ] = mid
dt.qvalue[ ind1 ] = myDiff.GRanges[ ind1 ]$qvalue
dt.meth.diff[ ind1 ] = myDiff.GRanges[ ind1 ]$meth.diff
dt.dist.to.gene.start[ ind1 ] = dist
}
dist.to.gene = data.table( dist.to.gene.start = dt.dist.to.gene.start,
  chr = dt.chr, start = dt.start, end = dt.end, mid = dt.mid,
  qvalue = dt.qvalue, meth.diff = dt.meth.diff )
dist.to.gene.sl = dist.to.gene[ ( dist.to.gene$dist.to.gene.start < 5000 ) &
  ( dist.to.gene$dist.to.gene.start > -5000 ) ]
save( dist.to.gene.sl, file = paste0( file_in, '.dist_to_gene_sl5000.qval', qval,

```

```

                                '.diff', diffval, '.Rdata' ) )
write.table( dist.to.gene.sl,
             file = paste0( file_in, '.dist_to_gene_sl5000.qval', qval,
                           '.diff', diffval, '.tsv' ),
             sep = '\t', row.name = FALSE, col.name = TRUE, quote = FALSE )
hist( dist.to.gene.sl$dist.to.gene.start, breaks = 50,
      xlim = c(-5000, 5000 ) ,
      xlab = 'Distance to start of gene',
      main = file_in )
png( filename = paste0( file_in, '.dist_to_gene_sl5000.qval', qval, '.diff', diffval,
                       '.png' ), width = 1000, height = 1000 )
hist( dist.to.gene.sl$dist.to.gene.start, breaks = 50,
      xlim = c(-5000, 5000 ) ,
      xlab = 'Distance to start of gene',
      main = file_in )
dev.off()

dist.to.gene.sl = dist.to.gene[ ( dist.to.gene$dist.to.gene.start < 1000 ) &
                               ( dist.to.gene$dist.to.gene.start > -1000 ) ]

save( dist.to.gene.sl,
      file = paste0( file_in, '.dist_to_gene_sl1000.qval', qval, '.diff', diffval,
                    '.Rdata' ) )
write.table( dist.to.gene.sl,
             file = paste0( file_in, '.dist_to_gene_sl1000.qval', qval, '.diff',
                           diffval, '.tsv' ),
             sep = '\t', row.name = FALSE, col.name = TRUE, quote = FALSE )
hist( dist.to.gene.sl$dist.to.gene.start, breaks = 50,
      xlim = c(-1000, 1000 ) ,
      xlab = 'Distance to start of gene',
      main = file_in )
png( filename = paste0( file_in, '.dist_to_gene_sl1000.qval', qval, '.diff', diffval,
                       '.png' ), width = 1000, height = 1000 )
hist( dist.to.gene.sl$dist.to.gene.start, breaks = 50,
      xlim = c(-1000, 1000 ) ,
      xlab = 'Distance to start of gene',
      main = file_in )
dev.off()
}
}
}

```

10.30 diff_meth_annotation.R

This custom R script compares the differentially methylated regions of the genome with known genes and other annotated features from Ensembl *Hordeum vulgare* v2.45 and the known transposable elements published by the IBSC (2012)

```

library( 'methylKit' )
library( 'genomation' )
library( 'GenomicFeatures' )
library( 'data.table' )
library( 'compiler' )
enableJIT( 3 )

if( exists( 'alreadyread' ) == FALSE )
{
  gff      =      gffToGRanges(      '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE )
  genes   =      gffToGRanges(      '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE, filter = 'gene' )
  filter.types = c( 'gene', 'ncrRNA_gene', 'lnc_RNA', 'tRNA', 'rRNA', 'snRNA', 'snoRNA',
                   'pre_miRNA' )
  gff.filtered = gff[ !is.na( match( gff$type, filter.types ) ) ]
  rm( gff )

  te.table =      fread(      file      =      '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Barley_TE_annotation_v2_18Aug16.tsv', sep = '\t', header = TRUE )
}

```

```

t2 = data.table( chr = te.table$seq_id, start = te.table$start, end = te.table$end,
                TE_code = te.table$TE_code )
te.chr1H = t2[ t2$chr == 'chr1H' ]
te.chr2H = t2[ t2$chr == 'chr2H' ]
te.chr3H = t2[ t2$chr == 'chr3H' ]
te.chr4H = t2[ t2$chr == 'chr4H' ]
te.chr5H = t2[ t2$chr == 'chr5H' ]
te.chr6H = t2[ t2$chr == 'chr6H' ]
te.chr7H = t2[ t2$chr == 'chr7H' ]
te.chrUn = t2[ t2$chr == 'chrUn' ]
rm( t2, te.table )

t1 = genes[ seqnames( genes ) == 'chrchr1H' ]
genes.chr1H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr2H' ]
genes.chr2H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr3H' ]
genes.chr3H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr4H' ]
genes.chr4H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr5H' ]
genes.chr5H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr6H' ]
genes.chr6H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchr7H' ]
genes.chr7H = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchrUn' ]
genes.chrUn = data.table( start = start( t1 ),
                        end = end( t1 ),
                        strand = as( strand( t1 ), 'vector' ) )
gff.chr1H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr1H' ]
gff.chr2H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr2H' ]
gff.chr3H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr3H' ]
gff.chr4H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr4H' ]
gff.chr5H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr5H' ]
gff.chr6H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr6H' ]
gff.chr7H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr7H' ]
gff.chrUn = gff.filtered[ seqnames( gff.filtered ) == 'chrchrUn' ]

}
alreadyread = 1

setwd( '/home/jason/Share/epiGBS_98_bismark/merged/' )
filebase = '/home/jason/Share/epiGBS_98_bismark/merged/'
for( context1 in c( 'CG', 'CHG', 'CHH' ) )
{
  for( lr in c( 'L', 'R' ) )
  {
    for( offset in c( 1500 ) )
    {
      ptm = proc.time()

      file_in = paste0( filebase, 'R.methylKit.diff.4reps.mincov5.windowed100.step100.',
                      context1, '.', lr, '.Rdata' )
      load( file = file_in )
      cat( paste0( file_in, '\n' ) )
      diffval = 10
      qval = 0.05
      myDiff_sl = getMethylDiff( myDiff10p, difference = diffval, qvalue = qval )
      myDiff.GRanges = as( myDiff_sl, 'GRanges' )
      len1 = NROW( myDiff.GRanges )
      prev_chr = 'chr1H'
      gene_chr = gff.chr1H
    }
  }
}

```

```

te_chr = te.chr1H
annotation = NULL
for( ind1 in 1:len1 )
{
  seqname = as( seqnames( myDiff.GRanges[ ind1 ] ), 'vector' )
  flag_change = 0
  if( seqname != prev_chr ){
    if( seqname == 'chr2H' ){
      prev_chr = 'chr2H'
      gene_chr = gff.chr2H
      te_chr = te.chr2H
      flag_change = 1
    }
    if( seqname == 'chr3H' ){
      prev_chr = 'chr3H'
      gene_chr = gff.chr3H
      te_chr = te.chr3H
      flag_change = 1
    }
    if( seqname == 'chr4H' ){
      prev_chr = 'chr4H'
      gene_chr = gff.chr4H
      te_chr = te.chr4H
      flag_change = 1
    }
    if( seqname == 'chr5H' ){
      prev_chr = 'chr5H'
      gene_chr = gff.chr5H
      te_chr = te.chr5H
      flag_change = 1
    }
    if( seqname == 'chr6H' ){
      prev_chr = 'chr6H'
      gene_chr = gff.chr6H
      te_chr = te.chr6H
      flag_change = 1
    }
    if( seqname == 'chr7H' ){
      prev_chr = 'chr7H'
      gene_chr = gff.chr7H
      te_chr = te.chr7H
      flag_change = 1
    }
    if( seqname == 'chrUn' ){
      prev_chr = 'chrUn'
      gene_chr = gff.chrUn
      te_chr = te.chrUn
      flag_change = 1
    }
  }
  temp1 = gene_chr[ ( ( end( myDiff.GRanges[ ind1 ] ) < end( gene_chr ) + offset ) &
                    ( end( myDiff.GRanges[ ind1 ] ) > start( gene_chr ) - offset ) ) |
                  ( ( start( myDiff.GRanges[ ind1 ] ) < end( gene_chr ) + offset ) &
                    ( start( myDiff.GRanges[ ind1 ] ) > start( gene_chr ) - offset ) ) ) ]
  mid = ( start( myDiff.GRanges[ ind1 ] ) + end( myDiff.GRanges[ ind1 ] ) ) / 2
  temp2 = te_chr[ ( ( mid < te_chr$end ) & ( mid > te_chr$start ) ) |
                ( ( start( myDiff.GRanges[ ind1 ] ) < te_chr$end ) &
                  ( start( myDiff.GRanges[ ind1 ] ) > te_chr$start ) ) |
                ( ( end( myDiff.GRanges[ ind1 ] ) < te_chr$end ) &
                  ( end( myDiff.GRanges[ ind1 ] ) > te_chr$start ) ) ) ]

  temp3 = data.frame( chr = character(), start = integer(), end = integer(),
                    TE_code = character(), stringsAsFactors = FALSE )

  # If everything is empty then the none flag is set
  if( ( NROW( temp1 ) == 0 ) & ( NROW( temp2 ) == 0 ) )
  {
    none = 1
  } else {
    none = 0
  }
  qv = myDiff.GRanges$qvalue[ ind1 ]
  dv = myDiff.GRanges$meth.diff[ ind1 ]
  # If it is a gene, then want to know the annotation
  if( as( sum( temp1$type == 'gene' ) > 0, 'integer' ) )
  {
    temp4 = temp1[ temp1$type == 'gene' ]
  }
}

```

```

    geneID = temp4$gene_id[ 1 ]
  } else {
    geneID = ''
  }
# If it is a gene at the same time as a TE then the TE label with 'g' at the start
# is used rather than the TE label
if( ( as( sum( temp1$type == 'gene' ) > 0, 'integer' ) == 1 ) &
    ( NROW( temp2 ) != 0 ) )
{
  temp1 = temp1[ 0, ]
  temp3 = temp2
  temp2 = data.frame( chr = character(), start = integer(), end = integer(),
                     TE_code = character(), stringsAsFactors = FALSE )
df1 = data.frame( chr = seqname,
                 start = start( myDiff.GRanges[ ind1 ] ),
                 end = end( myDiff.GRanges[ ind1 ] ),
                 gene = as( sum( temp1$type == 'gene' ) > 0, 'integer' ),
                 ncrRNA_gene = as( sum( temp1$type == 'ncrRNA_gene' ) > 0, 'integer' ),
                 lnc_RNA = as( sum( temp1$type == 'lnc_RNA' ) > 0, 'integer' ),
                 trRNA = as( sum( temp1$type == 'trRNA' ) > 0, 'integer' ),
                 rRNA = as( sum( temp1$type == 'rRNA' ) > 0, 'integer' ),
                 snRNA = as( sum( temp1$type == 'snRNA' ) > 0, 'integer' ),
                 snoRNA = as( sum( temp1$type == 'snoRNA' ) > 0, 'integer' ),
                 pre_miRNA = as( sum( temp1$type == 'pre_miRNA' ) > 0, 'integer' ),
                 DHH = as( sum( temp2$TE_code == 'DHH' ) > 0, 'integer' ),
                 DTA = as( sum( temp2$TE_code == 'DTA' ) > 0, 'integer' ),
                 DTC = as( sum( temp2$TE_code == 'DTC' ) > 0, 'integer' ),
                 DTH = as( sum( temp2$TE_code == 'DTH' ) > 0, 'integer' ),
                 DTM = as( sum( temp2$TE_code == 'DTM' ) > 0, 'integer' ),
                 DTT = as( sum( temp2$TE_code == 'DTT' ) > 0, 'integer' ),
                 DTX = as( sum( temp2$TE_code == 'DTX' ) > 0, 'integer' ),
                 DXX = as( sum( temp2$TE_code == 'DXX' ) > 0, 'integer' ),
                 RIX = as( sum( temp2$TE_code == 'RIX' ) > 0, 'integer' ),
                 RLC = as( sum( temp2$TE_code == 'RLC' ) > 0, 'integer' ),
                 RLG = as( sum( temp2$TE_code == 'RLG' ) > 0, 'integer' ),
                 RLX = as( sum( temp2$TE_code == 'RLX' ) > 0, 'integer' ),
                 RSX = as( sum( temp2$TE_code == 'RSX' ) > 0, 'integer' ),
                 RXX = as( sum( temp2$TE_code == 'RXX' ) > 0, 'integer' ),
                 XXX = as( sum( temp2$TE_code == 'XXX' ) > 0, 'integer' ),
                 gDHH = as( sum( temp3$TE_code == 'DHH' ) > 0, 'integer' ),
                 gDTA = as( sum( temp3$TE_code == 'DTA' ) > 0, 'integer' ),
                 gDTC = as( sum( temp3$TE_code == 'DTC' ) > 0, 'integer' ),
                 gDTH = as( sum( temp3$TE_code == 'DTH' ) > 0, 'integer' ),
                 gDTM = as( sum( temp3$TE_code == 'DTM' ) > 0, 'integer' ),
                 gDTT = as( sum( temp3$TE_code == 'DTT' ) > 0, 'integer' ),
                 gDTX = as( sum( temp3$TE_code == 'DTX' ) > 0, 'integer' ),
                 gDXX = as( sum( temp3$TE_code == 'DXX' ) > 0, 'integer' ),
                 gRIX = as( sum( temp3$TE_code == 'RIX' ) > 0, 'integer' ),
                 gRLC = as( sum( temp3$TE_code == 'RLC' ) > 0, 'integer' ),
                 gRLG = as( sum( temp3$TE_code == 'RLG' ) > 0, 'integer' ),
                 gRLX = as( sum( temp3$TE_code == 'RLX' ) > 0, 'integer' ),
                 gRSX = as( sum( temp3$TE_code == 'RSX' ) > 0, 'integer' ),
                 gRXX = as( sum( temp3$TE_code == 'RXX' ) > 0, 'integer' ),
                 gXXX = as( sum( temp3$TE_code == 'XXX' ) > 0, 'integer' ),
                 none = none,
                 geneID = geneID,
                 qvalue = qv,
                 diff = dv )
  annotation = rbind( annotation, df1 )
}
write.table( annotation, file = paste0( file_in,
                                     '.fixed.annotation.withTE.geneIDs.mod3.offset', offset, '.tsv' ),
            sep = '\t', row.name = FALSE, col.name = TRUE, quote = FALSE )
annotationsums = colSums( annotation[ , -c( 1, 2, 3, 43, 44, 45 ) ] )
check = rowSums( annotation[ , -c( 1, 2, 3, 43, 44, 45 ) ] )
write.table( annotationsums, file = paste0( file_in,
                                     '.fixed.annotationsums.withTE.mod3.offset', offset, '.tsv' ), sep = '\t',
            col.name = FALSE, row.name = TRUE, quote = FALSE )

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's,
     elapsed time = ', temp_time[ 3 ], 's\n', sep = ' ' )
}
}
}

```

10.31 random_genome_sampling.R

This custom R script generates 100,000 non-overlapping 100 bp random out of the sites methylation mapped sites. Since the methylation information does not map to the entire genome, there may be some biases in methylation mapping, so comparisons should be done with sites that the methylation sequence data maps to.

```
library( 'methylKit' )
library( 'genomation' )
library( 'GenomicFeatures' )
library( 'data.table' )
library( 'compiler' )
enableJIT( 3 )

if( exists( 'alreadyread' ) == FALSE )
{
  gff      =      gffToGRanges(      '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE )
  genes    =      gffToGRanges(      '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Hordeum_vulgare.IBSC_v2.45.gff3.gz', ensembl = TRUE, zero.based = FALSE, filter = 'gene' )
  filter.types = c( 'gene', 'ncrRNA_gene', 'lnc_RNA', 'tRNA', 'rRNA', 'snRNA', 'snoRNA',
'pre_miRNA' )
  gff.filtered = gff[ !is.na( match( gff$type, filter.types ) ) ]
  rm( gff )

  te.table = fread( file = '/home/jason/Share/Reference_Genome/Hvu_dl_2019-12-
05/Barley_TE_annotation_v2_18Aug16.tsv', sep = '\t', header = TRUE )
  t2 = data.table( chr = te.table$seq_id, start = te.table$start, end = te.table$end, TE_code
= te.table$TE_code )
  te.chr1H = t2[ t2$chr == 'chr1H' ]
  te.chr2H = t2[ t2$chr == 'chr2H' ]
  te.chr3H = t2[ t2$chr == 'chr3H' ]
  te.chr4H = t2[ t2$chr == 'chr4H' ]
  te.chr5H = t2[ t2$chr == 'chr5H' ]
  te.chr6H = t2[ t2$chr == 'chr6H' ]
  te.chr7H = t2[ t2$chr == 'chr7H' ]
  te.chrUn = t2[ t2$chr == 'chrUn' ]
  rm( t2, te.table )

  t1 = genes[ seqnames( genes ) == 'chrchr1H' ]
  genes.chr1H = data.table( start = start( t1 ),
end = end( t1 ),
strand = as( strand( t1 ), 'vector' ) )
  t1 = genes[ seqnames( genes ) == 'chrchr2H' ]
  genes.chr2H = data.table( start = start( t1 ),
end = end( t1 ),
strand = as( strand( t1 ), 'vector' ) )
  t1 = genes[ seqnames( genes ) == 'chrchr3H' ]
  genes.chr3H = data.table( start = start( t1 ),
end = end( t1 ),
strand = as( strand( t1 ), 'vector' ) )
  t1 = genes[ seqnames( genes ) == 'chrchr4H' ]
  genes.chr4H = data.table( start = start( t1 ),
end = end( t1 ),
strand = as( strand( t1 ), 'vector' ) )
  t1 = genes[ seqnames( genes ) == 'chrchr5H' ]
  genes.chr5H = data.table( start = start( t1 ),
end = end( t1 ),
strand = as( strand( t1 ), 'vector' ) )
  t1 = genes[ seqnames( genes ) == 'chrchr6H' ]
  genes.chr6H = data.table( start = start( t1 ),
end = end( t1 ),
strand = as( strand( t1 ), 'vector' ) )
  t1 = genes[ seqnames( genes ) == 'chrchr7H' ]
  genes.chr7H = data.table( start = start( t1 ),
end = end( t1 ),
```

```

        strand = as( strand( t1 ), 'vector' ) )
t1 = genes[ seqnames( genes ) == 'chrchrUn' ]
genes.chrUn = data.table( start = start( t1 ),
        end = end( t1 ),
        strand = as( strand( t1 ), 'vector' ) )
gff.chr1H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr1H' ]
gff.chr2H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr2H' ]
gff.chr3H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr3H' ]
gff.chr4H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr4H' ]
gff.chr5H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr5H' ]
gff.chr6H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr6H' ]
gff.chr7H = gff.filtered[ seqnames( gff.filtered ) == 'chrchr7H' ]
gff.chrUn = gff.filtered[ seqnames( gff.filtered ) == 'chrchrUn' ]

}
alreadyread = 1

setwd( '/home/jason/Share/epiGBS_98_bismark/merged/' )
lr = 'R'
#lr = 'L'

for( filebase in c( '/home/jason/Share/epiGBS_98_bismark/merged/',
'/home/jason/Share/epiGBS_93_varcall/merged/' ) )
{
  for( randsampling in c( 20000 ) )
  {
    for( context1 in c( 'CG', 'CHG', 'CHH' ) )
    {
      ptm = proc.time()

      file_in = paste0( filebase, 'randomised_bs_mapped_sampling_omit_some_na', randsampling,
'.'. , context1, '.tsv' )
      readin = read.csv( file_in, sep = '\t' )
      colnm = colnames( readin )
      colnames( readin ) = c( 'seqname', 'start', 'end' )
      cat( paste0( file_in, '\n' ) )
      len1 = NROW( readin )
      prev_chr = 'chr1H'
      gene_chr = gff.chr1H
      te_chr = te.chr1H
      annotation = NULL
      for( ind1 in 1:len1 )
      {
        # create a data.table with columns:
        # chr, start, end, qvalue, meth.diff, dist_to_gene_start, overlapping_genes,
overlapping_other

        seqname = readin$seqname[ ind1 ]
        flag_change = 0
        if( seqname != prev_chr ){
          if( seqname == 'chr2H' ){
            prev_chr = 'chr2H'
            gene_chr = gff.chr2H
            te_chr = te.chr2H
            flag_change = 1
          }
          if( seqname == 'chr3H' ){
            prev_chr = 'chr3H'
            gene_chr = gff.chr3H
            te_chr = te.chr3H
            flag_change = 1
          }
          if( seqname == 'chr4H' ){
            prev_chr = 'chr4H'
            gene_chr = gff.chr4H
            te_chr = te.chr4H
            flag_change = 1
          }
          if( seqname == 'chr5H' ){
            prev_chr = 'chr5H'
            gene_chr = gff.chr5H
            te_chr = te.chr5H
            flag_change = 1
          }
          if( seqname == 'chr6H' ){
            prev_chr = 'chr6H'
            gene_chr = gff.chr6H

```



```

        te_chr = te.chr6H
        flag_change = 1
    }
    if( seqname == 'chr7H' ){
        prev_chr = 'chr7H'
        gene_chr = gff.chr7H
        te_chr = te.chr7H
        flag_change = 1
    }
    if( seqname == 'chrUn' ){
        prev_chr = 'chrUn'
        gene_chr = gff.chrUn
        te_chr = te.chrUn
        flag_change = 1
    }
}
temp1 = gene_chr[ ( ( readin$end[ ind1 ] < end( gene_chr ) + offset ) & (
readin$end[ ind1 ] > start( gene_chr ) - offset ) ) |
( ( readin$start[ ind1 ] < end( gene_chr ) + offset ) & (
readin$start[ ind1 ] > start( gene_chr ) - offset ) ) ]

mid = ( readin$start[ ind1 ] + readin$end[ ind1 ] ) / 2

temp2 = te_chr[ ( ( mid < te_chr$end ) & ( mid > te_chr$start ) ) |
( ( readin$start[ ind1 ] < te_chr$end ) & ( readin$start[ ind1 ] >
te_chr$start ) ) |
( ( readin$end[ ind1 ] < te_chr$end ) & ( readin$end[ ind1 ] >
te_chr$start ) ) ]

temp3 = data.frame( chr = character(), start = integer(), end = integer(), TE_code =
character(), stringsAsFactors = FALSE )

# as( sum( temp1$type == '___' ) > 0, 'integer' ) so that it is either 0 or 1, not 2
or more which can occur.

# if everything is empty then the none flag is set
if( ( NROW( temp1 ) == 0 ) & ( NROW( temp2 ) == 0 ) )
{
    none = 1
} else {
    none = 0
}
# if it is a gene at the same time as a TE then the TE label with 'g' at the start is
used rather than the TE label
if( ( as( sum( temp1$type == 'gene' ) > 0, 'integer' ) == 1 ) & ( NROW( temp2 ) != 0 )
)
{
    temp1 = temp1[ 0, ]
    temp3 = temp2
    temp2 = data.frame( chr = character(), start = integer(), end = integer(), TE_code =
character(), stringsAsFactors = FALSE )
}

df1 = data.frame( chr = seqname,
start = readin$start[ ind1 ],
end = readin$end[ ind1 ],
gene = as( sum( temp1$type == 'gene' ) > 0, 'integer' ),
ncrRNA_gene = as( sum( temp1$type == 'ncrRNA_gene' ) > 0, 'integer' ),
lnc_RNA = as( sum( temp1$type == 'lnc_RNA' ) > 0, 'integer' ),
tRNA = as( sum( temp1$type == 'tRNA' ) > 0, 'integer' ),
rRNA = as( sum( temp1$type == 'rRNA' ) > 0, 'integer' ),
snRNA = as( sum( temp1$type == 'snRNA' ) > 0, 'integer' ),
snoRNA = as( sum( temp1$type == 'snoRNA' ) > 0, 'integer' ),
pre_miRNA = as( sum( temp1$type == 'pre_miRNA' ) > 0, 'integer' ),
DHH = as( sum( temp2$TE_code == 'DHH' ) > 0, 'integer' ),
DTA = as( sum( temp2$TE_code == 'DTA' ) > 0, 'integer' ),
DTC = as( sum( temp2$TE_code == 'DTC' ) > 0, 'integer' ),
DTH = as( sum( temp2$TE_code == 'DTH' ) > 0, 'integer' ),
DTM = as( sum( temp2$TE_code == 'DTM' ) > 0, 'integer' ),
DTT = as( sum( temp2$TE_code == 'DTT' ) > 0, 'integer' ),
DTX = as( sum( temp2$TE_code == 'DTX' ) > 0, 'integer' ),
DXX = as( sum( temp2$TE_code == 'DXX' ) > 0, 'integer' ),
RIX = as( sum( temp2$TE_code == 'RIX' ) > 0, 'integer' ),
RLC = as( sum( temp2$TE_code == 'RLC' ) > 0, 'integer' ),
RLG = as( sum( temp2$TE_code == 'RLG' ) > 0, 'integer' ),
RLX = as( sum( temp2$TE_code == 'RLX' ) > 0, 'integer' ),
RSX = as( sum( temp2$TE_code == 'RSX' ) > 0, 'integer' ),

```

```

RXX = as( sum( temp2$TE_code == 'RXX' ) > 0, 'integer' ),
XXX = as( sum( temp2$TE_code == 'XXX' ) > 0, 'integer' ),
gDHH = as( sum( temp3$TE_code == 'DHH' ) > 0, 'integer' ),
gDTA = as( sum( temp3$TE_code == 'DTA' ) > 0, 'integer' ),
gDTC = as( sum( temp3$TE_code == 'DTC' ) > 0, 'integer' ),
gDTH = as( sum( temp3$TE_code == 'DTH' ) > 0, 'integer' ),
gDTM = as( sum( temp3$TE_code == 'DTM' ) > 0, 'integer' ),
gDTT = as( sum( temp3$TE_code == 'DTT' ) > 0, 'integer' ),
gDTX = as( sum( temp3$TE_code == 'DTX' ) > 0, 'integer' ),
gDXX = as( sum( temp3$TE_code == 'DXX' ) > 0, 'integer' ),
gRIX = as( sum( temp3$TE_code == 'RIX' ) > 0, 'integer' ),
gRLC = as( sum( temp3$TE_code == 'RLC' ) > 0, 'integer' ),
gRLG = as( sum( temp3$TE_code == 'RLG' ) > 0, 'integer' ),
gRLX = as( sum( temp3$TE_code == 'RLX' ) > 0, 'integer' ),
gRSX = as( sum( temp3$TE_code == 'RSX' ) > 0, 'integer' ),
gRXX = as( sum( temp3$TE_code == 'RXX' ) > 0, 'integer' ),
gXXX = as( sum( temp3$TE_code == 'XXX' ) > 0, 'integer' ),
none = none )
annotation = rbind( annotation, df1 )
}

write.table( annotation, file = paste0( file_in, '.fixed.annotation.withTE.mod2.offset',
offset, '.tsv' ), sep = '\t', row.name = FALSE, col.name = TRUE, quote = FALSE )
annotationsums = colSums( annotation[ , -c( 1, 2, 3 ) ] )
check = rowSums( annotation[ , -c( 1, 2, 3 ) ] )
write.table( annotationsums, file = paste0( file_in,
'.fixed.annotationsums.withTE.mod2.offset', offset, '.tsv' ), sep = '\t', col.name = FALSE,
row.name = TRUE, quote = FALSE )

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's, elapsed
time = ', temp_time[ 3 ], 's\n', sep = ' ' )
}
}
}

```

10.32 demultiplex.py from the epiGBS suite

The following section of code needed to be modified in order to work. The original python code segment from the function `put_output` with errors is shown in Figure B.10.1 with the modified code in Figure B.10.2. The unedited code can be found at the github repository https://github.com/thomasvangurp/epiGBS/blob/epiGBS-Nature_methods/demultiplex.py. The errors and fixes were reported in repository issue #23 on 7 December 2018, but they remain un-actioned at the time of writing.

```

seq1_name = '%(code)s_(Flowcell)s_s_(lane)s_fastq.txt' % \
({ 'code': 'R1samplecode123', 'Flowcell': Flowcell, 'lane': Lane})
seq2_name = '%(code)s_(Flowcell)s_s_(lane)s_fastq.txt' % \
({ 'code': 'R2samplecode123', 'Flowcell': Flowcell, 'lane': Lane})
if not os.exists(os.path.join(dir_in, seq1_name)):
    seq1_name += ".gz"
    seq2_name += ".gz"

```

Figure B.10.1, Section of code from the `put_output` function in `demultiplex.py` with errors preventing correct operation.

```

seq1_name = '%(code)s_(Flowcell)s_s_(lane)s_fastq.txt' % \
({'code': 'R1_%s' % opts.output.split('/')[ -2], 'Flowcell': Flowcell, 'lane': Lane})
seq2_name = '%(code)s_(Flowcell)s_s_(lane)s_fastq.txt' % \
({'code': 'R2_%s' % opts.output.split('/')[ -2], 'Flowcell': Flowcell, 'lane': Lane})
if not os.path.exists(os.path.join(dir_in, seq1_name)):
    seq1_name += ".gz"
    seq2_name += ".gz"

```

Figure B.10.2, Section of code from the put_output function in demultiplex.py fixed for correct operation.

10.33 chromosome_mapping_sRNA_meth.R

This R script takes statistically significantly differentially expressed 23/24 nt sRNA and CHH differential methylation and counts the occurrences in each 1 Mbp window in the genome. The resulting information is saved in an image.

```

library( data.table )
library( png )

hvuGenome = list( 'chr1H' = 558535432,
                  'chr2H' = 768075024,
                  'chr3H' = 699711114,
                  'chr4H' = 647060158,
                  'chr5H' = 670030160,
                  'chr6H' = 583380513,
                  'chr7H' = 657224000 )

# The data set in the file below is already filtered for only significant reads
# Loads in to the variable myDiff10p
load(
'~/Share/epiGBS_98_bismark/merged/R.methylKit.diff.4reps.mincov5.windowed100.step100.CHH.L.Rdata' )

bigwin = 10000
ptm = proc.time()
chrn = matrix( data = NA, nrow = 786, ncol = 21 )
indexmul = 1

chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H' )

siRNAarray = array( data = 0, dim = c( 786, 7 ) )
metharray = array( data = 0, dim = c( 786, 7 ) )

lim = 5
for( chrind in 1:7 )
{
  siRNA1 = fread( paste0( '~/Share/Jason_SmallRNA_40_Leaf_aligned_23-
24nt/n1.mapped.counts.window.100.step.100.stats.qvals.chr', chrind, 'H.tsv' ) )
  siRNA2 = siRNA1[ !is.na( siRNA1$mean_c ) ]
  siRNA2 = siRNA2[ siRNA2$pvals < 0.05 ]
  for( c1 in 1:NROW( siRNA2 ) )
  {
    siRNAarray[ ceiling( siRNA2[ c1 ]$index / bigwin ), chrind ] = siRNAarray[ ceiling(
siRNA2[ c1 ]$index / bigwin ), chrind ] + 1
  }

  methvals = myDiff10p[ myDiff10p$chr == chromosomes[ chrind ] ]
  for( c2 in 1:NROW( methvals ) )
  {
    metharray[ ceiling( methvals[ c2 ]$start / ( bigwin * 100 ) ), chrind ] = metharray[
ceiling( methvals[ c2 ]$start / ( bigwin * 100 ) ), chrind ] + 1
  }
}

```

```

}
}

img1 = array( data = 1, dim = c( 786, 680, 3 ) )
siRNAscaling = max( siRNAarray )
methscaling = max( metharray )

for( chrind in 1:7 )
{
  xstart = 25 * 3 * chrind - 25
  for( pos in 1:floor( as.numeric( hvuGenome[ chrind ] ) / 1000000 ) )
  {
    img1[ pos + 1, xstart:( xstart + 23 ), 1 ] = 1 - array( data = siRNAarray[ pos, chrind ] /
siRNAscaling, dim = c( 1, 24, 1 ) )
    img1[ pos + 1, xstart:( xstart + 23 ), 2 ] = img1[ pos + 1, xstart:( xstart + 23 ), 1 ]
    img1[ pos + 1, xstart:( xstart + 23 ), 3 ] = img1[ pos + 1, xstart:( xstart + 23 ), 1 ]
  }
  xstart = 25 * 3 * chrind - 0
  for( pos in 1:floor( as.numeric( hvuGenome[ chrind ] ) / 1000000 ) )
  {
    img1[ pos + 1, xstart:( xstart + 23 ), 1 ] = 1 - array( data = metharray[ pos, chrind ] /
methscaling, dim = c( 1, 24, 1 ) )
    img1[ pos + 1, xstart:( xstart + 23 ), 2 ] = img1[ pos + 1, xstart:( xstart + 23 ), 1 ]
    img1[ pos + 1, xstart:( xstart + 23 ), 3 ] = img1[ pos + 1, xstart:( xstart + 23 ), 1 ]
  }
}

# ticks for chromosome size
for( y in ( 0:7 ) * 100 + 2 )
{
  img1[ y, 44:48, 1 ] = array( data = 0, dim = c( 1, 5, 1 ) )
  img1[ y, 44:48, 2 ] = array( data = 0, dim = c( 1, 5, 1 ) )
  img1[ y, 44:48, 3 ] = array( data = 0, dim = c( 1, 5, 1 ) )
}

# colour scale bar
scsize = 500
for( sc in ( 0:scsize ) )
{
  ys = 600
  xs = 625
  img1[ ys - sc, xs:( xs + 23 ), 1 ] = array( data = ( scsize - sc ) / scsize, dim = c( 1, 24,
1 ) )
  img1[ ys - sc, xs:( xs + 23 ), 2 ] = array( data = ( scsize - sc ) / scsize, dim = c( 1, 24,
1 ) )
  img1[ ys - sc, xs:( xs + 23 ), 3 ] = array( data = ( scsize - sc ) / scsize, dim = c( 1, 24,
1 ) )
}

# ticks for colour scale bar
for( yt in c( ys, round( ys - scsize / 3 ), round( ys - scsize * 2 / 3 ), round( ys - scsize )
) )
{
  cat( paste0( 'yt = ', yt, ' xs = ', xs, '\n' ) )
  img1[ yt, ( xs - 6 ):( xs - 2 ), 1 ] = array( data = 0, dim = c( 1, 5, 1 ) )
  img1[ yt, ( xs - 6 ):( xs - 2 ), 2 ] = array( data = 0, dim = c( 1, 5, 1 ) )
  img1[ yt, ( xs - 6 ):( xs - 2 ), 3 ] = array( data = 0, dim = c( 1, 5, 1 ) )
  img1[ yt, ( xs + 25 ):( xs + 29 ), 1 ] = array( data = 0, dim = c( 1, 5, 1 ) )
  img1[ yt, ( xs + 25 ):( xs + 29 ), 2 ] = array( data = 0, dim = c( 1, 5, 1 ) )
  img1[ yt, ( xs + 25 ):( xs + 29 ), 3 ] = array( data = 0, dim = c( 1, 5, 1 ) )
}
rasterImage( as.raster( img1 ), 0, 0, 1, 1 )

writePNG( img1, target = '~/Share/chromosome_mapping_siRNA_methylation_04_CHH_Leaf.png' )

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's, elapsed time =
', temp_time[ 3 ], 's\n', sep = ' ' )

```

10.34 sRNA_methylation_co-location.py

This python program compares differential sRNA expression mapped to the genome with differential methylation and determined the distance between the two.

```
#!/usr/bin/python
#
# Looks for co-locations of differential siRNA expression and differential methylation
#
# Jason Smith
# Last modified 04/06/2020 (d/m/y)

import sys
import gzip
import os
import time
import csv

q_cutoff_methylation = 0.05
q_cutoff_siRNA = 0.05
q_cutoff_siRNA_st = ''
win_size = 100
step_size = 100
n_windows = 100

starttime = time.time()
if( sys.platform[0] == 'w' ):
    basedir = 'C:/BaseSpace/'
else:
    basedir = '/home/jason/Share/'

dir_in_m = basedir + 'epiGBS_98_bismark/merged/'
dir_out = basedir + 'corr_04_siRNA_methylation/'

chromosomes = [ 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' ]
chr_sizes = [ 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000,
249774706 ]
s = [ 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6',
'R7', 'R8' ]
contexts = [ 'CHH', 'CG', 'CHG' ]

for siRNA_sz in [ '23', '24', '23-24' ]:
    print( siRNA_sz )
    for lr in [ 'L', 'R' ]:
        print( lr )
        if( lr == 'L' ):
            if( siRNA_sz == '24' ):
                dir_in_s = basedir + 'Jason_SmallRNA_40_Leaf_aligned_24nt/'
            elif( siRNA_sz == '23' ):
                dir_in_s = basedir + 'Jason_SmallRNA_40_Leaf_aligned_23nt/'
            else:
                dir_in_s = basedir + 'Jason_SmallRNA_40_Leaf_aligned_23-24nt/'
        else:
            if( siRNA_sz == '24' ):
                dir_in_s = basedir + 'Jason_SmallRNA_40_Root_aligned_24nt/'
            elif( siRNA_sz == '23' ):
                dir_in_s = basedir + 'Jason_SmallRNA_40_Root_aligned_23nt/'
            else:
                dir_in_s = basedir + 'Jason_SmallRNA_40_Root_aligned_23-24nt/'
        print( dir_in_s )
        current_m_index = 0
        current_s_index = 0
        for context in contexts:
            chromosome = chromosomes[ 0 ]
            fh_s = open( dir_in_s + 'nl_mapped_counts.window.100.step.100.stats.qvals.' +
                chromosome + '.tsv' )
            header_s = fh_s.readline().strip().split()
            sinram = []
            while( True ):
                next_s = fh_s.readline().strip().split()
                if( next_s == [] ):
                    break
```

```

        break
    if( float( next_s[ 1 ] ) < q_cutoff_siRNA ):
        sinram.append( next_s )
fh_s.close()

fh_m = gzip.open( dir_in_m + 'R.methylKit.diff.4reps.mincov5.windowed100.step100.'
                 + context + '.' + lr + '.tsv.gz', 'rb' )
header_m = fh_m.readline().strip().split()
next_m = fh_m.readline().strip().split()
fo0 = open( dir_out + 'corr.windowed100.step100.siRNA' + siRNA_sz + 'nt.' +
           context + '.' + lr + q_cutoff_siRNA_st + '.0.tsv', 'wb' )

fop = []
fon = []
for ctl in range( 0, n_windows ):
    fop.append( open( dir_out + 'corr.windowed100.step100.siRNA' + siRNA_sz +
                   'nt.' + context + '.' + lr + q_cutoff_siRNA_st + '.p%d.tsv'
                   % ( ctl + 1 ), 'wb' ) )
    fon.append( open( dir_out + 'corr.windowed100.step100.siRNA' + siRNA_sz +
                   'nt.' + context + '.' + lr + q_cutoff_siRNA_st + '.n%d.tsv'
                   % ( ctl + 1 ), 'wb' ) )
header = 'chr\tmeth_start\tmeth_end\tmeth_q\tmeth_diff\tsiRNA_index\tsiRNA_q\t' +
        'siRNA_mean_c\tsiRNA_mean_t\ttsame_dir\n'
fo0.write( header )
for ctl in range( 0, n_windows ):
    fop[ctl].write( header )
    fon[ctl].write( header )
print( 'Context ' + context )
print( 'Chromosome ' + chromosome )
while( next_m != [] ):
    if( float( next_m[ 5 ] ) < q_cutoff_methylation ):
        if( next_m[ 0 ] != chromosome ):
            chromosome = next_m[ 0 ]
            print( 'Chromosome ' + chromosome )
            fh_s = open( dir_in_s +
                       'nl.mapped.counts.window.100.step.100.stats.qvals.' +
                       chromosome + '.tsv' )
            header_s = fh_s.readline().strip().split()
            sinram = []
            while( True ):
                next_s = fh_s.readline().strip().split()
                if( next_s == [] ):
                    break
                if( float( next_s[ 1 ] ) < q_cutoff_siRNA ):
                    sinram.append( next_s )
            fh_s.close()
            sramind = 0
            index_m = int( next_m[ 1 ] ) / win_size
            index_s = int( float( sinram[ sramind ][ 0 ] ) )
            # deal with negative windows first (starting at - n_windows)
            for ct2 in range( - n_windows, 0 ):
                if( index_s < index_m + ct2 ):
                    # Need to skip forward
                    while( index_s < index_m + ct2 ):
                        sramind += 1
                        if( sramind > len( sinram ) - 1 ):
                            sramind = len( sinram ) - 1
                            index_s = int( 1000000000 )
                        else:
                            index_s = int( float( sinram[ sramind ][ 0 ] ) )
            # Now the siRNA index is either in the right place or doesn't exist
            if( index_s == index_m + ct2 ):
                #print( 'index_s %d, index_m %d, ' % ( index_s, index_m ) )
                if( float( sinram[ sramind ][ 1 ] ) < q_cutoff_siRNA ):
                    # This is a valid point, write to file
                    if( ( ( float( next_m[ 6 ] ) > 0 ) &
                          ( float( sinram[ sramind ][ 5 ] ) >
                            float( sinram[ sramind ][ 4 ] ) ) ) |
                        ( ( float( next_m[ 6 ] ) < 0 ) &
                          ( float( sinram[ sramind ][ 5 ] ) <
                            float( sinram[ sramind ][ 4 ] ) ) ) ) ):
                        same = '1'
                    else:
                        same = '0'
                    fon[ - ct2 - 1 ].write( chromosome + '\t' + \
                                           next_m[ 1 ] + '\t' + \
                                           next_m[ 2 ] + '\t' + \
                                           next_m[ 5 ] + '\t' + \

```

```

next_m[ 6 ] + '\t' +\
sinram[ sramind ][ 0 ] + '\t' +\
sinram[ sramind ][ 1 ] + '\t' +\
sinram[ sramind ][ 4 ] + '\t' +\
sinram[ sramind ][ 5 ] + '\t' +\
same + '\n' )

# deal with the overlapping window
if( index_s < index_m ):
    # Need to skip forward
    while( index_s < index_m ):
        sramind += 1
        if( sramind > len( sinram ) - 1 ):
            sramind = len( sinram ) - 1
            index_s = int( 1000000000 )
        else:
            index_s = int( float( sinram[ sramind ][ 0 ] ) )
# Now the siRNA index is either in the right place or doesn't exist
if( index_s == index_m ):
    #print( 'index_s %d, index_m %d, ' % ( index_s, index_m ) )
    if( float( sinram[ sramind ][ 1 ] ) < q_cutoff_siRNA ):
        # This is a valid point, write to file
        if( ( ( float( next_m[ 6 ] ) > 0 ) &
            ( float( sinram[ sramind ][ 5 ] ) >
              float( sinram[ sramind ][ 4 ] ) ) ) |
            ( ( float( next_m[ 6 ] ) < 0 ) &
              ( float( sinram[ sramind ][ 5 ] ) <
                float( sinram[ sramind ][ 4 ] ) ) ) ) ):
            same = '1'
        else:
            same = '0'
        fo0.write( chromosome + '\t' +\
            next_m[ 1 ] + '\t' +\
            next_m[ 2 ] + '\t' +\
            next_m[ 5 ] + '\t' +\
            next_m[ 6 ] + '\t' +\
            sinram[ sramind ][ 0 ] + '\t' +\
            sinram[ sramind ][ 1 ] + '\t' +\
            sinram[ sramind ][ 4 ] + '\t' +\
            sinram[ sramind ][ 5 ] + '\t' +\
            same + '\n' )
# Now deal with the positive windows
for ct2 in range( 1, n_windows + 1 ):
    if( index_s < index_m + ct2 ):
        # Need to skip forward
        while( index_s < index_m + ct2 ):
            sramind += 1
            if( sramind > len( sinram ) - 1 ):
                sramind = len( sinram ) - 1
                index_s = int( 1000000000 )
            else:
                index_s = int( float( sinram[ sramind ][ 0 ] ) )
# Now the siRNA index is either in the right place or doesn't exist
if( index_s == index_m + ct2 ):
    #print( 'index_s %d, index_m %d, ' % ( index_s, index_m ) )
    if( float( sinram[ sramind ][ 1 ] ) < q_cutoff_siRNA ):
        # This is a valid point, write to file
        if( ( ( float( next_m[ 6 ] ) > 0 ) &
            ( float( sinram[ sramind ][ 5 ] ) >
              float( sinram[ sramind ][ 4 ] ) ) ) |
            ( ( float( next_m[ 6 ] ) < 0 ) &
              ( float( sinram[ sramind ][ 5 ] ) <
                float( sinram[ sramind ][ 4 ] ) ) ) ) ):
            same = '1'
        else:
            same = '0'
        fop[ ct2 - 1 ].write( chromosome + '\t' +\
            next_m[ 1 ] + '\t' +\
            next_m[ 2 ] + '\t' +\
            next_m[ 5 ] + '\t' +\
            next_m[ 6 ] + '\t' +\
            sinram[ sramind ][ 0 ] + '\t' +\
            sinram[ sramind ][ 1 ] + '\t' +\
            sinram[ sramind ][ 4 ] + '\t' +\
            sinram[ sramind ][ 5 ] + '\t' +\
            same + '\n' )

next_m = fh_m.readline().strip().split()

```

```

fo0.close()
for ct1 in range( 0, n_windows ):
    fop[ ct1 ].close()
    fon[ ct1 ].close()

endtime = time.time()
print( "End time - start time = %f" % ( endtime - starttime ) )

```

10.35 plot_barley_sRNA_meth_co-locations.R

This R script takes the output from sRNA_methylation_co-location.py and generates information for plotting the distance between significantly different windows in sRNA expression and methylation levels.

```

# This R script looks at the correlation output from the pypy script
# sRNA_methylation_co-location.py

ptm = proc.time()

library( compiler )
library( data.table )

enableJIT( 3 )

chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H' )
chr_sizes = c( 558535432, 768075024, 699711114, 647060158, 670030160, 583380513,
               657224000, 249774706 )
s = c( 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5',
        'R6', 'R7', 'R8' )

dir_in = 'C:/BaseSpace/corr_04_siRNA_methylation/'

shift_nums = c( paste0( 'n', 100:1 ), '0', paste0( 'p', 1:100 ) )

df1 = data.frame( lr = NULL, siRNA_size = NULL, context = NULL, shift_num = NULL,
                  same_dir = NULL, count = NULL )
df3 = data.frame( lr = NULL, siRNA_size = NULL, context = NULL, shift_num = NULL,
                  same_dir = NULL, count = NULL )
for( lr in c( 'L', 'R' ) )
{
  for( siRNA_sz in c( '23', '24', '23-24' ) )
  {
    for( context in c( 'CHH', 'CG', 'CHG' ) )
    {
      df2 = data.frame( lr = NULL, siRNA_size = NULL, context = NULL, shift_num = NULL,
                        same_dir = NULL, count = NULL )
      for( shift_num in shift_nums )
      {
        filein = paste0( dir_in, 'corr.windowed100.step100.siRNA', siRNA_sz, 'nt.', context,
                          '.', lr, '.', shift_num, '.tsv' )
        dat1 = fread( filein, header = TRUE )
        same_dir = sum( dat1$same_dir )
        cat( paste0( paste( lr, siRNA_sz, context ),
                    ', shift number ', shift_num,
                    ', same direction ', sprintf( same_dir / NROW( dat1 ) * 100,
                                                fmt = '%#.1f' ),
                    '%, count ', NROW( dat1 ),
                    '\n' ) )
        df1 = rbind( df1,
                     data.frame( lr = lr, siRNA_size = siRNA_sz, context = context,
                                  shift_num = shift_num, same_dir = same_dir,
                                  count = NROW( dat1 ) ) )
        df2 = rbind( df2,
                     data.frame( lr = lr, siRNA_size = siRNA_sz, context = context,
                                  shift_num = shift_num, same_dir = same_dir,
                                  count = NROW( dat1 ) ) )
      }
    }
  }
}

```



```

    }
    y = df2$same_dir / df2$count
    t1 = df2[ seq( 100, 1 ), ]
    t2 = df2[ 102:201, ]
    for( t3 in 1:100 )
    {
      t2$same_dir[ t3 ] = ( t2$same_dir[ t3 ] + t1$same_dir[ t3 ] ) / 2
      t2$count[ t3 ] = ( t2$count[ t3 ] + t1$count[ t3 ] ) / 2
    }

    df3 = rbind( df3, df2[ 101, ], t2 )
  }
}

write.table( df3, file = paste0( dir_in, 'dist.between.siRNA.methylation.tsv' ), sep = '\t',
quote = FALSE, row.names = FALSE )

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's, elapsed time =
', temp_time[ 3 ], 's\n', sep = '' )

```

10.36 annotate_co-located_sRNA_methylation.R

This R script annotates the 100 bp genomic windows where there is co-location between differential sRNA expression and differential DNA methylation. Only sites where the co-location occurs within two windows are considered.

```

# This R script looks at the co-location output from the pypy script
# smallRNA_methylation_co-location.py

ptm = proc.time()

library( compiler )
library( data.table )

enableJIT( 3 )

chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
chr_sizes = c( 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000,
249774706 )
s = c( 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6',
'R7', 'R8' )
contexts = c( 'CG', 'CHG', 'CHH' )

#Just looking at siRNA 200bp around the methylation
shifts = '200bp'
shift_nums = c( 'n2', 'n1',
'0',
'p1', 'p2' )

dir_in_m = 'C:/BaseSpace/epiGBS_98_bismark/merged/'
dir_in_c = 'C:/BaseSpace/corr_01_siRNA_methylation/'

df1 = data.frame( lr = NULL, siRNA_size = NULL, context = NULL, shift_num = NULL, same_dir =
NULL, count = NULL )
df3 = data.frame( lr = NULL, siRNA_size = NULL, context = NULL, shift_num = NULL, same_dir =
NULL, count = NULL )
for( lr in c( 'L', 'R' ) )
{
  for( siRNA_sz in c( '23', '24', '23-24' ) )
  {
    filein_m = paste0( dir_in_m, 'R.methylKit.diff.4reps.mincov5.windowed100.step100.CHH.',
lr, '.Rdata.fixed.annotation.withTE.geneIDs.mod3.offset1500.tsv' )
    cat( filein_m, '\n' )

```

```

dat_m = fread( filein_m, header = TRUE )
shift_c = 1
dat_c = NULL
t3 = data.frame( NULL )
for( shift_num in shift_nums )
{
  filein_c = paste0( dir_in_c, 'corr.windowed100.step100.siRNA', siRNA_sz, 'nt.CHH.', lr,
'.siRNACutoff40pc.', shift_num, '.tsv' )
  t1 = fread( filein_c, header = TRUE )
  t2 = t1[ t1$same_dir == 1, ]
  t3 = rbind( t3, t2 )
  #dat_c[[ shift_c ]] = t1
  #shift_c = shift_c + 1
}
t4 = data.frame( NULL )
t3.1 = t3[ t3$chr == 'chr1H', ]
t3.2 = t3[ t3$chr == 'chr2H', ]
t3.3 = t3[ t3$chr == 'chr3H', ]
t3.4 = t3[ t3$chr == 'chr4H', ]
t3.5 = t3[ t3$chr == 'chr5H', ]
t3.6 = t3[ t3$chr == 'chr6H', ]
t3.7 = t3[ t3$chr == 'chr7H', ]
t3.U = t3[ t3$chr == 'chrUn', ]
for( ind_m in 1: NROW( dat_m ) )
{
  if( ( dat_m$chr[ ind_m ] == 'chr1H' ) &
      ( sum( as.numeric( t3.1$meth_start ) == as.numeric( dat_m$start[ ind_m ] ) ) > 0 ) )
  {
    t4 = rbind( t4, dat_m[ ind_m, ] )
  } else if( ( dat_m$chr[ ind_m ] == 'chr2H' ) &
             ( sum( as.numeric( t3.2$meth_start ) == as.numeric( dat_m$start[ ind_m ] ) )
> 0 ) )
  {
    t4 = rbind( t4, dat_m[ ind_m, ] )
  } else if( ( dat_m$chr[ ind_m ] == 'chr3H' ) &
             ( sum( as.numeric( t3.3$meth_start ) == as.numeric( dat_m$start[ ind_m ] ) )
> 0 ) )
  {
    t4 = rbind( t4, dat_m[ ind_m, ] )
  } else if( ( dat_m$chr[ ind_m ] == 'chr4H' ) &
             ( sum( as.numeric( t3.4$meth_start ) == as.numeric( dat_m$start[ ind_m ] ) )
> 0 ) )
  {
    t4 = rbind( t4, dat_m[ ind_m, ] )
  } else if( ( dat_m$chr[ ind_m ] == 'chr5H' ) &
             ( sum( as.numeric( t3.5$meth_start ) == as.numeric( dat_m$start[ ind_m ] ) )
> 0 ) )
  {
    t4 = rbind( t4, dat_m[ ind_m, ] )
  } else if( ( dat_m$chr[ ind_m ] == 'chr6H' ) &
             ( sum( as.numeric( t3.6$meth_start ) == as.numeric( dat_m$start[ ind_m ] ) )
> 0 ) )
  {
    t4 = rbind( t4, dat_m[ ind_m, ] )
  } else if( ( dat_m$chr[ ind_m ] == 'chr7H' ) &
             ( sum( as.numeric( t3.7$meth_start ) == as.numeric( dat_m$start[ ind_m ] ) )
> 0 ) )
  {
    t4 = rbind( t4, dat_m[ ind_m, ] )
  } else if( ( dat_m$chr[ ind_m ] == 'chrUn' ) &
             ( sum( as.numeric( t3.U$meth_start ) == as.numeric( dat_m$start[ ind_m ] ) )
> 0 ) )
  {
    t4 = rbind( t4, dat_m[ ind_m, ] )
  }
}

fileout = paste0( dir_in_c, 'annotate.corr.siRNA', siRNA_sz, '.methylation.', lr,
'.shifts', shifts, '.tsv' )
write.table( t4, file = fileout, sep = '\t', quote = FALSE, row.names = FALSE )
t5 = t4[ , -c( 1, 2, 3, 43, 44, 45 ) ]
t6 = colSums( t5 )
fileout = paste0( dir_in_c, 'annotatesums.corr.siRNA', siRNA_sz, '.methylation.', lr,
'.shifts', shifts, '.tsv' )
write.table( t6, file = fileout, sep = '\t', quote = FALSE, row.names = TRUE, col.names =
FALSE )

```

```

}
}

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's, elapsed time = ', temp_time[ 3 ], 's\n', sep = ' ' )

```

10.37 methylation_co-location_mRNA.R

This R script annotates and plots the differential methylation and mRNA differentially expressed sites.

```

# This R script looks at the co-location output from the pypy script
# smallRNA_methylation_co-location.py

ptm = proc.time()

library( compiler )
library( ggplot2 )
library( data.table )
library( GenomicRanges )

enableJIT( 3 )

dir1 = 'C:/BaseSpace/corr_02_methylation_mRNA/'
#dir1 = '~/Share/corr_02_methylation_mRNA/'
#lr = 'L'
lr = 'R'
context = 'CHH'
chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
chr_sizes = c( 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000, 249774706 )
s = c( 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6', 'R7', 'R8' )
contexts = c( 'CG', 'CHG', 'CHH' )

load( 'C:/BaseSpace/Reference_Genome/Hvu_dl_2019-12-05/Hordeum_vulgare.IBSC_v2.45.gff3.genes.Rdata' )
load( 'C:/BaseSpace/Reference_Genome/Hvu_dl_2019-12-05/Hordeum_vulgare.IBSC_v2.45.gff3.Rdata' )
)
gff$gene_id[ is.na( gff$gene_id ) ] = ''
extend = 2000
filein = paste0( dir1, 'methylation_location_correlation_with_mRNA_DE.tsv' )
cat( filein, '\n' )
dat = fread( filein, header = TRUE )
mask = !( is.na( dat$mRNA_log2fc ) )
dat = dat[ mask, ]

lr = 'L'
lr = 'R'
context = 'CHH'
if( lr == 'L' )
{
  tissue = 'leaf'
} else {
  tissue = 'root'
}

mask = dat$tissue == tissue
dat_lr = dat[ mask, ]
mask = dat_lr$context == context
dat_context = dat_lr[ mask, ]

udie = NULL
cl = NULL
for( i2 in 1:NROW( dat_context ) )
{

```

```

testgff = gff[ gff$gene_id == dat_context[ i2]$geneID ]
testgff2 = gff[ seqnames( gff ) == as.character( seqnames( testgff ) ) ]
testgff3 = testgff2[ ( start( testgff2 ) >= start( testgff ) ) & ( end( testgff2 ) <= end(
testgff ) ) ]
testgff4 = testgff3[ testgff3$type == 'exon' ]

if( NROW( testgff4 ) == 0 )
{
  cat( i2, dat_context[ i2]$geneID, 'no exons found\n' )
  l_introns = IRanges( start = NULL, end = NULL )
  l_exons = IRanges( start = NULL, end = NULL )
  testgff4 = testgff[ testgff$type == 'gene' ]
  if( NROW( testgff4 ) == 0 )
  {
    cat( 'Couldn\'t find the gene either \n' )

  } else {
    if( as.character( strand( testgff4 ) [ 1 ] ) == '+' )
    {
      upstream = IRanges( start = min( start( testgff4 [ 1 ] ) ) - extend, end = min( start(
testgff4 [ 1 ] ) ) - 1 )
      downstream = IRanges( start = max( end( testgff4 [ 1 ] ) ) + 1, end = max( end(
testgff4 [ 1 ] ) ) + extend )
    } else if( as.character( strand( testgff4 ) [ 1 ] ) == '-' ) {
      downstream = IRanges( start = min( start( testgff4 [ 1 ] ) ) - extend, end = min(
start( testgff4 [ 1 ] ) ) - 1 )
      upstream = IRanges( start = max( end( testgff4 [ 1 ] ) ) + 1, end = max( end( testgff4 [
1 ] ) ) + extend )
    }
  }
} else {
  cat( i2, 'of', NROW( dat_context ), 'finding', dat_context[ i2]$geneID, '\n' )
  l_exons = ranges( testgff4 )

  l_introns = IRanges( start = NULL, end = NULL )

  if( NROW( l_exons ) > 1 )
  {
    l_exons_ends = end( l_exons )
    l_exons_starts = start( l_exons )
    # remove duplicates
    if( NROW( l_exons_ends > 0 ) )
    {
      if( NROW( which( duplicated( l_exons_ends ) ) ) > 0 )
      {
        l_exons_ends = sort( l_exons_ends[ - which( duplicated( l_exons_ends ) ) ] )
      }
      if( NROW( which( duplicated( l_exons_starts ) ) ) > 0 )
      {
        l_exons_starts = sort( l_exons_starts[ - which( duplicated( l_exons_starts ) ) ] )
      }
    }
    ts = NULL
    te = NULL
    for( il in 1:( NROW( l_exons_ends ) - 1 ) )
    {
      if( sum( l_exons_starts >= l_exons_ends[ il ] + 1 ) > 0 )
      {
        ts = c( ts, l_exons_ends[ il ] + 1 )
        te = c( te, min( l_exons_starts[ l_exons_starts >= l_exons_ends[ il ] + 1 ] ) - 1 )
      }
    }
    l_introns = IRanges( start = ts, end = te )
  }

  if( as.character( strand( testgff4 ) [ 1 ] ) == '+' )
  {
    upstream = IRanges( start = min( start( l_exons ) ) - extend, end = min( start( l_exons
) ) - 1 )
    downstream = IRanges( start = max( end( l_exons ) ) + 1, end = max( end( l_exons ) ) +
extend )
  } else if( as.character( strand( testgff4 ) [ 1 ] ) == '-' ) {
    downstream = IRanges( start = min( start( l_exons ) ) - extend, end = min( start(
l_exons ) ) - 1 )
    upstream = IRanges( start = max( end( l_exons ) ) + 1, end = max( end( l_exons ) ) +
extend )
  }
}

```

```

# check:
#   is it in upstream
#   is it in downstream
#   is there any exon information, if not then leave blank
#   is it in the first exon
#   is it in any other exon
#   is it in the first intron
#   anywhere else is an intron

classification = ''
centre = ( dat_context[ i2]$start + dat_context[ i2]$end + 1 ) / 2
if( ( centre >= start( upstream ) ) & ( centre <= end( upstream ) ) )
{
  classification = 'upstream'
} else if( ( centre >= start( downstream ) ) & ( centre <= end( downstream ) ) )
{
  classification = 'downstream'
} else if( NROW( l_exons ) != 0 ){
  # it is either in an intron, an exon, or something went wrong
  if( ( centre >= start( l_exons[ 1 ] ) ) & ( centre <= end( l_exons[ 1 ] ) ) )
  {
    classification = 'firstexon'
  } else {
    mask = ( start( l_exons ) <= centre ) & ( end( l_exons ) >= centre )
    t1 = l_exons[ mask ]
    if( NROW( t1 ) != 0 )
    {
      classification = 'laterexon'
    } else {
      if( ( centre >= start( l_introns[ 1 ] ) ) & ( centre <= end( l_introns[ 1 ] ) ) )
      {
        classification = 'firstintron'
      } else {
        mask = ( start( l_introns ) <= centre ) & ( end( l_introns ) >= centre )
        t1 = l_introns[ mask ]
        if( NROW( t1 ) != 0 )
        {
          classification = 'laterintron'
        }
      }
    }
  }
}
cl = c( cl, classification )
cat( classification, '\n' )
}

}

new1 = cbind( dat_context, cl )
fname = paste0( dir1, 'classification_of_methylation_location_in_gene_', tissue, '_', context,
'.tsv' )
write.table( new1, file = fname, row.names = FALSE, col.names = TRUE, quote = FALSE, sep =
'\t' )

if( lr == 'L' )
{
  dat = fread( paste0( dir1, 'classification_of_methylation_location_in_gene_leaf_CHH.tsv' ) )
  tissue = 'Leaf'
} else {
  dat = fread( paste0( dir1, 'classification_of_methylation_location_in_gene_root_CHH.tsv' ) )
  tissue = 'Root'
}
context = 'CHH'
chromosomes = c( 'chr1H', 'chr2H', 'chr3H', 'chr4H', 'chr5H', 'chr6H', 'chr7H', 'chrUn' )
chr_sizes = c( 558535432, 768075024, 699711114, 647060158, 670030160, 583380513, 657224000,
249774706 )
s = c( 'L1', 'L2', 'L3', 'L4', 'L5', 'L6', 'L7', 'L8', 'R1', 'R2', 'R3', 'R4', 'R5', 'R6',
'R7', 'R8' )
contexts = c( 'CG', 'CHG', 'CHH' )

theme_set( theme_bw( base_size = 50 ) )
dset = 'all genes'
dsetf = 'all_genes'

```

```

simp = 'full'
gg = ggplot( dat, aes( x = meth_diff, y = mRNA_log2fc ) ) +
  geom_point( aes( col = cl ), alpha = 0.7, size = 5 ) + theme( legend.position = 'none' ) +
  labs( x = 'Methylation change (%)',
        y = expression("Messenger RNA "*log[2]*"(fold change)" ),
        title = paste( tissue, context, dset ),
        color = 'Gene region' ) +
  scale_colour_manual( name = 'Gene region',
                      values = c( 'downstream' = 'red',
                                   'firstintron' = 'black',
                                   'laterintron' = 'darkgrey',
                                   'firstexon' = 'green',
                                   'laterexon' = 'lightgreen',
                                   'upstream' = 'blue' ) )

plot( gg )
ggsave( filename = paste0( dir1, 'n1.scatterplot.nolegend.meth.mRNA.', tissue, '.', context,
                          '.', dsetf, '.', simp, '.png' ), device = 'png', dpi = 160, width = 20, height = 16)

cat( paste( tissue, dsetf, simp, '\n' ) )
features = levels( factor( dat$cl ) )
for( feature in features )
{
  ta = dat[ dat$cl == feature, ]
  ctest = cor.test( ta$mRNA_log2fc, ta$meth_diff, method = 'pearson' )
  cat( paste( feature, 'p-value', ctest$p.value, 'correlation', ctest$estimate, '\n' ) )
}

dat2 = dat
dat2$cl[ dat$cl == 'firstexon' ] = 'gene'
dat2$cl[ dat$cl == 'laterexon' ] = 'gene'
dat2$cl[ dat$cl == 'firstintron' ] = 'gene'
dat2$cl[ dat$cl == 'laterintron' ] = 'gene'

simp = 'simplified'
gg = ggplot( dat2, aes( x = meth_diff, y = mRNA_log2fc ) ) +
  geom_point( aes( col = cl ), alpha = 0.7, size = 5 ) + theme( legend.position = 'none' ) +
  labs( x = 'Methylation change (%)',
        y = expression("Messenger RNA "*log[2]*"(fold change)" ),
        title = paste( tissue, context, dset ),
        color = 'Gene region' ) +
  scale_colour_manual( name = 'Gene region',
                      values = c( 'downstream' = 'red', 'gene' = 'black', 'upstream' = 'blue'
                                   ) )
plot( gg )
ggsave( filename = paste0( dir1, 'n1.scatterplot.nolegend.meth.mRNA.', tissue, '.', context,
                          '.', dsetf, '.', simp, '.png' ), device = 'png', dpi = 160, width = 20, height = 16)

cat( paste( tissue, dsetf, simp, '\n' ) )
features = levels( factor( dat2$cl ) )
for( feature in features )
{
  ta = dat2[ dat2$cl == feature, ]
  ctest = cor.test( ta$mRNA_log2fc, ta$meth_diff, method = 'pearson' )
  cat( paste( feature, 'p-value', ctest$p.value, 'correlation', ctest$estimate, '\n' ) )
}

dset = 'targeting genes, not TEs'
dsetf = 'genes_no_TE'
datnoTE = dat[ dat$gene == 1, ]
dat2noTE = dat2[ dat2$gene == 1, ]

simp = 'full'
gg = ggplot( datnoTE, aes( x = meth_diff, y = mRNA_log2fc ) ) +
  geom_point( aes( col = cl ), alpha = 0.7, size = 5 ) + theme( legend.position = 'none' ) +
  labs( x = 'Methylation change (%)',
        y = expression("Messenger RNA "*log[2]*"(fold change)" ),
        title = paste( tissue, context, dset ),
        color = 'Gene region' ) +
  scale_colour_manual( name = 'Gene region',
                      values = c( 'downstream' = 'red',
                                   'firstintron' = 'black',
                                   'laterintron' = 'darkgrey',
                                   'firstexon' = 'green',
                                   'laterexon' = 'lightgreen',
                                   'upstream' = 'blue' ) )

```

```

plot( gg )
ggsave( filename = paste0( dir1, 'nl.scatterplot.nolegend.meth.mRNA.', tissue, '.', context,
'.'. , dsetf, '.', simp, '.png'), device = 'png', dpi = 160, width = 20, height = 16)

cat( paste( tissue, dsetf, simp, '\n' ) )
features = levels( factor( datnoTE$c1 ) )
for( feature in features )
{
  ta = datnoTE[ datnoTE$c1 == feature, ]
  ctest = cor.test( ta$mRNA_log2fc, ta$meth_diff, method = 'pearson' )
  cat( paste( feature, 'p-value', ctest$p.value, 'correlation', ctest$estimate, '\n' ) )
}

simp = 'simplified'
gg = ggplot( dat2noTE, aes( x = meth_diff, y = mRNA_log2fc ) ) +
  geom_point( aes( col = c1 ), alpha = 0.7, size = 5 ) + theme( legend.position = 'none' ) +
  labs( x = 'Methylation change (%)',
        y = expression("Messenger RNA "*log[2]*"(fold change)" ),
        title = paste( tissue, context, dset ),
        color = 'Gene region' ) +
  scale_colour_manual( name = 'Gene region',
                      values = c( 'downstream' = 'red', 'gene' = 'black', 'upstream' = 'blue' ) )
} )
plot( gg )
ggsave( filename = paste0( dir1, 'nl.scatterplot.nolegend.meth.mRNA.', tissue, '.', context,
'.'. , dsetf, '.', simp, '.png'), device = 'png', dpi = 160, width = 20, height = 16)

cat( paste( tissue, dsetf, simp, '\n' ) )
features = levels( factor( dat2noTE$c1 ) )
for( feature in features )
{
  ta = dat2noTE[ dat2noTE$c1 == feature, ]
  ctest = cor.test( ta$mRNA_log2fc, ta$meth_diff, method = 'pearson' )
  cat( paste( feature, 'p-value', ctest$p.value, 'correlation', ctest$estimate, '\n' ) )
}

dset = 'targeting genes and TEs'
dsetf = 'genes_with_TE'
datwithTE = dat[ dat$gene == 0, ]
dat2withTE = dat2[ dat2$gene == 0, ]

simp = 'full'
gg = ggplot( datwithTE, aes( x = meth_diff, y = mRNA_log2fc ) ) +
  geom_point( aes( col = c1 ), alpha = 0.7, size = 5 ) + theme( legend.position = 'none' ) +
  labs( x = 'Methylation change (%)',
        y = expression("Messenger RNA "*log[2]*"(fold change)" ),
        title = paste( tissue, context, dset ),
        color = 'Gene region' ) +
  scale_colour_manual( name = 'Gene region',
                      values = c( 'downstream' = 'red',
                                  'firstintron' = 'black',
                                  'laterintron' = 'darkgrey',
                                  'firstexon' = 'green',
                                  'laterexon' = 'lightgreen',
                                  'upstream' = 'blue' ) )
}
plot( gg )
ggsave( filename = paste0( dir1, 'nl.scatterplot.nolegend.meth.mRNA.', tissue, '.', context,
'.'. , dsetf, '.', simp, '.png'), device = 'png', dpi = 160, width = 20, height = 16)

cat( paste( tissue, dsetf, simp, '\n' ) )
features = levels( factor( datwithTE$c1 ) )
for( feature in features )
{
  ta = datwithTE[ datwithTE$c1 == feature, ]
  #cor( ta$mRNA_log2fc, ta$meth_diff, method = 'pearson' )
  ctest = cor.test( ta$mRNA_log2fc, ta$meth_diff, method = 'pearson' )
  cat( paste( feature, 'p-value', ctest$p.value, 'correlation', ctest$estimate, '\n' ) )
}

simp = 'simplified'
gg = ggplot( dat2withTE, aes( x = meth_diff, y = mRNA_log2fc ) ) +
  geom_point( aes( col = c1 ), alpha = 0.7, size = 5 ) + theme( legend.position = 'none' ) +
  labs( x = 'Methylation change (%)',
        y = expression("Messenger RNA "*log[2]*"(fold change)" ),
        title = paste( tissue, context, dset ),

```

```

        color = 'Gene region' ) +
    scale_colour_manual( name = 'Gene region',
                        values = c( 'downstream' = 'red', 'gene' = 'black', 'upstream' = 'blue'
    ) )
plot( gg )
ggsave( filename = paste0( dir1, 'nl.scatterplot.nolegend.meth.mRNA.', tissue, '.', context,
    '.', dsetf, '.', simp, '.png'), device = 'png', dpi = 160, width = 20, height = 16)

cat( paste( tissue, dsetf, simp, '\n' ) )
features = levels( factor( dat2withTE$c1 ) )
for( feature in features )
{
    ta = dat2withTE[ dat2withTE$c1 == feature, ]
    #cor( ta$mRNA_log2fc, ta$meth_diff, method = 'pearson' )
    ctest = cor.test( ta$mRNA_log2fc, ta$meth_diff, method = 'pearson' )
    cat( paste( feature, 'p-value', ctest$p.value, 'correlation', ctest$estimate, '\n' ) )
}

temp_time = proc.time() - ptm
cat( 'user time = ', temp_time[ 1 ], 's, system time = ', temp_time[ 2 ], 's, elapsed time =
', temp_time[ 3 ], 's\n', sep = '' )

```