Flinders University

Transcript Annotation and Seasonal Gene Expression in Kidney Tissue of the Australian Pygmy Bluetongue Lizard, *Tiliqua adelaidensis*

By

Carmel Margaret Maher
BSc (Hons)

Thesis
Submitted to Flinders University
For the degree of

Doctor of Philosophy
College of Science and Engineering
17th January 2024

Contents

Thesis Summary

Climate change is a key threatening process for endangered species world-wide. It is impacting habitat quality and potentially causing range shifts in many species as environmental conditions continue to change. The importance of assessing the genetic health of a population alongside ecological measures has been at the forefront of genetic studies in non-model organisms. Alongside other genetic measures of population diversity or measures of selection and local adaptation, RNA-seq can provide insights into which functional genes are present and being transcribed. Evidence for differential gene expression under certain conditions can also indicate how equipped an individual is to adapt to changing conditions in the shorter term and identify genes likely to be important in surviving such environmental stressors or those that may play a role in increased plasticity. Understanding what genomic regions may be involved in withstanding climate change may offer clues about a species' survival, either *in situ* or when moving animals to new environments as part of assisted dispersal or translocation efforts.

The Australian pygmy bluetongue lizard, *Tiliqua adelaidensis*, an endangered species endemic to South Australia, has an extremely low capacity for dispersal and exists in a highly fragmented habitat resulting from human modification of the landscape through development and cropping. Climate projections for South Australia include increasing aridity, higher-than-average maximum temperatures and lower rainfall, and the suitable range for this species has been projected to shift south under increasing impacts of climate change. Long-term conservation of the species in a warming and drying climate will ultimately rely on translocation efforts and the species' own ability to rapidly adapt to changing conditions.

The primary aim of this thesis was to identify and analyse genes that may play a role in these lizards surviving hotter, drier summers in the short term, and in the species' ability to survive future climate change by withstanding changes to its physical environment, such as low water availability and increased temperatures. Chapter 3 outlines two methods used to generate a transcript set: assembly of *de novo* transcripts; and processing of full-length long-read RNA-seq data. The *de novo* short-read assemblies generated a larger number of transcripts and scored higher on a measure of completeness, particularly in samples with more sequenced reads. However, the long-read transcripts expressed in the kidney tissue of *T. adelaidensis* provided full-length transcripts that could be used to annotate putative genes and conduct gene ontology analyses. These

annotated transcripts provide a reference for functional genes in this species and can be applied in future genomic research with a focus on renal function and drought resistance. A number of functional genes were implicated in renal processes and may determine how the species maintains water homeostasis in an arid climate; these were thus tested for differential gene expression.

The null hypothesis under test in Chapter 4 was that there would be no change in gene expression between groups of pygmy bluetongue individuals collected in different seasons and subject to different environmental factors. Kidney tissue samples were chosen as a focus for this study because of the involvement of kidneys in processes such as water retention. This makes them crucial organs considering our interest in genes associated with surviving the dry season. Gene expression was measured using RNA-seq technologies for eight individuals; four collected in September at the beginning of the Austral spring (the end of the Kaurna season Kudlila, the wet season) and four in late March/early April at the beginning of the Austral autumn (the end of the Kaurna season Warltati, the hot and dry season). Genes shown to be under differential expression included those for solute carriers, protein transporters, other protein binding and DNA or RNA transcription and repair factors, and a number of other receptors. These data will provide a much-needed reference for identifying and targeting other important adaptive genes expressed in the kidney of *T. adelaidensis* for future projects and will support concurrent studies into translocation viability and environmental response in pygmy bluetongues.

Declaration

I certify that this thesis does not incorporate, without acknowledgment, any material previously submitted for a degree or diploma in any university, and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University, and to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Carmel Maher

Acknowledgements

There are many people who have supported me throughout my PhD studies. I would like to thank my primary supervisor Mike Gardner for supporting me all this time, and acknowledge the huge amount of effort he put into the Lab of Evolutionary Genetics and Sociality (LEGS) after the passing of Professor Mike Bull. I also thank my adjunct supervisor Terry Bertozzi, who guided me through developing skills in coding and bioinformatics even before he signed on officially. Also, Mike Bull himself, who was a co-supervisor on my project in the beginning, and whose undergraduate teaching was one of the reasons I was set on this path in the first place. I would like to thank Mark Hutchinson of the SA Museum who assisted in sample collection for this study, the Nature Foundation of South Australia for providing access to our study site at Tiliqua Reserve and Tessa Bradford, an invaluable source of guidance in the lab at SARFMEE.

I am grateful for the support of the LEGS and all its members who have passed through during my time at Flinders university, but particularly Jess Clayton and Amy Slender, who as more senior members of the lab set the tone and provided invaluable support at Flinders and at the SARFMEE labs respectively. Thanks go to Ben Parslow, Olivia Davies, James Dorey and other members of the 'Bee lab' who helped foster a friendly collaborative environment and involvement in college societies. The studies of Lucy Clive and Bonnie Derne provided me an outlet and an opportunity to get out in the field. Last but not least, Robbie O'Reilly, as one of the few in our lab doing similar research on a similar time-frame became a source of solidarity and a sounding board while ironing out this final version; not forgetting Tara Daniell, René Campbell, Gerrut Norval, Dee Trewartha, Molly Stuart, Lucas Hearn and all the other members.

Outside university, I'm grateful to my parents and family for putting up with my perpetual student status. Sarah Churcher, Lottie Servin and Felicity Van Der Wijngaart, Ronan Ruggles and Tessa Hurley have all supported and grounded me in any number of ways. Thanks to Charlie, my galah, for always ensuring I got up at what he deemed was the appropriate time in the mornings, and Napoleon, our dog, and my hype-man, who is always keen to ensure I don't spend too many hours at the computer.

I acknowledge and pay respect to the Traditional Owners, past and present, on whose lands this study was conducted and facilitated. The Kaurna people are the Traditional Custodians of the Adelaide Plains and the land on which Flinders University, as well as SA Museum and the labs used during this study are built. I acknowledge the Traditional Owners and Custodians of the lands containing the Australian pygmy bluetongue's current and historical range: the Traditional Lands of the Kaurna and Ngadjuri people.

Two *T. adelaidensis* juveniles - Carmel Maher

Funding Support

The Australian Government Research Training Program Scholarship supported this study through tuition fees offset and a stipend for general living costs.

The Nature Foundation of South Australia Student Start-up Fund provided support for laboratory and sequencing costs 2016–18.

The Field Naturalists Society of South Australia Lirabenda Research Endowment provided support for laboratory and sequencing costs 2017–19.

Australian Research Council Linkage Projects LP120100426 and LP190100071 supported field work and tissue collection in the first years of this study.

The Holsworth Wildlife Research Endowment (now administrated by the Ecological Society of Australia) supported field work, tissue collection and initial lab costs in the first years of this study.

The Australian Mathematical Sciences Institute (AMSI) CHOOSEMATHS travel grant 2017 and Flinders University Students Association (FUSA) development travel grant 2018 supported attendance at bioinformatics conferences and associated professional development opportunities.

The Flinders University College of Science and Engineering supported attendance at local symposia and workshops.

***

This research was aided by use of the Nectar Research Cloud, a collaborative Australian research platform supported by the Australian Research Data Commons and National Collaborative Research Infrastructure Strategy

Animal Ethics

Governing body: Flinders University Animal Welfare Committee

Application number: E405 'Genomic diversity of pygmy bluetongues' approved for the period 1 September 2014–31 December 2017

The use of animals in this study adheres to the State Government of South Australia's *Animal Welfare act 1985* and the *Australian Code for the Care and use of Animals for Scientific Purposes, 8th edn, 2013*. As per the approved Flinders University College of Science and Engineering standard operating procedure for working with lizards, humane killing of sampled individuals was conducted using an intraperitoneal (intracoelomic) injection of pentobarbitone sodium (60 mg/ml) at a dose rate of 180 mg/kg (0.3 mL/100 g bodyweight).

Scientific Research Permit

Governing body: Government of South Australia, Department of Environment, Water and Natural Resources (now the Department for Environment and Water).

DEWNR Permit number: M26341-1

Native animals in this study were used in accordance with the *National Parks and Wildlife Act 1972* (SA). After tissue sampling, the five most intact carcasses were accessioned into the South Australian Museum Herpetology collections: voucher numbers SAMR68046, SAMR68380, SAMR68381, SAMR68963 and SAMR69488.

List of Conference Presentations Resulting from this Research

Mike Bull Symposium Presentation	06 October 2017
Flinders University Science and Engineering HDR Conference Presentation	29 November 2017
Australian Herpetological Society Conference (Perth) Presentation	22 June 2018
Field Naturalists of South Australia - Lirabenda Research Endowment; Grant Conditions Presentation	11 July 2018
Nature Foundation of South Australia, 2018 AGM; grant conditions Combined *T. adelaidensis* Presentation (With Lucy Clive and Bonnie Derne, Flinders University)	22 October 2018
Australian Herpetological Society Conference (Brisbane) Presentation	12 December 2018
9th World Congress of Herpetology Poster	05 January 2020 – 11 January 2020

1. Thesis Introduction

Tiliqua adelaidensis - C. Maher

1.1 Climate Change in Arid Australian Habitats

Australia has a rich endemic biodiversity, is home to around 10% of the world's 1,159 recognised reptile species and is one of the worlds hotspots for skink diversity (Uetz et al. 2020). The number of squamate species listed as near-threatened, Vulnerable or Endangered in Australia by the International Union for the Conservation of Nature (IUCN) is lower than the global average (even accounting for data deficient species; Tingley et al. 2019). However, a recent re-evaluation of the global conservation status of skinks (Scincidae) by the IUCN Species Survival Commission Skink Specialist Group—which covered more than 90% of described species—has found that approximately 20% of species are threatened with extinction according to IUCN criteria (Chapple et al. 2021). Many Australian species are listed as 'data deficient' with a number of species assessments based solely on habitat and predicted distribution trends in the absence of true population numbers or genetic diversity data (Tingley et al. 2019). Global climate change is one of the foremost threats to many species and conservation efforts, with predicted impacts on temperature, rainfall, sea level pH and oxygen concentrations, along with a number of other environmental factors (Hughes 2003; Intergovernmental Panel on Climate Change [IPCC] 2022). Understanding how species of conservation interest react to changing environmental conditions and their genetic potential for adaptation is of increasing importance (De Wit & Palumbi 2013).

While driven by anthropogenic factors, habitat loss and fragmentation—a key threatening process listed in the recovery plans of many Australian species—is also intimately linked to climate change and is strongly affected by changes in mean maximum temperature, and yearly mean rainfall changes (Mantyka-pringle, Martin & Rhodes 2012). Australasian land areas warmed by approximately 1.4ºC between 1910 and 2020, and southern Australia is projected to see an increase in aridity, a decrease in rainfall and an increase in the occurrence of droughts (IPCC 2022). The impact of a changing climate is a key threatening process listed in the species recovery plan for *T. adelaidensis,* and for many other Australian species where habitat destruction and contractions in suitable range are already a concern. Projections for the species of interest in this thesis, the Australian pygmy bluetongue *Tiliqua adelaidensis*, indicate that suitable habitat in their current range will contract significantly under various climate change scenarios but that suitable habitat will likely remain in the southern part of their historical range (Delean et al. 2013; Fordham et al. 2012).

Species respond to environmental changes at different rates. These variable rates of adaptability will affect species survival, especially as we see the effects of climate change increasing in severity and frequency (IPCC 2022). For example, the lizard genus *Anolis* has a large number of species with an equally large distribution in areas with vastly different environments, and shows indications of the ability to rapidly adapt and evolve with changing conditions (Thorpe et al. 2015), whereas *T. adelaidensis* has a very low capacity for dispersal and range shifting (Ebrahimi & Bull 2014b). Due to the projected rapid pace of climate change, many species will not have the short generational time or fecundity for species-level evolution to occur, and understanding of mechanisms at a smaller timescale driven by plasticity may provide insight into how individuals or populations are able to respond to changes in the short term. Locally adapted populations within a species may also respond to these pressures differently and have variable susceptibility to climate change. Meier et al. (2014) showed that brown trout collected from populations that are normally exposed to variable temperatures nonetheless showed different gene expression responses when an F2 generation was reared at two consistent experimentally set temperatures. A species' ability to adapt to changing environmental conditions in its existing range, within individuals and in the short term, is therefore of key interest when assessing the impact of changing conditions on their overall survival. Many Australian species are already adapted to semi-arid range conditions; however with the projections of increasing aridity and decreasing rainfall in the future, coupled with greater unpredictability in rainfall and an increase in extreme weather events, understanding the response of key genes necessary for survival in a changing environment is paramount.

1.2 Conservation and the Study of Genomics

Conservation genetics, the study of demographic processes of endangered species through the use of genetic markers, has become an invaluable resource for assessing populations, local adaptation, gene flow, mating systems and dispersal (Ouborg et al. 2010). However, most research has focused on neutral genetic markers and often does not inform about selective pressures or functional diversity (Holderegger & Wagner 2006; Rivers et al. 2014). These studies are not always able to provide information on the response of a species to environmental factors, such as climate change. Many studies also assume that adaptive areas of the genome will adhere to patterns observed in neutral genetic data (Kohn et al. 2006). This assumption is unreliable, as functional genes (often simplistically considered to be genes that code for a protein or other regulatory factor, as opposed to non-coding neutral markers) unlinked to the neutral markers in question may be under various

selective pressures that do not affect the neutral markers (Cammen et al. 2011; Herdegen, Babik & Radwan 2014; Nielsen, EE et al. 2009).

Differences between allele frequencies at neutral and adaptive markers are also often used as evidence for selection (Bos et al. 2008). While use of a combination of functional and neutral markers strengthens conclusions about population dynamics and selective pressures (Clozato et al. 2015), adaptive genetic data are required to assess a species' ability to respond to changing conditions and to predict these changes, particularly when considering translocation, breeding and other conservation measures (Ekblom & Galindo 2011; Radwan, Biedrzycka & Babik 2010). Genomics can provide insight into the mechanisms behind the interaction between selectively important genetic variation and environmental conditions (Kohn et al. 2006; Ouborg et al. 2010), as well of the biology of species, which can improve conservation management processes (Hogg et al. 2022).

Advances in sequencing technology and analysis techniques have allowed genetic studies to focus on functional and selectively important genes and responses to environmental conditions or stimuli. Investigation of the expression of these functional genes has been made possible by high-throughput sequencing techniques such as RNA-seq (Conesa et al. 2016; Van den Berge et al. 2019); and now long-read transcript sequencing such as the Pacific Biosciences Iso-Seq (Minio et al. 2019; Rhoads & Au 2015; Workman et al. 2018) and Oxford Nanopore Technologies direct RNA and cDNA sequencing (Branton et al. 2008; Byrne et al. 2019), which allow cost-effective retrieval of entire transcripts (Amarasinghe et al. 2020). Increasing use of transcriptomics in conservation requires approaches to assemble and analyse these data in threatened organisms for which there may be little to no previous genomic information and a lack of reference annotations. Methods for *de novo* assembly of transcripts and identification of gene functions continue to be revised and improved (Grabherr et al. 2011; Hara et al. 2015; Kumar & Blaxter 2010; Lighten, van Oosterhout & Bentzen 2014; Marchet et al. 2019; Ren et al. 2012; Surget-Groba & Montoya-Burgos 2010; Ungaro et al. 2017; Wang, SF & Gribskov 2017; Zhao et al. 2011). While complete accuracy and coverage cannot be guaranteed in species without reference genomes, the genetic data obtained allow functional genetic diversity to be measured and used to inform conservation efforts.

The value of functional genomic information in conservation genetics is often debated or compared to the value of neutral genetic and population markers (Teixeira & Huber, CD 2021). Regardless, such information has proved exceptionally informative in many species, to the point

that many initiatives have goals to generate reference genomes for study species (Rhie et al. 2021). Genetic diversity is an important factor to consider when assessing threatened populations and species but it may not be a reliable sole indicator of conservation status (Wilii et al. 2022). Rather than considering conservation genomics as the 'silver bullet' to solving conservation problems, it should be considered one of many useful tools to be used in a holistic approach, often tailored to the problem at hand. Creation of the reference genome for the Tasmanian devil (*Sarcophilus harrisii*) was an important milestone that facilitated a detailed understanding of the devil population and directly informed conservation management practices aimed at both population and disease management (Hogg et al. 2017). The genome information allowed development of single nucleotide polymorphism (SNP) analysis, characterisation of devil facial tumour strains, characterisation of immune genes and genome-wide association studies (Brandies et al. 2019). Although *S. harrisii* remains threatened, high-quality genomic data have proven their value in the response to the key threatening disease and population decline and prompted discussions with governments and stakeholders to expand the framework for an integrated approach to generating genomic data on threatened species as part of conservation management (Hogg et al. 2022).

1.2.1 Transcriptomics as a Genomic Tool

The transcriptome is the complete set of messenger RNA (mRNA) transcripts present in a cell or specific sample of cells (Wang, Z, Gerstein & Snyder 2009). Transcriptomic methods allow for the specific analysis of genes that are being actively expressed at the time of collection and stabilisation of such cells. Gene expression provides a measure of the transcription of mRNA used to code for a protein product or other regulatory factor. A variety of studies may be conducted by selecting for different RNA during the methodological process. RNA studies can also detect non-coding genes, including transcription factors and regulatory factors such as long non-coding RNAs (lncRNAs), which are epigenetic regulators affecting gene transcription and with a primary role in acetylation and methylation of genes. While not directly transcribed to proteins, factors such as lncRNAs have been shown to play a role in disease pathogenesis (Yin et al. 2020). Gene expression analysis of mRNA normally aims to eliminate detection of micro-RNA and other small RNAs through size selection, and remove rRNA and tRNAs by selecting for mRNA with poly(A) tails (Zhao et al. 2018). Depending on the methods used, transcriptomics allows researchers to measure genes that are expressed, characterise transcribed isoforms and annotate full-length transcripts. Commonly, analysis of transcripts is achieved by using reverse transcription methods to synthesise

complementary DNA sequences based on mRNAs present in the sample, which are then sequenced (Ekblom & Galindo 2011; Marioni et al. 2008).

Gene expression is controlled by a number of molecular processes; therefore, care must be taken when interpreting gene expression results. A large amount of research has gone into predicting regulation networks in medicine for purposes such as monitoring breast cancer, and predictive models must consider a number of factors including expression, methylation, expression of other genes in related pathways, and transcription factors (Regondi et al. 2021). Chromatin structure has been shown to coordinate expression of genes in major histocompatibility complex, Hox and globin genes where the sharing of regulatory elements and changes to the histone structure can affect expression of the gene cluster (Sproul et al. 2005). Transcription factors such as SOX proteins control gene expression by combining with other factors and interacting with various regulatory areas on DNA (Remenyi et al. 2004). Expression can also be influenced by biological mechanisms such as growth hormones (Rotwein 2020). Interpretation of expression analyses should take these into account, particularly in studies such as the current one, where samples were collected *in situ* with minimal control over factors other than the sampled individual and season of collection.

Transcriptomics also allows the study of organisms' gene expression activity as a function of environmental or biological changes (Ouborg et al. 2010). These applications have continued to improve rapidly as a consequence of cost-effective ultra-high-throughput sequencing methods (Liu, L et al. 2012), which allow the capture of entire transcriptomes with high accuracy; for example, RNA-seq (Marioni et al. 2008). These sequencing methods have enabled the analysis of large amounts of highly repeatable sequencing data, and quickly became the preferred method for transcriptome sequencing over previous analysis methods such as cDNA microarrays used to measure transcript expression (Schena et al. 1995) and SAGE (serial analysis of gene expression) (Velculescu et al. 1995). Gene expression analyses and transcriptomics are a central part of disease research, where large-scale prediction of potential cancer-susceptible regions and disease in humans is achieved by identifying regions with altered expression levels (Alisoltani et al. 2014; Cloonan et al. 2008). Transcriptomics has since become a fixture in population genomics studies, particularly in the conservation of non-model organisms (Cahais et al. 2012). Transcriptomics can provide the data necessary to draw lines of causation between evidence for selection and specific functional genes and environmental factors.

1.2.2 Conservation and other Applications

Transcriptomics has innumerable applications in conservation studies. Understanding gene expression in different environmental seasons, life stages, tissues, individuals and populations allows insight into gene functionality and potential selective pressure and adaptation. Transcriptomics has been used to identify current and historical selection (Gleason & Burton 2015; Meier et al. 2014), local adaptation (Yang, Y et al. 2015), adaptation within lineages (Backstrom, Zhang & Edwards 2013) and specific selection pressures and gene functions. It has also been used in the isolation of specific genes of interest through the annotation of the whole transcriptome (Guttman et al. 2010), as well as to gauge responses to pollution and other changes in environmental conditions (Whitehead et al. 2010). Understanding functional genes and their effect on the survivability of individuals and populations is crucial for application of appropriate conservation measures, particularly translocation (Sommer 2005; Weeks et al. 2011; Wright et al. 2014; Ralls et al. 2020).

Species of conservation concern often have a slew of ecological data collected but few genomic resources (Hornett & Wheat 2012). Even species considered model species for particular ecosystems, such as the Aleppo pine (*Pinus halepensis*) for drought and fire regimes, may have limited genomic data available because of the nature of conservation research (Pinosio et al. 2014). Fortunately, the number of non-model organisms with genomic data is increasing and transcriptome analysis and annotation is being used as a resource for future studies involving functional genes in threatened species (Ekblom & Galindo 2011; Patnaik et al. 2016).

Gene expression can be used to identify a species' reaction to various kinds of stimuli, including stressors. Fraser et al. (2011) identified male-specific expressed genes in the guppy (*Poecilia reticulata*) and stress responses in juveniles to the presence of predators. Gene expression can also be used to investigate factors driving immunological host–parasite interactions; for example, a study by Liu, Y et al. (2014) identified differentially expressed immune genes that play a role in response to infection in the Asian corn borer *Ostrinia furnacalis*—a crop pest species—which may help inform biological control approaches. Transcriptomics can also measure molecular stress responses and an organism's tolerance to anthropogenic effects; for example, estuarine populations of killifish (*Fundulus heteroclitus*) on the North American Atlantic coast were found to have evolved a higher tolerance to pollution (Whitehead et al. 2010). Changes in gene expression can also be used as a measure of the effects of translocation or changing environment conditions.

Xu et al. (2015) found concurrent increases in gene expression as genetic diversity decreased in *Miscanthus lutarioriparius* grasses when transplanted into a hostile environment. Gene expression plasticity may be as important as genetic diversity when considering the impact of changes to environmental conditions.

Transcriptomics also plays an important role in agriculture; for example, analysis of transcripts from various *Oryza* species found evidence of severe bottlenecking during domestication (Nabholz et al. 2014). Tissue-specific expression has been identified in yews of the genus *Taxus* (Hao et al. 2012) and the bark of commercially valuable rubber trees (*Hevea brasiliensis*) (Li, DJ et al. 2012; Xia et al. 2011), information which will now inform breeding practices. These results may be used to assist in breeding these crops. The same methods can be applied to any population to quantify functional genetic diversity and applied in other breeding programs.

Transcriptomics can be used to both assess population genetic diversity and identify specific genes. Understanding population structure and potential local adaptation (Yang, WZ, Qi & Fu 2014) is essential before commencing conservation programs. Local adaptation is evident in the population structure and gene flow patterns of red abalone (*Haliotis rufescens*) (De Wit & Palumbi 2013) and Atlantic herring (*Clupea harengus*) (Limborg et al. 2012). Genetic health of populations is a key factor to consider alongside other ecological factors and relationships when designing conservation strategies (Ouborg et al. 2010). With smaller fragmented population or range distributions there is a risk of loss of genetic diversity (Weeks et al. 2011). Genetic diversity equips individuals with the potential innate immune response to a wide number of pathogens and provides populations with greater potential to adapt to changing conditions (Herdegen, Babik & Radwan 2014; Piertney & Oliver 2006; Shamiminoori 2015; Xu et al. 2015). Exploring the functional genetic diversity of threatened species enhances our understanding and offers a more comprehensive view of their adaptive potential.

Identification of lineage-specific adaptations and genes with a high ratio of nonsynonymous to synonymous mutations is often used as an indicator of local adaptation; for example, this ratio has been used to indicate a high rate of positive selection in passerine lineages (Backstrom, Zhang & Edwards 2013). However, transcript studies now also provide insight into adaptation using gene expression responses. Parasite-driven adaptation is evident in stickleback fishes (*Gasterosteus aculeatus*): individuals from different populations experimentally infected with parasites showed changes in gene expression as an indication of local adaptation to pathogen pressures in different

habitats (Lenz et al. 2013). Temperature is a significant environmental variable, particularly for marine organisms, and has been studied at length in many species due to changes predicted by climate change models. Local adaptation based on thermal stress and environmental conditions has been observed and measured in many species including copepods (Schoville et al. 2012), the intertidal snail *Chlorostoma funebralis* (Gleason & Burton 2015), brown trout (*Salmo trutta*) (Meier et al. 2014) and Australian barramundi (*Lates calcarifer*) (Newton, Zenger & Jerry 2013). When planning conservation programs, particularly where manipulation of population structure or available breeding individuals is artificially altered—such as with translocation or assisted dispersal (Seddon et al. 2015)—it is important to consider functionally important genetic diversity (Limborg et al. 2012).

Environmental conditions and selective pressures that result in significant changes in expression of certain genes at certain times are relevant to conservation to increase understanding of the effects that habitat changes may have on a species. Conservation studies on non-model species and species *in situ* often mean that variables that can be controlled for in other species in a lab setting, for example by using inbred or clonal strains, or specific lab environment conditions (Strickler, Bombarely & Mueller 2012), cannot be controlled or eliminated. As a result, conservation transcriptome studies are often exploratory and must account for a much wider potential effect of variables. The number of studies of genomic data for squamate species has been steadily increasing, including studies of transcripts and gene expression (Appendix 4.2, Figure 6-14). An example of this is studies of the Australian central bearded dragon (*Pogona vitticeps*) prior to the annotation of a genome, where research focus was on functional sex-determining genes (Georges et al. 2015). Transcript data have been used to assess environmental stressors on the desert tortoise (*Gopherus agassizii*) and transcript analysis has produced evidence for positive selection and local adaptation to high altitudes in the agamid lizard *Phrynocephalus erythrurus* with evidence that similar mechanisms may have evolved in related species (Yang et al. 2015). Adaptation to changing environmental conditions often may include short-term phenotypic changes, or changes driven by altered gene expression, as well as longer-term changes due to genetic changes and subsequent selection and evolution. These short-term changes, while allowing a species to persist in stressful environmental conditions or potentially expand their range, can have unexpected effects on subsequent evolutionary processes by masking genetic diversity or traits, or 'softening' selective pressures (Koch & Guillaume 2020). Studies on adaptation in squamates are often retrospective; for example, in the above studies, as populations had already been subjected to

different conditions, the influence of gene regulation on adaptation is difficult to assess. However, identifying gene expression changes and areas of the genome that respond to specific stressors is an important first step to identifying genes that may play an important role in coping with that stressor.

With the growing focus on taking advantage of the ever-increasing availability and affordability of genomic sequencing to build genomic tools that may be used by conservation biologists, resources are needed to make sure these tools are accessible to biologists. A prime example of this is the integrated effort to manage devil facial tumour in the critically endangered Tasmanian devil, which demanded a highly collaborative approach among scientific disciplines and government agencies and management (Hogg et al. 2015). This will ensure the best tools are used to inform management decisions and has recently led to initiatives like the Bioplatforms Australia-funded Australian Amphibian and Reptile Genome initiative (AusARG), which will include genomes from Australian skink species (AusARG 2020).

1.2.3 Transcript Analysis and Gene Expression Studies

Analysis and assembly of the abundance of data generated using next-generation sequencing technologies poses bioinformatic challenges that depend on study aims and chosen species (Martin & Wang 2011). Transcripts and genes can more easily be aligned and annotated in species such as mice and humans because of the existence of comprehensive genomic databases and reference genome data. The development of assembly techniques that do not require a reference genome, as well as informative methods for dealing with such large amounts of data (De Wit et al. 2012), has allowed broad application of transcriptomics analysis in non-model species. A number of tools for RNA-seq transcriptome assembly and analysis rely on aligning reads to a reference genome; these include TopHat (Trapnell, Pachter & Salzberg 2009), Scripture (Guttman et al. 2010) and Cufflinks (Trapnell et al. 2012; Trapnell et al. 2010). Assemblies that map to a reference are also superior to *de novo* methods when isoforms are present as the reference can help identify splice junctions. Directly aligning transcripts to a related genome usually results in a higher number of identified putative genes; however, increasing divergence between the target species and the genomic reference is a source of bias and error, particularly for genes with high rates of change. When transcripts were mapped to a genome with 15% divergence, performance in transcript recovery was better than with *de novo* assembly methods (Vijay et al. 2013); however, when divergence increased to 30%, *de novo* assemblies produced better results. This shows that there is a

degree of divergence at which references for guided assembly impede gene identification. A similar comparison was run by Hornett and Wheat (2012), who determined that using a related species with less than 100 million years since divergence as a genome reference will have little effect on the amount of each gene covered by the reference material.

De novo approaches aim to assemble transcripts directly from short reads without a reference. Transcriptome assembly guided by genome mapping is consistently more accurate than *de novo* assembly (Grabherr et al. 2011; Vijay et al. 2013), but in non-model species, *de novo* assembly is often the only option. A number of approaches for *de novo* transcriptome assembly have been used, including ABySS (Birol et al. 2009; Simpson et al. 2009), Trans-ABySS (Robertson et al. 2010), SOAPdenovo2 (Luo et al. 2012), Oases (Schulz et al. 2012) and Trinity, a three-step program that also includes packages that measure transcript abundance and identify differential expression across multiple samples (Haas et al. 2013). Assembly methods that produce a transcriptome from scratch usually involve identifying an assembly combination where transcript k-mers overlap in an n-dimensional De Bruijn graph; nodes are connected if they overlap perfectly by length n-1 nucleotides (Haas et al. 2013). The length of k-mer used for assembly greatly influences results. Assembly can also be achieved using variable k-mer lengths (Surget-Groba & Montoya-Burgos 2010). Assemblies using shorter k-mers result in longer, more complete transcripts; however if assembly can be achieved with longer k-mers it will often result in better accuracy when compared to a reference, and greater potential for gene identification (Grabherr et al. 2011; Haas et al. 2013). Comparative studies have demonstrated that Trinity outperforms older programs under various conditions (Zhao et al. 2011). Guided assembly-based assembly is better for contiguity of known transcripts, whereas *de novo* transcriptome assembly when coupled with deep sequencing may produce more putative transcripts. However, the latter is less reliable than direct genome mapping overall, resulting in higher error rates, alignment errors and shifts in representation of functional gene annotation (Grabherr et al. 2011; Hornett & Wheat 2012).

Transcriptomes constructed using *de novo* methods are still able to be annotated using the genomes or gene databases of related species after assembly (Colgan et al. 2011; Collins et al. 2008; Crawford et al. 2010; Esteve-Codina et al. 2011; Kawahara-Miki et al. 2011; Kunstner et al. 2010; Toth et al. 2007; Wolf et al. 2010). Genomes from more phylogenetically diverged species that are not suitable for direct transcriptome mapping can still be used to identify similar orthologues or gene ontologies through alignment or mapping after *de novo* assembly; however

results should not be considered comprehensive or exhaustive as any genes that do not align may be missed.

1.2.4 Challenges

Common problems with transcriptome assembly (in both short- and long-read data) include varying transcript coverage (low *v*. very highly expressed transcripts), which affects rare transcript detection; sequencing errors; difficulties in differentiating isoforms with short-read data, which may not span splicing regions; chimeras resulting from fused loci and splicing of loci; and sequences that are repeated in multiple genes, introducing ambiguity to transcriptome assembly (Grabherr et al. 2011). Analyses can be standardised across samples to reduce false positive expression in certain transcripts that occurs through variable expression rates and numbers of isoforms (Trapnell et al. 2013). However solutions are often tailored to specific study goals or constraints, with read cut-off points and assembly guidelines that are suitable for one analysis but may be inappropriate for the next (Ouborg et al. 2010). Methods for *de novo* assembly of transcriptomes are accurate enough to draw conclusions and identify expressed genes (Cahais et al. 2012; Li, ZJ et al. 2019). Species with large and complex genomes, such as polyploids (containing more than two paired sets of chromosomes), often do not have a reference genome (Martin & Wang 2011). Nonetheless, transcriptome assembly is still possible, if a little more involved.

Global transcriptome profiling is another popular way to measure gene expression in many cells or tissue samples. Multiple samples can be collected to compare treatments or cell lines, and combination of these 'snap-shots' provides a detailed picture of gene expression patterns. Single-call RNA sequencing of embryos at specific time points during development has provided information about the developmental origin of vertebrate tissues and timing of differentiation (Yu et al. 2014; Briggs et al. 2018; Wang et al. 2020). However, diversity among individuals can obscure changes that may be attributed to the treatment or stimulus (Bayne et al. 2006). Even clonal lines of rainbow trout with no statistically different transcription profiles showed different transcript abundance in hepatic and pronephric tissue (Bayne et al. 2006). Integration of single-cell data across datasets relies on commonalities and can be challenging to achieve without loss of individual variation (Butler et al. 2018). Increasing throughput and barcoding or microchip technologies means that more cells may be analysed, facilitating the analysis of cell heterogeneity in a sample, spatial transcriptomics and cell communication, and allowing the dissection of different functional cells in complex tissues (Shen et al. 2022).

Studies on non-model species of conservation interest are also usually performed with as few individuals as possible to minimise impact on the sampled population. Experimental parameters further restrict the inclusion of individuals in replicate groups, and a larger number of variables tested or controlled for will require a larger number of sampled animals. One advantage of transcriptomic sequencing methods is the large amount of data obtained, allowing transcriptome studies to use very small sample sizes. For example, a study on the critically endangered Chinese alligator (*Alligator sinensis*) (Sun et al. 2018) included a number of tissue samples but from only two individuals. Schurch et al. (2016) showed that while transcript studies should ideally aim for biological replicates of n =12, methods such as Edge R or DESeq2 can identify significant differentially expressed genes with much lower replicate numbers due to sensitivity and a controlled false discovery rate (FDR). Edge R is suitable for as few as n =3 replicates per comparison group with a highly sensitive FDR (Schurch et al. 2016), although initial sampling should still aim for a larger number, determined for each individual study, allowing for factors that may influence each stage of sample processing and analysis.

Long-read sequencing methods facilitate the sequencing of whole transcripts in a single read. This eliminates the error introduced by post-sequencing transcript assembly. As the generated transcripts are sequenced in their entirety, they allow for the identification of gene isoforms (Liu, X et al. 2017) and do not result in the same kind of errors or transcript chimeras that can be created by *de novo* assembly of short reads, such as Trinity (Haas et al. 2013). Long-read sequencing may help gene annotation in cases where accurate full-length assemblies or isoform detection allows for better alignment to reference sequences. Where *de novo* transcripts may not completely align to orthologues, long-transcript reads provide the degree of certainty needed to identify whether assembled transcripts are natural variations or real errors due to mismatches in the assembly process (Amarasinghe et al. 2020). As such, long-read sequences have the potential to act as references or a source of validation for short reads when attempting to assemble short-read transcriptomic sequencing outputs.

While long-read sequencing of transcripts increases the accuracy of the transcripts produced relative to short-read assemblies—especially where there is not a closely related reference genome on which to base *de novo* assemblies (Vijay et al. 2013)—short-read sequencing is still important for gene expression analyses. Because cDNA is transcribed from the 3′ to 5′ end by reverse transcriptase, long reads can show a bias towards the 3′ end of reads, as fragmentation or

truncation of reads will affect the 5′ end being read at sequencing (Amarasinghe et al. 2020). The sequencing protocol for short-read sequencing outputs can also have a better base-level accuracy than that for long reads (Amarasinghe et al. 2020) and provides estimates of expression counts (Liu, L et al. 2012). By combining both methods, gene expression studies can take advantage of the strengths of each.

The number of reference genomes for squamate species is limited. A search on the National Centre for Biotechnology Information (NCBI) database for submitted genome information by organism (NCBI, last accessed 13/10/2022) returned 84 eukaryote genomic entries listed for Squamata with genome information; among these were only 16 organisms with identified chromosomes. Most of the other entries represent mitochondrial genome sequences or other incomplete scaffold assemblies. Complete chromosome annotation and annotation of the structure and function of assemblies in genome studies improves accuracy and can help identify splice variants and pseudogenes (Steward et al. 2017). There are eight reptile genomes listed on the Ensembl Rapid Release database: six squamates and two turtles (EMBL-EBI 2022). For order Squamata, containing 11,430 entries in the Reptile Database (last accessed 13/11/2022), both databases showed an extremely small representation of genomic references available (Uetz et al. 2019). Relative to scaffold genome assemblies, assemblies with more detailed chromosome-level information allow heterozygosity estimation and visualisation, and provide a more comprehensive means to assess genetic diversity, which is of particular interest in the study of threatened species (Totikov et al. 2021). Initiatives to increase the number of genomes for novel species have seen an increase in the availability of genetic data, including Genome 10K (Haussler et al. 2009), AusARG, Bat1K—an effort to sequence and annotate chromosome-level genomes for threatened bat species (Teeling et al. 2017)—and the Oz Mammals Genomics initiative (Eldridge et al. 2020). Recently, a chromosome-level genome for the desert horned lizard (*Phrynosoma platyrhinos*) was assembled and chromosomal changes were explored between this and 12 other reptile species with chromosome-level genomes available (Koochekian et al. 2022). Species divergence times for *T. adelaidensis* from its two closest phylogenetically related species with annotated genomes—the Australian central bearded dragon and the Carolina anole *Anolis carolinensis*—have been estimated by the tool TimeTree5 as approximate median times of 174 MYA. None of the abovementioned species are sufficiently phylogenetically related to the species of interest in this thesis, *T. adelaidensis*, to represent a genome reference to enable the use of direct transcript mapping assembly methods

(Hornett & Wheat 2012; Vijay et al. 2013), therefore *de novo* assembly methods are required, and may be just as effective at transcript identification as direct mapping against the above genomes.

1.3 Species of Focus: *Tiliqua adelaidensis,* an Endangered Australian Lizard

Tiliqua adelaidensis (Peters, 1863) is a solitary skink in the Egerniinae (Tiliquini) group (Shea 2021), endemic to South Australia (Duffy et al. 2012; Souter et al. 2007). It is found in highly fragmented habitat patches of native grassland in the state's mid-north where cropping and human land use impacts have altered the landscape (Souter et al. 2007; Duffy et al. 2012) and is classified as Endangered by the IUCN (Australasian Reptile and Amphibian Specialist Group 1996). Presumed extinct until 1992, *T. adelaidensis* was rediscovered when a specimen was found in the gut contents of a deceased brown snake (Armstrong & Reid 1992; Armstrong, Reid & Hutchinson 1993). Despite the narrow distribution and cryptic nature of this species, scientific knowledge has gradually built over the years.

The pygmy bluetongue is a solitary species that lives in abandoned wolf spider (Lycosid) and trapdoor spider (Mygalomorphae) burrows (Souter et al. 2007). It eats a diet of mostly arthropods ambushed within a small radius of the burrow entrance (Fenner, Bull & Hutchinson 2007). The species is viviparous and gives birth to live young (Nielsen, TP & Bull 2017; Shamiminoori et al. 2015) that eventually leave the natal burrow to find their own. This initial dispersal, as well as male-based movement during the mating season (Schofield et al. 2012), is the extent of dispersal in this species. Individuals spend a large portion of their time in and around their burrow, and undergo a period of brumation (winter torpor, or hibernation in ectotherms) (Cloudsley-Thompson 1999; McKenzie et al. 2020) between autumn and spring until increased temperatures are followed by an increase in activity again. Brumation has been recorded in all reptile taxa, and temperature is one of the key factors determining the timing of brumation periods (Cloudsley-Thompson 1999). Frequent recaptures within short distances in field studies, as well as population genetic data, indicates that individuals will travel distances of 20–200 metres from their burrow but rarely further (Milne et al. 2003; Smith et al. 2009). Suitable soil quality (Clayton & Bull 2016), vegetation cover, livestock grazing habits (Pettigrew & Bull 2012) and the presence of prey arthropods as well as lycosid or mygalomorph spider species that construct burrows utilised by the lizards (Clayton et al. 2020) are all key to their survival in an area (Souter et al. 2007). Habitat fragmentation and loss further limit the species range, and changing climactic conditions presents the greatest challenge to

conservation efforts for the pygmy bluetongue (Chapple et al. 2021; Delean et al. 2013; Fenner & Bull 2007; Hughes 2003).

Assessments of the Australian pygmy bluetongue lizard as Endangered under each of the IUCN red list, *Environment Protection and Biodiversity Conservation Act 1999* (Cth) and Schedule 7 of the *National Parks and Wildlife Act 1972* (SA) are strongly based on habitat fragmentation and assessments of fragmented population areas which is estimated at less than 500km^2 (Australasian Reptile and Amphibian Specialist Group 1996): no full population estimate is available (Australasian Reptile and Amphibian Specialist Group 1996). The species is one of the 110 'priority species' for Australia under the new *Threatened Species Action Plan 2022–2032* (Department of Climate Change, Energy, the Environment and Water 2022), along with two other species from the Egerniinae: the great desert skink *Liopholis kintorei* (Stirling & Zietz, 1893) and the Arnhem Land Gorges skink *Bellatorias obiri* (Wells & Wellington 1985). *Tiliqua adelaidensis'* already-limited range is expected to contract further under various climate change scenarios, disappearing completely in worst-case scenario models (Delean et al. 2013).

Mating strategies, population structure and gene flow in *T. adelaidensis* have been assessed using microsatellites (Schofield et al. 2012; Schofield et al. 2014; Smith et al. 2009). While these studies show some genetic sub-structuring by distance both between and within populations, genetic diversity has been maintained despite the highly fragmented habitat and low capacity for dispersal, potentially through population sizes which are not accurately known, promiscuous mating, and multiple paternity of offspring (Schofield et al. 2014). There is little published data on functional genes or potential local adaptation in this species, or its capacity to adapt to rapidly changing environmental conditions (Appendix 4.1). Genetic information is not just important for assessing overall population health and observing dispersal trends (Schofield et al. 2012) or the ability to adapt to changing conditions, but can also be used to estimate effective population sizes and population viability (Frankham, Bradshaw & Brook 2014). A greater understanding of these would provide a vital resource in evaluating genetic diversity and potentially adaptive genes and has been shown to be valuable when assessing the genetic health of a species (Gugger et al. 2016). Therefore, functional genomic data must be studied to improve the overall assessment of population health and conservation management.

This species and its remaining habitat undergo ongoing monitoring; however the *Recovery plan for Pygmy Bluetongue Lizard (Tiliqua adelaidensis) 2012*, released by the Department of Environment

and Natural Resources SA (Duffy et al. 2012) does not include the explicit status of population genetic health or benchmarks through which to measure it. In line with the IUCN listing, the recovery plan and newly released 'Conservation Advice for *Tiliqua adelaidensis'* heavily focus on land management and habitat conservation through sustainable grazing practices and community involvement, of the known 31 population sites, along with public outreach programs to local communities with a goal of improving the species listing to Conservation Dependent (Duffy et al. 2012; Department of Climate Change, Energy, the Environment and Water, 2023). Ongoing research into pygmy bluetongue lizards focuses on investigating the potential for local adaptation in the face of their highly fragmented population. Because of extremely low dispersal, research into the plasticity and ability of this species to persist long enough to quickly adapt to changing conditions, as well as options for translocation are key focus research areas (Weeks et al. 2011).

Projections of current habitat suitability suggest that translocation may become necessary to the conservation of this species (Delean et al. 2013; Fordham et al. 2012). Poor habitat quality has been shown to have negative impacts on body condition, basking times and dispersal rates (Ebrahimi & Bull 2015; Ebrahimi & Bull 2014a). Models of climate change impacts often indicate shifts in areas of suitable habitat or changes in species' ranges (Boyle et al. 2016). Translocation or 'assisted colonisation' to aid dispersal to areas that may remain suitable habitat under changing environmental conditions is a powerful tool in biodiversity conservation and one that must be considered for species with low capacity for natural dispersal (Travis et al. 2013). While rates of translocation success for reptiles and amphibians are often very low (Germano & Bishop 2009), with cautious execution including genetic management, the method has proven successful in some cases (Michaelides, Cole & Funk 2014).

Translocation was recommended by Fordham et al. (2012) as perhaps the only long-term option for the pygmy bluetongue lizard because of its low dispersal potential. Recent studies (Clive et al. 2019) have shown that *T. adelaidensis* individuals will successfully breed with conspecifics from different localities, and that population augmentation does not appear to have a negative impact on resident lizards. While these observable measures are important to study in these individuals and their offspring, genetic consequences of translocation must also be considered. Measuring gene expression and identifying genes that may be beneficial in founders or individuals supplemented into populations in the future will help ensure that genetic health of fragmented populations is maintained alongside other ecological measures. Analysis of functional genes may

also be used to determine whether isolated population patches have become locally adapted. To improve success rates for conservation management, genetic information needs to be considered when designing conservation management models.

The main goal of this thesis was to create a set of gene transcript references for *T. adelaidensis* renal tissue and identify genes that may be beneficial to surviving dry periods in an arid environment. This was achieved through the analysis and annotation of transcripts, and exploration of differential expression between two seasons in which lizards were subjected to different environmental pressures. This allowed identification of genes expressed in the kidney tissue and genes differentially expressed between spring and autumn, indicating genes that may be of particular interest in relation to the effect of climate change on these lizards.

While individual pygmy bluetongues have been observed drinking from morning condensation or rainfall from vegetation around their immediate burrow area (Nielsen, TP, Ebrahimi & Bull 2016), they do not travel long distances, relying on weather conditions facilitating dew or rain within the radius of their burrow. Water availability also governs the activity and distribution of the closely related sleepy lizard (*T. rugosa*; Kearney et al. 2018). It is not only water availability that is important for the pygmy bluetongue's behaviour; so is the appropriate habitat, such as vegetation cover, which enables this water to be collected. Both water availability and vegetation cover throughout pygmy bluetongue habitat are at risk under future climate change projections (Delean et al. 2013). These habitats are vulnerable to changing environmental conditions and experience temperature extremes during summer and highly variable rainfall over seasons and years. Identifying and annotating genes in the pygmy bluetongue lizard that may play a functional role in the renal system and potentially allow for different responses to summer conditions is of interest, particularly because of increasingly hot and dry summer periods. Once genomic references are developed, less invasive samples from individuals such as blood on FTA paper could be used to target key genes of interest or compare genomic profiles to the data produced here.

1.4 The Importance of Kidney Function in Arid-adapted Species

The kidneys play a major role in maintaining fluid retention, blood volume and blood pressure, all of which can be regulated by changes in osmotic pressure and the renin–angiotensin–aldosterone system (RAAS) (Kobayashi & Takei 1996; Takei et al. 1993). The RAAS regulates Na^+ reabsorption in the kidney and therefore the osmotic gradient and water loss (Sjaastad, Hove & Sand 2016), and

also has systemic effects on vasoconstriction and dilation, and blood pressure. The pathway of fluid flow within the kidney begins at the glomerulus, a small bundle of tiny vessels fed by the right and left renal arteries, where blood is filtered and fluid will flow into the Bowman's capsule, from there fluid flows through the proximal convoluted tubule, the descending and ascending loop of Henle (when present), the distal convoluted tubule, and then the collecting duct where urine is funnelled through the ureter to the bladder. In most vertebrate species, the greatest portion of osmotic exchange in the kidney to maintain homeostasis is conducted through passive reabsorption of solutes and water, predominantly in the proximal tubules of the kidney. The distal tubules of the kidney are lined with solute transporters and is where active transport occurs to regulate water, electrolyte and ion balance. Active transporters are also then present in the collecting duct such as *aquaporin 1*, and *aquaporin 2*. These active transporters are often under hormonal control—in the case of aquaporins, from antidiuretic hormone (vasopressin/vasotocin) (Golosova, Shakhmatova & Natochin 2019)—and may display variation in gene expression. Mammal kidney nephrons have a loop of Henle structure that generates a counter-osmotic current and allows concentration of urine fluid (Singh & Dyce 2018). This structure is lacking or poorly developed in birds (Sjaastad, Hove & Sand 2016) and absent in reptiles (99% of reptile nephrons are unlooped) (Holz 2020). As a result, reptiles cannot produce hypertonic urine and rely on processes in the liver, which is involved in the production of solid uric acid to remove wastes without losing water.

An exploration of potential genes of interest was conducted through an NCBI gene database search using key terms for genes with functions relating to renal function, homeostasis and osmoregulation, among others, published for reptiles (Chapter 3). Some genes expected to be found in this study include aquaporin (*aqp*) genes, which facilitate water movement in and out of cells and are crucial in renal tissues (Nielsen, S et al. 2002; Virkki et al. 2002); vasopressin (vasotocin in reptiles) (Golosova, Shakhmatova & Natochin 2019) as a regulator (Nielsen et al. 1995); and solute carriers (*slc*), which are an extremely large group of trans-membrane proteins that play an important role in homeostasis (Schaller & Lauschke 2019). Other active transporters for glucose and other substances were also incorporated (e.g., sglt1) that move glucose up the concentration gradient in secondary active transport and are usually expressed in the small intestine, but are also found in kidney renal tubule tissue. Other relevant genes often studied in reptiles were also included; for example heat shock proteins (HSPs) have been widely examined in reptiles in studies focused on heat and frost resistance (Costa-Martins et al. 2018), and relationships with other exothermic processes (Harry, Williams & Briscoe 1990). These genes were a focus in this study

because the semi-arid habitat occupied by *T. adelaidensis* can reach the high 40s (°C) during summer days and endures long periods with little to no rainfall (Bureau of Meteorology 2022).

This thesis explored gene expression in renal tissue of *T. adelaidensis* using both short- and long-read sequencing data, and used long-read sequencing data to facilitate transcript and isoform identification and annotation. This initial annotation allowed the identification of an extensive list of genes for potential further study, and the gene expression data provide context regarding which ones are of specific note from a renal homeostasis perspective under changing climate conditions for this species. Kidney tissue was chosen as a focus for the research because of the fundamental role kidneys play in maintaining homeostasis, along with their involvement in a variety of cellular transport functions that affect water and toxin retention or excretion.

1.5 Chapter Outline

With the threat of climate change and increasing interest in translocation as a mitigation tool for this species' survival, the overarching goal of this thesis was to identify any genes that may assist these lizards in surviving hotter, drier summers in the short term and gain an understanding of which genes may play a role in the species' ability to survive future climate change. Data from this study will be used to identify and annotate genes expressed during two seasonal collection periods that may be involved in surviving dry environmental periods. Identifying the pygmy bluetongue lizard's capacity to adapt to changing conditions will be crucial for assessing conservation priorities.

The study goal was pursued through the following objectives:

1. Sequence and assemble transcripts from RNA isolated from *T. adelaidensis* kidney tissues.
2. Compare the transcripts generated through *de novo* assembly of short-read RNA-seq and long-read Iso-Seq techniques and discuss these methods in the context of *T. adelaidensis* gene annotation.
3. Conduct a literature and gene database search to establish a reference list of genes that have been identified in other reptiles and may be of interest regarding kidney function.
4. Annotate assembled transcripts from RNA isolated from *T. adelaidensis* kidney tissues to identify putative genes present that may play a role in kidney function, including homeostasis and water retention.
5. Conduct a differential gene expression analysis to compare genes expressed in different seasons (early spring [September] and autumn [March/April]) using kidney tissue of *T. adelaidensis*.

This thesis has the following format:

Chapter 2 describes sample collection of whole tissues from eight *T. adelaidensis* individuals and subsequent wet-lab methodologies for RNA isolation and sequencing. Preliminary bioinformatics methods and a description of the computing resources used are also presented.

Chapter 3 details the *de novo* assembly of transcripts from short-read Illumina RNA-seq data and the processing of PacBio Iso-Seq data. Assemblies are compared and contrasted between individual samples, and completeness is estimated through a comparison to Benchmarking Universal Single-Copy Ortholog (BUSCO) databases. An assessment is also made of the utility of generating a reference transcriptome from the short-read data compared to the long-read Iso-Seq data used to generate a list of annotated transcripts for RNA expressed in *T. adelaidensis* kidney tissue. A systematic approach was used to identify genes that have been annotated previously in birds and reptiles and to identify specific genes of interest. This list of genes focused on those involved in homeostasis, water retention, ion channels and other renal functions in line with this study's overall focus.

Chapter 4 builds on the previous analyses by exploring gene expression in kidney tissue and the differential expression of genes at two seasonal extremes. Any transcripts with expression differences between season groups had putative annotations explored with a focus on genes involved in water homeostasis processes.

Chapter 5 provides an overall discussion of all chapters with regard to the common aims of this project and the analysis of functional genes in *T. adelaidensis*. The identification of genes of potential adaptive interest will serve to further ongoing research into the conservation of this species and provide resources for use in assessing genetic health or translocation options, as well as this species' potential responses to climate change.

1.6 References

Alisoltani, A, Fallahi, H, Ebrahimi, M, Ebrahimi, M & Ebrahimie, E 2014, 'Prediction of potential cancer-risk regions based on transcriptome data: towards a comprehensive view', *Plos One*, vol. 9, no. 5, p. e96320.

Amarasinghe, SL, Su, S, Dong, X, Zappia, L, Ritchie, ME & Gouil, Q 2020, 'Opportunities and challenges in long-read sequencing data analysis', *Genome Biology*, vol. 21, no. 1, p. 30.

Armstrong, G & Reid, J 1992, 'The re-discovery of the Adelaide pygmy bluetongue, *Tiliqua adelaidensis* (Peters, 1863)', *Herpetofauna*, vol. 22, no. 2, pp. 3-6.

Armstrong, G, Reid, JRW & Hutchinson, MN 1993, 'Discovery of a population of the rare scincid lizard *Tiliqua adelaidensis* (Peters)', *Records of the South Australian Museum*, vol. 36, pp. 153-5.

Australian Government, Department of Climate Change, Energy, the Environment and Water (DCCEEW) 2022, 'Priority Species: Threatened Species Action Plan 2022-2032', https://www.dcceew.gov.au/environment/biodiversity/threatened/publications/priority-species

Australasian Reptile and Amphibian Specialist Group 1996, '*Tiliqua adelaidensis*', *The IUCN Red List of Threatened Species*, vol. Version 2014.3.

Backstrom, N, Zhang, Q & Edwards, SV 2013, 'Evidence from a house finch (*Haemorhous mexicanus*) spleen transcriptome for adaptive evolution and biased gene conversion in passerine birds', *Mol Biol Evol*, vol. 30, no. 5, pp. 1046-50.

Bayne, CJ, Gerwick, L, Wheeler, PA & Thorgaard, GH 2006, 'Transcriptome profiles of livers and kidneys from three rainbow trout (Oncorhynchus mykiss) clonal lines distinguish stocks from three allopatric populations', Comparative Biochemistry and Physiology D-Genomics & Proteomics, vol. 1, no. 4, pp. 396-403.

Bureau of Meteorology 2022, 'Climate Data Online, Weather Station Directory', *Australian Government, Bureau of Meteorology*, http://www.bom.gov.au/climate/data/stations/ [Accessed 21 Mar. 2022].

Butler, A, Hoffman, P, Smibert, P, Papalexi, E & Satija, R 2018, 'Integrating single-cell transcriptomic data across different conditions, technologies, and species', Nature Biotechnology, vol. 36, no. 5, p. 411.

Birol, I, Jackman, SD, Nielsen, CB, Qian, JQ, Varhol, R, Stazyk, G, Morin, RD, Zhao, Y, Hirst, M, Schein, JE, Horsman, DE, Connors, JM, Gascoyne, RD, Marra, MA & Jones, SJ 2009, 'De novo transcriptome assembly with ABySS', *Bioinformatics*, vol. 25, no. 21, pp. 2872-7.

Bos, DH, Gopurenko, D, Williams, RN & Dewoody, JA 2008, 'Inferring population history and demography using microsatellites, mitochondrial DNA, and major histocompatibility complex (MHC) genes', *Evolution*, vol. 62, no. 6, pp. 1458-68.

Boyle, M, Schwanz, L, Hone, J & Georges, A 2016, 'Dispersal and climate warming determine range shift in model reptile populations', *Ecological Modelling*, vol. 328, pp. 34-43.

Branton, D, Deamer, DW, Marziali, A, Bayley, H, Benner, SA, Butler, T, Di Ventra, M, Garaj, S, Hibbs, A, Huang, XH, Jovanovich, SB, Krstic, PS, Lindsay, S, Ling, XSS, Mastrangelo, CH, Meller, A, Oliver, JS, Pershin, YV, Ramsey, JM, Riehn, R, Soni, GV, Tabard-Cossa, V, Wanunu, M, Wiggin, M & Schloss, JA 2008, 'The potential and challenges of nanopore sequencing', *Nature Biotechnology*, vol. 26, no. 10, pp. 1146-53.

Briggs, JA, Weinreb, C, Wagner, DE, Megason, S, Peshkin, L, Kirschner, MW & Klein, AM 2018, 'The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution', Science, vol. 360, no. 6392, p. 980.

Byrne, A, Cole, C, Volden, R & Vollmers, C 2019, 'Realizing the potential of full-length transcriptome sequencing', *Philosophical Transactions of the Royal Society B-Biological Sciences*, vol. 374, no. 1786, p. 9.

Cahais, V, Gayral, P, Tsagkogeorga, G, Melo-Ferreira, J, Ballenghien, M, Weinert, L, Chiari, Y, Belkhir, K, Ranwez, V & Galtier, N 2012, 'Reference-free transcriptome assembly in non-model animals from next-generation sequencing data', *Mol Ecol Resour*, vol. 12, no. 5, pp. 834-45.

Cammen, K, Hoffman, JI, Knapp, LA, Harwood, J & Amos, W 2011, 'Geographic variation of the major histocompatibility complex in Eastern Atlantic grey seals (*Halichoerus grypus*)', *Mol Ecol*, vol. 20, no. 4, pp. 740-52.

Chapple, DG, Roll, U, Böhm, M, Aguilar, R, Amey, AP, Austin, CC, Baling, M, Barley, AJ, Bates, MF, Bauer, AM, Blackburn, DG, Bowles, P, Brown, RM, Chandramouli, SR, Chirio, L, Cogger, H, Colli, GR, Conradie, W, Couper, PJ, Cowan, MA, Craig, MD, Das, I, Datta-Roy, A, Dickman, CR, Ellis, RJ, Fenner, AL, Ford, S, Ganesh, SR, Gardner, MG, Geissler, P, Gillespie, GR, Glaw, F, Greenlees, MJ, Griffith, OW, Grismer, LL, Haines, ML, Harris, DJ, Hedges, SB, Hitchmough, RA, Hoskin, CJ, Hutchinson, MN, Ineich, I, Janssen, J, Johnston, GR, Karin, BR, Keogh, JS, Kraus, F, LeBreton, M, Lymberakis, P, Masroor, R, McDonald, PJ, Mecke, S, Melville, J, Melzer, S, Michael, DR, Miralles, A, Mitchell, NJ, Nelson, NJ, Nguyen, TQ, de Campos Nogueira, C, Ota, H, Pafilis, P, Pauwels, OSG, Perera, A, Pincheira-Donoso, D, Reed, RN, Ribeiro-Júnior, MA, Riley, JL, Rocha, S, Rutherford, PL, Sadlier, RA, Shacham, B, Shea, GM, Shine, R, Slavenko, A, Stow, A, Sumner, J, Tallowin, OJS, Teale, R, Torres-Carvajal, O, Trape, J-F, Uetz, P, Ukuwela, KDB, Valentine, L, Van Dyke, JU, van Winkel, D, Vasconcelos, R, Vences, M, Wagner, P, Wapstra, E, While, GM, Whiting, MJ, Whittington, CM, Wilson, S, Ziegler, T, Tingley, R & Meiri, S 2021, 'Conservation status of the world's skinks (Scincidae): Taxonomic and geographic patterns in extinction risk', *Biological Conservation*, vol. 257, p. 109101.

Clayton, J & Bull, CM 2016, 'The impact of sheep grazing on the depth of spider burrows and of burrows selected by the pygmy bluetongue lizard (*Tiliqua adelaidensis*)', *Wildlife Research*, vol. 43, no. 8, pp. 691-703.

Clayton, J, Gardner, MG, Fenner, AL & Bull, M 2020, 'Co-occupancy of spider-engineered burrows within a grassland community changes temporally', *Austral Ecology*, vol. 45, no. 4, pp. 454-9.

Clive, LFR, Gardner, MG, Clayton, J, Baring, R, Hutchinson, MN, Fenner, A & Bull, CM 2019, 'Population augmentation had no effect on the abundance or body condition of conspecifics and co-occurring lizard species in a native grassland community', *Austral Ecology*.

Cloonan, N, Forrest, AR, Kolle, G, Gardiner, BB, Faulkner, GJ, Brown, MK, Taylor, DF, Steptoe, AL, Wani, S, Bethel, G, Robertson, AJ, Perkins, AC, Bruce, SJ, Lee, CC, Ranade, SS, Peckham, HE, Manning, JM, McKernan, KJ & Grimmond, SM 2008, 'Stem cell transcriptome profiling via massive-scale mRNA sequencing', *Nat Methods*, vol. 5, no. 7, pp. 613-9.

Cloudsley-Thompson, JL 1999, 'Daily and Seasonal Cycles, Hibernation, Aestivation and Migration', in JL Cloudsley-Thompson (ed.), The Diversity of Amphibians and Reptiles: An Introduction, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 157-76.

Clozato, CL, Mazzoni, CJ, Moraes-Barros, N, Morgante, JS & Sommer, S 2015, 'Spatial pattern of adaptive and neutral genetic diversity across different biomes in the lesser anteater (*Tamandua tetradactyla*)', *Ecol Evol*, vol. 5, no. 21, pp. 4932-48.

Colgan, TJ, Carolan, JC, Bridgett, SJ, Sumner, S, Blaxter, ML & Brown, MJ 2011, 'Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*', *Bmc Genomics*, vol. 12, p. 623.

Collins, LJ, Biggs, PJ, Voelckel, C & Joly, S 2008, 'An approach to transcriptome analysis of non-model organisms using short-read sequences', in J Arthur & SK Ng (eds), *Genome Informatics 2008, Vol 21*, Imperial College Press, Covent Garden, vol. 21, pp. 3-14.

Conesa, A, Madrigal, P, Tarazona, S, Gomez-Cabrero, D, Cervera, A, McPherson, A, Szczesniak, MW, Gaffney, DJ, Elo, LL, Zhang, X & Mortazavi, A 2016, 'A survey of best practices for RNA-seq data analysis', *Genome Biol*, vol. 17, p. 13.

Costa-Martins, AG, Lima, L, Alves, JMP, Serrano, MG, Buck, GA, Camargo, EP & Teixeira, MMG 2018, 'Genome-wide identification of evolutionarily conserved Small Heat-Shock and eight other proteins bearing alpha-crystallin domain-like in kinetoplastid protists', *Plos One*, vol. 13, no. 10, p. 28.

Crawford, JE, Guelbeogo, WM, Sanou, A, Traore, A, Vernick, KD, Sagnon, N & Lazzaro, BP 2010, 'De Novo Transcriptome Sequencing in *Anopheles funestus* Using Illumina RNA-Seq Technology', *Plos One*, vol. 5, no. 12, p. 12.

De Wit, P & Palumbi, SR 2013, 'Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation', *Mol Ecol*, vol. 22, no. 11, pp. 2884-97.

De Wit, P, Pespeni, MH, Ladner, JT, Barshis, DJ, Seneca, F, Jaris, H, Therkildsen, NO, Morikawa, M & Palumbi, SR 2012, 'The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis', *Mol Ecol Resour*, vol. 12, no. 6, pp. 1058-67.

Delean, S, Bull, CM, Brook, BW, Heard, LMB & Fordham, DA 2013, 'Using plant distributions to predict the current and future range of a rare lizard', *Diversity and Distributions*, vol. 19, no. 9, pp. 1125-37.

Department of Climate Change, Energy, the Environment and Water 2023. 'Conservation Advice for *Tiliqua adelaidensis* (pygmy blue-tongue lizard), (Environment Protection and Biodiversity Conservation Act 1999)'. Australian Government.

Duffy, A, Pound, L & How, T 2012, 'Recovery Plan for the pygmy bluetongue lizard *Tiliqua adelaidensis*', *Department of Environment and Natural Resources, South Australia*.

Ebrahimi, M & Bull, C 2015, 'Behavioural changes in an endangered grassland lizard resulting from simulated agricultural activities', *Journal of Arid Environments*, vol. 113, pp. 102-7.

Ebrahimi, M & Bull, CM 2014a, 'Resources and their distribution can influence social behaviour at translocation sites: Lessons from a lizard', *Applied Animal Behaviour Science*, vol. 156, pp. 94-104.

EMBL-EBI 2022, 'Ensembl Rapid Release', https://rapid.ensembl.org/index.html [Last Updated 29 July 2022]

Ebrahimi, M & Bull, CM 2014b, 'Short-Term Dispersal Response of an Endangered Australian Lizard Varies with Time of Year', *Plos One*, vol. 9, no. 8, p. 7.

Eldridge, MDB, Deakin, JE, MacDonald, AJ, Byrne, M, Fitzgerald, A, Johnson, RN, Moritz, C, Palmer, S & Young, A 2020, 'The Oz Mammals Genomics (OMG) initiative: Developing genomic

resources for mammal conservation at a continental scale', Australian Zoologist, vol. 40, no. 3, pp. 505-9.

Ekblom, R & Galindo, J 2011, 'Applications of next generation sequencing in molecular ecology of non-model organisms', *Heredity (Edinb)*, vol. 107, no. 1, pp. 1-15.

Esteve-Codina, A, Kofler, R, Palmieri, N, Bussotti, G, Notredame, C & Perez-Enciso, M 2011, 'Exploring the gonad transcriptome of two extreme male pigs with RNA-seq', *Bmc Genomics*, vol. 12, p. 552.

Fenner, AL & Bull, CM 2007, 'Short-term impact of grassland fire on the endangered pygmy bluetongue lizard', *Journal of Zoology*, vol. 272, no. 4, pp. 444-50.

Fenner, AL, Bull, CM & Hutchinson, MN 2007, 'Omnivorous diet of the endangered pygmy bluetongue lizard, *Tiliqua adelaidensis*', *Amphibia-Reptilia*, vol. 28, no. 4, pp. 560-5.

Fordham, DA, Watts, MJ, Delean, S, Brook, BW, Heard, LM & Bull, CM 2012, 'Managed relocation as an adaptation strategy for mitigating climate change threats to the persistence of an endangered lizard', *Global Change Biology*, vol. 18, no. 9, pp. 2743-55.

Frankham, R, Bradshaw, CJA & Brook, BW 2014, 'Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses', *Biological Conservation*, vol. 170, pp. 56-63.

Fraser, BA, Weadick, CJ, Janowitz, I, Rodd, FH & Hughes, KA 2011, 'Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome', *Bmc Genomics*, vol. 12, p. 202.

Georges, A, Li, QY, Lian, JM, O'Meally, D, Deakin, J, Wang, ZJ, Zhang, P, Fujita, M, Patel, HR, Holleley, CE, Zhou, Y, Zhang, XW, Matsubara, K, Waters, P, Graves, JAM, Sarre, SD & Zhang, GJ 2015, 'High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*', *Gigascience*, vol. 4, p. 10.

Germano, JM & Bishop, PJ 2009, 'Suitability of amphibians and reptiles for translocation', *Conserv Biol*, vol. 23, no. 1, pp. 7-15.

Gleason, LU & Burton, RS 2015, 'RNA-seq reveals regional differences in transcriptome response to heat stress in the marine snail *Chlorostoma funebralis*', *Mol Ecol*, vol. 24, no. 3, pp. 610-27.

Golosova, DV, Shakhmatova, EI & Natochin, YV 2019, 'Differences between Arginine-Vasotocin and Arginine-Vasopressin Effects on the Rat Kidney in Evolution of Osmoregulation in Vertebrates', *Journal of Evolutionary Biochemistry and Physiology*, vol. 55, no. 1, pp. 71-9.

Grabherr, MG, Haas, BJ, Yassour, M, Levin, JZ, Thompson, DA, Amit, I, Adiconis, X, Fan, L, Raychowdhury, R, Zeng, Q, Chen, Z, Mauceli, E, Hacohen, N, Gnirke, A, Rhind, N, di Palma, F, Birren, BW, Nusbaum, C, Lindblad-Toh, K, Friedman, N & Regev, A 2011, 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nat Biotechnol*, vol. 29, no. 7, pp. 644-52.

Gugger, PF, Cokus, SJ & Sork, VL 2016, 'Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*)', *Tree Genetics & Genomes*, vol. 12, no. 2, p. 14.

Guttman, M, Garber, M, Levin, JZ, Donaghey, J, Robinson, J, Adiconis, X, Fan, L, Koziol, MJ, Gnirke, A, Nusbaum, C, Rinn, JL, Lander, ES & Regev, A 2010, 'Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs', *Nat Biotechnol*, vol. 28, no. 5, pp. 503-10.

Haas, BJ, Papanicolaou, A, Yassour, M, Grabherr, M, Blood, PD, Bowden, J, Couger, MB, Eccles, D, Li, B, Lieber, M, MacManes, MD, Ott, M, Orvis, J, Pochet, N, Strozzi, F, Weeks, N, Westerman, R, William, T, Dewey, CN, Henschel, R, Leduc, RD, Friedman, N & Regev, A 2013, 'De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis', *Nature Protocols*, vol. 8, no. 8, pp. 1494-512.

Hao, DC, Ma, P, Mu, J, Chen, SL, Xiao, PG, Peng, Y, Huo, L, Xu, LJ & Sun, C 2012, 'De novo characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum*', *Science China-Life Sciences*, vol. 55, no. 5, pp. 452-66.

Hara, Y, Tatsumi, K, Yoshida, M, Kajikawa, E, Kiyonari, H & Kuraku, S 2015, 'Optimizing and benchmarking de novo transcriptome sequencing: From library preparation to assembly evaluation', *Bmc Genomics*, vol. 16, no. 1.

Harry, JL, Williams, KL & Briscoe, DA 1990, 'Sex determination in loggerhead turtles: Differential expression of two hnRNP proteins', *Development*, vol. 109, no. 2, pp. 305-12.

Haussler, D, O'Brien, SJ, Ryder, OA, Barker, FK, Clamp, M, Crawford, AJ, Hanner, R, Hanotte, O, Johnson, WE, McGuire, JA, Miller, W, Murphy, RW, Murphy, WJ, Sheldon, FH, Sinervo, B, Venkatesh, B, Wiley, EO, Allendorf, FW, Amato, G, Baker, CS, Bauer, A, Beja-Pereira, A, Bermingham, E, Bernardi, G, Bonvicino, CR, Brenner, S, Burke, T, Cracraft, J, Diekhans, M, Edwards, S, Ericson, PGP, Estes, J, Fjelsda, J, Flesness, N, Gamble, T, Gaubert, P, Graphodatsky, AS, Graves, JAM, Green, ED, Green, RE, Hackett, S, Hebert, P, Helgen, KM, Joseph, L, Kessing, B, Kingsley, DM, Lewin, HA, Luikart, G, Martelli, P, Moreira, MAM,

Nguyen, N, Orti, G, Pike, BL, Rawson, DM, Schuster, SC, Seuanez, HN, Shaffer, HB, Springer, MS, Stuart, JM, Sumner, J, Teeling, E, Vrijenhoek, RC, Ward, RD, Warren, WC, Wayne, R, Williams, TM, Wolfe, ND, Zhang, YP, Graph-Odatsky, A, Johnson, WE, Felsenfeld, A, Turner, S, Genome, KCS, Mammals, G, Birds, G, Amphibians Reptiles, G, Fishes, G, General Policy, G & Anal, G 2009, 'Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species', Journal of Heredity, vol. 100, no. 6, pp. 659-74.

Herdegen, M, Babik, W & Radwan, J 2014, 'Selective pressures on MHC class II genes in the guppy (*Poecilia reticulata*) as inferred by hierarchical analysis of population structure', *J Evol Biol*, vol. 27, no. 11, pp. 2347-59.

Hogg, CJ, Grueber, CE, Pemberton, D, Fox, S, Lee, AV, Ivy, JA & Belov, K 2015, '"Devil Tools & Tech": A Synergy of Conservation Research and Management Practice', *Conservation Letters*, vol. 10, no. 1, pp. 133-8.

Hogg, C, J., Ottewell, K, Latch, P, Rossetto, M, Biggs, J, Gilbert, A, Richmond, S & Belov, K 2022, 'Threatened Species Initiative: Empowering conservation action using genomic resources', *Proceedings of the National Academy of Sciences*, vol. 119, no. 4, p. e2115643118.

Holderegger, R & Wagner, HH 2006, 'A brief guide to Landscape Genetics', *Landscape Ecology*, vol. 21, no. 6, pp. 793-6.

Holz, P 2020, 'Anatomy and Physiology of the Reptile Renal System', *The veterinary clinics of North America. Exotic animal practice*, vol. 23, pp. 103-14.

Hornett, EA & Wheat, CW 2012, 'Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species', *Bmc Genomics*, vol. 13, p. 16.

Hughes, L 2003, 'Climate change and Australia: Trends, projections and impacts', *Austral Ecology*, vol. 28, no. 4, pp. 423-43.

IPCC 2022, 'Climate Change 2022: Impacts, Adaptation, and Vulnerability' Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)], *Cambridge University Press*, Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844

Kawahara-Miki, R, Wada, K, Azuma, N & Chiba, S 2011, 'Expression profiling without genome sequence information in a non-model species, Pandalid shrimp (*Pandalus latirostris*), by next-generation sequencing', *Plos One*, vol. 6, no. 10, p. e26043.

Kearney, MR, Munns, SL, Moore, D, Malishev, M & Bull, CM 2018, 'Field tests of a general ectotherm niche model show how water can limit lizard activity and distribution', *Ecological Monographs*, vol. 88, no. 4, pp. 672-93.

Koch, EL & Guillaume, F 2020, 'Additive and mostly adaptive plastic responses of gene expression to multiple stress in Tribolium castaneum', *Plos Genetics*, vol. 16, no. 5, p. 27.

Kobayashi, H & Takei, Y 1996, *The renin-angiotensin system : comparative aspects*, 1st ed. 1996. edn, Zoophysiology ; Volume 35, Springer Verlag, Berlin ;.

Kohn, MH, Murphy, WJ, Ostrander, EA & Wayne, RK 2006, 'Genomics and conservation genetics', *Trends Ecol Evol*, vol. 21, no. 11, pp. 629-37.

Koochekian, N, Ascanio, A, Farleigh, K, Card, DC, Schield, DR, Castoe, TA & Jezkova, T 2022, 'A chromosome-level genome assembly and annotation of the desert horned lizard, Phrynosoma platyrhinos, provides insight into chromosomal rearrangements among reptiles', Gigascience, vol. 11, p. 14.

Kumar, S & Blaxter, ML 2010, 'Comparing de novo assemblers for 454 transcriptome data', *Bmc Genomics*, vol. 11, p. 12.

Kunstner, A, Wolf, JB, Backstrom, N, Whitney, O, Balakrishnan, CN, Day, L, Edwards, SV, Janes, DE, Schlinger, BA, Wilson, RK, Jarvis, ED, Warren, WC & Ellegren, H 2010, 'Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species', *Mol Ecol*, vol. 19 Suppl 1, pp. 266-76.

Lenz, TL, Eizaguirre, C, Rotter, B, Kalbe, M & Milinski, M 2013, 'Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis', *Mol Ecol*, vol. 22, no. 3, pp. 774-86.

Li, DJ, Deng, Z, Qin, B, Liu, XH & Men, ZH 2012, 'De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.)', *Bmc Genomics*, vol. 13, p. 14.

Li, ZJ, Meng, MM, Li, SS & Deng, B 2019, 'The transcriptome analysis of *Protaetia brevitarsis Lewis ldarvae*', *Plos One*, vol. 14, no. 3, p. 16.

Lighten, J, van Oosterhout, C & Bentzen, P 2014, 'Critical review of NGS analyses for de novo genotyping multigene families', *Mol Ecol*, vol. 23, no. 16, pp. 3957-72.

Limborg, MT, Helyar, SJ, De Bruyn, M, Taylor, MI, Nielsen, EE, Ogden, R, Carvalho, GR, Consortium, FPT & Bekkevold, D 2012, 'Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*)', *Mol Ecol*, vol. 21, no. 15, pp. 3686-703.

Liu, L, Li, Y, Li, S, Hu, N, He, Y, Pong, R, Lin, D, Lu, L & Law, M 2012, 'Comparison of next-generation sequencing systems', *J Biomed Biotechnol*, vol. 2012, p. 251364.

Liu, X, Mei, W, Soltis, PS, Soltis, DE & Barbazuk, WB 2017, 'Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome', *Mol Ecol Resour*, vol. 17, no. 6, pp. 1243-56.

Liu, Y, Shen, D, Zhou, F, Wang, G & An, C 2014, 'Identification of immunity-related genes in *Ostrinia furnacalis* against entomopathogenic fungi by RNA-seq analysis', *Plos One*, vol. 9, no. 1, p. e86436.

Luo, R, Liu, B, Xie, Y, Li, Z, Huang, W, Yuan, J, He, G, Chen, Y, Pan, Q, Liu, Y, Tang, J, Wu, G, Zhang, H, Shi, Y, Liu, Y, Yu, C, Wang, B, Lu, Y, Han, C, Cheung, DW, Yiu, SM, Peng, S, Xiaoqian, Z, Liu, G, Liao, X, Li, Y, Yang, H, Wang, J, Lam, TW & Wang, J 2012, 'SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler', *Gigascience*, vol. 1, no. 1, p. 18.

Mantyka-pringle, CS, Martin, TG & Rhodes, JR 2012, 'Interactions between climate and habitat loss effects on biodiversity: a systematic review and meta-analysis', *Global Change Biology*, vol. 18, no. 4, pp. 1239-52.

Marchet, C, Lecompte, L, Da Silva, C, Cruaud, C, Aury, JM, Nicolas, J & Peterlongo, P 2019, 'De novo clustering of long reads by gene from transcriptomics data', *Nucleic Acids Research*, vol. 47, no. 1, p. 13.

Marioni, JC, Mason, CE, Mane, SM, Stephens, M & Gilad, Y 2008, 'RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays', *Genome Res*, vol. 18, no. 9, pp. 1509-17.

Martin, JA & Wang, Z 2011, 'Next-generation transcriptome assembly', *Nat Rev Genet*, vol. 12, no. 10, pp. 671-82.

McKenzie, CM, Oesterle, PT, Stevens, B, Shirose, L, Mastromonaco, GF, Lillie, BN, Davy, CM, Jardine, CM & Nemeth, NM 2020, 'Ophidiomycosis in Red Cornsnakes (Pantherophis guttatus): Potential Roles of Brumation and Temperature on Pathogenesis and Transmission', Veterinary Pathology, vol. 57, no. 6, pp. 825-37.

Meier, K, Hansen, MM, Normandeau, E, Mensberg, KL, Frydenberg, J, Larsen, PF, Bekkevold, D & Bernatchez, L 2014, 'Local adaptation at the transcriptome level in brown trout: evidence from early life history temperature genomic reaction norms', *Plos One*, vol. 9, no. 1, p. e85171.

Michaelides, S, Cole, N & Funk, SM 2014, 'Translocation retains genetic diversity of a threatened endemic reptile in Mauritius', *Conservation Genetics*, vol. 16, no. 3, pp. 661-72.

Milne, T, Bull, CM & Hutchinson, MN 2003, 'Use of burrows by the endangered pygmy blue-tongue lizard, *Tiliqua adelaidensis* (Scincidae)', Wildlife Research, vol. 30, no. 5, pp. 523-8.

Minio, A, Massonnet, M, Figueroa-Balderas, R, Vondras, AM, Blanco-Ulate, B & Cantu, D 2019, 'Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development', *G3 Genes|Genomes|Genetics*, vol. 9, no. 3, pp. 755-67.

Nabholz, B, Sarah, G, Sabot, F, Ruiz, M, Adam, H, Nidelet, S, Ghesquiere, A, Santoni, S, David, J & Glemin, S 2014, 'Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*)', *Mol Ecol*, vol. 23, no. 9, pp. 2210-27.

Newton, JR, Zenger, KR & Jerry, DR 2013, 'Next-generation transcriptome profiling reveals insights into genetic factors contributing to growth differences and temperature adaptation in Australian populations of barramundi (*Lates calcarifer*)', *Mar Genomics*, vol. 11, pp. 45-52.

Nielsen, EE, Hemmer-Hansen, J, Larsen, PF & Bekkevold, D 2009, 'Population genomics of marine fishes: identifying adaptive variation in space and time', *Mol Ecol*, vol. 18, no. 15, pp. 3128-50.

Nielsen, S, Chou, CL, Marples, D, Christensen, EI, Kishore, BK & Knepper, MA 1995, 'Vasopressin increases water permeability of kidney collecting duct by inducing translocation of aquaporin-CD water channels to plasma membrane', *Proc Natl Acad Sci U S A*, vol. 92, no. 4, pp. 1013-7.

Nielsen, S, Frokiaer, J, Marples, D, Kwon, TH, Agre, P & Knepper, MA 2002, 'Aquaporins in the kidney: from molecules to medicine', *Physiol Rev*, vol. 82, no. 1, pp. 205-44.

Nielsen, TP & Bull, CM 2017, 'The impact of sheep grazing on the fecundity and timing of reproduction in the endangered pygmy bluetongue lizard, *Tiliqua adelaidensis*', *Amphibia-Reptilia*, vol. 38, no. 4, pp. 401-9.

Nielsen, TP, Ebrahimi, M & Bull, CM 2016, 'A thirsty little lizard: drinking by the pygmy bluetongue lizard', *Transactions of the Royal Society of South Australia*, vol. 140, no. 1, pp. 1-6.

Ouborg, NJ, Pertoldi, C, Loeschcke, V, Bijlsma, RK & Hedrick, PW 2010, 'Conservation genetics in transition to conservation genomics', *Trends Genet*, vol. 26, no. 4, pp. 177-87.

Patnaik, BB, Wang, TH, Kang, SW, Hwang, HJ, Park, SY, Park, EB, Chung, JM, Song, DK, Kim, C, Kim, S, Lee, JS, Han, YS, Park, HS & Lee, YS 2016, 'Sequencing, De Novo Assembly, and Annotation of the Transcriptome of the Endangered Freshwater Pearl Bivalve, *Cristaria plicata*, Provides Novel Insights into Functional Genes and Marker Discovery', *Plos One*, vol. 11, no. 2, p. e0148622.

Pettigrew, M & Bull, CM 2012, 'The response of pygmy bluetongue lizards to simulated grazing in the field during three drought years', *Wildlife Research*, vol. 39, no. 6, pp. 540-5.

Piertney, SB & Oliver, MK 2006, 'The evolutionary ecology of the major histocompatibility complex', *Heredity (Edinb)*, vol. 96, no. 1, pp. 7-21.

Pinosio, S, Gonzalez-Martinez, SC, Bagnoli, F, Cattonaro, F, Grivet, D, Marroni, F, Lorenzo, Z, Pausas, JG, Verdu, M & Vendramin, GG 2014, 'First insights into the transcriptome and development of new genomic tools of a widespread circum-Mediterranean tree species, *Pinus halepensis Mill*', *Mol Ecol Resour*, vol. 14, no. 4, pp. 846-56.

Radwan, J, Biedrzycka, A & Babik, W 2010, 'Does reduced MHC diversity decrease viability of vertebrate populations?', *Biological Conservation*, vol. 143, no. 3, pp. 537-44.

Ralls, K, Sunnucks, P, Lacy, RC & Frankham, R 2020, 'Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation', *Biological Conservation*, vol. 251, p. 8.

Regondi, C, Fratelli, M, Damia, G, Guffanti, F, Ganzinelli, M, Matteucci, M & Masseroli, M 2021, 'Predictive modeling of gene expression regulation', Bmc Bioinformatics, vol. 22, no. 1, p. 21.

Remenyi, A, Scholer, HR & Wilmanns, M 2004, 'Combinatorial control of gene expression', Nature Structural and Molecular Biology, vol. 11, p. 812+.

Ren, XW, Liu, T, Dong, J, Sun, LL, Yang, J, Zhu, YF & Jin, Q 2012, 'Evaluating de Bruijn Graph Assemblers on 454 Transcriptomic Data', *Plos One*, vol. 7, no. 12, p. 9.

Rhoads, A & Au, KF 2015, 'PacBio Sequencing and Its Applications', *Genomics Proteomics Bioinformatics*, vol. 13, no. 5, pp. 278-89.

Rivers, MC, Brummitt, NA, Nic Lughadha, E & Meagher, TR 2014, 'Do species conservation assessments capture genetic diversity?', *Global Ecology and Conservation*, vol. 2, pp. 81-7.

Robertson, G, Schein, J, Chiu, R, Corbett, R, Field, M, Jackman, SD, Mungall, K, Lee, S, Okada, HM, Qian, JQ, Griffith, M, Raymond, A, Thiessen, N, Cezard, T, Butterfield, YS, Newsome, R, Chan, SK, She, R, Varhol, R, Kamoh, B, Prabhu, AL, Tam, A, Zhao, Y, Moore, RA, Hirst, M, Marra, MA, Jones, SJ, Hoodless, PA & Birol, I 2010, 'De novo assembly and analysis of RNA-seq data', *Nat Methods*, vol. 7, no. 11, pp. 909-12.

Rotwein, P 2020, 'Regulation of gene expression by growth hormone', Molecular and Cellular Endocrinology, vol. 507, p. 7.

Schaller, L & Lauschke, VM 2019, 'The genetic landscape of the human solute carrier (SLC) transporter superfamily', *Human genetics*, vol. 138, no. 11-12, pp. 1359-77.

Schena, M, Shalon, D, Davis, RW & Brown, PO 1995, 'Quantitative monitoring of gene expression patterns with a complementary DNA microarray', *Science*, vol. 270, no. 5235, pp. 467-70.

Schofield, JA, Fenner, AL, Pelgrim, K & Bull, CM 2012, 'Male-biased movement in pygmy bluetongue lizards: implications for conservation', *Wildlife Research*, vol. 39, no. 8, pp. 677-84.

Schofield, JA, Gardner, MG, Fenner, AL & Bull, CM 2014, 'Promiscuous mating in the endangered Australian lizard *Tiliqua adelaidensis*: a potential windfall for its conservation', *Conservation Genetics*, vol. 15, no. 1, pp. 177-85.

Schoville, SD, Barreto, FS, Moy, GW, Wolff, A & Burton, RS 2012, 'Investigating the molecular basis of local adaptation to thermal stress: population differences in gene expression across the transcriptome of the copepod *Tigriopus californicus*', *BMC Evol Biol*, vol. 12, p. 170.

Schulz, MH, Zerbino, DR, Vingron, M & Birney, E 2012, 'Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels', *Bioinformatics*, vol. 28, no. 8, pp. 1086-92.

Schurch, NJ, Schofield, P, Gierliński, M, Cole, C, Sherstnev, A, Singh, V, Wrobel, N, Gharbi, K, Simpson, GG, Owen-Hughes, T, Blaxter, M & Barton, GJ 2016, 'How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?', *RNA*, vol. 22, no. 6, pp. 839-51.

Seddon, PJ, Moro, D, Mitchell, NJ, Chauvenet, AL & Mawson, PR 2015, 'Proactive conservation or planned invasion? Past, current and future use of assisted colonisation', *Advances in reintroduction biology of Australian and New Zealand fauna*, vol. 105.

Shamiminoori, L 2015, 'Impacts of environmental variation on the fitness of the pygmy bluetongue lizards, *Tiliqua adelaidensis*', *Adelaide: Flinders University*.

Shamiminoori, L, Fenner, AL, Schofield, JA & Bull, CM 2015, 'Variation in size and condition of neonate pygmy bluetongue lizards, *Tiliqua adelaidensis*', *Transactions of the Royal Society of South Australia*, vol. 139, no. 2, pp. 131-44.

Shea, GM 2021, 'Nomenclature of supra-generic units within the Family Scincidae (Squamata)', *Zootaxa*, vol. 5067, no. 3, pp. 301-51.

Shen, XH, Zhao, YC, Wang, Z & Shi, QH 2022, 'Recent advances in high-throughput single-cell transcriptomics and spatial transcriptomics', Lab on a Chip, vol. 22, no. 24, pp. 4774-91.

Simpson, JT, Wong, K, Jackman, SD, Schein, JE, Jones, SJ & Birol, I 2009, 'ABySS: a parallel assembler for short read sequence data', *Genome Res*, vol. 19, no. 6, pp. 1117-23.

Singh, B & Dyce, KM 2018, *Dyce, Sack, and Wensing's textbook of veterinary anatomy*, Fifth edition. edn, Textbook of veterinary anatomy, Elsevier, St. Louis, Missouri.

Sjaastad, OV, Hove, K & Sand, O 2016, *Physiology of domestic animals*, Third edition. edn, Scandinavian Veterinary Press, Oslo, Norway.

Smith, AL, Gardner, MG, Fenner, AL & Bull, CM 2009, 'Restricted gene flow in the endangered pygmy bluetongue lizard (*Tiliqua adelaidensis*) in a fragmented agricultural landscape', *Wildlife Research*, vol. 36, no. 6, pp. 466-78.

Sommer, S 2005, 'The importance of immune gene variability (MHC) in evolutionary ecology and conservation', *Front Zool*, vol. 2, p. 16.

Souter, NJ, Bull, CM, Lethbridge, MR & Hutchinson, MN 2007, 'Habitat requirements of the endangered pygmy bluetongue lizard, *Tiliqua adelaidensis*', *Biological Conservation*, vol. 135, no. 1, pp. 33-45.

Sproul, D, Gilbert, N & Bickmore, WA 2005, 'The role of chromatin structure in regulating the expression of clustered genes', Nature Reviews Genetics, vol. 6, p. 775+.

Steward, CA, Parker, APJ, Minassian, BA, Sisodiya, SM, Frankish, A & Harrow, J 2017, 'Genome annotation for clinical genomic diagnostics: strengths and weaknesses', Genome Medicine, vol. 9, no. 1, p. 49.

Strickler, SR, Bombarely, A & Mueller, LA 2012, 'Designing a transcriptome next-generation sequencing project for a nonmodel plant species', *Am J Bot*, vol. 99, no. 2, pp. 257-66.

Sun, H, Zuo, X, Sun, L, Yan, P, Zhang, F, Xue, H, Li, E, Zhou, Y, Wu, R & Wu, X 2018, 'Insights into the seasonal adaptive mechanisms of Chinese alligators (*Alligator sinensis*) from transcriptomic analyses', *Australian Journal of Zoology*, vol. 66, no. 2, pp. 93-102.

Surget-Groba, Y & Montoya-Burgos, JI 2010, 'Optimization of de novo transcriptome assembly from next-generation sequencing data', *Genome Research*, vol. 20, no. 10, pp. 1432-40.

Takei, Y, Silldorff, EP, Hasegawa, Y, Watanabe, TX, Nakajima, K, Stephens, GA & Sakakibara, S 1993, 'New angiotensin I isolated from a reptile, *Alligator mississippiensis*', *Gen Comp Endocrinol*, vol. 90, no. 2, pp. 214-9.

Teeling, EC, Vernes, SC, Davalos, LM, Ray, DA, Gilbert, MTP, Myers, E & Bat, KC 2018, 'Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species', in HA Lewin & RM Roberts (eds), Annual Review of Animal Biosciences, Vol 6, Annual Reviews, Palo Alto, vol. 6, pp. 23-46.

The Australian Amphibians and Reptile Genomics initiative 2020, https://ausargenomics.com/

Thorpe, RS, Barlow, A, Malhotra, A & Surget-Groba, Y 2015, 'Widespread parallel population adaptation to climate variation across a radiation: implications for adaptation to climate change', *Mol Ecol*, vol. 24, no. 5, pp. 1019-30.

Tingley, R, Macdonald, SL, Mitchell, NJ, Woinarski, JCZ, Meiri, S, Bowles, P, Cox, NA, Shea, GM, Böhm, M, Chanson, J, Tognelli, MF, Harris, J, Walke, C, Harrison, N, Victor, S, Woods, C, Amey, AP, Bamford, M, Catt, G, Clemann, N, Couper, PJ, Cogger, H, Cowan, M, Craig, MD, Dickman, CR, Doughty, P, Ellis, R, Fenner, A, Ford, S, Gaikhorst, G, Gillespie, GR, Greenlees, MJ, Hobson, R, Hoskin, CJ, How, R, Hutchinson, MN, Lloyd, R, McDonald, P, Melville, J, Michael, DR, Moritz, C, Oliver, PM, Peterson, G, Robertson, P, Sanderson, C, Somaweera, R, Teale, R, Valentine, L, Vanderduys, E, Venz, M, Wapstra, E, Wilson, S & Chapple, DG 2019, 'Geographic and taxonomic patterns of extinction risk in Australian squamates', *Biological Conservation*, vol. 238, p. 108203.

Toth, AL, Varala, K, Newman, TC, Miguez, FE, Hutchison, SK, Willoughby, DA, Simons, JF, Egholm, M, Hunt, JH, Hudson, ME & Robinson, GE 2007, 'Wasp gene expression supports an

evolutionary link between maternal behavior and eusociality', *Science*, vol. 318, no. 5849, pp. 441-4.

Totikov, A, Tomarovsky, A, Prokopov, D, Yakupova, A, Bulyonkova, T, Derezanin, L, Rasskazov, D, Wolfsberger, WW, Koepfli, KP, Oleksyk, TK & Kliver, S 2021, 'Chromosome-Level Genome Assemblies Expand Capabilities of Genomics for Conservation Biology', Genes, vol. 12, no. 9, p. 17.

Trapnell, C, Hendrickson, DG, Sauvageau, M, Goff, L, Rinn, JL & Pachter, L 2013, 'Differential analysis of gene regulation at transcript resolution with RNA-seq', *Nat Biotechnol*, vol. 31, no. 1, pp. 46-53.

Trapnell, C, Pachter, L & Salzberg, SL 2009, 'TopHat: discovering splice junctions with RNA-Seq', *Bioinformatics*, vol. 25, no. 9, pp. 1105-11.

Trapnell, C, Roberts, A, Goff, L, Pertea, G, Kim, D, Kelley, DR, Pimentel, H, Salzberg, SL, Rinn, JL & Pachter, L 2012, 'Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks', *Nat Protoc*, vol. 7, no. 3, pp. 562-78.

Trapnell, C, Williams, BA, Pertea, G, Mortazavi, A, Kwan, G, van Baren, MJ, Salzberg, SL, Wold, BJ & Pachter, L 2010, 'Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation', *Nat Biotechnol*, vol. 28, no. 5, pp. 511-5.

Travis, JMJ, Delgado, M, Bocedi, G, Baguette, M, Bartoń, K, Bonte, D, Boulangeat, I, Hodgson, JA, Kubisch, A, Penteriani, V, Saastamoinen, M, Stevens, VM & Bullock, JM 2013, 'Dispersal and species' responses to climate change', *Oikos*, vol. 122, no. 11, pp. 1532-40.

Uetz, P., Freed, P. & Hošek, J. (eds.) 2020, 'The Reptile Database', http://www.reptile-database.org, accessed [Accessed 13/10/2021]

Ungaro, A, Pech, N, Martin, JF, McCairns, RJS, Mevy, JP, Chappaz, R & Gilles, A 2017, 'Challenges and advances for transcriptome assembly in non-model species', *Plos One*, vol. 12, no. 9, p. 21.

Van den Berge, K, Hembach, KM, Soneson, C, Tiberi, S, Clement, L, Love, MI, Patro, R & Robinson, MD 2019, 'RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis', *Annual Review of Biomedical Data Science*, vol. 2, no. 1, pp. 139-73.

Velculescu, VE, Zhang, L, Vogelstein, B & Kinzler, KW 1995, 'Serial analysis of gene expression', *Science*, vol. 270, no. 5235, pp. 484-7.

Vijay, N, Poelstra, JW, Kunstner, A & Wolf, JB 2013, 'Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments', *Mol Ecol*, vol. 22, no. 3, pp. 620-34.

Virkki, LV, Franke, C, Somieski, P & Boron, WF 2002, 'Cloning and functional characterization of a novel aquaporin from *Xenopus laevis* oocytes', *Journal of Biological Chemistry*, vol. 277, no. 43, pp. 40610-6.

Wang, SF & Gribskov, M 2017, 'Comprehensive evaluation of de novo transcriptome assembly programs and their effects on differential gene expression analysis', *Bioinformatics*, vol. 33, no. 3, pp. 327-33.

Wang, Z, Gerstein, M & Snyder, M 2009, 'RNA-Seq: a revolutionary tool for transcriptomics', *Nat Rev Genet*, vol. 10, no. 1, pp. 57-63.

Wang, J, Chen, XW, He, FN, Song, X, Huang, S, Yue, WC, Chen, YP, Su, ZX & Wang, CH 2020, 'Global Analysis of Gene Expression Profiles Provides Novel Insights into the Development and Evolution of the Large Crustacean Eriocheir sinensis', Genomics Proteomics & Bioinformatics, vol. 18, no. 4, pp. 443-54.

Weeks, AR, Sgro, CM, Young, AG, Frankham, R, Mitchell, NJ, Miller, KA, Byrne, M, Coates, DJ, Eldridge, MD, Sunnucks, P, Breed, MF, James, EA & Hoffmann, AA 2011, 'Assessing the benefits and risks of translocations in changing environments: a genetic perspective', *Evol Appl*, vol. 4, no. 6, pp. 709-25.

Whitehead, A, Triant, DA, Champlin, D & Nacci, D 2010, 'Comparative transcriptomics implicates mechanisms of evolved pollution tolerance in a killifish population', *Mol Ecol*, vol. 19, no. 23, pp. 5186-203.

Willi, Y, Kristensen, TN, Sgrò, CM, Weeks, AR, Ørsted, M & Hoffmann, AA 2022, 'Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species', Proceedings of the National Academy of Sciences, vol. 119, no. 1, p. e2105076119.

Wolf, JB, Bayer, T, Haubold, B, Schilhabel, M, Rosenstiel, P & Tautz, D 2010, 'Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow', *Mol Ecol*, vol. 19 Suppl 1, pp. 162-75.

Workman, RE, Myrka, AM, Wong, GW, Tseng, E, Welch, KC, Jr. & Timp, W 2018, 'Single-molecule, full-length transcript sequencing provides insight into the extreme metabolism of the ruby-throated hummingbird *Archilochus colubris*', *Gigascience*, vol. 7, no. 3, pp. 1-12.

Wright, DJ, Spurgin, LG, Collar, NJ, Komdeur, J, Burke, T & Richardson, DS 2014, 'The impact of translocations on neutral and functional genetic diversity within and among populations of the Seychelles warbler', *Mol Ecol*, vol. 23, no. 9, pp. 2165-77.

Xia, Z, Xu, H, Zhai, J, Li, D, Luo, H, He, C & Huang, X 2011, 'RNA-Seq analysis and de novo transcriptome assembly of *Hevea brasiliensis*', *Plant Mol Biol*, vol. 77, no. 3, pp. 299-308.

Xu, Q, Xing, S, Zhu, C, Liu, W, Fan, Y, Wang, Q, Song, Z, Yang, W, Luo, F, Shang, F, Kang, L, Chen, W, Yan, J, Li, J & Sang, T 2015, 'Population transcriptomics reveals a potentially positive role of expression diversity in adaptation', *J Integr Plant Biol*, vol. 57, no. 3, pp. 284-99.

Yang, WZ, Qi, Y & Fu, JZ 2014, 'Exploring the Genetic Basis of Adaptation to High Elevations in Reptiles: A Comparative Transcriptome Analysis of Two Toad-Headed Agamas (Genus *Phrynocephalus*)', *Plos One*, vol. 9, no. 11, p. 8.

Yang, Y, Wang, L, Han, J, Tang, X, Ma, M, Wang, K, Zhang, X, Ren, Q, Chen, Q & Qiu, Q 2015, 'Comparative transcriptomic analysis revealed adaptation mechanism of *Phrynocephalus erythrurus*, the highest altitude Lizard living in the Qinghai-Tibet Plateau', *BMC Evolutionary Biology*, vol. 15, no. 1.

Yin, J, Zeng, X, Ai, Z, Yu, M, Wu, Yo & Li, S 2020, 'Construction and analysis of a lncRNA-miRNA-mRNA network based on competitive endogenous RNA reveal functional lncRNAs in oral cancer', *BMC Medical Genomics*, vol. 13, no. 1, p. 84.

Yu, Y, Fuscoe, JC, Zhao, C, Guo, C, Jia, MW, Qing, T, Bannon, DI, Lancashire, L, Bao, WJ, Du, TT, Luo, H, Su, ZQ, Jones, WD, Moland, CL, Branham, WS, Qian, F, Ning, BT, Li, Y, Hong, HX, Guo, L, Mei, N, Shi, TL, Wang, KY, Wolfinger, RD, Nikolsky, Y, Walker, SJ, Duerksen-Hughes, P, Mason, CE, Tong, WD, Thierry-Mieg, J, Thierry-Mieg, D, Shi, LM & Wang, C 2014, 'A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages', Nature Communications, vol. 5, p. 11.

Zhao, S, Zhang, Y, Gamini, R, Zhang, B & von Schack, D 2018, 'Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: poly(A)+ selection versus rRNA depletion', *Scientific Reports*, vol. 8, no. 1, p. 4781.

Zhao, QY, Wang, Y, Kong, YM, Luo, D, Li, X & Hao, P 2011, 'Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study', *Bmc Bioinformatics*, vol. 12 Suppl 14, p. S2.

2. General Methods

Tiliqua adelaidensis in an artificial burrow – C. Maher

2.1 Preface

Chapter 2 provides an outline of sampling, wet-lab and sequencing protocols that are relevant to Chapters 3 and 4. They are included here in full to avoid repetition throughout the thesis. Any additions or alterations to the methods here, and further bioinformatic analysis are described in later chapters where relevant.

2.2 Sample Collection

Gene expression analysis requires tissue samples to be collected and snap frozen or fixed in RNA-later at the time of collection to adequately preserve mRNA. As *T. adelaidensis* is a threatened species (Australasian Reptile and Amphibian Specialist Group 1996), the number of biological replicates sampled and dissected was extremely restricted.

Sampling was conducted over multiple years because of different researchers being involved in this project. I took over the project and completed the sampling in 2016. Permits were acquired for collection of eight individuals, four from each experimental group (season of collection); sampling was limited by the threatened status of the species. Capture of lizards was completed by using an optiscope in burrows to identify occupied burrows which were than marked. At a later time or date to the disturbance from the optiscope, mealworms were then used to patiently lure lizards out of the burrow for hand capture. Lizards were processed quickly, and one at a time, and not all lizards for a season were collected on the same day. This was to minimise capture time and stress, as well as to facilitate tissues being immediately placed in liquid nitrogen for each collection. The aim was to collect two females and two males from each of two seasons (autumn and spring); however, because of limited sexual dimorphism in this species and (at the time) the lack of a genetic assay to determine sex chromosome which in the *Egernia* group is XX/XY (Bouffet-Halle et al. 2022), sex identification could not be confirmed until after euthanasia and dissection. While sex of a *T. adelaidensis* individual may often be indicated by a larger skull size or overall body size of full-grown males, there is a wide range of overlap in morphological measurements of each sex, and ages of captured lizards may vary (Shamiminoori et al. 2014). Derne et al. (2018) found that observation of hemipenes when taking cloacal samples could confirm a male, but lack of observation should not be used to assume female. The lack of a reliable live sexing method meant that three males and five females were collected. Thus, the seasonal groups were represented by four individuals each, but the autumn group comprises of three females and one male (Table 2-1).

As this is the first study of this type on *T. adelaidensis* the potential extent of different gene expression between for females and males of each season group is unknown and additional analyses were conducted separating males from the analysis at a later stage.

The distribution area in Figure 2-1 is intentionally obfuscated because of a history of illegal pet trade in this genus (Heinrich et al. 2022). Eight individuals were collected (Table 2-1) from the Nature Foundation of South Australia's *Tiliqua* Reserve in South Australia between 2014 and 2016. Four lizards were collected in early Austral spring (September) during the wet period leading into summer, and four at the beginning of the Austral autumn (March and April) at the end of the dry period but before the rains of winter.

Table 2-1: *Tiliqua adelaidensis* samples collected from *Tiliqua* Reserve by season and sex. Lizard ID is shown in bold and collection date in parentheses.

	Male	**Male**	**Female**	**Female**	**Female**
Spring	**G5** (27-Sep-15)	**G6** (27-Sep-15)	**G1** (25-Sep-14)	**G2** (25-Sep-14)	-
Autumn	**G4** (26-Mar-15)	-	**G3** (25-Mar-15)	**G7** (13-Apr-16)	**G8** (13-Apr-16)

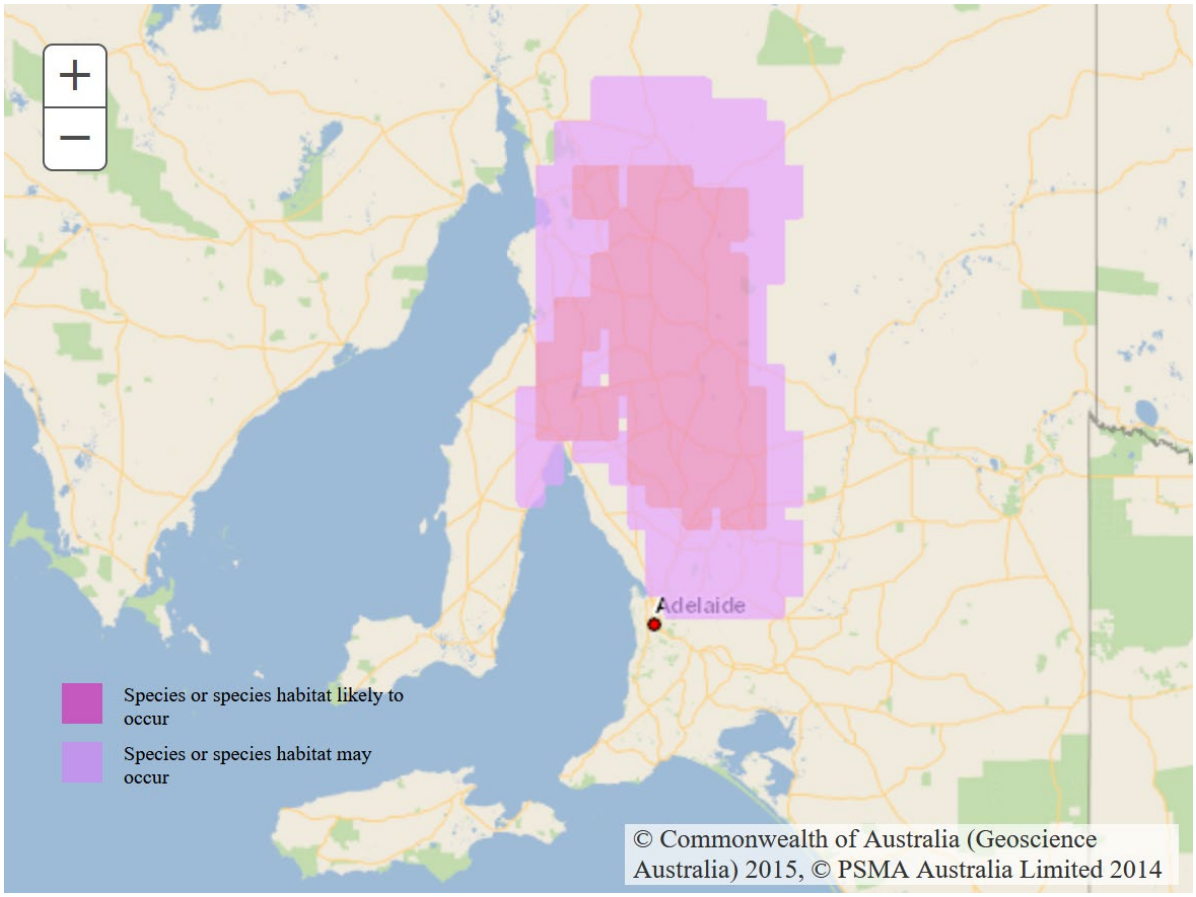


Figure 2-1: *Tiliqua adelaidensis* distribution, Australian Government; Department of Climate Change, Energy, the Environment and Water (2023) generalised from the Departments Species of National Environmental Significance dataset, Approximate range is shown in dark pink, previous predicted or likely range is shown in light pink (Previously available through DAWE 2012). Licensed under CC-BY 4.0.

This species has only recently been successfully bred in captivity at Monarto Zoo SA, and using controlled environments to manipulate any factors expected to influence gene expression—such as cohorts, diet or temperature—was not possible. As the focus of this study was identifying individual responses to environmental changes, samples were collected from wild populations during naturally occurring seasons.

Temperature measurements from Clare High School Station were used in Figure 2-2 as it is the nearest synoptic station with a full dataset, despite being over 40 km distant from the field site. Burra Community School Station, however, collects rainfall data closer to the site—approximately 15 kilometres from the field site—and is more consistently quality checked for rainfall data integrity (Figure 2-3). Missing data for some months prevented a total yearly rainfall being calculated for 2014 and 2016. The annual total rainfall in 2015 was 354.6 mm, while the annual average rainfall was 424.3mm. At the time of sampling in 2014–16 an El Niño climate event was promoting warmer and drier conditions (Bureau of Meteorology 2022[b]). Temperature and rainfall data during the years of this study are shown in Figure 2-2 and Figure 2-3.

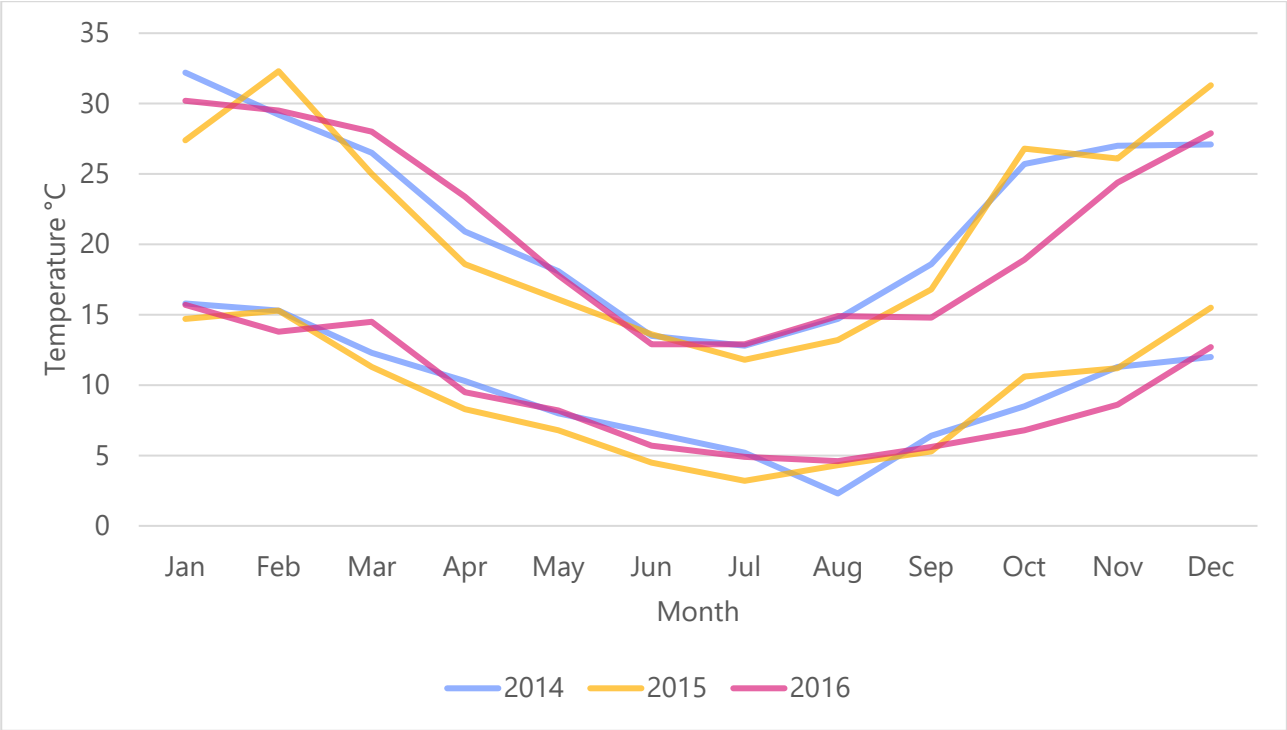


Figure 2-2: Monthly Mean Maximum and Minimum Temperature (Clare High School Station) for 2014-2016. Station number 021131, Station purpose Synoptic and Aeronautical, Latitude -33.8226, Longitude 138.5933, Elevation 395m. (Bureau of Meteorology 2022[a])

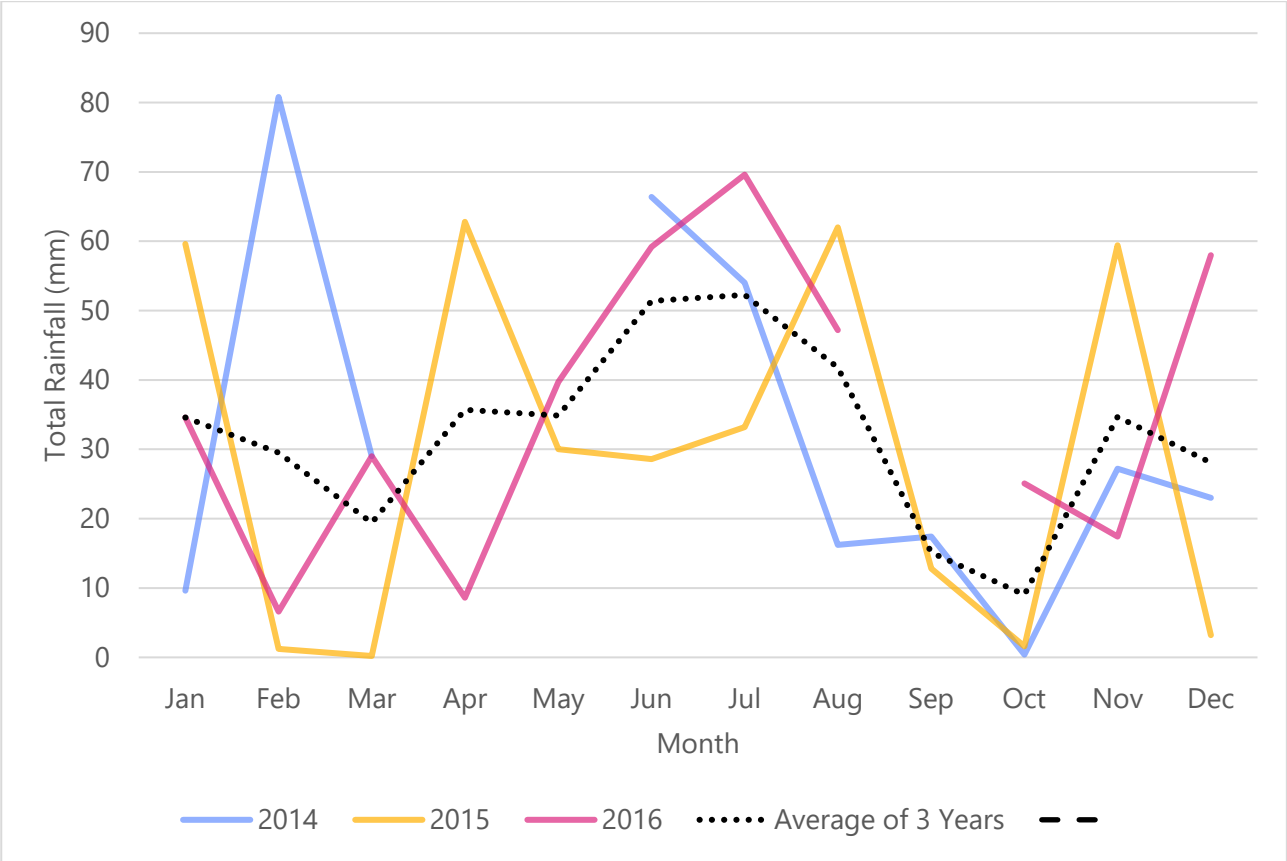


Figure 2-3: Monthly total rainfall and mean monthly rainfall for select years (mm) (Burra Community School Station). Station number 021077, Station purpose Telegraphic Rain, Latitude -33.6781, Longitude 138.9371, Elevation 470m. Breaks in 2014 and 2016 are a result of incomplete data. (Bureau of Meteorology 2022[a])

Pygmy bluetongue individuals were humanely killed on site in the field shortly after capture, as outlined previously in the ethics declaration, and tissue samples were dissected and then stored in liquid nitrogen *in situ* to preserve RNA. A total of 71 frozen tissue samples consisting of blood, spleen, epididymis, eye, whole snout (vomeronasal), tongue, brain, heart, lung, ovary, testes, kidney, liver, muscle, stomach and hindgut were collected (full list: Table 2-2). Blood samples stored on FTA (Flinders Technology Associates) cards (Smith & Burgoyne 2004), blood smears, cloacal swabs and opportunistic scat samples were also collected from some individuals. Tissue samples were subsequently stored at –80 °C at the South Australian Museum in Adelaide.

Table 2-2: Tissues collected for eight *Tiliqua adelaidensis* lizards sampled at the *Tiliqua* Reserve, Burra S.A. in September, and March/April of 2014 2015, and 2016. Specimens are stored at -80 °C in the South Australian Museum. Entries marked with a red 0 were collected but have been exhausted. Entries marked with a black 0 were lost in transit. Museum IDs are SA Museum Registration numbers denoted by the SAMA prefix, or tissue numbers denoted by the ABTC prefix.

Sample information							Tissues Collected								
Specimen	Collected	Code	Sex	Snout-Vent Length	Museum ID	Kidney	Spleen	Epididymis	Eye	Snout (Vomeronasal)	Brain	Testis	Blood	Muscle	Ovary
1	25-Sep-14	G1	F	-	ABTC113950	1	-	-	-	-	0	-	-	-	1
2	25-Sep-14	G2	F	-	SAMAR68046	1	-	-	-	-	-	-	-	-	1
3	25-Mar-15	G3	F	101mm	SAMAR68380	1	-	-	-	-	-	-	-	-	1
4	26-Mar-15	G4	M	93mm	SAMAR68381	1	0	-	-	-	-	1	-	-	-
5	27-Sep-15	G5	M	93.5mm	ABTC113954	1	-	1	1	1	1	1	1	1	-
6	27-Sep-15	G6	M	80.0mm	SAMAR68963	0	-	1	-	-	-	1	1	0	-
7	13-Apr-16	G7	F	96mm	SAMAR69488	0	-	-	-	-	-	-	1	1	1
8	13-Apr-16	G8	F	85mm	ABTC113957	1	-	-	1	1	1	-	1	1	1

Sample information				Tissues taken						Other samples				
Specimen	Collected	Code	Sex	Lung	Stomach	Tongue	Heart	Hindgut	Liver	Blood FTA	Blood Smear	Scat	Cloacal swab	notes
1	25-Sep-14	G1	F	-	-	1	1	1	0	1	1	-	-	FTA Smear poor
2	25-Sep-14	G2	F	-	-	1	1	0	1	1	1	-	-	
3	25-Mar-15	G3	F	1	1	1	1	1	1	1	1	-	-	RHS ovary
4	26-Mar-15	G4	M	1	1	1	1	1	1	1	1	-	-	
5	27-Sep-15	G5	M	1	0	1	1	1	1	1	1	1	1	including skin, LHS testis, scat in SAF
6	27-Sep-15	G6	M	1	1	-	1	1	1	1	1	-	1	including skin, RHS testis, FTA poor
7	13-Apr-16	G7	F	1	1	-	1	1	1	1	1	1	-	tail/leg, scat in ethanol
8	13-Apr-16	G8	F	1	1	1	1	1	1	1	1	-	-	tail/leg, whole snout

2.3 Sample Preparation and Sequencing

2.3.1 RNA Isolation

Laboratory work was conducted in the South Australian Regional Facility for Molecular Ecology and Evolution (SARFMEE) facility at the University of Adelaide. RNA was extracted from kidney tissue from each of the eight individuals using the QIAGEN RNeasy mini extraction kit according to the manufacturer's instructions. Use of this kit results in the extraction of total RNA from the sample and includes a step to eliminate genomic DNA from the extraction. The supernatant was run through the spin column twice at the elution step to maximise yield.

Total RNA samples were gel visualised against a 1kb Promega ladder, using a bleach-agarose gel preparation (Aranda, LaJoie & Jorcyk 2012) consisting of 1.0% bleach added to a preparation of 1.0% agarose solution. Bromophenol blue and glycerol solution was mixed 1:1 with samples as a loading buffer and gels were dyed using MIDORI ^{Green} (Nippon Genetics) after electrophoresis in 1X TBE buffer at 100 V for 35 minutes. All RNA extracts were checked for RNA concentration and quality using an Agilent Bioanalyser high-sensitivity RNA assay (nanochip) at SA Pathology (Table 2-3). Bioanalyser reports for the six used samples and summary data for all extractions of other tissues are provided in Appendix 1 (Table 6-1, Figure 6-6). Ideally for long-read sequencing a high RNA integrity number (RIN) quality score is desirable; hence the RNA extraction with the best quality score on the Bioanalyser report (sample G6) was used for both long- and short-read sequencing. Some samples show low RIN scores (e.g., sample G1 has the lowest score of 5.9; Table 2-3), which may be due to a large amount of short-length degraded RNA evident in the peak at ~25 bp in Figure 6-1 and a low 28s/86s ratio of 0.6. However, these peaks are still distinguishable, and because of highly restricted sample availability, all of these samples were used for short-read sequencing.

The liver RNA extractions consistently showed low sample integrity, characterised by such low quality that they were not assigned a RIN score, and neither the 28s nor 18s peaks were discernible on the electropherogram. These samples, sought after for their relevance to the liver's involvement in uric acid excretion in reptiles and its contribution to water conservation, unfortunately did not yield extractions of the requisite quality for further analysis (Table 6-1).

Table 2-3: Summary Table of quality scoring using an Agilent Bioanalyser for total RNA extractions from *Tiliqua adelaidensis* kidney tissue samples used for sequencing. Full table of other unused tissue extractions available (Appendix 1, Table 6-1).

Sample ID #	Tissue	Date extracted	Date on report	Assay Class	Sample tube name	RIN	RNA Area	Concentration (ng/µL)	Ratio 28s/18s	µg RNA eluted
G1	Kidney	7/11/2016	23/11/2018	Eucaryote Total RNA Nano	G1-Kidney	5.9	78.2	52	0.6	1.56
G2	Kidney	14/11/2016	23/11/2018	Eucaryote Total RNA Nano	G2-Kidney	6.7	174.3	116	0.5	3.48
G5	Kidney	7/11/2016	23/11/2018	Eucaryote Total RNA Nano	G5-Kidney	7	299.6	200	0.7	6
G6	Kidney	5/06/2018	14/06/2018	Eucaryote Total RNA Nano	G6-Kidney	8.6	1958.4	530	1.1	15.9
G3	Kidney	14/11/2016	23/11/2018	Eucaryote Total RNA Nano	G3-Kidney	7	675.3	451	0.6	13.53
G4	Kidney	14/11/2016	23/11/2018	Eucaryote Total RNA Nano	G4-Kidney	7.1	192.4	128	3.4	3.84
G7	Kidney	5/06/2018	14/06/2018	Eucaryote Total RNA Nano	G7-Kidney	7.9	566.7	153	0.9	4.59
G8	Kidney	14/11/2016	23/11/2018	Eucaryote Total RNA Nano	G8-Kidney	6.8	347.2	232	0.6	6.96

2.3.2 Short-read Sequencing

Sequencing was completed over two runs because of extraction difficulties and timeframes, and to allow analysis of the first run to begin while remaining samples were processed in the lab. Splitting these samples over two Illumina HiSeq sequencing runs was also expected to achieve greater read depth per sample, because of the resulting fewer samples pooled per sequencing lane. A minimum of 10 million reads per sample available on the sequencing plate is recommended (Todd, Black & Gemmell 2016) but greater read depth is preferable, particularly in species with limited genomic data where the expected outputs are unknown, as is the case here. Six samples (G1, G2, G3, G4, G5 and G8) were chosen for the first sequencing run based on concentration and quality scores (Appendix 1, Table 6-1) and to provide the most equal representation of each experimental group at the time. These were two females and one male from each sampled season.

Complementary DNA (cDNA) was synthesised using the Clontech SMARTer cDNA synthesis kit and the Advantage2 PCR kit was used for amplification, according to the manufacturer's protocols to generate a non-stranded library. The SMARTer II A Oligonucleotide
5'– AAGCAGTGGTATCAACGCAGAGTACXXXXX –3' (X = undisclosed base in the proprietary, SMARTer oligo sequence), and 3' SMART CDS Primer II A
5'– AAGCAGTGGTATCAACGCAGAGTACT$_{(30)}$N$_{-1}$N –3' (N = A, C, G or T; N-1 = A, G or C) was used for first-strand cDNA synthesis, followed by the 5' PCR Primer II A for the second-step PCR reaction. Second-step PCR amplification cycles were optimised by observing 28s and 18s bands on a bleach gel (Aranda, LaJoie & Jorcyk 2012) after samples were subjected to different numbers of cycles. The

final PCR program was 95°C 1 min, followed by an optimised number of cycles of 95°C 15 sec, 65°C 30 sec and 68°C 6 min.

Samples G1–G5 were optimised at 17 cycles for amplification and sample G8 was optimised at 16 cycles. Dual-stranded cDNA for each sample was visualised on 1.5% agarose against a 1 kb ladder. Electrophoresis was again performed using 1:1 bromophenol blue and glycerol solution as a loading buffer in 1 X TBE buffer at 105 V for one hour. Products were cleaned using the MO-Bio Laboratories UltraClean PCR Clean-Up Kit according to the manufacturer's instructions, except that samples were spun at 12,000 x g at steps 4, 8, 10 and 13, and the supernatant was eluted through the spin filter membrane twice at steps 12–13.

Samples were checked for RNA quality, concentration and length (to determine degradation) on an Agilent Technologies 220Tapestation on a High Sensitivity D1000 ScreenTape before submission to the Australian Genomics Research Facility (AGRF) where further quality checks were also performed before sequencing. Between 2200 ng and 2500 ng of each cDNA extract at a concentration of >34 ng/µL was provided for sequencing. Multiplexed library preparation and size sonication were performed by AGRF before 150 bp PE sequencing on the HiSeq platform. This read length was chosen after initial sonication aiming for longer read inserts resulted in poor library representation due to inefficient shearing and enrichment of much smaller cDNA fragments consisting of Clontech PCR primer dimer.

The two remaining kidney samples (G6 and G7) were sequenced on a separate Illumina HiSeq run to a target length of 125 bp. For these two samples, synthesis of cDNA was not completed as above; rather total RNA in RNase-free water was sent to AGRF where synthesis of cDNA, stranded library preparation and 125 bp PE HiSeq sequencing on an Illumina HiSeq were completed (the six samples above were non-stranded due to the nature of the kit). This shorter sequencing length is the result of changes to AGRF protocols for the HiSeq system and was unavoidable.

The sequencing data from these kidney tissues were used for assembly and gene expression analysis presented in Chapters 3 and 4 respectively.

2.3.3 Long-read Sequencing

Sample G6 was also sequenced using long-read Iso-Seq library preparation. This sample was chosen based on the total RNA concentration and a RIN score of 8.6 for extractions quantified on

the Agilent Bioanalyser (Appendix 1 Table 6-1 & Figure 6-6). Sample G6 is from a male individual collected in spring. Template preparation and cDNA synthesis for this sample were carried out at SARFMEE using the Pacific Biosciences SMRTBell Template Prep Kit 1.0 SPv3 and Sequel Sequencing Kit v3.0, following the Pacific Biosciences Iso-Seq Template Preparation for Sequel Systems protocol (version 5) as intended for a target size of <4 kb (i.e., no size selection). This protocol was followed to completion of the 'large-scale optimised PCR cycle' step as per the manufacturer's protocols, to synthesise 16 reactions of 50 µL. This 800 µL of PCR product was sent on dry ice to the Ramaciotti Centre for Genomics at the University of NSW, Sydney, Australia, where the final Iso-Seq library preparation and long-read sequencing on a Pacific Biosciences Sequel system was completed (Table 3-3). Sequencing was conducted on 1 SMRT cell with a movie length of 600 minutes on a PacBio Sequel I instrument using Sequel Sequencing Kit v3.0 chemistry and the SMRT software version Sequel v6.0 (Rhoads & Au 2015).

The sequencing data from this single long-read sample were used for transcript annotation as described in Chapter 3 and forms the pseudo-reference for short-read mapping presented in Chapter 4.

2.4 Data Analysis - Computing Resources

Assembly of short-read transcripts, annotation of long-read transcripts and clustering of putative gene families is discussed in Chapter 3, and differential gene expression analysis between seasonal sample groups is discussed in Chapter 4.

The analyses used a combination of high-performance computing (HPC) and cloud virtual machines with the following specifications.

2.4.1 Short-read Data

Short-read QC, trimming, initial Trinity assembly and alignments were completed using Flinders University's HPC 'Deep Thought' with 16 x 256 Gb (4 Tb RAM), 1024 AMD x86 CPU cores at 2.0 GHz across 14 standard compute and 2 data science nodes, and 100 TiB of usable storage via the Dell EMC Network Storage reference architecture.

2.4.2 Long-read Data

Long-read cleaning, read frame prediction and BLAST searches were run on the Australian National eResearch Collaboration Tools and Resources project (NECTAR) cloud, using an m1.xlarge flavour allocation with 8 AMD Opteron 63xx class CPUs, 32 Gb RAM, 10 Gb root disk, and 240 Gb of secondary disk storage.

A single step of the ANGEL program, which required a higher allocation of short-term memory, was run on a Dell PowerEdge R910 server with 40 cores and 512 Gb RAM at the South Australian Museum.

2.5 References

Aranda, PS, LaJoie, DM & Jorcyk, CL 2012, 'Bleach gel: A simple agarose gel for analyzing RNA quality', *Electrophoresis*, vol. 33, no. 2, pp. 366-9.

Australian Government; Department of Climate Change, Energy, the Environment and Water 2023, '*Tiliqua adelaidensis*—Pygmy Blue-tongue Lizard, Adelaide Blue-tongue Lizard', https://www.environment.gov.au/cgi-bin/sprat/public/publicspecies.pl?taxon_id=1270

Australasian Reptile and Amphibian Specialist Group 1996, '*Tiliqua adelaidensis*', *The IUCN Red List of Threatened Species*, vol. Version 2014.3.

Bouffet-Halle, A, Yang, W, Gardner, MG, Whiting, MJ, Wapstra, E, Uller, T & While, GM 2022, 'Characterisation and cross-amplification of sex-specific genetic markers in Australasian Egerniinae lizards and their implications for understanding the evolution of sex determination and social complexity', Australian Journal of Zoology, vol. 69, no. 2, pp. 33-40, 8.

Bureau of Meteorology 2022[a], 'Climate Data Online, Weather Station Directory', *Australian Government, Bureau of Meteorology*, http://www.bom.gov.au/climate/data/stations/ [Accessed 21 Mar. 2022].

Bureau of Meteorology 2022[b], 'ENSO Outlook', Australian Government, Bureau of Meteorology, http://www.bom.gov.au/climate/enso/outlook/ [Accessed 21 Mar. 2022].

Derne, B, Hutchinson, M, Weinstein, P, Gardner, M & Halliday, B 2018, 'Parasite in peril? A new species of mite in the genus Ophiomegistus Banks (Parasitiformes: Paramegistidae) on an endangered host, the pygmy bluetongue lizard Tiliqua adelaidensis (Peters) (Squamata: Scincidae)', *Austral Ecology*, vol. 44.

Heinrich, S, Toomes, A, Shepherd, CR, Stringham, OC, Swan, M & Cassey, P 2022, 'Strengthening protection of endemic wildlife threatened by the international pet trade: The case of the Australian shingleback lizard', *Animal Conservation*, vol. 25, no. 1, pp. 91-100.

Rhoads, A & Au, KF 2015, 'PacBio Sequencing and Its Applications', *Genomics Proteomics Bioinformatics*, vol. 13, no. 5, pp. 278-89.

Shamiminoori, L, Fenner, AL & Bull, CM 2014, 'Weight watching in burrows: variation in body condition in pygmy bluetongue lizards', Australian Journal of Zoology, vol. 62, no. 4, pp. 284-93.

Smith, LM & Burgoyne, LA 2004, 'Collecting, archiving and processing DNA from wildlife samples using FTA databasing paper', *BMC Ecol*, vol. 4, p. 4.

Todd, EV, Black, MA & Gemmell, NJ 2016, 'The power and promise of RNA-seq in ecology and evolution', *Mol Ecol*, vol. 25, no. 6, pp. 1224-41.

3. *De novo* Assembly and Annotation of genes expressed in *Tiliqua adelaidensis* Kidneys

Tiliqua adelaidensis - C. Maher

3.1 Abstract

The absence of full-length, accurate, reference sequences is a common hurdle to be overcome in genomic studies of non-model species. Differences in sequencing methods, or choice of bioinformatics tools often result in differences in transcript assembly accuracy or completeness. However, with the increasing accuracy provided by newer long-read sequencing technologies such as Pacific Biosciences Iso-Seq, and their increasing accessibility and lower costs, full-length transcript sequencing from a study species is now possible. Gene expression studies, in which short-read RNA-seq is used to quantify expression of transcripts, rely heavily on accurate reference sequences. Ideally an annotated genome of the study species would be used, but references are rarely available for many non-model organisms. Since there is currently no closely related published full genome data available for *T. adelaidensis*, other options for references include use of genomes of less related species or existing transcriptomes using *de novo* assembly of the same short-read data; or use of alternative technologies to produce comparative transcript sequences. The analyses reported in this chapter used both short- and long-read RNA-seq data for poly(A) selected RNA isolated from the kidney tissue of *T. adelaidensis* individuals via sequencing on an Illumina HiSeq and Pacific Biosciences Iso-Seq Sequel system, respectively. Initial plans for this study involved using a pseudo-reference generated through *de novo* assembly of multiple short-read samples into a single list of transcripts. The number of resulting transcripts and completeness of the library for this approach are compared to directly sequenced transcripts of a single sample using long-read sequencing. Scores for assembly completeness were assessed using BUSCO facilitated sample comparison and showed that while short-read assemblies had greater transcript discovery, the long-read data had less incomplete or fragmented transcripts, and—because of clustering—much less duplication. The full-length Iso-seq transcripts were annotated using BLASTx and the UniProt SwissProt database. A database search of NCBI to identify genes of further interest was also conducted to identify genes with putative annotations using gene symbols previously identified in other reptile species, to cross-check against putative annotations for the *T. adelaidensis* transcripts generated here. Here I have produced a transcriptome of genes expressed in the kidney of *T. adelaidensis* using PacBio full-length cDNA sequencing, clustered into gene families and given putative annotations using NCBI BLASTx. A gene orthology analysis was also conducted to summarise the functions and pathways involving genes in this set. These results stem from the first full analysis of transcripts in this species and provide valuable reference sequences for

genes expressed in the kidney. The bioinformatic pipeline used herein also provides an example of analysis for a novel species.

3.2 Introduction

Despite it being a highly diverse taxon that makes up 25% of all known reptile species (Hedges 2014), there is no high-quality, fully annotated genome published for a member of the Scincidae family (Appendix 3; Appendix 4.2). Indeed, there are currently few annotated, high-quality genome references available for any squamate species, and many available sequence data are mitochondrial in origin. A working understanding of the genetic diversity and adaptive capability of species is necessary for an effective understanding of conservation threats and management (Ralls et al. 2020). Two of the most phylogenetically related lizard genomes to *T. adelaidensis* available are the Carolina green anole *Anolis carolinensis* (Alfoldi et al. 2011) and the Australian central bearded dragon *Pogona vitticeps* (Georges et al. 2015). However, these represent species from two families—Iguanidae and Agamidae—that are more phylogenetically similar to each other than to any member of the Scincidae and are therefore of limited use as reference genomes for *T. adelaidensis*. Annotated references are a valuable tool in many genetic studies and gene expression analyses; facilitate identification of alleles and gene isoforms; and following detailed analysis, can inform breeding or translocation programs.

The current approach in transcript and gene expression studies is to align reads to a reference genome or transcriptome of the same or closely related species as this improves transcript recovery and identification of isoforms (Van den Berge et al. 2019; Vijay et al. 2013). However, conservation biology is often focused on organisms where such references are unavailable, and this lack of reference genomes has led to the emergence of numerous methods for *de novo* analysis of RNA sequencing data (Singhal 2013). Assembly, correction and methods for estimating expression levels in these scenarios are rapidly and continually improving. However, *de novo* assembly methods are still inferior to what can be achieved when a reference is available (Fu et al. 2019; Grabherr et al. 2011; Roscito et al. 2018; Vijay et al. 2013), including transcript references generated using long-read technologies (Minio et al. 2019).

Transcriptomes provide valuable genomic resources, particularly for gene annotation. For example, they were the first step towards annotating a large number of genes in the tuatara (*Sphenodon punctatus*; Miller et al. 2012). For gene expression studies, assembled transcriptomes often become

the only available reference resource for the quantification of expressed transcripts in species that lack other reference genomic data—sometimes with transcript assembly and then expression quantification completed on the same datasets; for example using programs such as Corset (Davidson & Oshlack 2014). Ideally, separate datasets should be used to generate a transcript reference (or train a prediction model) and run the analysis to avoid introducing a confirmation bias into the model. Continual improvements on these approaches have been made, such as incorporating transcriptome supplementation into the reference set (Davidson & Oshlack 2018). Ensuring the accuracy of this initial assembly is paramount when the assembled transcripts are intended to be used as a reference.

Many assembly programs use variations of k-mer (a read fragment of length k) selection and scaffolding (Rana et al. 2016; Wang & Gribskov 2017). The program Trinity, for example, relies heavily on de Bruijn graphs, which aim to line up k-mers into predicted transcript strings (Grabherr et al. 2011). This can introduce inaccuracies such as artifactual repeats into the assembly (Lee et al. 2017). The predictive algorithms used for *de novo* assembly have the capacity to create a large number of putative transcripts, but without a biological reference there is no way to tell if any are artefacts of random assembly (Holzer & Marz 2019). Where short reads do not span gene exon junctions in a transcript, identifying likely alternative splicing is not possible and reads from isoforms of a gene may become assembled into the wrong isoform. Genes with a variety of transcript isoforms may have these different isomer fragments combined into a longer, chimeric sequence or be missed entirely in the assembly, although use of larger k-mer sizes has been shown to produce fewer chimeric sequences (Wang & Gribskov 2017). Parsing isoform expression from gene-level expression has been a goal of statistical analysis since the inception of RNA-seq (Jiang & Wong 2009; Zhang et al. 2017) and can be controlled with long-read sequencing (Liu et al. 2017).

A plethora of bioinformatic methodologies are available to tackle *de novo* assembly, and biologists are often spoiled for choice when it comes to selecting the appropriate (and often, uniquely tailored) method for particular study designs, datasets and sequencing platforms (Amarasinghe et al. 2020; Wang & Gribskov 2017). Initial plans for the current project included *de novo* assembly of Illumina short-read data with no reference. However, long-read sequencing is becoming more accurate and more accessible. Here Pacific Biosciences Iso-seq long-read data are processed and compared to short-read data generated through *de novo* assembly methods on Illumina HiSeq to examine the benefits of combining and complementing sequencing approaches when working with

squamate organisms without a reference genome. The use of both short- and long-read sequencing methods allows a degree of self-correction through comparison of the outputs.

Many genetic studies rely on alignment to gene references to determine the identity or function of sequences. The availability of phylogenetically close reference sequences allows for greater accuracy and a better chance of identifying genes correctly, compared to *de novo* assembly (O'Keeffe & Jones 2019). In the case of Scincidae, genetic studies requiring reference alignment must often rely on orthologues from phylogenetically distant species. The limitations of these methods become evident when dealing with highly variable or large gene families, and genes with a number of transcript isoforms. In these cases, distinct paralogues may be misaligned to one representative orthologue despite being sequences of similar genes or gene isoforms obtained from the study species (Vijay et al. 2013).

Annotation of genes underpins analysis of genomic and transcriptomic data, providing essential context and allowing comparison of orthologues across species—as well as inferences about gene functions—to be made. Annotation of transcripts in RNA-seq studies where direct alignment to a genome is not possible can help identify evolutionary features and compare protein coding regions of sequences (Raghavan et al. 2022). Understanding the context in which genes are expressed, or the annotation of genes under differential expression, is essential for beginning to interpret RNA-seq results. Annotation of genes allows comparison between populations and has been used to measure gene expression and selection of genes at low temperatures and different elevations of *Phrynocephalus* lizards (Jin et al. 2021). Annotation has also identified a number of venom proteins through transcript isoform analysis (Tan & Tan 2021; Wong et al. 2021), and transcripts can provide the sequences necessary to identify splice isoforms and coding regions in genome assembly projects, supplementing other data (Eckalbar et al. 2013; Roscito et al. 2018).

There is no accepted single best-practice method for genotyping multigene families even when only a single gene family can be targeted by sequencing technologies (Lighten, van Oosterhout & Bentzen 2014; Amarasinghe et al. 2020), and there is even less consensus in regard to best practice when annotating entire transcriptomes. Errors in sequencing data can be difficult to pick up in organisms where deviations from a distant reference genome cannot be easily identified as error or natural variation. Ensuring high-quality initial data (Zhou & Rokas 2014), continued comparison and testing of gene identification consistency in a multi-method approach and utilising alignments

to references where available is necessary to ensure that annotations are completed with the highest accuracy.

Annotation of transcriptomes and identification of specific functions of genes can be achieved by comparison to annotated genomes of related species and existing protein and gene databases (Fraser et al. 2011). Gene ontology (GO) is the primary means of classifying functional genes using set attributes for their molecular function, biological process and cellular component (Fruzangohar et al. 2013). Classifying genes using this system allows comparison and understanding of relationships between genes. GO analysis facilitates the investigation of large amounts of functional genomic data through the categorisation of numerous sequences into functional groups. Such an analysis is used in this chapter to provide an overview of the kinds of transcript expressed in the pygmy bluetongue kidney, and their functions.

In this chapter I compare sequencing outputs and data analysis for short-read Illumina RNA-seq and long-read Pacific Biosciences Iso-Seq sequencing, with their respective mRNA transcripts isolated and sequenced from kidney tissue of *T. adelaidensis*. I annotated the genes expressed in the kidney tissue of *T. adelaidensis* and shortlisted those that may play a role in helping lizards survive dry periods. A broad list of potential genes of interest for further comparison and validation of annotations was constructed using a systematic NCBI gene database search (Appendix 3). I have focused on aquaporin genes as a case study for further annotation as this is a well-studied gene family in other species and is known to have roles in water permeability and renal processes. These genes have previously been identified in squamates using the Carolina green anole genome (Agre et al. 2002; Babonis, Miller & Evans 2011; Finn et al. 2014; Nielsen et al. 1995; Nielsen et al. 2002); however, no skink orthologues were found in the database search conducted here (Appendix 3).

3.3 Methods

3.3.1 Overview

3.3.1.a Datasets

'**Short-read**': Sequencing data from 150PE Illumina reads (non-stranded library) or 125PE Illumina reads (stranded library, for two samples only) generated from RNA isolated from eight individual *T. adelaidensis* kidneys (four from each of two experimental periods) sequenced on an Illumina HiSeq (Chapter 2).

'**Long-read**': Transcripts sequenced on one smart cell with a movie length of 600 minutes on a Pacific Biosciences Sequel I platform, from RNA isolated from a *T. adelaidensis* kidney (Chapter 2). On the basis of extracted RNA quality, the G6 sample was used to generate both datasets.

3.3.2 Transcript Assembly

3.3.2.a Short-read Data

Data Cleaning

Short-read sequencing data were assessed for quality prior to adapter or quality trimming, using FastQC with default parameters (Andrews 2018). The FastQC reports for each sample were summarised together using ngsReports (Ward & Pederson 2018) in the program R Studio (R Core Team 2019; R Studio Team 2020). Data analysis was conducted on Flinders University's HPC machine Deep Thought. Full analysis notes and scripts are provided in repositories supplied in Appendix 2 analysis files, and quality reports are available for download from GitHub.

Illumina TruSeq adapters were removed from the 3' ends of the sequencing reads using the program Cutadapt version 4.1 with the 'regular 3' paired adapters option to keep only read pairs. Reads were quality trimmed at the same time to a score of 5 on the phred-33 scale (a base call accuracy of 68%). This conservative quality trimming paired with a minimum read length cut-off was chosen to minimise sequence loss or potential effects of trimming on downstream expression analyses (Williams et al. 2016). Also, when aligning to long-read transcripts at a later stage any unmatching poor-quality sequence ends will not be mapped or counted (Yang et al. 2019). SMARTer PCR tags used to generate the cDNA were subsequently trimmed using the 'linked adapters' function with no filtering for 'anchored adapters', meaning that sequences were not required to have the 5' adapter to be retained. Reads of less than 30 bases were discarded, as the length assessment is completed on all reads regardless of adapter presence; it was applied only at this last stage. By default, any read pair that did not meet length or quality cut-offs was removed as a pair so that the cleaned data contained no singleton reads. Trimmed reads were subsequently assessed using FastQC version 0.11.9 (Andrews 2018) as detailed above, and reports compiled using ngsReports (Ward & Pederson 2018).

Transcript Assembly

Preparation of the cDNA library for the first of these sequencing runs resulted in a non-stranded library, whereas a stranded library was synthesised for the second sequencing run. As these two

datasets were analysed together, all data were treated as non-stranded in Trinity. *De novo* assembly of transcripts was conducted using Trinity version 2.14.0 (Grabherr et al. 2011; Haas et al. 2013) using a singularity container on the Deep Thought computer. Trinity was run on the paired Fastq files from each sample independently (Figure 3-1), using 16 CPU cores and a maximum of 128 Gb of memory.

The forward and reverse reads for each sample were also concatenated into single forward and reverse read file containing data from all eight individuals, and assembled again into one larger transcript assembly using Trinity with the same parameters (Figure 3-1). This assembly was run using 16 CPU cores and a maximum of 256 Gb of memory.

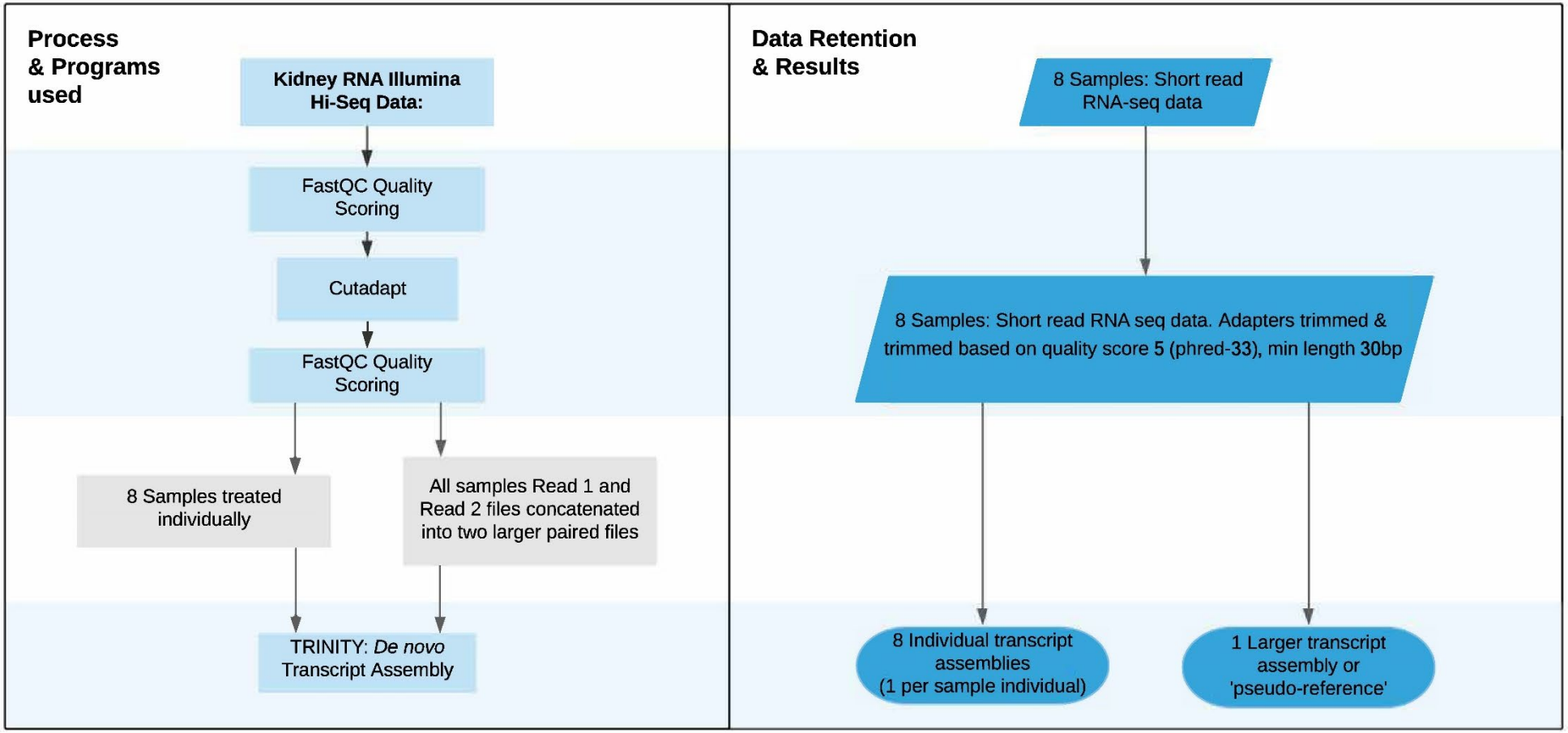


Figure 3-1: Flowchart showing data processing of poly(A) selected mRNA, isolated from kidney tissue from eight *Tiliqua adelaidensis* skinks, four collected in autumn and four collected in spring, sequenced using Illumina Hi-Seq PE short-read sequencing (two were then removed after batch effects could not be resolved). Transcripts assembled using the program Trinity. Respective Processes and outcomes are displayed in line, with bioinformatics workflow (left) and specific data values and retention of reads/unique transcripts (right).

3.3.2.b Long-read Data

Sampling, RNA purification and sequencing were completed as outlined in Chapter 2, on *T. adelaidensis* individual G6.

Transcript Cleaning

The Isoseq3 program (Pacific Biosciences 2018) and dependencies were installed on the NECTAR research cloud allocation within a conda environment, created using miniconda2 version 4.7.12 (Anaconda Software Distribution 2017). Sequencing output was processed as recommended by Pacific Biosciences, using the Isoseq3 program version 3.0.0 (commit v3.0.0-7-gcc6cddd, a subset of the program package smrttools-release_6.0.0.47835) (Pacific Biosciences 2018). Reads were collapsed (a step that corrects for this platform's base substitution error rate) (Figure 3-2) into one representative sequence for each zero-mode waveguide (ZMW) using the program ccs (installed with pbccs package version pbccs-3.4.1), which generated consensus sequences and lima version 1.9.0 that removed primers (Pacific Biosciences 2018). The 'cluster' tool in Isoseq3 was used to cluster circular consensus reads and generate unpolished transcripts that were then polished using the 'polish' tool, and a summary of transcript isoforms was generated by the 'summarize' tool. The Isoseq3 pipeline produces high-quality and low-quality transcript datasets. Only the high-quality transcript dataset was used in downstream analysis.

Within a new virtual Python environment, conda was used to install the program Cogent version 3.3 (Hun & Tseng 2019), which was used to cluster sequences into *de novo* isoform contigs by identifying transcript 'families'. Cogent uses k-mer profiles to calculate pairwise distances between sequences and to cluster isoforms into folders based on similarity, which are then reconstructed into an 'artificial genome' reference that is a list of canonical transcripts in cases where putative isoforms contain different putative exons. Minimap2 version 2.11-r797 (Li 2018) was then used to align the artificial genome to the high-quality transcripts generated in Isoseq3 sorted by k-mer length (longest to shortest) to create a sorted SAM file. Identical (or redundant) isoforms within these 'family' sets were then collapsed using cDNA_Cupcake version 8.3 (Tseng et al. 2021) (Figure 3-2). This created a fasta file listing the longest full sequences of all non-redundant, putative full-length transcript isoforms.

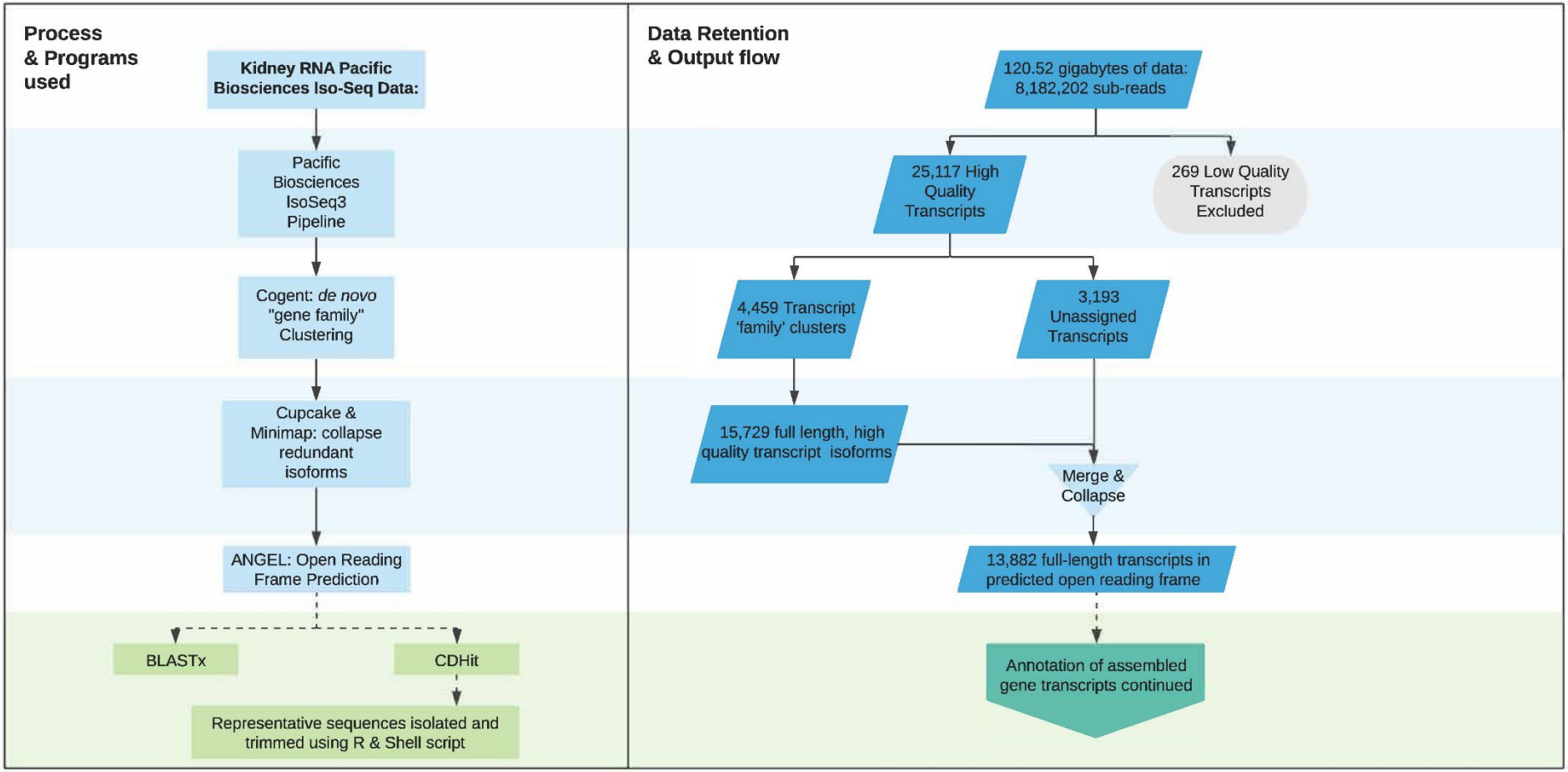


Figure 3-2: Flowchart showing data processing of long-read Iso-Seq sequencing of poly(A) selected RNA extracted from *Tiliqua adelaidensis* kidney tissue to a complete list of unique transcript isomers in predicted open reading frame. Respective processes and outcomes are displayed in line, with bioinformatics workflow (left) and specific data values and retention of reads/unique transcripts (right).

3.3.3 Transcript Annotation

3.3.3.a Long-read Data

Coding Region Prediction

The fasta file of non-redundant transcript isoforms was piped through the program ANGEL version 2.7 (Tseng et al. 2019) in the same Python environment to predict the protein coding open read frames. ANGEL also returns the 5′ and 3′ untranslated regions of transcripts as well as the coding sequence and translated peptide sequence in separate fasta files with scores for completeness and prediction confidence.

Putative Gene Clustering

The translated peptide sequences predicted by ANGEL were clustered into putative gene clusters using the CD-Hit online portal (Huang et al. 2010; Li & Godzik 2006; Li, Jaroszewski & Godzik 2001) with a similarity threshold of 0.99.

The CD-Hit output was imported into R Studio (R Core Team 2019; R Studio Team 2020) and manipulated using stringr (Wickham 2019) and dplyr (Wickham et al. 2021) to select the list of transcript IDs for only the representative canonical sequence for each putative isoform cluster, indicated by CD-Hit with an asterisk. Transcript ID names were appended with additional information at each stage of analysis; therefore, the fasta headers of the open read frame sequences after ANGEL processing contained more information than the original headers for the full-length transcripts. For this reason, the list of transcript identifiers for clustered sequences was truncated to remove additional unique information and match the original full-length transcript headers. The representative longest coding sequence identified by CD-Hit per cluster could then be cross-referenced to the file of non-redundant full-length transcript isoforms (as output by Cogent). This allowed the representative sequence for each cluster to be subset into a new file containing the full-length transcript; not only the predicted open read frame (the input for CD-Hit, to allow analysis based on protein translations). Care was taken to ensure the manual manipulation of the list resulted in no loss of data or mis-categorisation of transcripts, with counts and checks taken after every step. There were some instances of one full-length transcript resulting in two predicted open read frame sequences at different positions. Where more than one open read frame referenced back to the same full-length transcript, duplicates were removed from the final list. Although 9,907 putative clusters were generated by CD-Hit, 94 of the coding regions predicted by ANGEL that were assigned as representative sequences for these clusters were seeded from the

same transcript. Thus, the final list of representative transcripts included 9,813 full-length sequences without duplication. The transcripts identified as the representative sequence for each cluster in CD-Hit were subset into a fasta file and are referred to as putative genes from now on. A custom Perl command (Pers. comm. Bertozzi 2020) was then used to remove poly(A) tails from sequences in this file as some sequences still had long tails that persisted past the Isoseq3 pipeline stage. All scripts are available in Appendix 2.

BLAST Analyses

BLASTx searches on the coding sequence output by ANGEL for all transcript isoforms as the query were conducted using BLAST 2.9.0 (NCBI Resource Coordinators 2016) against the UniProt Swiss-Prot protein database (The UniProt Consortium 2018) downloaded on 28 May 2019 and *A. carolinensis* annotated protein databases *AnoCar2.0.pep.all* and *AnoCar2.0.pep.abinitio* (Alfoldi et al. 2011; Eckalbar et al. 2013) downloaded on 12 November 2019. The output format *-outfmt 6* was used for analysis with the parameters *-e-value = 1x10^{-5}, -max target seqs = 1* and *-max hsps = 1* (BLASTx parameters compared in initial data exploration included e-values of 1x10^{-5}and 1x10^{-10}, and variations of 1 and 5 max target seqs and max hsps enabled).

BLASTx searches for predicted coding regions of transcripts returned identifiers based on the format of the database searched. BLASTx searches against *A. carolinensis* genome data returned Ensembl Protein IDs, while BLASTx searches against the UniProt Swiss-Prot database returned UniProt IDs. Both sets of identifiers required further steps to obtain gene symbols for the matching transcript sequences. Further analysis was conducted only on the results from the UniProt Swiss-Prot database.

UniProt IDs were exported as a single column text file and translated using the online UniProt tool (The UniProt Consortium, T 2018) (Accessed 15 June 2020), which converts IDs to 'gene names'. Despite its designation, the UniProt 'gene name' field provides the same format of identifier as the 'gene symbol' field in the NCBI gene search (i.e., the abbreviated gene name ID e.g., '*slc1*' for 'solute carrier 1'). This ensured that annotated gene symbols could be compared across files using dplyr.

All of the above was completed with the goal of annotating as many isoforms as possible and including all data in the GO analysis. The reference set of 9,813 putative genes clustered using CD-

Hit was then compared to the full list of non-redundant transcripts using the dplyr filter function in R Studio to subset the BLAST results and include only hits for the canonical transcript for these putative genes.

Gene Ontology Analysis

GO analysis was conducted using Blast2GO within the OmicsBox suite version 1.2.4 (Conesa et al. 2005; Gotz et al. 2008). A BLASTx search using BLAST 2.9.0 on the predicted coding sequences was conducted against the *A. carolinensis* protein database AnoCar2.0.pep.all and the UniProt Swiss-Prot database with parameters *-evalue 1x10^{-5}, -max_target_seqs 5* and *-max_hsps 5*, using BLASTx output format *-outfmt 5* to produce an .xml library required for import into the OmicsBox program. The difference in max targets and max hits per sequence from the above query compared with initial BLAST analyses results from Omics-Box being capable of parsing multiple target hits per sequence and returning GO annotations (Gotz et al. 2008), which requires a differently formatted output (format 5 rather than 6). Based on these BLASTx hits and default Blast2GO options, gene symbols were used to infer GO terms and pathways. GO analysis of the BLASTx hits from the UniProt Swiss-Prot database was used here; however initial data exploration of GO terms produced by an analysis of the *A. carolinensis* genome was also used to supplement search terms in the genes of interest database search outlined in Appendix 3.

GO output was analysed for all three categories (P = biological process, F = molecular function, C = cellular component), and levels were explored based on distribution of the most representative sequences. Summary data for terms in each category across all levels were also generated.

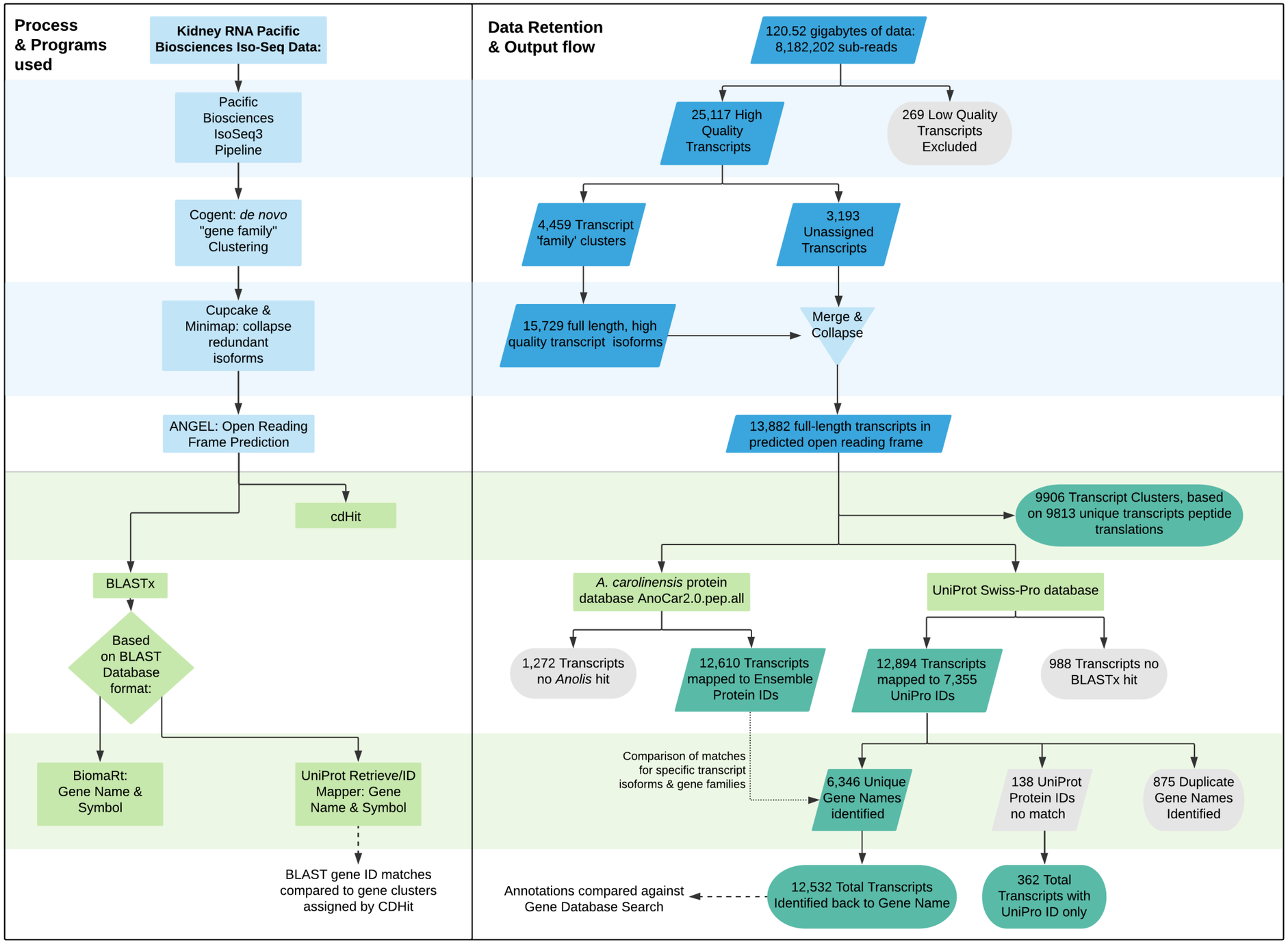


Figure 3-3: Bioinformatics workflow (left) and specific data values and retention through filtering of these same steps (right) showing pipeline from Pacific Biosciences Iso-Seq reads of sequenced *Tiliqua adelaidensis* kidney RNA to a complete list of unique transcript isomers in predicted open reading frame, assigned gene names based on BLASTx hits. This is an expanded version of (Figure 3-2) for a full depiction of bioinformatic analysis. Continued annotation workflow is indicated in green.

3.3.4 Selection of Genes of Interest

One of the ways these data were analysed was through the creation of and comparison to a list of 'genes of interest' (Appendix 3). A guided search of the NCBI gene database was conducted to create a shortlist of genes to ensure any genes not identified through a literature review might also be considered for further manual exploration. This list of genes was used to compare sequenced transcripts for identification purposes and to guide analysis of genes with an expected renal function that have been annotated in other reptiles. Key terms for the gene database search were initially selected based on common terms and themes found in the literature review (Appendix 4.2) and tailored to include genes likely to be affected by environmental conditions (Gust et al. 2018) or show differential expression in kidney tissue of reptiles (Ramsey & Crews 2007); in particular squamates (Babonis, Miller & Evans 2011). Initial literature reviews were used to inform final NCBI search terms to ensure capture of genes with important roles in the kidney based on physiological pathways in other reptiles.

A search on the NCBI gene database was conducted to identify orthologues of the 'genes of interest' in related taxa (Appendix 3). The search for gene candidates conducted here included all of the Sauria clade (a group encompassing Archosauria, Testudines, Sphenodontia and Squamata used in NCBI's taxonomy database [Schoch et al. 2020]). The gene and protein symbol conventions commonly used differ between human genes, mouse genes and genes of other vertebrate species. Thus, for consistency, in this thesis all protein identifiers are listed in upper case and are not italicised, while gene symbols are lower case and italicised except where abbreviations are used.

The terms used for this search were selected in multiple ways (Appendix 3). Initial key terms were added using common terms collected during a review of literature (Appendix 4.2). After searching in the AmiGO database for these search terms and other kidney and renal pathways, some other common GO terms and results were added to the list of search terms. As a GO analysis had been completed on the kidney transcripts at this stage, common terms from these results were also considered and added to the search as necessary. An NCBI gene search only recognises the first 600 matches in the database to a string of terms for any wildcard '*' while returning the true list of terms considered (Bethesda(MD) 2006). For example, in the search string below, 'Diffus*', 'Filt*' and 'Osmo*' query the database for more than one term with different suffixes. In most cases a wildcard search was initially used to explore the suffixes provided by the database, which were then changed

to more specific terms as wildcards were either unnecessary, or to reduce the number of irrelevant results.

The final search of the NCBI gene database used to populate the table of genes was conducted in the NCBI gene database on 14 June 2020. Search results were exported as a .csv file and de-duplicated in excel taking care to manually preserve data fields and gene symbols. The final search string entered was as below:

> (Sauria[Organism]) AND (Aquaporin OR Bile OR Dehydration OR Diffus* OR Excretion OR Filt* OR Fluid OR (Heat AND Stress) OR Heat Shock OR Heat Stress OR Hibernation OR Homeostasis OR Ion Channel OR Ion Transport OR Kidney OR Membrane And (Potential OR Permeability OR Pore) OR Metabolic OR Osmo* OR Permeability OR Ph Balance OR Renal OR Solute OR Stress OR Temperature OR Thermal OR Uric OR Water OR Water Retention OR Water Permeability OR Water Transport)

The goal of this search was to generate a list of gene symbols of further interest to guide the analysis of annotated transcripts and enable enumeration and comparison of genes relevant to renal function that are annotated and available for birds and reptiles with those identified in *T. adelaidensis.* As this list provides IDs and gene names for comparison only, the representative organism or specific corresponding sequence saved for each instance is of no consequence; but has been retained for reference; in other words the BLAST analysis of transcripts is not limited by the sequences returned here as only gene symbols were compared across lists.

Candidate genes were sorted based on gene symbol and duplicate gene symbols were removed keeping only the first instance of each gene. Although only gene symbols were used in the final analysis, filtering was also completed by organism to assess the current availability of reference sequences among squamates for the genes identified. The preferred order of squamate organisms was based on phylogenetic divergence from *T. adelaidensis*, or genome quality (Appendix 3). The final order of sorting here determined which gene example was kept (i.e., the first instance) when duplicate genes were removed. The species in the resulting list were highly phylogenetically diverged from *T. adelaidensis* (Pyron, Burbrink & Wiens 2013; Reeder et al. 2015) and therefore some species were arranged based on presence or quality of genome assembly; for example *A. carolinensis* (Alfoldi et al. 2011) and the Australian bearded dragon (Georges et al. 2015) were given precedence over other squamates based on genome completeness and quality. Schlegel's Japanese gecko *Gekko japonicus* (Liu et al. 2015), the common wall lizard *Podarcis muralis* (Andrade

et al. 2019) and the Asian water monitor *Varanus salvator* (Lind et al. 2019) also have genome information available. Birds and other reptiles, including crocodilians and turtles are phylogenetically diverged from squamates (Chiari et al. 2012) and sorting was arbitrary at this level for gene duplicate removal; however, included species are listed in their respective taxonomic groups for ease of interpretation (Appendix 3). The chicken, *Gallus gallus domesticus* is a commonly studied species with an annotated genome (Hillier et al. 2004); therefore, it was given precedence above all other birds for duplicate removal.

The total number of gene search results obtained and categorised for each taxon was recorded (Figure 3-4) and totals were calculated and compared to ensure no data were misplaced during this manipulation. After sorting by gene symbol and by an assigned taxon category, duplicates were removed from the dataset, leaving the 'first' result for each gene symbol for the most desired taxon group. Gene ID synonyms were checked by entering the list of filtered gene IDs into the HUGO Gene Nomenclature Committee's online multi-symbol checker (Seal et al. 2023). Results labelled 'LOC#' (n = 7,009) were returned as unique identifiers in the initial sorting and duplicate removal because gene identifiers for these loci were listed as the unique NCBI gene symbol preceded by 'LOC'. These entries included inconsistent annotation data or additional information in the description fields only and were manually removed.

Comparisons between the annotated dataset and the dataset from the genes of interest database were based on gene symbols. The list of identified genes of interest through the above search and the BLASTx gene results were compared in excel using unix shell commands. Comparative counts of total genes identified and genes identified in *T. adelaidensis*, as well as the NCBI shortlist, were generated.

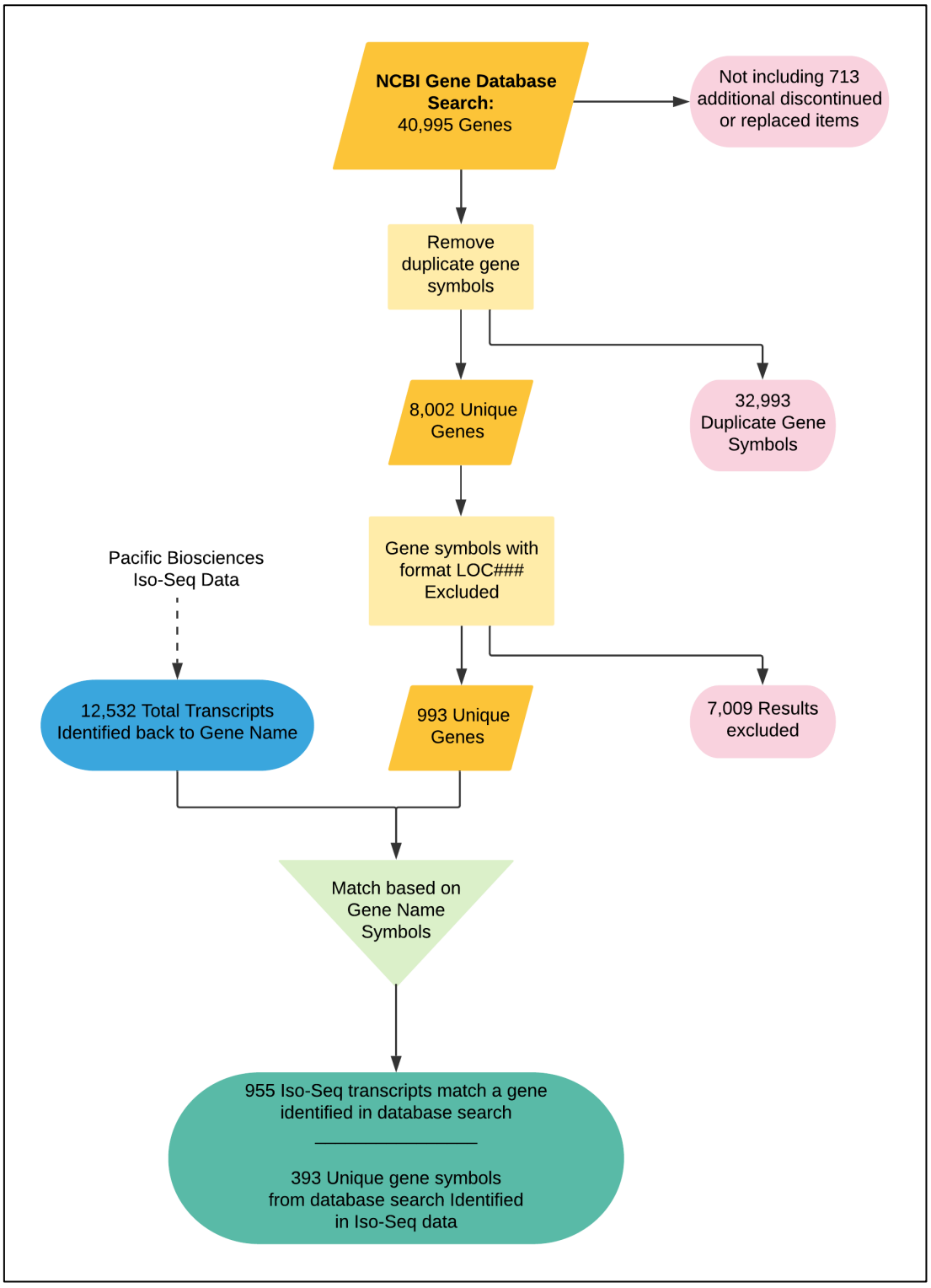


Figure 3-4: Selection of gene list based on a systematic search of the NCBI gene database using keywords tailored to return genes related to renal function and homeostasis, for further focus in annotations and application to putative annotations (Figure 3-3). Search returns for genes are represented by 104 different species, which are retained representing at least one gene throughout filtering. 7,009 Gene symbols with unconfirmed locus IDs with the name format "LOC" followed by the NCBI barcode number, were also excluded.

3.3.5 Transcriptome Comparison

3.3.5.a BUSCO

Assemblies were compared using BUSCO (Manni et al. 2021), which uses a database of near-universal orthologues to create a measure of completeness of gene assemblies based on conserved 'housekeeping' genes that are expected to be present. Samples were compared to the database 'vertebrata_odb10' using BUSCO version 5.4.20 on the Flinders Deep Thought machine. 'Completeness' scores (the number of sequences in the BUSCO set that aligned to sequences in the provided transcript files) were parsed through a Python script and visualised using R Studio.

3.3.5.b Burrows–Wheeler Alignment

To further compare the larger combined Trinity *de novo* assembly file made from all short-reads from all samples (including the later excluded G6 and G7) the assembled transcripts were aligned to the file of long-read Iso-Seq reference transcripts after clustering using the Burrows–Wheeler alignment (BWA) algorithm through BWA_MEM2 version 2.2.1 (Li & Durbin 2009; Li 2013).

3.3.5.c BLASTn comparison

A BLASTn search (NCBI Resource Coordinators 2016) through BLAST+ module version 2.12 was also conducted using the Trinity assembled transcripts as query sequences against the long-read sequences as the reference database. The parameters used were output format 6, e-value $1x10^{-5}$, maximum target sequences 5 and maximum hit per sequence 5. The tab delimited output format file was filtered using unix command line sort functions to select the result with the lowest e-value for each query returning one or more hits. The BLASTn output (with max one result per query) was then sorted by target sequence so that the transcript IDs for all hit sequences could be manually extracted and used to subset these sequences into a smaller fasta file for alignment visualisation. Examples of these alignments were exported using Seqtk version 1.4 and visualised manually to explore sequence similarities. The BLASTn option within MEGA-X version 10.0.5 was used for manual BLAST analysis of individual sequences against the NCBI database during visualisation. Trinity appends *de novo* transcript headers based on grouping and similarity when assembling, and includes <cluster number> (a result of the analysis indicating the initial cluster of reads parsed through Trinity at phase 1), <gene number> and <isoform number>. Therefore, after sorting by target sequence, groups of aligning query sequences to be visualised manually were selected based on exploring examples of a number of Trinity 'gene' or 'isoform' hits to the target sequence and instances where queries with very different numbering hit the same sequence.

3.4 Results

3.4.1 Transcript Assembly

3.4.1.a Short-read Data

Short-read sequencing for six samples returned 15–17 million reads per sample before adapter and quality trimming (Table 3-1). Samples G6 and G7 (the second sequencing run) had 70,947,779 and 73,164,031 paired-end reads respectively (Table 3-1) because of the smaller number of samples included in the run (Table 3-1). Normalisation factors were calculated and batch effects examined in downstream analyses (Chapter 4).

Table 3-1: Sequencing yield for eight *Tiliqua adelaidensis* kidney cDNA synthesised on poly(A) selected mRNA, isolated from Kidney tissue. Sequenced to 150 bp Paired End or 125 bp Paired End on an Illumina Hi-Seq RNA-seq (over two separate sequencing runs -with G6 and G7 separate).

Sample Name	Insert Size	Paired Reads	Data Yield
G1_KI	150 bp PE	17,054,742	5.12 Gb
G2_KI	150 bp PE	16,648,601	4.99 Gb
G3_KI	150 bp PE	15,335,728	4.60 Gb
G4_KI	150 bp PE	15,908,204	4.77 Gb
G5_KI	150 bp PE	15,437,712	4.63 Gb
G8_KI	150 bp PE	16,924,712	5.08 Gb
G6K	125 bp PE	70,947,779	17.88 Gb
G7K	125 bp PE	73,164,031	18.44 Gb

The comparison of FastQC reports before and after trimming shows that adapters were successfully removed, and that quality was improved across most samples. Other values in the FastQC report flagged with 'Fail' or 'Warning' were not unexpected in RNA-seq data; for example, the introduction of warning flags in all samples for sequence length distribution is expected as sequences were trimmed based on primer presence and quality to any length down to a minimum of 30 base pairs. Trimming reads successfully removed adapter content and improved per-tile sequence quality (Figure 3-5; Figure 3-6).

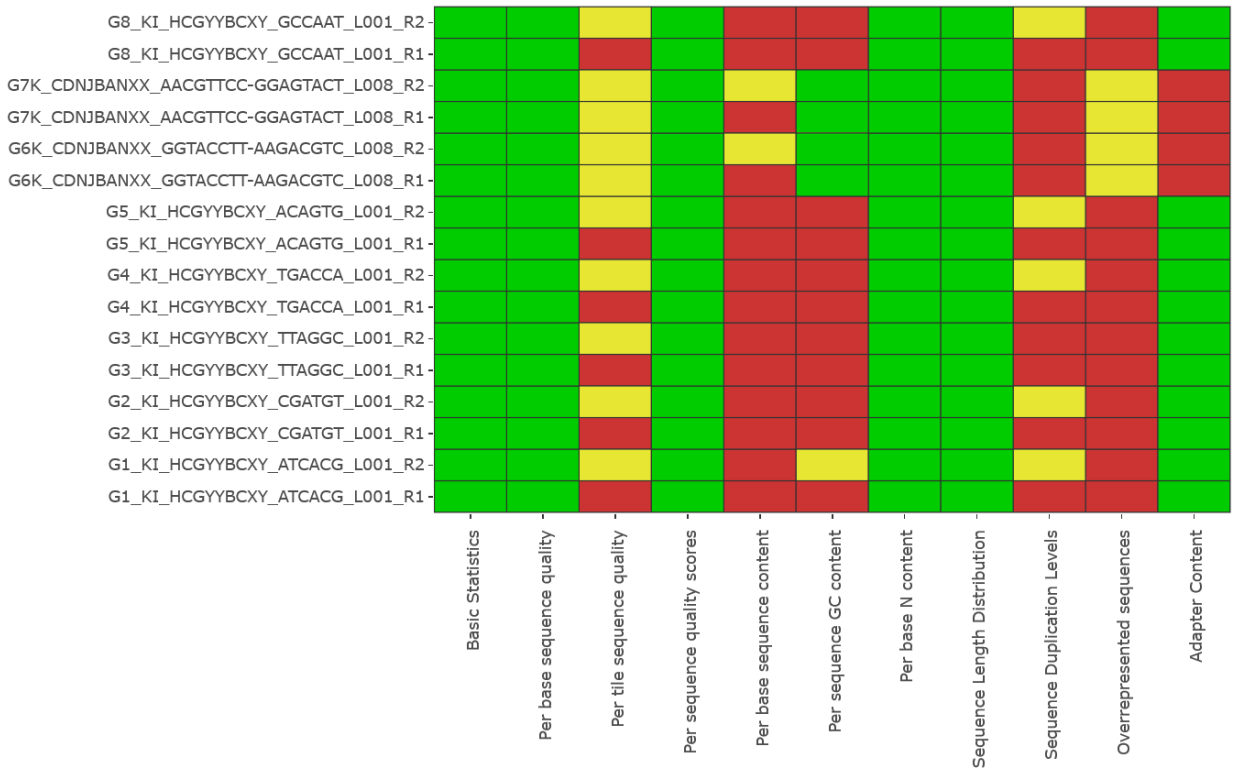


Figure 3-5: Summary of FastQC assessment of short-read samples prior to adapter removal or quality trimming. Colours indicate assessment of scores by FastQC for each overall category: Green=Pass, Yellow=Warning, Red=Fail. Poly(A) selected mRNA, isolated from kidney tissue from six *Tiliqua adelaidensis*.

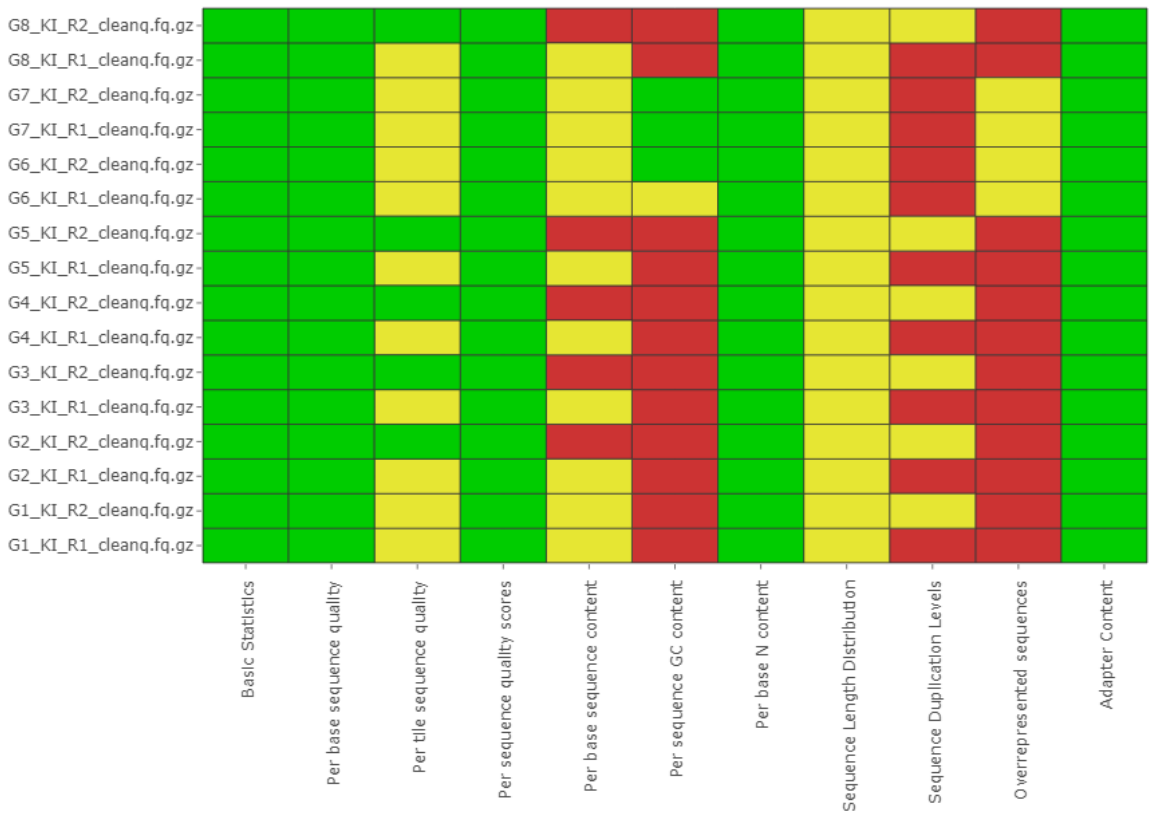


Figure 3-6: Summary of FastQC assessment of short-read samples after adapter removal and quality trimming. Colours indicate assessment of scores by FastQC for each overall category: Green=Pass, Yellow=Warning, Red=Fail. Poly(A) selected mRNA, isolated from kidney tissue from six *Tiliqua adelaidensis* skinks, three collected in autumn and three collected in spring, sequenced using Illumina Hi-Seq PE short-read sequencing.

The concatenated dataset containing reads from all samples and run as a single Trinity assembly (Figure 3-1) contained the most transcripts; approximately twice as many as for individual assemblies for G1, G2, G3, G4, G5 and G8 (Table 3-2).

Table 3-2: Poly(A) selected mRNA, isolated from kidney tissue from six *Tiliqua adelaidensis* skinks, three collected in autumn and three collected in spring, sequenced using Illumina Hi-Seq PE short-read sequencing. Total number of transcripts assembled by Trinity for each sample *note G6 and G7 are from a separate sequencing batch run.

Sample	Concatenated Trinity run	G1	G2	G3	G4	G5	G6*	G7*	G8
Number of Transcripts	359,197	48,161	45,075	37,039	40,893	38,719	185,868	229,145	43,621

3.4.1.b Long-read Data

Sequencing of the RNA from *T. adelaidensis* kidney tissue on the Pacific Biosciences Sequel system produced 120.52 gigabytes of data, containing 8,182,202 sub reads. A total of 14,139,224,811 base pairs were sequenced (Table 3-3). Initial filtering and collapsing transcript sequences with the Isoseq3 pipeline produced 25,117 high-quality and 269 low-quality consensus transcripts.

Table 3-3: Poly(A) selected mRNA, isolated from kidney tissue from one *Tiliqua adelaidensis* skink ID: G6, a representative individual for Pacific Biosciences Iso-Seq long-read transcript sequencing and annotation to generate full-length reference transcripts. Sequencing output — Long-read sequencing: Pacific Biosciences Iso-Seq Sequel system, no size selection (Sample G6)

1 smart cell, movie length = 600 minutes on sequel v6.0	
Filename	20190204_MAH6260A1.tar
Number of Sub reads	8,182,202
Total length of Sub reads (bp)	14,139,224,811
Data yield (Gb)	120.52

Clustering the high-quality transcripts resulted in 4,459 transcript isoform ‘families’, each of which contain transcript isoforms likely to originate from the same putative locus; and 3,193 singleton transcripts, which were not grouped into any ‘family’ cluster. Identical isoforms within these clusters were collapsed to provide a list of 15,729 unique, full-length, high-quality transcripts. Note that

unique isoforms within each cluster are retained and this analysis does not perform a gene-level collapse (if only one representative of each isoform family cluster and unassigned transcripts were retained there would only be around 7,000 resulting transcripts).

These isoforms were then further collapsed into representative transcripts of putative gene families based on ANGEL predicted open read frames and CD-Hit protein clustering. This resulted in a list of 13,882 unique, non-redundant, full-length 'canonical transcripts' in the predicted open reading frame (Figure 3-2), which further clustered into 9,813 putative gene loci based on protein translation (Figure 3-2).

3.4.2 Annotation

3.4.2.a Long-read Data

BLAST Analyses

The coding sequences of 13,882 transcripts in predicted open reading frame (Figure 3-3), when compared to the UniProt Swiss-Prot protein database, provided hits to 7,355 unique UniProtKB protein IDs for 12,532 transcript isoforms. These included 6,346 unique gene symbols and 138 protein IDs that did not map to gene symbol. Note that this BLAST analysis was conducted on the entire list of full-length transcripts in predicted read frame, before clustering into putative genes using CD-Hit. This was done to create the largest possible reference file for all transcript isoforms. When the full set of BLASTx hits were compared with the clustered list of putative genes, it was determined that 8,861 out of 9,813 gene sequences produced a BLASTx hit and was able to be annotated to gene name.

Gene Ontology Analysis

Omics-Box uses BLASTx hits to map GO terms and assign GO annotation. Of the 13,882 full-length transcripts in open read frame, 12,877 had BLAST hits. Of these, 19 had BLAST hits but no GO mapping; 186 had GO mapping but were not assigned GO annotation; and 12,672 transcripts were assigned GO mapping and then GO annotation, based on one or more BLAST result (Figure 3-7).

The top GO terms over all levels are summarised in Figure 3-8, Figure 3-9 and Figure 3-10, for biological process, molecular function and cellular component, respectively, to show the greatest number of terms based on frequency in the annotated transcript dataset. A much larger number of GO terms were returned in the biological process category overall, many of which are related to

metabolic processes, responses to stimuli and cell communication. Many of the GO terms included in the molecular function category are related to genetic binding and ion binding activity. 'Membrane' is the seventh top cellular component GO term, and more than half of the others are related to cellular or organelle membrane structure—whether directly 'organelle membrane', 'plasma membrane' or other membrane bound (or specifically non-membrane bounded) structures.

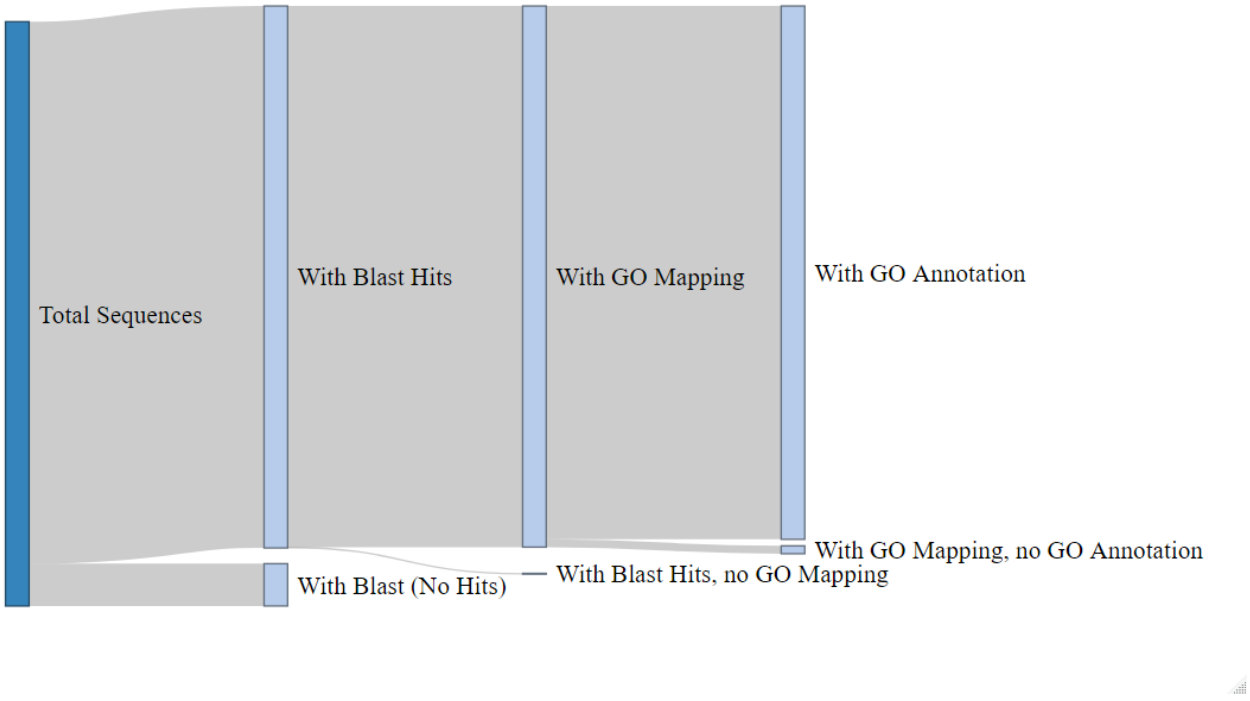


Figure 3-7: Sankey chart summary data distribution of Blast2GO analysis using the UniProt Swiss-Prot protein BLAST database. BLAST hits are required for GO mapping, which is required for GO annotation. All GO annotated transcripts therefore have both BLAST hits and GO mapping. The file of predicted *Tiliqua adelaidensis* transcripts in open reading frame contained 13,882 total unique sequences, 12672 of these sequences produced >1 BLASTx hit with Mapping and GO Annotation, 186 mapping but no GO annotation, 19 produced a BLAST hit but no GO mapping, and 1005 did not produce BLASTx hits.

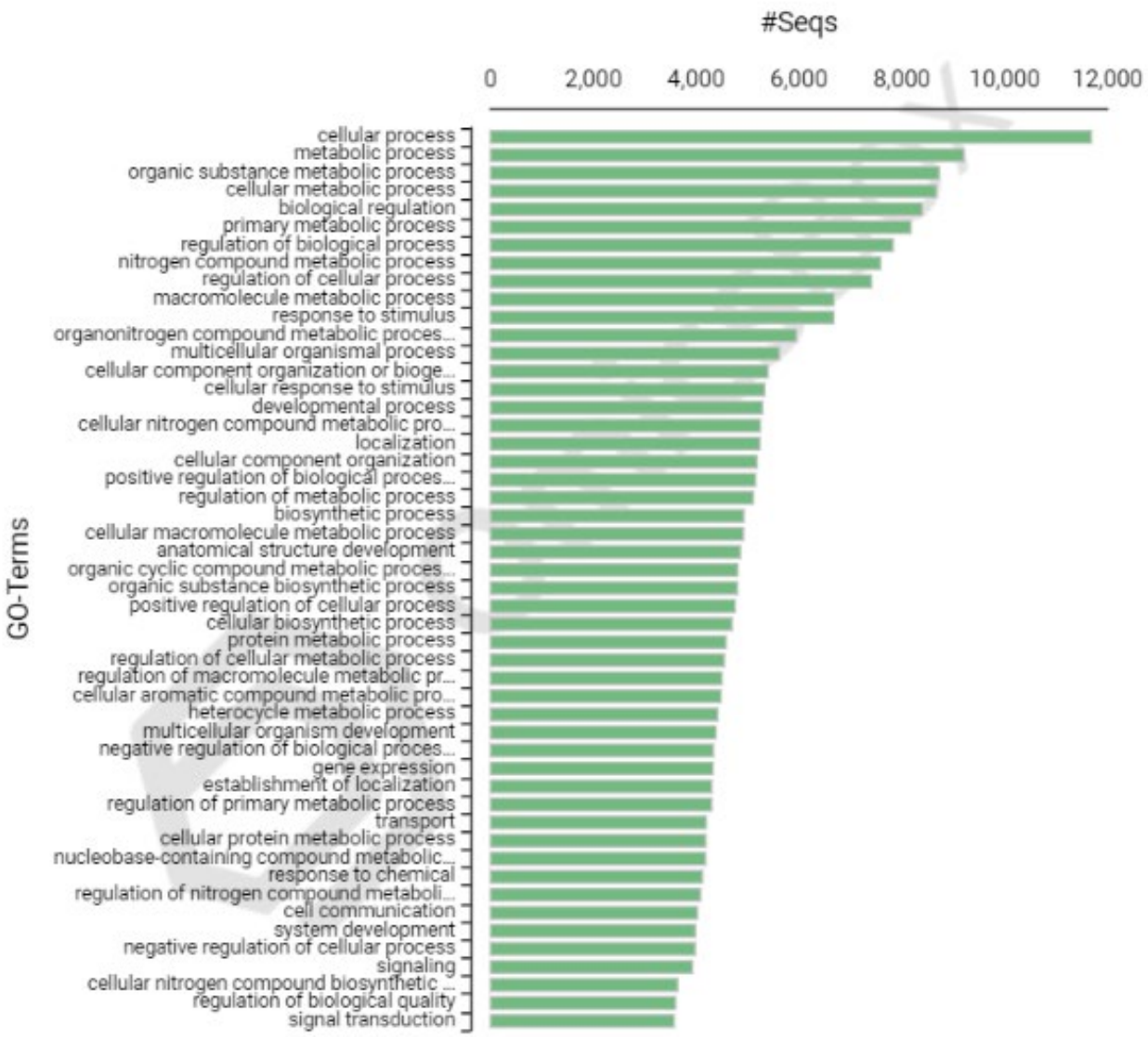


Figure 3-8: Poly(A) selected mRNA, isolated from kidney tissue from one *Tiliqua adelaidensis* skink ID: G6, a representative individual for Pacific Biosciences Iso-Seq long-read transcript sequencing and annotation to generate full-length reference transcripts. Gene ontology conducted using B2GO. Top gene ontology terms by number of annotated sequences for all GO levels in the biological process category for predicted *Tiliqua adelaidensis* transcripts in open reading frame.

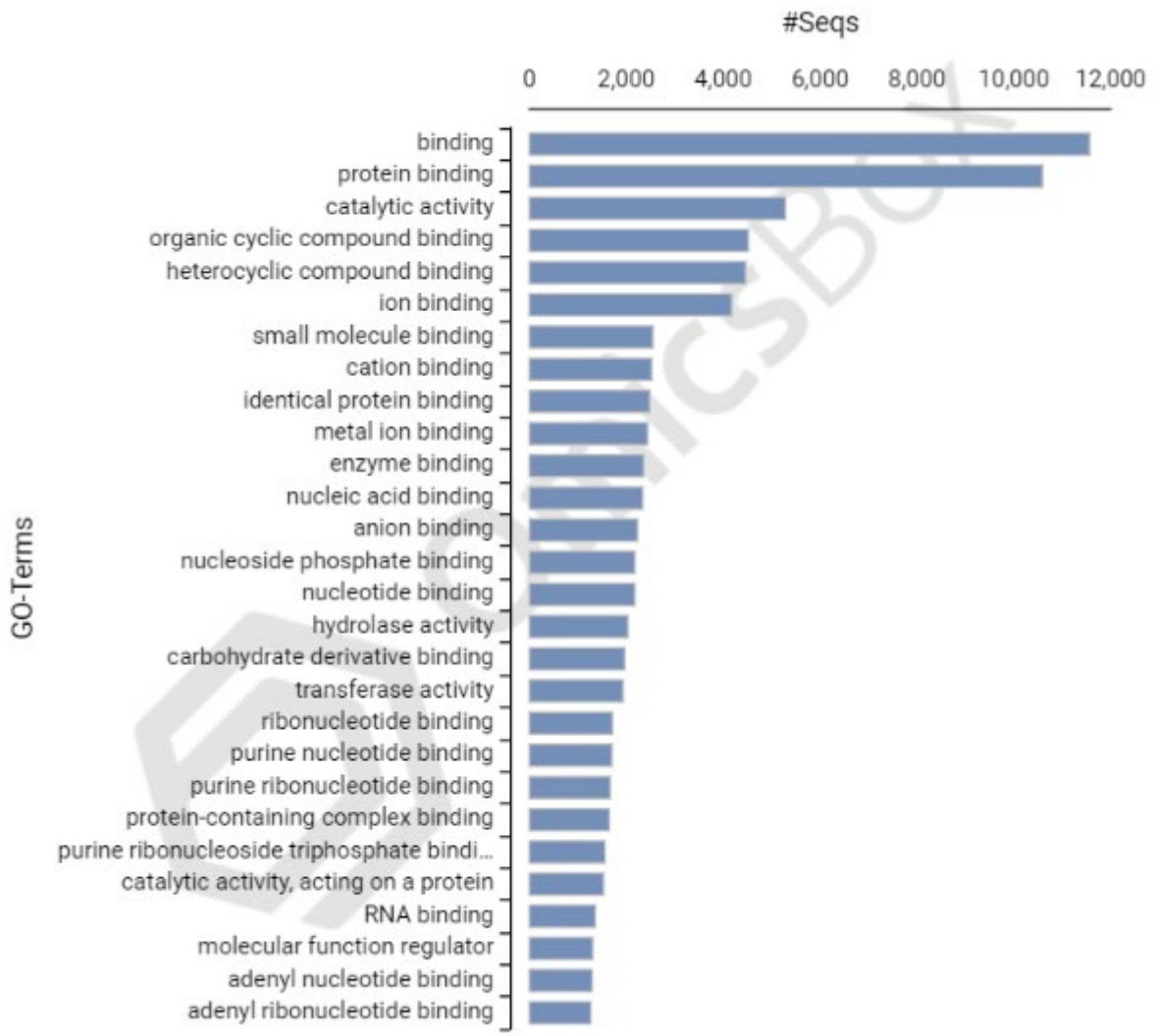


Figure 3-9: Poly(A) selected mRNA, isolated from kidney tissue from one *Tiliqua adelaidensis* skink ID: G6, a representative individual for Pacific Biosciences Iso-Seq long-read transcript sequencing and annotation to generate full-length reference transcripts. Gene ontology conducted using B2GO. Top gene ontology terms by number of annotated sequences for all GO levels in the molecular function category for predicted *Tiliqua adelaidensis* transcripts in open reading frame.

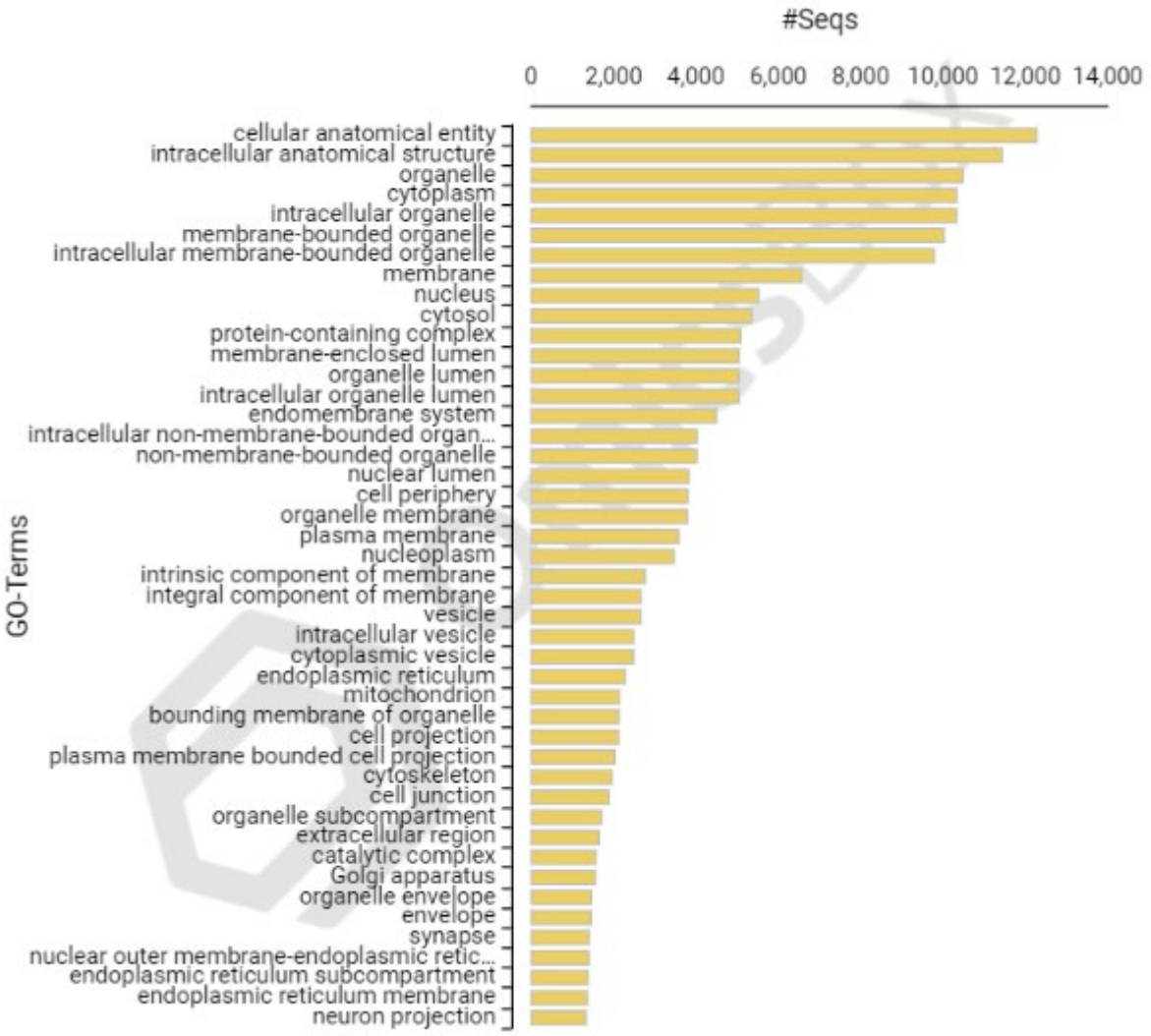


Figure 3-10: Poly(A) selected mRNA, isolated from kidney tissue from one *Tiliqua adelaidensis* skink ID: G6, a representative individual for Pacific Biosciences Iso-Seq long-read transcript sequencing and annotation to generate full-length reference transcripts. Gene ontology conducted using B2GO. Top gene ontology terms by number of annotated sequences for all GO levels in the cellular component category for predicted *Tiliqua adelaidensis* transcripts in open reading frame.

3.4.3 Selection of Genes of Interest

An NCBI gene search using key terms to focus on genes annotated in reptile species and with functions relevant to kidney function or surviving dry periods (Appendix 3) returned an initial 40,995 gene results for 104 taxa. Removal of duplicates resulted in 8,002 unique genes from 102 taxa, and removal of unique NCBI locus identifiers for putative sequences that were not clearly annotated to specific gene names further narrowed this gene list to 993 unique genes (Figure 3-11). There were many gene duplicates for genes recorded in multiple taxa in the original search, with the retained 993 genes representing only a small proportion of the initial database result (Figure 3-11). Only 11 of the taxa represented in the NCBI search results were squamates (Table

3-4), none of which were from the family Scincidae. Sixty-four % of genes identified were represented only in birds (Figure 3-12).

Table 3-4: Squamate species identified in NCBI gene search for genes represented by reptile species (including birds) with roles in water homeostasis and suspected renal functions; scientific and common names. Eleven squamate species were identified. Six turtles, four crocodilians, and 77 avian species were also found (Appendix 3).

Squamate name	**Common name**
Anolis carolinensis	Carolina anole
Pogona vitticeps	Australian Central bearded dragon
Podarcis muralis	Common wall lizard
Lacerta agilis	Sand lizard
Gekko japonicus	Schlegel's Japanese gecko
Python bivittatus	Burmese python
Notechis scutatus	Tiger snake
Thamnophis elegans	Common garter snake
Thamnophis sirtalis	Common garter snake
Pseudonaja textilis	Eastern Brown snake
Protobothrops mucrosquamatus	Brown-spotted pit viper or Pointed-scaled pit viper

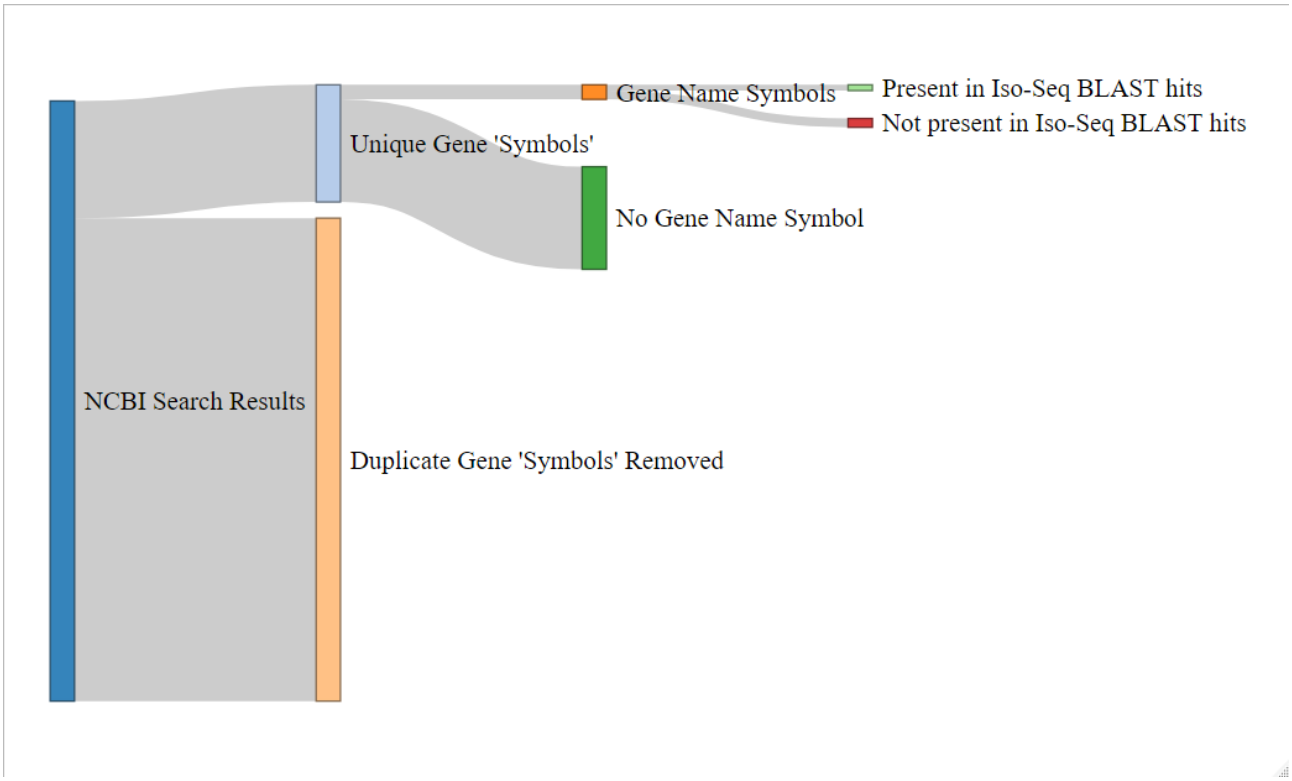


Figure 3-11: Sankey chart showing proportion of duplication (gene representation in numerous taxa) in the NCBI gene search. The search returned a total of 40995 sequences for all taxa included in the search. Nine hundred and ninety three unique gene symbols were retained as candidate genes of interest, of which 393 matched a gene identified through BLASTx hits in the final Iso-Seq transcript set.

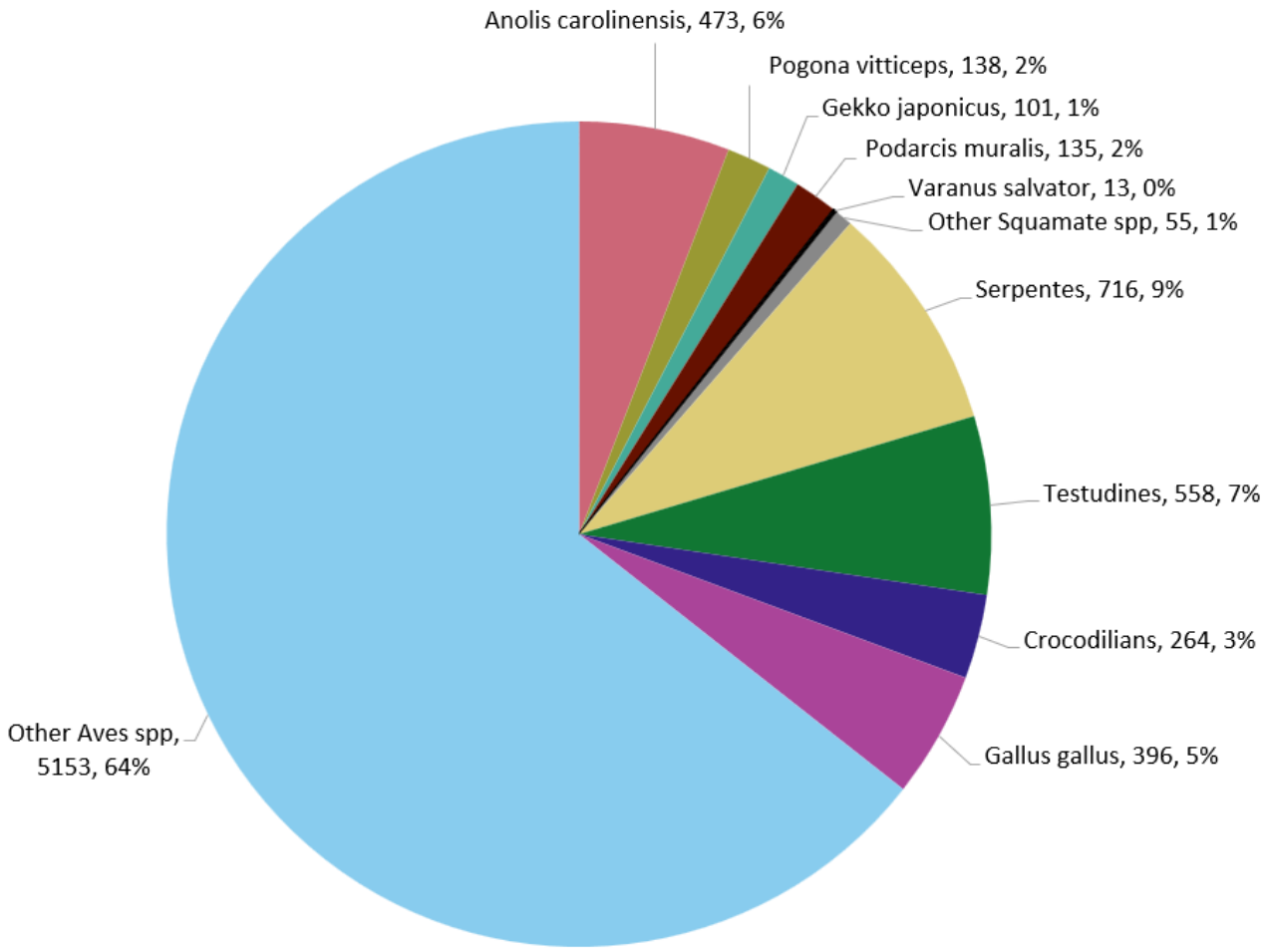


Figure 3-12: Pie chart showing number of unique gene symbols distributed taxa for unique gene symbols from NCBI search for genes of interest. Total genes =.8,002. Total Taxa = 102 (link to full list Appendix 3)

Of the initial 13,882 full-length non-redundant mRNA transcript isoforms identified in *T. adelaidensis*, 12,532 were mapped to a protein ID. Using BLASTx search outputs, 7,218 of 7,355 protein identifiers from UniProtKB AC/ID were successfully mapped to 6,813 unique 'gene name' IDs (gene symbols). Among these 6,813 unique genes identified, 393 represent genes also detected in the systematic NCBI gene search. Based on gene symbols, 955 *T. adelaidensis* transcript isoforms match one of these genes of interest in other reptiles (Appendix 3, Table 6-6). When these BLASTx hits were correlated with the putative genes predicted by CD-Hit clustering in Chapter 3, 8,906 of 9,813 genes were successfully annotated to a gene symbol.

Annotation was completed based on the translated proteins of putative coding regions of transcripts compared to the reviewed UniProt Swiss Prot database using BLASTx. This BLASTx annotation was conducted on the full list of *T. adelaidensis* transcript isoforms before clustering

with CD-Hit and the annotation information was kept for the subset-clustered transcript file. The full list of transcript coding sequences was assigned a BLASTx hit ~90% of the time with around 10% of all predicted peptide translations of transcripts not returning a BLASTx hit. A comparison of the annotated transcripts with genes published for other reptile species on the NCBI database was also conducted to identify orthologues in reptile species. Among transcripts that returned a gene name ID, 393 of the symbols were associated with genes in the genes of interest list. Some of the putative genes (8,906 gene transcript sequences with BLAST hits) predicted by CD-Hit returned BLASTx hits for the same gene symbols (6,813 unique gene symbols).

The full list including transcripts from the BLASTx hit and associations of gene symbols with the genes of interest list is uploaded to GitHub: rows are sorted by log fold change (ascending), and then by gene symbol favouring presence on the genes of interest list, followed by transcripts with only a BLASTx hit. Transcripts without expression data are transcript isoforms, which were not included in the clustered file of putative genes.

3.4.4 Transcriptome Comparison

3.4.4.a BUSCO

Transcript assemblies created with Trinity, which included samples G6 and G7 (both individual assemblies and the larger concatenated assembly), resulted in a much higher number of BUSCO orthologues being identified compared to the long-read transcripts (Table 3-2). However, these orthologues also showed a much higher level of duplication (Figure 3-13). Other transcript assemblies (G1, G2, G3, G4, G5 and G8) created from short-read data resulted in a larger discovery rate of single-copy, full orthologues than in the long-read assemblies and a much higher number of fragmented or partial matches (Figure 3-13).

The effect of clustering of the long-read data is evident at each stage of filtering in the BUSCO output (Figure 3-13) with the clustered list of reference transcripts containing the lowest number of duplicated BUSCOs. The full dataset at the first stage of Isoseq3 processing, which includes all full-length, high-quality, non-redundant isoforms (High-quality Iso-seq Transcripts), had higher orthologue discovery than other long-read sets and also the highest rate of duplication. The file that contained only the predicted coding sequence for these transcripts (Unique Isoform CDS) reduced duplicated sequences and had similar total BUSCOs aligned, but a higher number of fragmented orthologues, which were only partially identified. The full-length transcripts of coding

sequences clustered into putative genes using CD-Hit (Clustered Reference Transcripts) shows a much smaller number of duplicated BUSCOs, and fragmented BUSCO alignments are reduced again to the original levels of the full-transcript dataset.

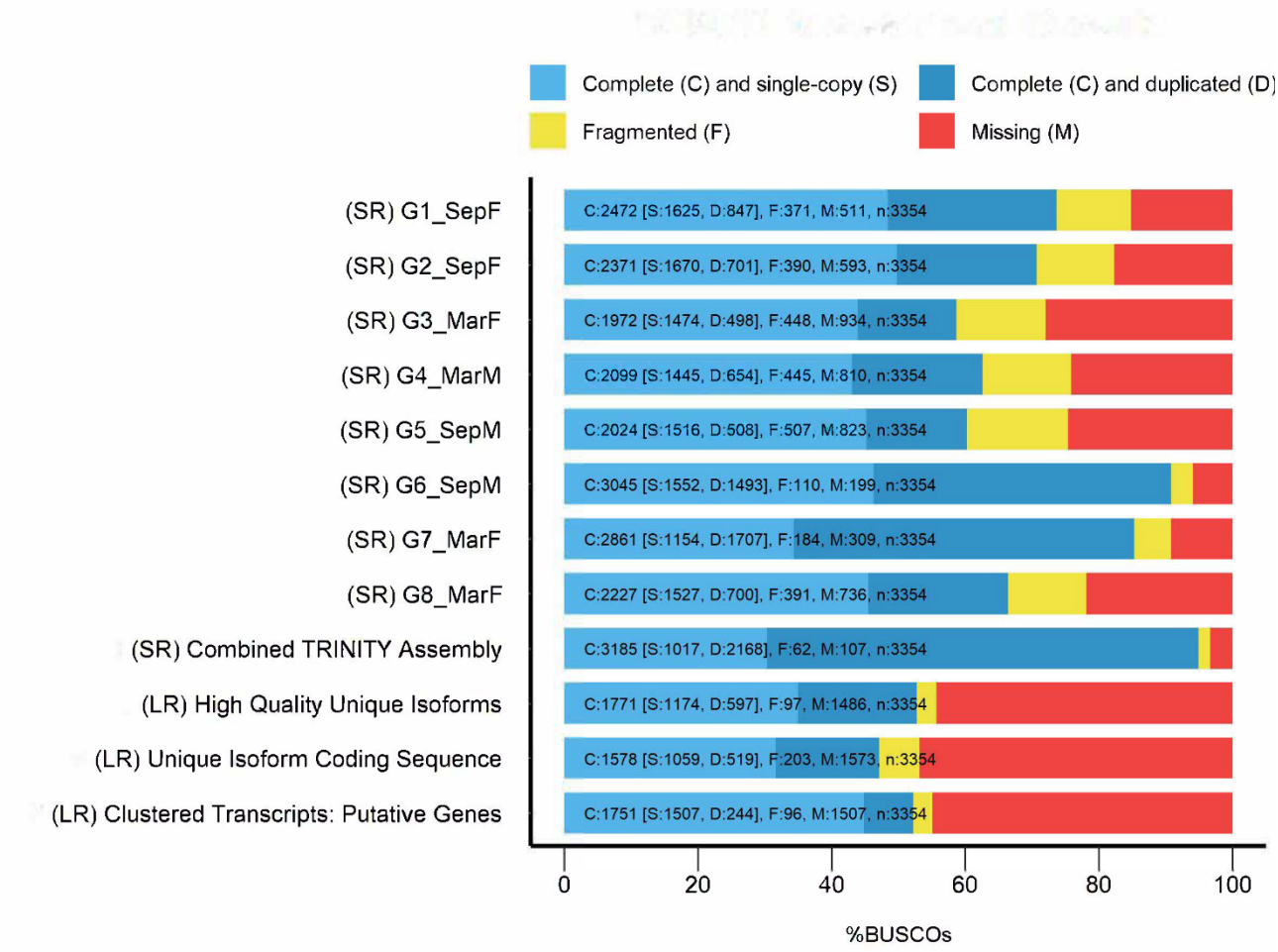


Figure 3-13: Benchmarking Universal Single-Copy Orthologs (BUSCO) as a measure of completeness for transcript assemblies of *Tiliqua adelaidensis* kidney tissue mRNA. Bar proportions are shown in % of BUSCO library, true values are provided in text. Transcripts compared against lineage dataset vertebrata_odb10 which includes 3354 BUSCOs for vertebrate species. Short-read *de novo* Trinity RNA transcript assemblies are indicated by (SR) and long-read data by (LR). BUSCO sequence scores C=completed, S=single, D=duplicated, F=fragmented, M=missing, n=total number of BUSCOs in database. Short-read individual samples are identified based on lizard ID, Season, and Sex: Sep=September (spring), Mar=March (autumn).

3.4.4.b Burrows–Wheeler Alignment

Alignment of the transcripts *de novo* assembled from short reads of all samples G1–G8 aligned using the BWA-mem tool resulted in a total of 185,795 transcripts mapping to an Iso-Seq long-read transcript, which is less than half (46.11%) of the total reference long-read transcripts in the provided reference file. The BWA-mem tool reported that 402,933 sequences had passed the

quality control check and 0 had failed, while 43,736 of these sequences were flagged as supplementary.

3.4.4.c BLASTn comparison

Prior to filtering for the best hit per query, 457, 579 BLASTn hits were returned. This was filtered down to 128,185 *de novo* transcripts, which returned a hit against a long-read transcript. Of the total 9,813 long-read transcripts in the reference target sequence database, 9,805 produced a hit. Some of these resulting groups of transcripts were visualised manually by selecting and extracting groups of query sequences that hit a common target sequence. An example is that target sequence 'PB.100.1-049515-p5:1-1563(+)-t/19243' returned hits for the query sequences 'Trinity_DN10_c0_g1_i10' and a number of other sequences with similar IDs but assigned a different isoform number by Trinity (_i## at the end of the header ID). The exception was query 'Trinity_DN6789_c0_g2_i1', which matched with the highest e-score and 100% identity but only for a length of 70 base pairs. When this sequence was run through a BLASTn search of the entire NCBI database via MEGA, the highest percent identity (pident) match was 11% of the query.

3.5 Discussion

The lack of a high-quality reference genome for many species has been a hinderance to conservation genomics research. However, both long- and short-read sequencing methodologies can be used to recover transcript sequences from cDNA sequencing libraries. While long reads resulted in fewer incomplete transcripts, assembled short reads achieved a higher level of overall transcriptome completeness (indicated by a higher number of BUSCO mappings) when both methods are assessed through BUSCO analysis. The use of long-read technologies to generate a reference has benefitted transcript profiling and gene expression studies in recent years (Ali, Thorgaard & Salem 2021; Minio et al. 2019).

Higher accuracy in long reads but higher transcript discovery in short reads was an expected outcome and has been shown in simulated and repeated analyses of data from other species such as the zebra finch (*Taeniopygia guttata*) (Vijay et al. 2013). The zebra finch has a high-quality genome available and is used as a model for studies of the avian brain, which allowed Vijay et al. (2013) to conduct assemblies with and without the aid of genomic alignments for comparison. The roughly eight million ZMWs that contain one read in the pygmy bluetongue long-read dataset (each containing a number of subreads used to generate the consensus sequence) would be less

likely to detect rare transcripts than would the nearly 300 million reads of the short-read data. This is unsurprising as library insert size and read length has been shown to impact assembly completeness in short-read data (Hara et al. 2015). This is a result of the relationship between sequencing volume and transcript abundance, and if more sequencing is completed (allowing more read coverage per sample), transcripts with lower abundance are more likely to be detected. The benefit of long-read data is shown with the extremely low number of fragmented BUSCO sequences indicating that there are fewer fragmented or incomplete transcripts. Sequencing full-length transcripts provides fast detection of alternatively spliced transcript isoforms and the identification of alternative splicing and gene locus (Liu et al. 2017).

The short-read assemblies, understandably, had a much higher number of BUSCO sequences as fragmented matches, showing that many of the assembled transcripts are incomplete; however, particularly with extremely large read depths, they produced greater discovery of expected BUSCO benchmarking sequences. Short-read transcript assembly using Trinity generated the most *de novo* assembled transcripts when all sample reads were combined. Read depth of the original sequencing and the number of assembled transcripts were associated with a greater percentage of BUSCOs aligned, indicating a more complete transcript assembly. The second sequencing batch when samples G6 and G7 were sequenced resulted in a much greater read depth per sample and assembled nearly six times as many transcripts per sample as the first short-read sequencing batch, which included samples G1, G2, G3, G4, G5 and G8. All *de novo* assemblies also had a much higher level of duplication than the long-read transcript data at each stage of analysis with nearly half of the BUSCO sequences aligning to duplicate transcripts. BUSCO completeness scores less than 80%, and even as low as 50–60% may also be explained because these data represent the sequencing of only one tissue. Similarly low completeness was shown in transcripts generated using similar methods for the ruby-throated hummingbird *Archilochus colubris* (Workman et al. 2018) generated using only liver tissue.

It is important to note that the BUSCO completeness scores are percentages of the BUSCO library provided and that the long-read data achieved a completeness score only slightly lower than samples G3, G4 and G5 from the short-read sets, despite providing more than 30 times fewer sequences to the program BUSCO. BUSCO sequences are selected because they are universal and should be detected if the correct database for the target species is being used (e.g., the vertebrata database), and a quality BUSCO score could be expected to be over 80% (Raghavan et al. 2022).

However, Schaarschmidt et al. (2020) showed a marked increase in BUSCO score from around 70% for partial assemblies to over 90% when a PacBio sequencing run included multiple pooled tissues. The fact that this study focused on kidney tissue alone may mean that some genes were simply not captured in that tissue. Rather than a consequence of stemming from a novel species, a low BUSCO score is more likely to be a result of lower gene capture or low sequence depth. Although Workman et al. (2018) did not compare short-read datasets, they also observed a diminishing return in the number of complete BUSCOs obtained with increasing sequencing depth. The short-read data show a higher rate of complete and duplicated sequences (G6 and G7), when read depth is greatly increased. *De novo* assemblies often have high levels of BUSCO duplication because of splice isoforms (Raghavan et al. 2022). It is possible the shorter read inserts result in programs using shorter k-mers and allow *de novo* assembly programs more flexibility in generating transcripts because less overlapping sequence is required (Vijay et al. 2013). It is also possible the deeper read depth of the short-read sequencing has identified more transcripts with lower expression levels (Chow et al. 2014). The large number of fragmented BUSCO matches in the short-read data indicates a large number of incomplete transcripts that have been assembled; with no reliable reference for comparison the accuracy of short-read assemblies cannot be determined (Ockendon et al. 2016). The results here support the current literature that short-read sequencing with more reads allocated per sample is desirable to recover the highest number of transcripts possible (Chow et al. 2014)—although with diminishing return (Todd, Black & Gemmell 2016)—and suggest that a minimum depth of 17–18 Gb of data per sample may be desirable for *T. adelaidensis* for future short-read sequencing to achieve a BUSCO completeness score above 80%.

The Pacific Biosciences Iso-Seq pipeline (Rhoads & Au 2015) returned transcript sequences that were subjected to further clustering and isoform identification. This additional collapsing, which had been completed, removed redundant isoforms and resulted in a much lower number of fragmented and duplicated BUSCOs compared with the original full long-read transcript library (Schaarschmidt et al. 2020). Short-read assemblies cannot always be accurately assigned to isoforms and may result in multi mapping, while long-read data allow isoforms to be identified and sequenced whole (Ali, Thorgaard & Salem 2021). The Pacific Biosciences Iso-Seq technology base substitution error rate is <1% (Amarasinghe et al. 2020). The most common error in Iso-seq sequencing is indels (random insertions or deletions of bases in the sequence) introduced into reads, which affect accuracy (Carneiro et al. 2012). However, this is corrected through the collapse of multiple subreads present in the circular polymerase read into one representative consensus

sequence per ZMW. The conformation of full-length isoforms is overall more accurate than predicted short-read *de novo* assembly transcripts and is continually improving with new technology releases (Amarasinghe et al. 2020).

The final CD-Hit clustering step used in this chapter was implemented so that the resulting reference transcript file consists of putative genes identified by clustering proteins from predicted open read frames. In species where direct genome mapping is not possible and transcript annotation may be incomplete, transcript analysis (and gene expression studies) is heavily reliant on the reference file provided (Grabherr et al. 2011; Holzer & Marz 2019). Gene expression counts may not be distinguishable from transcript isoform counts and—if a file with many isoforms with similar regions is supplied as a reference—can result in counts mapping to multiple places. This can cause anomalies in gene expression analysis where certain genes or families with multiple isoforms may show different expression levels depending on how many representative transcripts are identified (Deschamps-Francoeur, Simoneau & Scott 2020). Greater accuracy in the assembly is provided through long-read sequencing and the ability to identify isoforms. Pacific Biosciences has improved assemblies in a number of species including the rainbow trout (*Oncorhynchus mykiss*; Ali, Thorgaard & Salem 2021), and facilitated the identification of alternatively spliced transcript isoforms in the venom proteins of the Habu snake (*Protobothrops flavoviridis*; Ogawa et al. 2019). Long-read technologies may also allow older data to be revisited and corrected using newer assembly or alignment methods, particularly if the focus is on more polymorphic genes. The addition of comprehensive transcript data is valuable even in studies where whole genomes are being sequenced (Beiki et al. 2019; Eckalbar et al. 2013), and complements annotation of genes and functional or comparative studies.

Clustering of long-read transcripts as discussed above was completed in this study to reduce the impact of genes with varying number of isoforms on gene expression counts or potential multimapping (Soneson et al. 2016). The strict cut-off for the clustering step was chosen for this reason; most variation in these data from one individual may be attributed to sequencing errors, truncation or allelic variation. On inspection, most of the variation in clustered transcripts is in the untranslated regions at the 5′ and 3′ ends. This was expected, as these sequences were clustered based on protein coding sequence of the predicted open read frame. The goal of this clustering was to reduce the reference transcripts to the gene level, and depending on sequence similarity,

heterozygous alleles of the same gene may also have been collapsed. However, higher levels of variation in some genes may mean they are still represented by multiple canonical transcripts.

In the comparison of the combined assembled short reads to the long-read full-length transcripts, the low alignment score using the BWA tool is unsurprising. Successful alignment of fewer than half of the sequences is consistent with the BUSCO score reported for these long-read sequences. This may be a result of the data being extracted from a single tissue (kidney) for each individual and may also be a result of sequencing depth. The number of 'supplementary sequences' identified by the BWA statistics flag indicate sequences that do not represent a linear alignment (Li et al. 2009; Peter et al. 2015; Danecek et al. 2021) and may be referred to as chimeric reads. These sequences may indicate mis-assemblies of short-read sequences, or may simply indicate similar alternative splicing isoforms to the reference transcript to which they aligned. Based on further exploration of the combined *de novo* assembled transcripts conducted using a BLASTn search and manual visualisation, another mis-assembly may have been identified. The long-read transcript 'PB.100.1-049515-p5:1-1563(+)-t/19243' received hits from sequences assigned similar gene and isoform IDs by Trinity 'Trinity_DN10_c0_g1_i#', but also from the sequence 'Trinity_DN6789_c0_g2_i1' which has been assigned different identifiers by TRINITY. The best hit for this query sequence was a very small portion of the query, which was also true of a BLASTn search of the NCBI nucleotide database. This extremely small region of the sequence providing a hit to reference sequences indicates it may be an incomplete, fragmented assembly or possibly a mis-assembly. The long-read data represent a single individual, while the assembled short-read data represent eight individuals. There is a possibility of 16 times as many variations of alleles in the combined short-read assembly and it may be that not all individual variation will result in a match to the reference transcripts, particularly for highly variable genes. The inability to tease out which *de novo* transcripts are accurate, which are mis-assemblies, and which do not provide a strong database hit simply because of the lack of a closely related orthologue being available is why long-read data of full-length transcripts are preferable when transcript accuracy is the goal (Raghavan et al. 2022).

In a similar way to how single cell RNA-seq (scRNA-seq) complications arise from individual variation (Butler et al. 2018), individual variation and SNPs will also exist in the data presented in this thesis. In this case the long-read sequences were clustered based on protein coding sequence and even in the case of heterozygous alleles, may have become clustered together, resulting in loss

of some diversity from the dataset. However, the short-read assembly has not been filtered or clustered and includes information from six individuals. Some of the duplication seen in the BUSCO analysis of the combined Trinity assembly may be an indication of allelic variants of similar genes represented in the six individuals. Where alleles are very similar or only differ by a single SNP, it is also possible that Trinity may create mis-assemblies and combine sequence output from multiple individuals into one sequence. Analysis of SNP data can be performed *de novo* (Valenzuela-Miranda et al. 2014) using programs such as Soapsnp (He et al. 2017), which can also be used in conjunction with a genome (Li et al. 2009). A combination of Trinity, KisSplice and transdecoder (Lopez-Maestre et al. 2016), and many programs such as SUPPA2 can compare Iso-Seq assemblies regardless of whether they were initially assembled using a genome reference (Wen et al, 2023). However, more accurate results would be gained by using these programs with a reference genome, such as by aligning to a reference genome with BWA and running TopHat and R DEXSeq to find SNPs (Devisetty et al. 2014; Zhao 2021). Insertions and deletions may also be assessed using programs such as IndelMINER (Ratan et al 2015). As the long-read transcripts in this study are from a single individual and have been clustered for the purposes of gene expression analysis, the reference list is unsuitable as a reference for SNP detection without potentially losing a large amount of data. Comparing the short-read assemblies to each other or using another genomic reference would be most effective.

The use of long-read sequencing to generate a more accurate list of canonical transcripts from a representative kidney sample provides a reference that was previously unavailable for further analysis of *T. adelaidensis,* an endangered skink with no reference genome assembly to directly map reads or conduct a guided assembly of transcripts. The *de novo* assemblies here have higher BUSCO scores than the long read transcriptomes, while the long-read transcripts have a much lower incidence of fragmented or incomplete matches than the *de novo* assemblies. As each of these methods has its own strengths, the use of both can supplement the other. Although deeper sequencing is likely to detect genes with lower expression levels and may be achieved faster and at lower cost using short-read sequencing, using long-read sequencing of transcripts to generate the reference sequences for transcript analysis reduces the risk of mis-assembly, and full-length transcripts allow for the identification of isoforms and splice regions. In this project this represented a trade-off between completeness (or detection of rare transcripts) and accuracy; however, Pacific Biosciences's new machine Revio, announced on 25 October 2022, is designed to generate 25 million ZMW subreads over four SMRT cells in parallel, greatly increasing the depth

available for long-read sequencing (PacBio 2022) and likely making this the more desired option if transcript assembly is the goal.

Annotation of transcripts is necessary for effective analysis and interpretation of results. The purpose of the analysis in this chapter was to identify and annotate genes expressed in kidney tissue of *T. adelaidensis* using long-read sequencing, and create a reference list of these genes with a focus on genes established through literature and database searching (Appendix 3: 'genes of interest') that have been found to be associated with ion transport, water retention or excretion, and renal functions, as well as stress responses with a focus on temperature or homeostasis. I identified genes relevant to kidney function such as aquaporins, solute carriers, ion transporters and HSPs in *T. adelaidensis* that may play a role in the species surviving hot or dry periods. The gene symbols associated with genes in the genes of interest list indicate they may be of particular interest for further analysis regarding their function. Some of the putative genes clustered by CD-Hit, which returned BLASTx hits for the same gene symbols, are partially a result of the strict cut-off used in the initial clustering step, but may also indicate that loci have a high level of variation in *T. adelaidensis* for these genes resulting in a number of variants. This may also be a case of the best representative BLASTx hit mapping with higher or lower percent identity matches for multiple genes in a gene family when a more accurate gene orthologue was unavailable in the SwissProt database.

Despite finding genes which are potentially orthologous to *T. adelaidensis* genes, relevant to kidney function in a variety of species, there were many instances where specific genes for skinks were not identified in the database. This is consistent with the paucity of genomes available for species in the family Scincidae. However, numerous other squamates were represented in the genes of interest list (Appendix 3). The number was greatly reduced when search term parameters were set to extract specific genes that may be expressed in the kidney or during times of environmental stress related to water availability and homeostasis during hot, dry periods. This may have limited the number of squamates represented in the genes returned. This indicates that even squamate reptiles with genome data are not as comprehensively annotated as other species that are routinely used in genomic research (model species). This lack of comprehensive annotation is exemplified by the large number of genes represented by *Gallus gallus* in the final list (Appendix 3).

Aquaporins are membrane proteins that serve as water and small solute channels and are present in the kidney tubules. I identified four aquaporins in *T. adelaidensis* (*aqp1, aqp2, aqp9* and *aqp11*).

In comparison, 14 aquaporin genes with 16 paralogues have been identified in *A. carolinensis* (Finn et al. 2014). It is likely that because only kidney tissue was sequenced, aquaporin genes expressed in other tissues such as intestines (Wang et al. 2015) or liver (Gregoire et al. 2015) were not detected. All aquaporins identified in *T. adelaidensis* matched aquaporins returned in the genes of interest list and these have all been shown to be expressed in kidneys (Su et al. 2019). Aquaporin 9 is an aquaglyceroporin that plays a role in the transport of water, urea and glycerol, and aquaporin 11 has been shown to be expressed in the proximal tubule of humans (Su et al. 2019). Two transcripts were annotated as aquaporin 1, which is expressed in the distal kidney tubules and six transcripts were annotated as aquaporin 2, which is expressed in the collecting duct. Both play a crucial role in the reabsorption of water (Golosova, Shakhmatova & Natochin 2019). It is possible these identified aquaporin genes play a role in the resilience of *T. adelaidensis* to dry conditions.

Six HSP genes from the genes of interest list were also identified in *T. adelaidensis*, with 12 HSP genes identified in other squamate species. However, a much larger number of transcripts returned a BLASTx hit to a HSP gene including those not on the genes of interest list, suggesting the transcripts assembled and annotated here represent a broader variety of HSP genes than are published for squamates—'heat shock' was specifically included in the NCBI gene search term—or that *T. adelaidensis* potentially has a larger number of HSP paralogues. HSPs are a large gene family and have primarily been studied in squamates with reference to embryological development or temperature-dependent sex determination. Heat shock protein 70 (*hsp70*) and *hsp90*, both identified here in *T. adelaidensis*, protect cells from proteins damaged by heat or stress, but can also influence the impact of other mutations and affect phenotypes (Chen, Feder & Kang 2018). Chen, Feder and Kang (2018) also postulate that high base levels of *hsp70* expression may indicate heat and desiccation stress, and that HSP genes should be considered when evaluating the impacts of changing climatic conditions.

Solute carriers (*slc*), another extremely large gene family involved in solute and water transport (Schaller & Lauschke 2019) were represented by hundreds of genes in the genes of interest list. The number of *T. adelaidensis* transcripts annotated to a *slc* gene was four times that of the unique genes identified, and not all of the genes identified in *T. adelaidensis* were present in the genes of interest list, indicating they have not been annotated in other reptiles and birds in the NCBI database. In such a large and diverse gene family, there is potential for partial hits among more similar genes, but in this case the genes of interest list has failed to include an orthologue for

another bird or reptile in the database, for at least four *slc* genes identified in *T. adelaidensis*, where one does exist in a less related species present in the UniProt SwissProt database. This may be the first example of these *slc* genes sequenced in a squamate reptile. Solute carriers are expressed throughout the body, but play a distinct role in the kidney nephron, and many genes have housekeeping roles as well as their involvement in solute transport (Lewis et al. 2021). A large proportion of *slc* genes is distributed in the thin limb of the loop of Henle in mice (Lewis et al. 2021), the absence of which in reptiles (Holz 2020) would mean that most other *slc* genes would be expected in the proximal tubule followed by the collecting duct. This absence of the loop of Henle nephron structures in reptiles may also be a factor in the lower number of *slc* genes annotated from transcripts. Gene families such as the solute carriers are responsible for a large number of the duplicate BLASTx matches to the genes of interest list in Appendix 3, due to large variation in the gene family.

Because of the different BLASTx searches (to generate output format 5 and 6) and more max hits per sequence being allowed in the input, a larger number of hits was included in the GO analysis. GO analysis helps identify and summarise gene functions in a large dataset and can also allow a comparison of pathways or pathway enrichment between related organisms (Workman et al. 2018), and identify sexual variation (Valenzuela-Miranda et a;. 2014) or tissue-specific expression in a single organism (Macrander, Broe & Daly 2016).

Many GO functions identified through Blast2GO are unsurprising for kidney tissue samples, including many metabolic process terms such as 'cellular process', 'regulation of cellular process', various metabolic processes and metabolic regulation, 'biological regulation', 'response to stimulus', 'gene expression' and 'transport'. Many of the terms most commonly found in the biological process category are related to cellular function and metabolic processes, while many of the highly represented GO terms in the molecular functions category play a role in regulation of other processes or are directly involved with DNA and RNA binding. The top two GO categories with more than double the number of sequences than any other GO category are 'binding' and 'protein binding' activity. These binding factors with GO terms 'adenyl nucleotide binding' and 'adenyl ribonucleotide binding', as well as 'purine binding', 'nucleotide binding', 'adenyl ribonucleotide binding' and 'RNA binding and nucleotide binding' may play a number of regulatory roles in DNA or RNA maintenance and transcription, as well as regulation (Hudson & Ortlund 2014). The high level of 'protein binding' activity is also essential in the kidney beyond

these functions. Larger molecules such as plasma proteins are too large for passive diffusion and must be actively transported to be reabsorbed. Because reptiles lack the counter-current multiplication of osmolarity gradient that the loop of Henle propagates in mammals, protein binding is even more important for maintaining lower plasma levels of secreted solutes (Meyer & Hostetter 2021). A number of the top represented GO terms relate to membrane structures, which are a core aspect of kidney function because solutes and water are reabsorbed or excreted via osmolarity gradients or active transporters.

While potentially resulting from environmental factors differing between experimental groups—including rainfall, seasonal temperatures, food availability and the active season of this species—some GO terms that relate to stress responses may be present as a result of handling of lizards during sampling, rather than the sampling season period. 'Response to stress' and 'response to chemical' may be indicative of acute reactions to being handled and exposed to euthanasia drugs during this study, which are well known to affect short-term biochemical measurements in many species and likely to also affect gene expression in the same way (Young et al. 2019). Stress response genes are notoriously difficult to isolate given these effects (Aballai et al. 2017).

Annotated references are a vital tool in genomic studies and having full-length transcripts for genes relevant to experimental goals solves many bioinformatic issues that arise in the study of non-model organisms, including gene annotation and splice isoform identification (Amarasinghe et al. 2020). Guided assembly methods (Ungaro et al. 2017), or simply the ability to align or cross-reference gene identification in RNA-seq studies, greatly improves accuracy of results. Identification of isoforms also allows informed interpretation where gene expression data may multi-map or be assigned to one gene by gene count programs.

RNA sequencing and transcript assembly can benefit genome assemblies by allowing the identification of large or complex gene families using transcript analysis. Livernois et al. (2016) used 100 bp Illumina HiSeq reads to identify additional interleukin immune genes that had not previously been annotated in the Australian bearded dragon genome. Poly(A)-selected long-read Iso-seq datasets and short-read Illumina RNA sequencing have also been combined in other studies such as Beiki et al. (2019), where additional data were used to supplement domestic pig transcript annotations and expression data. These methods often involve a high degree of manual curation and validation of *de novo* assembled sequences, and the accuracy of long-read

sequencing can increase confidence of transcript assembly, enhancing understanding of gene family diversity in organisms with less complete genomes (Larsen, Heilman & Yoder 2014).

While thresholds of quality control were applied to this dataset during sample extraction, purification and sequencing, as well as throughout the bioinformatic analysis, this annotation does rely on a single individual representative, and kidney tissue only. This is due in part to the available resources and limited sampling capacity when working on an endangered species, which is not an uncommon restriction. For example, Sun et al. (2018) sampled only two Chinese alligators (*Alligator sinensis*) because of the species' threatened status. Error rates of Iso-Seq have declined since the technology's inception in 2011 (Amarasinghe et al. 2020) but comparison between multiple samples and sequencing methods improves the reliability of results even in model organisms (Tardaguila et al. 2018). Broader gene annotation may be achieved by sequencing a greater variety of tissues.

The genes identified in this chapter provide reference transcripts for *T. adelaidensis*. These may be used as the reference index for gene expression analysis in the future, and potentially as a guide for primer development to target these genes. Further analysis of isoforms or comparative genomics between the pygmy bluetongue skink and other reptiles could be completed in future. Using the genes identified here, gene family diversity and isoforms could be analysed from a phylogenetic perspective in future work. The generated transcripts are assumed to be full length and allow for the identification of gene isoforms and coding regions, and aid in overall gene annotation. Transcriptomic data have been used in these ways to supplement whole-genome assemblies by providing validation for prediction of coding genes when constructing genomic assemblies such as in the tegu lizard (*Salvator merianae*; Roscito et al. 2018). Even in highly studied species, transcriptomic data have the potential to identify coding regions or entirely new transcripts, as was the case with the bovine reference genome assembly (Weikard et al. 2013). The references generated here for the pygmy bluetongue will also allow us to compare genes with existing databases of genes and proteins in other related species and provide a framework for future studies of these genes in the pygmy bluetongue. If primers were developed for specific regions of interest, identified genes could further studied using real-time PCR methods (Guttman et al. 2010). Primers developed using *T. rugosa* transcriptome data to amplify genes of the major histocompatibility complex successfully amplify in this species (Ansari et al. 2014). Therefore, there

is potential for transcripts identified through this study to reciprocally assist as references in any number of studies on genes identified for skink species in the Tiliquini group (Shea 2021).

3.6 References

Aballai, V, Aedo, JE, Maldonado, J, Bastias-Molina, M, Silva, H, Meneses, C, Boltana, S, Reyes, A, Molina, A & Valdes, JA 2017,''RNA-seq analysis of the head-kidney transcriptome response to handling-stress in the red tusk-eel (*Genypterus chilensis*'', *Comparative Biochemistry and Physiology D-Genomics & Proteomics*, vol. 24, pp. 111-7.

Agre, P, King, LS, Yasui, M, Guggino, WB, Ottersen, OP, Fujiyoshi, Y, Engel, A & Nielsen, S 2002,''Aquaporin water channel—-from atomic structure to clinical medicin'', *J Physiol*, vol. 542, no. Pt 1, pp. 3-16.

Alfoldi, J, Di Palma, F, Grabherr, M, Williams, C, Kong, L, Mauceli, E, Russell, P, Lowe, CB, Glor, RE, Jaffe, JD, Ray, DA, Boissinot, S, Shedlock, AM, Botka, C, Castoe, TA, Colbourne, JK, Fujita, MK, Moreno, RG, ten Hallers, BF, Haussler, D, Heger, A, Heiman, D, Janes, DE, Johnson, J, de Jong, PJ, Koriabine, MY, Lara, M, Novick, PA, Organ, CL, Peach, SE, Poe, S, Pollock, DD, de Queiroz, K, Sanger, T, Searle, S, Smith, JD, Smith, Z, Swofford, R, Turner-Maier, J, Wade, J, Young, S, Zadissa, A, Edwards, SV, Glenn, TC, Schneider, CJ, Losos, JB, Lander, ES, Breen, M, Ponting, CP & Lindblad-Toh, K 2011, 'The genome of the green anole lizard and a comparative analysis with birds and mammals', *Nature*, vol. 477, no. 7366, pp. 587-91.

Ali, A, Thorgaard, GHH & Salem, M 2021, 'PacBio Iso-Seq Improves the Rainbow Trout Genome Annotation and Identifies Alternative Splicing Associated With Economically Important Phenotypes', *Frontiers in Genetics*, vol. 12, pp. 683408-.

Amarasinghe, SL, Su, S, Dong, X, Zappia, L, Ritchie, ME & Gouil, Q 2020, 'Opportunities and challenges in long-read sequencing data analysis', *Genome Biology*, vol. 21, no. 1, p. 30.

Anaconda Software Distribution 2017, 'Computer software. Vers. 4.6.14. Anaconda', https://anaconda.com

Andrade, P, Pinho, C, Pérez i de Lanuza, G, Afonso, S, Brejcha, J, Rubin, C-J, Wallerman, O, Pereira, P, Sabatino, SJ, Bellati, A, Pellitteri-Rosa, D, Bosakova, Z, Bunikis, I, Carretero, MA, Feiner, N, Marsik, P, Paupério, F, Salvi, D, Soler, L, While, GM, Uller, T, Font, E, Andersson, L & Carneiro, M 2019,''Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizar'', *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, p. 5633.

Andrews, S. 2018, 'FastQC: a quality control tool for high throughput sequence data' http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ansari, TH, Bertozzi, T, Hacking, J, Cooper, SJB & Gardner, MG 2014,''Random non-coding fragments of lizard DNA: anonymous nuclear loci for the Australian skink, *Tiliqua rugosa*, and their utility in other Egernia-group specie'', *Australian Journal of Zoology*, vol. 62, no. 6.

Babonis, LS, Miller, SN & Evans, DH 2011,''Renal responses to salinity change in snakes with and without salt gland'', *Journal of Experimental Biology*, vol. 214, no. 13, pp. 2140-56.

Beiki, H, Liu, H, Huang, J, Manchanda, N, Nonneman, D, Smith, TPL, Reecy, JM & Tuggle, CK 2019, 'Improved annotation of the domestic pig genome through integration of Iso-Seq and RNA-seq data', *Bmc Genomics*, vol. 20, no. 1, p. 344.

Bethesda(MD) 2006, *Entrez Help [Internet]*, National Center for Biotechnology Information, US, <https://www.ncbi.nlm.nih.gov/books/NBK3837/>.

Bureau of Meteorology 2022, 'Climate Data Online, Weather Station Directory', Australian Government, Bureau of Meteorology, http://www.bom.gov.au/climate/data/stations/ [Accessed 21 Mar. 2022].

Butler, A, Hoffman, P, Smibert, P, Papalexi, E & Satija, R 2018, 'Integrating single-cell transcriptomic data across different conditions, technologies, and species', Nature Biotechnology, vol. 36, no. 5, p. 411.

Carneiro, MO, Russ, C, Ross, MG, Gabriel, SB, Nusbaum, C & DePristo, MA 2012, 'Pacific biosciences sequencing technology for genotyping and variation discovery in human data', *Bmc Genomics*, vol. 13, no. 1, p. 375.

Chen, B, Feder, ME & Kang, L 2018,''Evolution of heat-shock protein expression underlying adaptive responses to environmental stres'', *Mol Ecol*, vol. 27, no. 15, pp. 3040-54.

Chow, KS, Ghazali, AK, Hoh, CC & Mohd-Zainuddin, Z 2014, 'RNA sequencing read depth requirement for optimal transcriptome coverage in *Hevea brasiliensis*', *BMC Research Notes*, vol. 7, no. 1.

Chiari, Y, Cahais, V, Galtier, N & Delsuc, F 2012,''Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria'', *Bmc Biology*, vol. 10, no. 1, p. 65.

Conesa, A, Götz, S, García-Gómez, JM, Terol, J, Talón, M & Robles, M 2005,''Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics researc'', *Bioinformatics*, vol. 21, no. 18, pp. 3674-6.

Danecek, P, Bonfield, JK, Liddle, J, Marshall, J, Ohan, V, Pollard, MO, Whitwham, A, Keane, T, McCarthy, SA, Davies, RM & Li, H 2021, 'Twelve years of SAMtools and BCFtools', Gigascience, vol. 10, no. 2.

Davidson, NM & Oshlack, A 2014, 'Corset: enabling differential gene expression analysis for de novo assembled transcriptomes', *Genome Biol*, vol. 15, no. 7, p. 410.

Davidson, N. M., Oshlack, A. 2018, 'Necklace: combining reference and assembled transcriptomes for more comprehensive RNA-Seq analysis', *Gigascience*, vol. 7, no. 5, p. 6.

Deschamps-Francoeur, G, Simoneau, J & Scott, MS 2020, 'Handling multi-mapped reads in RNA-seq', *Computational and Structural Biotechnology Journal*, vol. 18, pp. 1569-76.

Devisetty, UK, Covington, MF, Tat, AV, Lekkala, S & Maloof, JN 2014, 'Polymorphism Identification and Improved Genome Annotation of Brassica rapa Through Deep RNA Sequencing', G3-Genes Genomes Genetics, vol. 4, no. 11, pp. 2065-78.

Eckalbar, WL, Hutchins, ED, Markov, GJ, Allen, AN, Corneveaux, JJ, Lindblad-Toh, K, Di Palma, F, Alfoldi, J, Huentelman, MJ & Kusumi, K 2013, 'Genome reannotation of the lizard *Anolis carolinensis* based on 14 adult and embryonic deep transcriptomes', *Bmc Genomics*, vol. 14, p. 10.

Fenner, A., Hutchinson, M., McDonald, P. & Robertson, P. 2018, '*Tiliqua adelaidensis*', *The IUCN Red List of Threatened Species 2018: e.T21902A101743579,* https://dx.doi.org/10.2305/IUCN.UK.2018-1.RLTS.T21902A101743579.en.

Finn, RN, Chauvigne, F, Hlidberg, JB, Cutler, CP & Cerda, J 2014,''The lineage-specific evolution of aquaporin gene clusters facilitated tetrapod terrestrial adaptatio'', *Plos One*, vol. 9, no. 11, p. e113686.

Flinders U 2021, 'Deep Thought (HPC)', Flinders University.

Fu, SH, Chang, PL, Friesen, ML, Teakle, NL, Tarone, AM & Sze, SH 2019, 'Identifying similar transcripts in a related organism from de Bruijn graphs of RNA-Seq data, with applications to the study of salt and waterlogging tolerance in *Melilotus*', *Bmc Genomics*, vol. 20, p. 14.

Fraser, BA, Weadick, CJ, Janowitz, I, Rodd, FH & Hughes, KA 2011,''Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptom'', *Bmc Genomics*, vol. 12, p. 202.

Fruzangohar, M, Ebrahimie, E, Ogunniyi, AD, Mahdi, LK, Paton, JC & Adelson, DL 2013,''Comparative GO: a web application for comparative gene ontology and gene ontology-based gene selection in bacteri'', *Plos One*, vol. 8, no. 3, p. e58759.

Georges, A, Li, QY, Lian, JM, O'Meally, D, Deakin, J, Wang, ZJ, Zhang, P, Fujita, M, Patel, HR, Holleley, CE, Zhou, Y, Zhang, XW, Matsubara, K, Waters, P, Graves, JAM, Sarre, SD & Zhang, GJ 2015, 'High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*', *Gigascience*, vol. 4, p. 10.

Golosova, DV, Shakhmatova, EI & Natochin, YV 2019,''Differences between Arginine-Vasotocin and Arginine-Vasopressin Effects on the Rat Kidney in Evolution of Osmoregulation in Vertebrate'', *Journal of Evolutionary Biochemistry and Physiology*, vol. 55, no. 1, pp. 71-9.

Gotz, S, Garcia-Gomez, JM, Terol, J, Williams, TD, Nagaraj, SH, Nueda, MJ, Robles, M, Talon, M, Dopazo, J & Conesa, A 2008,''High-throughput functional annotation and data mining with the Blast2GO suit'', *Nucleic Acids Res*, vol. 36, no. 10, pp. 3420-35.

Grabherr, MG, Haas, BJ, Yassour, M, Levin, JZ, Thompson, DA, Amit, I, Adiconis, X, Fan, L, Raychowdhury, R, Zeng, Q, Chen, Z, Mauceli, E, Hacohen, N, Gnirke, A, Rhind, N, di Palma, F, Birren, BW, Nusbaum, C, Lindblad-Toh, K, Friedman, N & Regev, A 2011, 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nat Biotechnol*, vol. 29, no. 7, pp. 644-52.

Gregoire, F, Lucidi, V, Zerrad-Saadi, A, Virreira, M, Bolaky, N, Delforge, V, Lemmers, A, Donckier, V, Deviere, J, Demetter, P, Perret, J & Delporte, C 2015,''Analysis of aquaporin expression in liver with a focus on hepatocyte'', *Histochem Cell Biol*, vol. 144, no. 4, pp. 347-63.

Gust, KA, Chaitankar, V, Ghosh, P, Wilbanks, MS, Chen, XF, Barker, ND, Pham, D, Scanlan, LD, Rawat, A, Talent, LG, Quinn, MJ, Vulpe, CD, Elasri, MO, Johnson, MS, Perkins, EJ & McFarland, CA 2018,''Multiple environmental stressors induce complex transcriptomic responses indicative of phenotypic outcomes in Western fence lizar'', *Bmc Genomics*, vol. 19, p. 20.

Haas, BJ, Papanicolaou, A, Yassour, M, Grabherr, M, Blood, PD, Bowden, J, Couger, MB, Eccles, D, Li, B, Lieber, M, MacManes, MD, Ott, M, Orvis, J, Pochet, N, Strozzi, F, Weeks, N, Westerman, R, William, T, Dewey, CN, Henschel, R, Leduc, RD, Friedman, N & Regev, A 2013, 'De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis', *Nature Protocols*, vol. 8, no. 8, pp. 1494-512.

Hara, Y, Tatsumi, K, Yoshida, M, Kajikawa, E, Kiyonari, H & Kuraku, S 2015, 'Optimizing and benchmarking de novo transcriptome sequencing: From library preparation to assembly evaluation', *Bmc Genomics*, vol. 16, no. 1.

He, B, Li, YG, Ni, ZX & Xui, LA 2017, 'Transcriptome sequencing and SNP detection in Phoebe chekiangensis', PeerJ, vol. 5, p. 14.

Hedges, SB 2014,''The high-level classification of skinks (Reptilia, Squamata, Scincomorpha'', *Zootaxa*, vol. 3765, no. 4, pp. 317-38.

Hillier, LW, Miller, W, Birney, E, Warren, W, Hardison, RC, Ponting, CP, Bork, P, Burt, DW, Groenen, MAM, Delany, ME, Dodgson, JB, Chinwalla, AT, Cliften, PF, Clifton, SW, Delehaunty, KD, Fronick, C, Fulton, RS, Graves, TA, Kremitzki, C, Layman, D, Magrini, V, McPherson, JD, Miner, TL, Minx, P, Nash, WE, Nhan, MN, Nelson, JO, Oddy, LG, Pohl, CS, Randall-Maher, J, Smith, SM, Wallis, JW, Yang, S-P, Romanov, MN, Rondelli, CM, Paton, B, Smith, J, Morrice, D, Daniels, L, Tempest, HG, Robertson, L, Masabanda, JS, Griffin, DK, Vignal, A, Fillon, V, Jacobbson, L, Kerje, S, Andersson, L, Crooijmans, RPM, Aerts, J, van der Poel, JJ, Ellegren, H, Caldwell, RB, Hubbard, SJ, Grafham, DV, Kierzek, AM, McLaren, SR, Overton, IM, Arakawa, H, Beattie, KJ, Bezzubov, Y, Boardman, PE, Bonfield, JK, Croning, MDR, Davies, RM, Francis, MD, Humphray, SJ, Scott, CE, Taylor, RG, Tickle, C, Brown, WRA, Rogers, J, Buerstedde, J-M, Wilson, SA, Stubbs, L, Ovcharenko, I, Gordon, L, Lucas, S, Miller, MM, Inoko, H, Shiina, T, Kaufman, J, Salomonsen, J, Skjoedt, K, Wong, GK-S, Wang, J, Liu, B, Wang, J, Yu, J, Yang, H, Nefedov, M, Koriabine, M, deJong, PJ, Goodstadt, L, Webber, C, Dickens, NJ, Letunic, I, Suyama, M, Torrents, D, von Mering, C, Zdobnov, EM, Makova, K, Nekrutenko, A, Elnitski, L, Eswara, P, King, DC, Yang, S, Tyekucheva, S, Radakrishnan, A, Harris, RS, Chiaromonte, F, Taylor, J, He, J, Rijnkels, M, Griffiths-Jones, S, Ureta-Vidal, A, Hoffman, MM, Severin, J, Searle, SMJ, Law, AS, Speed, D, Waddington, D, Cheng, Z, Tuzun, E, Eichler, E, Bao, Z, Flicek, P, Shteynberg, DD, Brent, MR, Bye, JM, Huckle, EJ, Chatterji, S, Dewey, C, Pachter, L, Kouranov, A, Mourelatos, Z, Hatzigeorgiou, AG, Paterson, AH, Ivarie, R, Brandstrom, M, Axelsson, E, Backstrom, N, Berlin, S, Webster, MT, Pourquie, O, Reymond, A, Ucla, C, Antonarakis, SE, Long, M, Emerson, JJ, Betrán, E, Dupanloup, I, Kaessmann, H, Hinrichs, AS, Bejerano, G, Furey, TS, Harte, RA, Raney, B, Siepel, A, Kent, WJ, Haussler, D, Eyras, E, Castelo, R, Abril, JF, Castellano, S, Camara, F, Parra, G, Guigo, R, Bourque, G, Tesler, G, Pevzner, PA, Smit, A, Fulton, LA, Mardis, ER, Wilson, RK, International Chicken Genome Sequencing, C, Overall, c, Genome fingerprint map, s, assembly, Mapping, c, DNAs, Other, s, libraries, Analysis, annotation & Project, m 2004,''Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolutio'', *Nature*, vol. 432, no. 7018, pp. 695-716.

Holz, P 2020,''Anatomy and Physiology of the Reptile Renal Syste'', *The veterinary clinics of North America. Exotic animal practice*, vol. 23, pp. 103-14.

Holzer, M & Marz, M 2019, 'De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers', *Gigascience*, vol. 8, no. 5, p. 16.

Huang, Y, Niu, B, Gao, Y, Fu, L & Li, W 2010, 'CD-HIT Suite: a web server for clustering and comparing biological sequences', *Bioinformatics*, vol. 26, no. 5, pp. 680-2.

Hudson, WH & Ortlund, EA 2014,''The structure, function and evolution of proteins that bind DNA and RN'', *Nat Rev Mol Cell Biol*, vol. 15, no. 11, pp. 749-60.

Hun, H., Tseng, E. 2019, 'Cogent: Coding GENome reconstruction Tool', https://github.com/Magdoll/Cogent/wiki

Jiang, H & Wong, WH 2009, 'Statistical inferences for isoform expression in RNA-Seq', *Bioinformatics*, vol. 25, no. 8, pp. 1026-32.

Jin, YT, Brandt, DYC, Li, JS, Wo, YB, Tong, HJ & Shchur, V 2021,''Elevation as a selective force on mitochondrial respiratory chain complexes of the *Phrynocephalus* lizards in the Tibetan platea'', *Current Zoology*, vol. 67, no. 2, pp. 191-9.

Larsen, PA, Heilman, AM & Yoder, AD 2014,''The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organism'', *Bmc Genomics*, vol. 15, p. 15.

Lee, WK, Zainuddin, NAM, Teh, HY, Lim, YY, Jaafar, MU, Khoo, JS, Ghazali, AK, Namasivayam, P, Abdullah, JO & Ho, CL 2017, 'Reducing the number of artifactual repeats in de novo assembly of RNA-Seq data by optimizing the assembly pipeline', *Gene Reports*, vol. 9, pp. 7-12.

Lewis, S, Chen, L, Raghuram, V, Khundmiri, SJ, Chou, CL, Yang, CR & Knepper, MA 2021,'""SLC-omic"" of the kidney: solute transporters along the nephro'', *Am J Physiol Cell Physiol*, vol. 321, no. 3, pp. C507-c18.

Li H. 2013, 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', arXiv:1303.3997v2 [q-bio.GN].

Li, H 2018, 'Minimap2: pairwise alignment for nucleotide sequences', *Bioinformatics*, vol. 34, no. 18, pp. 3094-100.

Li, H & Durbin, R 2009, 'Fast and accurate short read alignment with Burrows-Wheeler transform', Bioinformatics, vol. 25, no. 14, pp. 1754-60.

Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N, Marth, G, Abecasis, G & Durbin, R 2009, 'The Sequence Alignment/Map format and SAMtools', Bioinformatics, vol. 25, no. 16, pp. 2078-9.

Li, RQ, Li, YR, Fang, XD, Yang, HM, Wang, J, Kristiansen, K & Wang, J 2009, 'SNP detection for massively parallel whole-genome resequencing', Genome Research, vol. 19, no. 6, pp. 1124-32.

Li, W & Godzik, A 2006, 'Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences', *Bioinformatics*, vol. 22, no. 13, pp. 1658-9.

Li, W, Jaroszewski, L & Godzik, A 2001, 'Clustering of highly homologous sequences to reduce the size of large protein databases', *Bioinformatics*, vol. 17, no. 3, pp. 282-3.

Lighten, J, van Oosterhout, C & Bentzen, P 2014,''Critical review of NGS analyses for de novo genotyping multigene familie'', *Mol Ecol*, vol. 23, no. 16, pp. 3957-72.

Lind, AL, Lai, YYY, Mostovoy, Y, Holloway, AK, Iannucci, A, Mak, ACY, Fondi, M, Orlandini, V, Eckalbar, WL, Milan, M, Rovatsos, M, Kichigin, IG, Makunin, AI, Johnson Pokorná, M, Altmanová, M, Trifonov, VA, Schijlen, E, Kratochvíl, L, Fani, R, Velenský, P, Rehák, I, Patarnello, T, Jessop, TS, Hicks, JW, Ryder, OA, Mendelson, JR, Ciofi, C, Kwok, P-Y, Pollard, KS & Bruneau, BG 2019,''Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizard'', *Nature Ecology & Evolution*, vol. 3, no. 8, pp. 1241-52.

Liu, X, Mei, W, Soltis, PS, Soltis, DE & Barbazuk, WB 2017, 'Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome', *Mol Ecol Resour*, vol. 17, no. 6, pp. 1243-56.

Liu, Y, Zhou, Q, Wang, YJ, Luo, LH, Yang, J, Yang, LF, Liu, M, Li, YR, Qian, TM, Zheng, Y, Li, MY, Li, J, Gu, Y, Han, ZJ, Xu, M, Wang, YJ, Zhu, CL, Yu, B, Yang, YM, Ding, F, Jiang, JP, Yang, HM & Gu, XS 2015,''*Gekko japonicus* genome reveals evolution of adhesive toe pads and tail regeneratio'', *Nature Communications*, vol. 6, p. 11.

Livernois, A, Hardy, K, Domaschenz, R, Papanicolaou, A, Georges, A, Sarre, SD, Rao, S, Ezaz, T & Deakin, JE 2016,''Identification of interleukin genes in *Pogona vitticeps* using a de novo transcriptome assembly from RNA-seq dat'', *Immunogenetics*, vol. 68, no. 9, pp. 719-31.

Lopez-Maestre, H, Brinza, L, Marchet, C, Kielbassa, J, Bastien, S, Boutigny, M, Monnin, D, El Filali, A, Carareto, CM, Vieira, C, Picard, F, Kremer, N, Vavre, F, Sagot, MF & Lacroix, V 2016, 'SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence', Nucleic Acids Research, vol. 44, no. 19, p. 13.

Macrander, J, Broe, M & Daly, M 2016,''Tissue-Specific Venom Composition and Differential Gene Expression in Sea Anemone'', *Genome Biology and Evolution*, vol. 8, no. 8, pp. 2358-75.

Manni, M, Berkeley, MR, Seppey, M, Simão, FA & Zdobnov, EM 2021, 'BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes', *Molecular biology and evolution*, vol. 38, no. 10, pp. 4647-54.

Meyer, TW & Hostetter, TH 2021,''Why Is the GFR So High?: Implications for the Treatment of Kidney Failur'', *Clinical Journal of the American Society of Nephrology*, vol. 16, no. 6, pp. 980-7.

Miller, HC, Biggs, PJ, Voelckel, C & Nelson, NJ 2012, 'De novo sequence assembly and characterisation of a partial transcriptome for an evolutionarily distinct reptile, the tuatara (*Sphenodon punctatus*)', *Bmc Genomics*, vol. 13, p. 12.

Minio, A, Massonnet, M, Figueroa-Balderas, R, Vondras, AM, Blanco-Ulate, B & Cantu, D 2019, 'Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development', *G3 Genes|Genomes|Genetics*, vol. 9, no. 3, pp. 755-67.

Nielsen, S, Chou, CL, Marples, D, Christensen, EI, Kishore, BK & Knepper, MA 1995,''Vasopressin increases water permeability of kidney collecting duct by inducing translocation of aquaporin-CD water channels to plasma membran'', *Proc Natl Acad Sci U S A*, vol. 92, no. 4, pp. 1013-7.

Nielsen, S, Frokiaer, J, Marples, D, Kwon, TH, Agre, P & Knepper, MA 2002,''Aquaporins in the kidney: from molecules to medicin'', *Physiol Rev*, vol. 82, no. 1, pp. 205-44.

Ockendon, NF, O'Connell, LA, Bush, SJ, Monzon-Sandoval, J, Barnes, H, Szekely, T, Hofmann, HA, Dorus, S & Urrutia, AO 2016, 'Optimization of next-generation sequencing transcriptome annotation for species lacking sequenced genomes', *Mol Ecol Resour*, vol. 16, no. 2, pp. 446-58.

Ogawa, T, Oda-Ueda, N, Hisata, K, Nakamura, H, Chijiwa, T, Hattori, S, Isomoto, A, Yugeta, H, Yamasaki, S, Fukumaki, Y, Ohno, M, Satoh, N & Shibata, H 2019, 'Alternative mRNA splicing in three venom families underlying a possible production of divergent venom proteins of the habu snake, *protobothrops flavoviridis*', *Toxins*, vol. 11, no. 10.

O'Keeffe, KR & Jones, CD 2019, 'Challenges and solutions for analysing dual RNA-seq data for non-model host-pathogen system', *Methods in Ecology and Evolution*, vol. 10, no. 3, pp. 401-14.

Pacific Biosciences 2018, 'SMRT® Tools Reference Guide, Pacific Biosciences of California, Inc. Version 04', https://www.pacb.com/support/software-downloads/

Pacific Biosciences 2022, 'PacBio Announces Revio, a Revolutionary New Long Read Sequencing System Designed to Provide 15 Times More HiFi Data and Human Genomes at Scale for Under $1,000, Pacific Biosciences of California, Inc.', https://www.pacb.com/press_releases/pacbio-announces-revio-a-revolutionary-new-long-read-sequencing-system-designed-to-provide-15-times-more-hifi-data-and-human-genomes-at-scale-for-under-1000/

Peter, JAC, James, B, Bastien, C & Heng, L 2015, 'SAM/BAM format v1.5 extensions for de novo assemblies', bioRxiv, p. 020024.

Pyron, RA, Burbrink, FT & Wiens, JJ 2013,''A phylogeny and revised classification of Squamata, including 4161 species of lizards and snake'', *BMC Evolutionary Biology*, vol. 13, no. 1, p. 93.

R Core Team 2019, 'R: A language and environment for statistical computing. R Foundation for Statistical Computing', *Vienna, Austria,* https://www.R-project.org/

R Studio Team 2020, 'Rstudio: Integrated Development for R. Rstudio', *PBC, Boston, MA.*, http://www.rstudio.com/

Raghavan, V, Kraft, L, Mesny, F & Rigerte, L 2022,''A simple guide to de novo transcriptome assembly and annotatio'', *Briefings in Bioinformatics*, vol. 23, no. 2, p. bbab563.

Ralls, K, Sunnucks, P, Lacy, RC & Frankham, R 2020,''Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variatio'', *Biological Conservation*, vol. 251, p. 8.

Ramsey, M & Crews, D 2007,''Adrenal-kidney-gonad complex measurements may not predict gonad-specific changes in gene expression patterns during temperature-dependent sex determination in the red-eared slider turtle (*Trachemys scripta elegans*'', *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, vol. 307A, no. 8, pp. 463-70.

Rana, SB, Zadlock, FJ, Zhang, ZP, Murphy, WR & Bentivegna, CS 2016, 'Comparison of De Novo Transcriptome Assemblers and k-mer Strategies Using the Killifish, *Fundulus heteroclitus*', *Plos One*, vol. 11, no. 4, p. 16.

Ratan, A, Olson, TL, Loughran, TP & Miller, W 2015, 'Identification of indels in next-generation sequencing data', Bmc Bioinformatics, vol. 16, no. 1, p. 42.

Reeder, TW, Townsend, TM, Mulcahy, DG, Noonan, BP, Wood, PL, Jr., Sites, JW, Jr. & Wiens, JJ 2015,''Integrated Analyses Resolve Conflicts over Squamate Reptile Phylogeny and Reveal Unexpected Placements for Fossil Tax'', *Plos One*, vol. 10, no. 3, p. e0118199.

Roscito, JG, Sameith, K, Pippel, M, Francoijs, KJ, Winkler, S, Dahl, A, Papoutsoglou, G, Myers, G & Hiller, M 2018, 'The genome of the tegu lizard *Salvator merianae*: combining Illumina, PacBio, and optical mapping data to generate a highly contiguous assembly', *Gigascience*, vol. 7, no. 12, p. 13.

Schaarschmidt, S, Fischer, A, Lawas, LMF, Alam, R, Septiningsih, EM, Bailey-Serres, J, Jagadish, SVK, Huettel, B, Hincha, DK & Zuther, E 2020, 'Utilizing PacBio Iso-Seq for Novel Transcript and Gene Discovery of Abiotic Stress Responses in Oryza sativa L', Int J Mol Sci, vol. 21, no. 21.

Schaller, L & Lauschke, VM 2019,''The genetic landscape of the human solute carrier (SLC) transporter superfamil'', *Human genetics*, vol. 138, no. 11-12, pp. 1359-77.

Schoch, CL, Ciufo, S, Domrachev, M, Hotton, CL, Kannan, S, Khovanskaya, R, Leipe, D, McVeigh, R, ''Neill, K, Robbertse, B, Sharma, S, Soussov, V, Sullivan, JP, Sun, L, Turner, S & Karsch-Mizrachi, I 2020,''NCBI Taxonomy: a comprehensive update on curation, resources and tool'', *Database : the journal of biological databases and curation*, vol. 2020.

Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, Bruford EA 2023, 'Genenames.org: the HGNC resources in 2023', *Nucleic Acids Res*. PMID: 36243972 DOI: 10.1093/nar/gkac888

Shea, GM 2021,''Nomenclature of supra-generic units within the Family Scincidae (Squamata'', *Zootaxa*, vol. 5067, no. 3, pp. 301-51.

Singhal, S 2013, 'De novo transcriptomic analyses for non-model organisms: an evaluation of methods across a multi-species data set', *Molecular Ecology Resources*, vol. 13, no. 3, pp. 403-16.

Soneson, C, Matthes, KL, Nowicka, M, Law, CW & Robinson, MD 2016, 'Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage', Genome Biology, vol. 17, no. 1, p. 12.

Su, W, Cao, R, Zhang, X-y & Guan, Y 2019,''Aquaporins in the kidney: physiology and pathophysiolog'', *American Journal of Physiology-Renal Physiology*, vol. 318, no. 1, pp. F193-F203.

Sun, H, Zuo, X, Sun, L, Yan, P, Zhang, F, Xue, H, Li, E, Zhou, Y, Wu, R & Wu, X 2018,''Insights into the seasonal adaptive mechanisms of Chinese alligators (*Alligator sinensis*) from transcriptomic analyse'', *Australian Journal of Zoology*, vol. 66, no. 2, pp. 93-102.

Tan, CH & Tan, KY 2021,''De novo venom-gland transcriptomics of spine-bellied sea snake (*Hydrophis curtus*) from penang, malaysia—next-generation sequencing, functional annotation and toxinological correlatio'', *Toxins*, vol. 13, no. 2.

Tardaguila, M, de la Fuente, L, Marti, C, Pereira, C, Pardo-Palacios, FJ, Del Risco, H, Ferrell, M, Mellado, M, Macchietto, M, Verheggen, K, Edelmann, M, Ezkurdia, I, Vazquez, J, Tress, M, Mortazavi, A, Martens, L, Rodriguez-Navarro, S, Moreno-Manzano, V & Conesa, A 2018,''SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantificatio'', *Genome Res*.

The Australian Amphibians and Reptile Genomics initiative 2020, https://doi.org/10.25953/xer9-2e12

Todd, EV, Black, MA & Gemmell, NJ 2016, 'The power and promise of RNA-seq in ecology and evolution', *Mol Ecol*, vol. 25, no. 6, pp. 1224-41.

Tseng, E., Ripp, F., Botsford, D., Dunn, C. 2019, 'ANGEL: Robust Open Reading Frame prediction', *Pacific Biosciences,* https://github.com/PacificBiosciences/ANGEL

Tseng, E, Underwood, JG, Evans Hutzenbiler, BD, Trojahn, S, Kingham, B, Shevchenko, O, Bernberg, E, Vierra, M, Robbins, CT, Jansen, HT & Kelley, JL 2021, 'Long-read isoform sequencing reveals tissue-specific isoform expression between active and hibernating brown bears (Ursus arctos)', *G3 Genes|Genomes|Genetics*, vol. 12, no. 3.

Uetz, P., Freed, P. & Hošek, J. (eds.) 2020, 'The Reptile Database', http://www.reptile-database.org, accessed [Accessed 13/10/2021]

Ungaro, A, Pech, N, Martin, JF, McCairns, RJS, Mevy, JP, Chappaz, R & Gilles, A 2017,''Challenges and advances for transcriptome assembly in non-model specie'', *Plos One*, vol. 12, no. 9, p. 21.

The UniProt Consortium 2018, 'UniProt: the universal protein knowledgebase', Nucleic Acids Research, vol. 46, no. 5, p. 2699.

The UniProt Consortium 2021, 'UniProt: the universal protein knowledgebase in 2021', Nucleic Acids Research, vol. 49, no. D1, pp. D480-D489.

The UniProt Consortium 2023, 'UniProt: the Universal Protein Knowledgebase in 2023', Nucleic Acids Research, vol. 51, no. D1, pp. D523-D531.

Valenzuela-Miranda, D, Gallardo-Escárate, C, Valenzuela-Muñoz, V, Farlora, R & Gajardo, G 2014, 'Sex-dependent transcriptome analysis and single nucleotide polymorphism (SNP) discovery in the brine shrimp Artemia franciscana', Marine Genomics, vol. 18, pp. 151-4.

Van den Berge, K, Hembach, KM, Soneson, C, Tiberi, S, Clement, L, Love, MI, Patro, R & Robinson, MD 2019, 'RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis', *Annual Review of Biomedical Data Science*, vol. 2, no. 1, pp. 139-73.

Vijay, N, Poelstra, JW, Kunstner, A & Wolf, JB 2013, 'Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments', *Mol Ecol*, vol. 22, no. 3, pp. 620-34.

Wang, YH, Liu, TT, Kung, WM, Chen, CC, Wen, YT, Lin, IC, Huang, CC & Wei, L 2015,''Expression of aquaporins in intestine after heat strok'', *Int J Clin Exp Pathol*, vol. 8, no. 8, pp. 8742-53.

Wang, SF & Gribskov, M 2017, 'Comprehensive evaluation of de novo transcriptome assembly programs and their effects on differential gene expression analysis', *Bioinformatics*, vol. 33, no. 3, pp. 327-33.

Ward, C., To, T., Pederson, S. 2018, 'ngsReports: An R Package for managing FastQC reports and other NGS related log files', *bioRxiv,* 313148. https://doi.org/10.1101/313148

Wen, HM, Chen, W, Chen, Y, Wei, G & Ni, T 2023, 'Integrative analysis of Iso-Seq and RNA-seq reveals dynamic changes of alternative promoter, alternative splicing and alternative polyadenylation during Angiotensin II-induced senescence in rat primary aortic endothelial cells', Frontiers in Genetics, vol. 14, p. 16.

Wickham, H. 2019, 'stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0.', https://CRAN.R-project.org/package=stringr

Wickham, H., François, R., Henry, L. & Müller, K. 2021, 'dplyr: A Grammar of Data Manipulation. R package version 1.0.6.', https://CRAN.R-project.org/package=dplyr

Williams, CR, Baccarella, A, Parrish, JZ & Kim, CC 2016, 'Trimming of sequence reads alters RNA-Seq gene expression estimates', *Bmc Bioinformatics*, vol. 17, no. 1, p. 103.

Wong, KY, Tan, KY, Tan, NH, Gnanathasan, CA & Tan, CH 2021,''Elucidating the venom diversity in Sri Lankan spectacled cobra (*Naja naja*) through de novo venom gland transcriptomics, venom proteomics and toxicity neutralizatio'', *Toxins*, vol. 13, no. 8.

Workman, RE, Myrka, AM, Wong, GW, Tseng, E, Welch, KC, Jr. & Timp, W 2018, 'Single-molecule, full-length transcript sequencing provides insight into the extreme metabolism of the ruby-throated hummingbird *Archilochus colubris'*, *Gigascience*, vol. 7, no. 3, pp. 1-12.

Yang, SF, Lu, CW, Yao, CT & Hung, CM 2019, 'To Trim or Not to Trim: Effects of Read Trimming on the De Novo Genome Assembly of a Widespread East Asian Passerine, the Rufous-Capped Babbler (*Cyanoderma ruficeps Blyth*)', *Genes (Basel)*, vol. 10, no. 10.

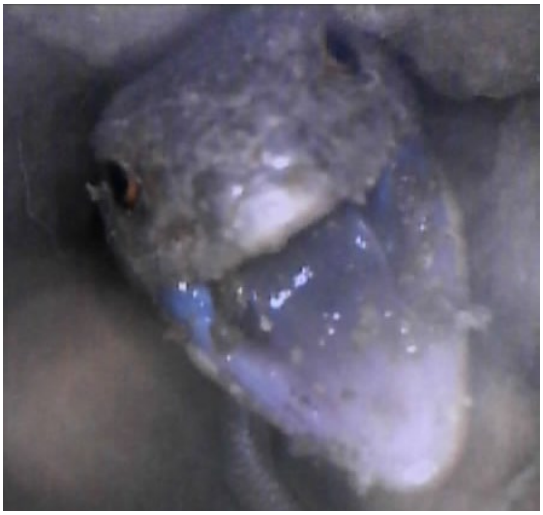
Young, T, Walker, S, Alfaro, A, Fletcher, L, Murray, JS, Lulijwa, R & Symonds, J 2019,''Impact of acute handling stress, anaesthesia and euthanasia on fish plasma biochemistry: Implications for veterinary screening and metabolomic samplin'', *Fish Physiology and Biochemistry*, vol. 45.

Zhao, Z, Elsik, CG, Hibbard, BE & Shelby, KS 2021, 'Detection of alternative splicing in western corn rootworm (Diabrotica virgifera virgifera LeConte) in association with eCry3.1Ab resistance using RNA-seq and PacBio Iso-Seq', Insect Molecular Biology, vol. 30, no. 4, pp. 436-45.

Zhang, C, Zhang, B, Lin, L-L & Zhao, S 2017, 'Evaluation and comparison of computational tools for RNA-seq isoform quantification', *Bmc Genomics*, vol. 18, no. 1, p. 583

Zhou, X & Rokas, A 2014,''Prevention, diagnosis and treatment of high-throughput sequencing data pathologie'', *Mol Ecol*, vol. 23, no. 7, pp. 1679-700.

4. Seasonal Variation in Gene Expression in *Tiliqua adelaidensis* Kidney Tissue



Tiliqua adelaidensis viewed in burrow with endoscope - C. Maher

4.1 Preface

Gene expression studies provide insight into changes in transcription levels in individuals of a species in response to changing environmental factors or stimuli, and allow the identification of genes that may be of functional interest. Here I compared gene expression of RNA extracted from kidney tissue from eight *T. adelaidensis* individuals; four collected in Austral autumn (March or April) and the end of the hot dry season, and four collected in Austral spring (September), at the start of the activity period for lizards following the wet winter season. Renal tissue was selected for comparison to focus on genes likely to play a role in lizards surviving the dry period and with roles in homeostasis and water retention. Cleaned short-read sequencing data were used to quantify estimated transcript expression counts using the program Kallisto, and the transcript list from long-read data generated in Chapter 3. The output was then analysed in R Studio using Edge R to calculate differential expression of genes between seasonal groups and visualise these results. A total of 214 genes was identified to be differentially expressed between these seasons: 25 were upregulated, and 189 were downregulated in autumn compared to spring. Of these, 118 were annotated to putative gene name, including genes that function in water and ion transport, homeostasis and cell signalling in the kidney of *T. adelaidensis*.

4.2 Introduction

Australia has the largest reptile diversity of any country, and climate change is the fourth most important key threatening process for Australia's 20 most threatened squamates, behind the impacts of other invasive species, agriculture and natural system changes (Geyle et al. 2020). Examining changing environmental conditions and species' capacity for adaptation is crucial in conservation management regimes going forwards. In ectotherms—squamate reptiles in particular—extreme variation in thermal extremes coupled with rainfall are key environmental indicators of the critical thermal tolerance levels of resident species (Clusella-Trullas, Blackburn & Chown 2011). Genetic responses to these factors have been particularly studied in reptiles with temperature-dependent sex determination, such as the common snapping turtle (*Chelydra serpentina*; Das et al. 2020).

Climate change is also one of the key threatening processes predicted to cause increasing aridity and environmental extremes, which will have severe impacts on the endangered Australian skink *T. adelaidensis*. Genetic measures of population health are now routinely measured alongside

ecological measures when assessing the viability of species and the likely impacts of climate change (Razgour et al. 2019). RNA-seq analysis in ecology and with a conservation focus has become increasingly accessible with the continual development of analysis tools and increasing affordability and quality of sequencing options available (Todd, Black & Gemmell 2016; Van den Berge et al. 2019). Heat shock proteins Hsp70 and Hsp90 have been identified as important biomarkers of the effect of heat shock on loggerhead turtle (*Caretta caretta*) embryos due to gene expression regulation under environmental stress (Bentley et al. 2017; Chow et al. 2021; Tedeschi et al. 2015), and this variation in expression levels has been shown to be heritable (Tedeschi et al. 2016).

Gene expression studies provide a snapshot of genes being actively transcribed at the time of sampling. Thus, these studies have been used to measure an animal's response to various stimuli, including stress (Williams et al. 2003), immune challenges (Fassbinder-Orth 2014; Xu et al. 2020), environmental factors such as elevation (Hofmann et al. 2019; Jin et al. 2021) and temperature-dependent sex development (Feiner et al. 2018; Fleming et al. 1999; Harry, Williams & Briscoe 1990; Tedeschi et al. 2015; Torres-Maldonado et al. 2001). Some studies have limited their focus to expression of specific genes to further characterise protein synthesis, including research on venom and toxins (Junqueira-de-Azevedo et al. 2016; Tan & Tan 2021; Viala et al. 2015), viviparity (Brandley et al. 2012; Griffith et al. 2016) and colour variation (Fulgione et al. 2014; McLean et al. 2019).

The aim of the study described in this chapter was to identify genes of pygmy bluetongues that were differentially expressed during Austral spring and autumn, to assess these lizards' gene expression response to seasonal conditions. Kidney tissue was chosen to measure gene expression responses that may affect water retention or renal efficiency, and reflect other effects of low water availability on renal function (Babonis, Miller & Evans 2011; Nielsen et al. 1995).

4.3 Methods

Short-read Illumina RNA sequencing reads (as described in Chapter 3) were quantified using the program Kallisto (Bray et al. 2016), using the clustered, annotated canonical transcript list generated in Chapter 3 as a reference. Kallisto estimates transcript counts per million reads (CPM) using pseudoalignment methods to align the short reads to a reference. Its use quantified the cleaned paired-end short-reads (as sequenced in Chapter 2 and cleaned in Chapter 3) against the

list of 9,813 full-length gene nucleotide sequences. Kallisto version 0.48.0 was used in a Miniconda version 4.9.2 environment to index the clustered file of putative genes and quantify the trimmed short reads using 100 bootstraps to estimate the accuracy of the estimated abundances. These abundance values generated by Kallisto were then imported into R Studio (R Studio Team 2020) with R version 3.6.1 (R Core Team 2019), using the EdgeR package (V. 3.28.1) (Chen, Y, Lun, A & Smyth, G 2016; Chen, Y, Lun, AT & Smyth, GK 2016; McCarthy, Chen & Smyth 2012; Robinson, McCarthy & Smyth 2010), and the contained catchKallisto function.

Initial data analysis was done on all eight samples, as well as on the six samples completed on the same sequencing run (excluding samples G6 and G7), to explore potential batch effects (Appendix 2). Before applying calculations to experimental groups, summary statistics for samples were plotted—including library size and counts per reference sequence within samples—and samples were compared using a multidimensional scaling (MDS) plot. Analyses were run using season (spring *v*. autumn) as the variable to test expression differences, and then also using sex (male *v*. female) to explore potential sources of bias in the data. Seasonal gene expression was also calculated for the four female samples alone because of observed differences for the male sample G5 (see Section 4.4)). All these analyses (i.e., of all eight samples; only the four female samples; and the main six samples) were conducted on differential expression data as outlined below. As well as an overall larger sequencing depth and resulting transcript number, samples G6 and G7 showed extensive differences from all other samples and each other on an MDS plot. In light of evidence for a large sequencing batch effect, which was not resolved by EdgeR's normalisation functions, the full final analysis was performed on the data excluding G6 and G7: three samples from each season group, each with two female and one male member. Short-read sample IDs were generated using lizard ID, season group and sex: Sep = September (spring), Mar = March (autumn). For example, "G1_SepF" is Lizard G1 collected in September and female.

A heatmap was created to show the most variable genes (based on the difference in counts from the mean across all samples) across samples and visualise gene expression, and is representative only of total log CPM for each gene in each sample. The top 250 most variable genes were identified in RStudio based on variance values calculated among rows using the log counts matrix. Because of the extremely small sample size in comparator groups in this dataset, genes with fewer than five reads in two samples were filtered from the dataset. This was the only filtering step. Filtering gene counts before analysis has been shown to improve the FDR, particularly for genes

with a dominant isoform (Soneson et al. 2016). It also prevents noisy or erroneous splicing contributing to low-abundance isoforms (Soneson et al. 2016), which in this case with so few individuals may affect observed seasonal group differences. Filtering can also be conducted directly using the FDR or CPM, value which can allow for library differences in the data (Ren et al. 2023). However, Aceto et al. (2014) showed that manual filtering based on FDR is more likely to introduce filtering-induced bias than it is to improve the power of an analysis. Filtering may also be used in studies where multiple long-read samples are being compared, using transcripts per million or count values (Ren et al. 2023) and is often applied to counts in fewer than three samples. In EdgeR, autumn was input as the second variable and EdgeR compared these factors as 'Comparison of groups: 2-1'. This means that when a gene is said to have a higher fold change in expression counts—or is 'upregulated'—it has higher expression counts in autumn (group 2) than in spring (group 1).

EdgeR was then used to calculate normalisation factors, estimate common and tagwise dispersion and run an exact test (which uses quantile-adjusted conditional maximum likelihood [qCML] methods) to produce fold-change values for changes in expression between specified experimental group factors (spring *v*. autumn) and calculate *p* values and FDRs. The top 25 differentially expressed genes based on *p* value were output using the 'topTags' EdgeR function and this table was joined to the table of annotation data produced in Chapter 3 using the column of transcript IDs as the common identifying factor, to create a table with putative annotations and gene expression statistics for each transcript ID. Summary information about the number of upregulated (higher differential expression levels) or downregulated (lower differential expression levels) genes was generated using the summary function and 'decideTests' within EdgeR using default parameters (*p* value cut-off = 0.05). The full table produced by decideTests includes each gene transcript ID and a column with a value of –1, 0 or 1 indicating significantly downregulated, no significant change and significantly upregulated genes respectively.

Using stringr, dplyr (Wickham 2019; Wickham et al. 2021) and stringi (Gagolewski 2021), the dataframe including all expression data was then joined to the dataframe containing annotation information using transcript IDs as the common value to join rows. This dataframe including all joined expression and annotation data was then filtered based on upregulation and downregulation values to produce summary tables on the differentially expressed genes. The tables of differentially expressed genes, as well as the full dataframe, was also filtered for missing

annotation data and compared with the genes of interest list explored in Chapter 3 to explore putative annotations for expressed genes. This filtering was also used to count how many genes were annotated, how many were on the genes of interest list and which of the genes in these categories were differentially expressed, and to create summary tables.

Visualisation of differentially expressed genes was done using a plotSmear with log fold change on the *y* axis and average log CPM reads on the *x* axis. The negative log of the *p* value was also compared to the log fold change using a volcano plot (Raghavan et al. 2022) displaying genes with a significant *p* value (<0.05) and log fold change >2.

Full analysis notes and scripts are provided in repositories supplied in Appendix 2 analysis files; quality reports are available for download from GitHub; and the full analysis and R script is available in Appendix 2. This appendix was written using R Markdown (Allaire et al. 2019) with additional formatting completed using Rmdformats (Barnier 2021) and FormatR (Xie 2021); presentation of figures was edited using gplots (Warnes et al. 2020). RColourBrewer (Neuwirth 2014) was used to specify a colour-blind suitable colour palate using hex colour codes that was kept consistent among samples throughout the various factor analyses explored.

Over-representation of significantly expressed genes (both upregulated and downregulated in autumn) was analysed using the Panther GO enrichment analysis tool (PANTHER 18.0, release 20230705). As recommended, the lists of differentially expressed genes were uploaded to the online portal using the gene ID, and the tool was run against a custom reference list of all unique gene IDs identified in the entire transcript dataset. The tool was run with default options to use a Fisher's exact test and calculate the FDR as the error correction method. Note that not all differentially expressed genes were included in this analysis as it only includes genes with a BLASTx hit.

4.4 Results

Most genes had median $\log_2$ CPM reads of slightly less than 5 per sample; in other words, for every million reads, the average transcript has slightly less than 2^5 (or 32) counts. However, every sample had some genes with an extremely large number of CPM. The count distributions per million reads for samples G1–G5 and G8 are similar to each other (Figure 4-1).

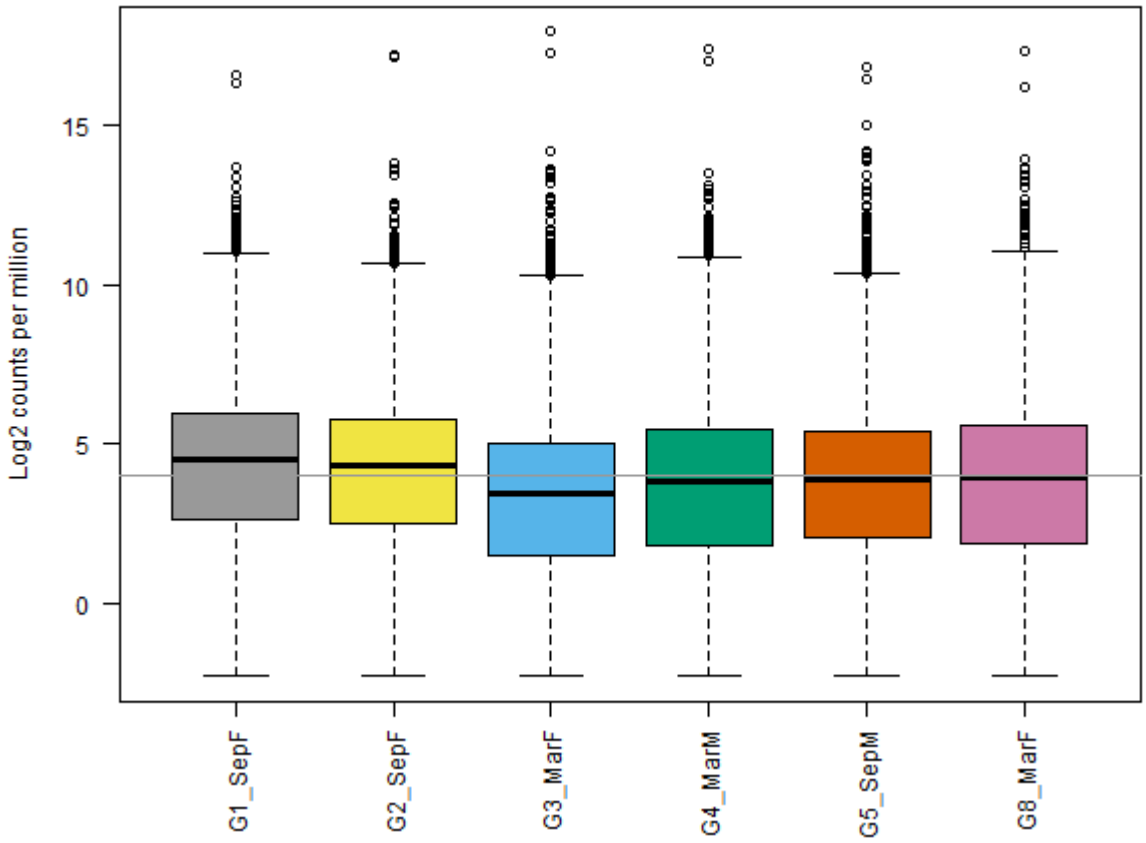


Figure 4-1: Boxplot showing Log2 of estimated counts per million for each putative gene transcript estimated by Kallisto for gene expression of mRNA sequenced from kidney tissue of six *Tiliqua adelaidensis* individuals, three collected in autumn and three collected in spring. Season and Sex are indicated in labels: Sep = September (spring), Mar = March (autumn), F = female, M =male.

Exploration of expression fold changes and differences among samples are shown using MDS plots. Samples G1 and G2 are both females collected in spring, while samples G3, G8 and G4 are two females and a male collected in autumn. While the latter three samples are not tightly clustered, they are closer to each other than to any other samples. Sample G5 is the only male collected in spring and is the most removed from all other samples (Figure 4-2). A similar plot to this one in Appendix 2 shows samples G6 and G7 farther removed from all other samples and each other, forming part of the rationale for their exclusion. The difference here in G5 is why there is a section in Appendix 2 dedicated to comparing only the female samples (G1 and G2, to G3 and G8), which as in this graph show clear differences between seasons. In the heatmap generated from raw counts of the top 250 most variable genes, G5 was again starkly different from all other samples (Figure 4-3).

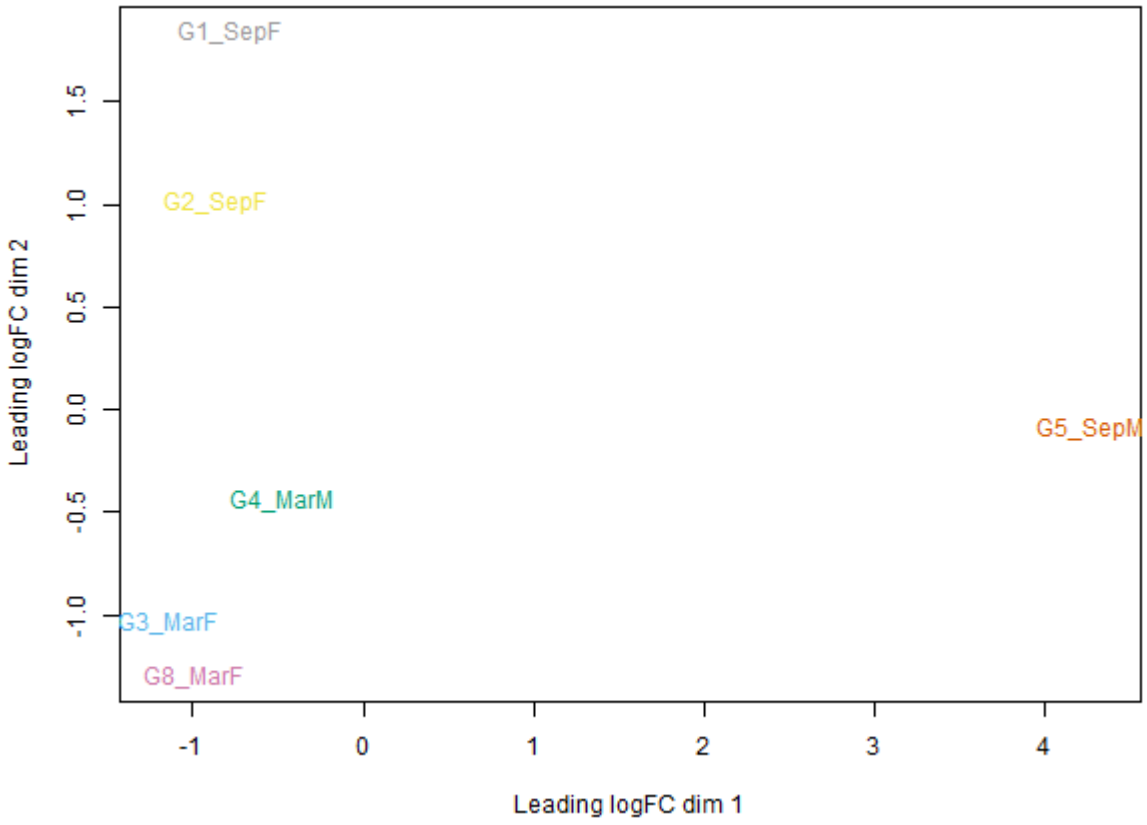


Figure 4-2: Multidimensional scaling plot showing variation using leading log fold changes among samples' gene expression values of mRNA, isolated from kidney tissue from six *Tiliqua adelaidensis* skinks, three collected in autumn and three collected in spring. G5 is the only September Male in this dataset (Season and Sex are indicated in labels: Sep = September, Mar = March, F =Female, M=Male)

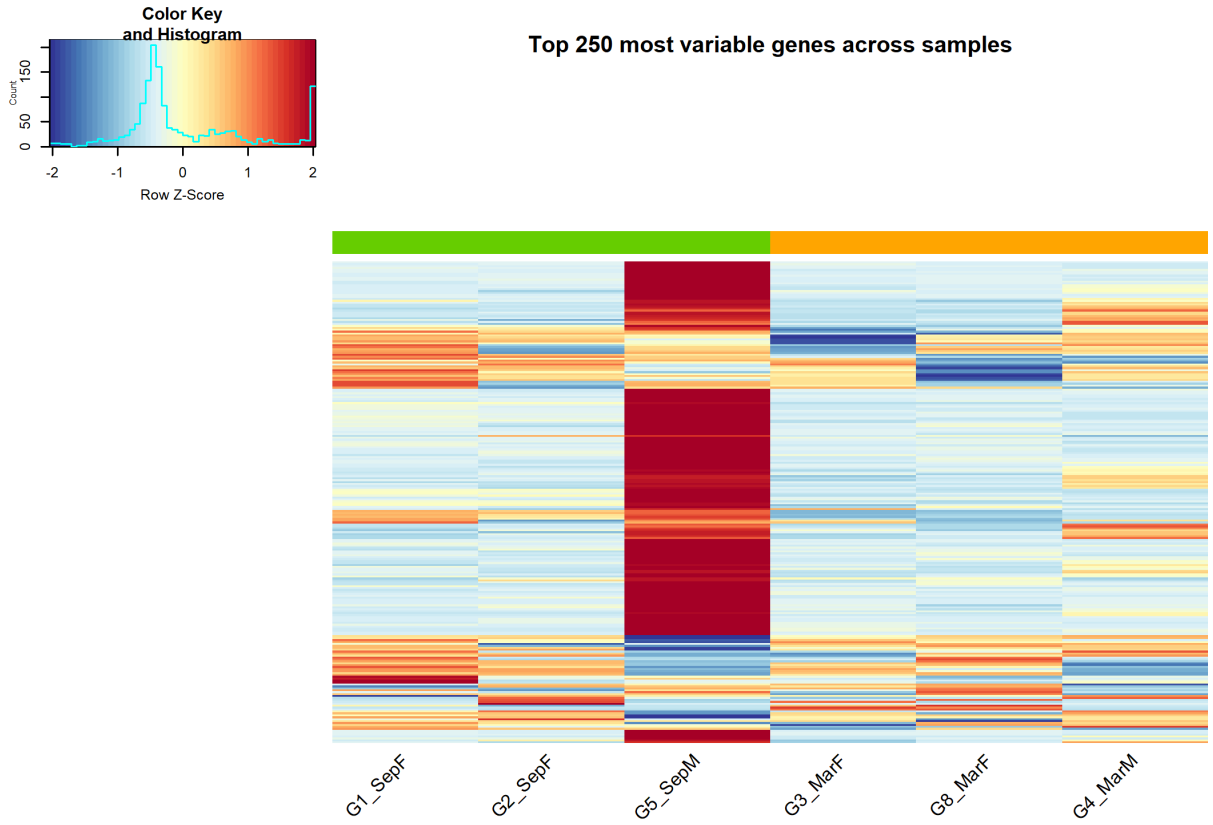


Figure 4-3: Top 250 most variable genes (based on log counts among all samples before seasonal expression calculations). Gene expression counts of mRNA isolated from kidney tissue from six *Tiliqua adelaidensis* skinks, three collected in autumn and three collected in spring The colours indicate relative count levels among individual samples scaled by the Row Z-score function which was calculated as: (Gene expression value in sample of interest-Mean expression across all samples)/Standard Deviation of all counts across all samples. Colours indicate how far a gene in a particular sample deviates from the mean (0) lower expression values are blue and higher expression values are red. (Season and Sex are indicated in labels: Sep = September, Mar = March, F =Female, M=Male)

The common dispersion factor —the mean dispersion across all genes— calculated for all genes across seasons was 0.2581349. A larger number of genes showed downregulation in autumn than in spring. This is evident in Figure 4-4 by the skewing of data points to the left of the *x*-axis, indicating that more canonical genes have a negative log fold change (logFC).

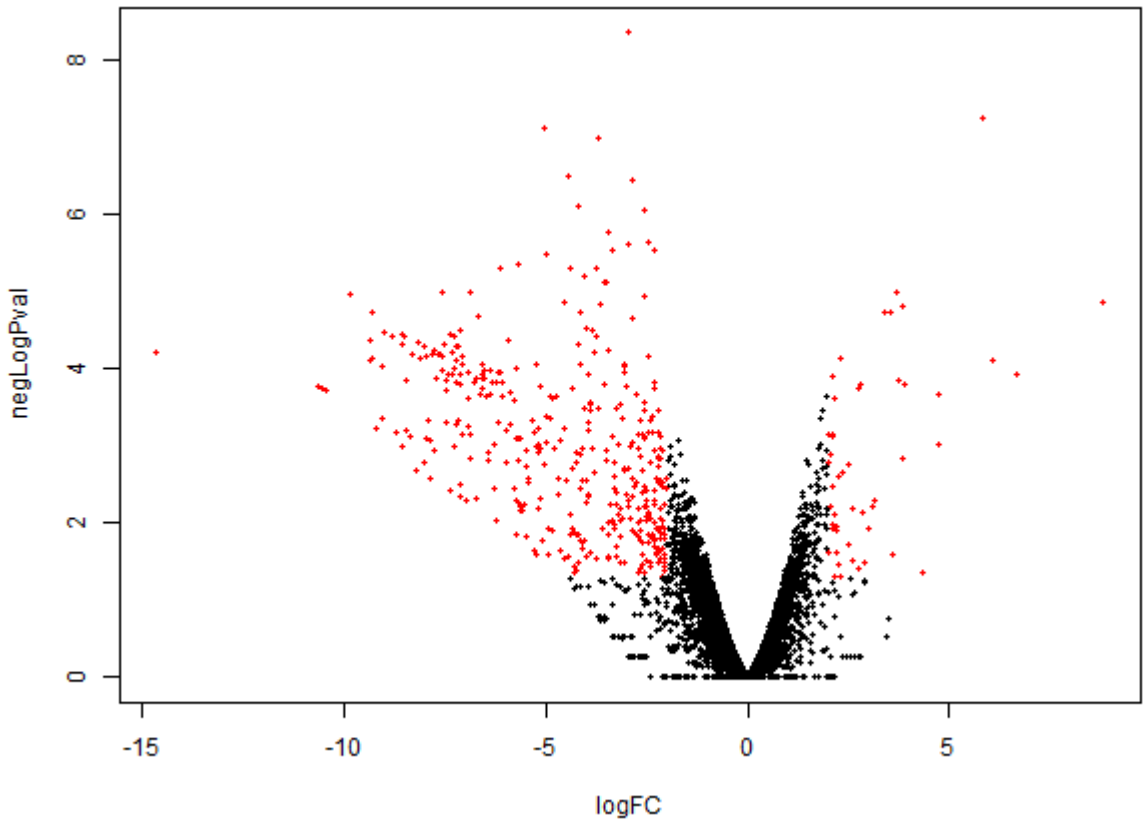


Figure 4-4: Gene expression comparison of mRNA isolated from kidney tissue from six *Tiliqua adelaidensis* skinks, three collected in autumn and three collected in spring. Negative Log of the p value against Log Fold Change in expression level comparing March to September season variables. Red dots indicate genes with a significant change over season, with p-values <0.05 and also a log fold change >2. Data points higher on the y-axis have a lower p value, and data points farther from 0 (in either direction) on the x-axis indicate a greater fold change in counts per million. Positive fold change indicates upregulation while negative fold change indicates downregulation.

Of the total 9,813 putative genes used as a reference, 9,599 did not have significantly different expression between the autumn and spring collection periods. Twenty-five genes were significantly upregulated in autumn, with 19 of these able to be annotated using BLASTx. However, only one of these, transcript 'PB.6136.2|ea3739|path0:15-2762(+)|transcript/5251', identified as gene *gja*, was also present on the genes of interest list generated in Chapter 3 (Table 4-1, Table 4-4). A total of 189 genes were significantly downregulated in autumn compared to spring expression levels, and 99 of these could be annotated with BLASTx; seven of these were present on the genes of interest list (Table 4-1, Table 4-3). A list of all significantly differentially expressed genes that did return a BLAST hit (19 upregulated and 99 downregulated) is provided in Appendix 1 (Table 6-2), which includes their full name and a brief description of function as per the UniProtKB database.

Table 4-1: Summary of gene expression regulation data of mRNA sequenced from kidney tissue of six *Tiliqua adelaidensis* individuals over two seasonal experimental groups (1=Spring {September}, 2=Autumn {March/April}). This '2-1' comparison means that upregulated genes indicate a higher expression was observed for samples collected in Autumn compared to Spring, and downregulated genes indicates comparatively lower counts in Autumn.

	Number of genes	**Number with BLASTx hits**	**Number present on Genes of Interest list**
Upregulated	25	19	1
Not Significant	9599		
Downregulated	189	99	7

Of the top 25 *most* differentially expressed genes (based on *p* value) only three were present in the genes of interest list: *slc25a15,* minichromosome maintenance complex component 3 (*mcm3*) and thymidine kinase (*tk1*). Not all the significantly differentially expressed genes were identified in the BLASTx UniProt SwissProt database (The UniProt Consortium 2018). Two genes on this list were upregulated in autumn and only one was assigned a BLASTx hit and subsequent gene symbol to *f52c9.6*, an uncharacterised transposon-derived protein. The top two differentially expressed genes failed to return a BLAST hit and were not annotated. All other genes that did return a BLASTx hit on the list of top 25 differentially expressed genes were downregulated. Genes identified through annotation using BLASTx that were in the top ten most differentially expressed genes were *pbk* (PDZ binding kinase), *tk1*, *mcm3*, *mad2l1* (MAD2 mitotic arrest deficient-like 1), *gatm* (glycine amidinotransferase (L-arginine:glycine amidinotransferase)), and *ostc* (oligosaccharyltransferase complex subunit). (Table 4-2).

Table 4-2: Top 25 most differentially expressed genes comparing March to September based on P value, of mRNA sequenced from kidney tissue of six *Tiliqua adelaidensis* individuals (2F 1M per season group) with corresponding BLASTx hits for Gene symbol and presence/absence in the genes of interest list generated in Chapter 3. Blue rows are genes that were assigned a BLASTx hit but were not present on the 'genes of interest list', yellow rows were assigned a BLASTx hit which is present in the genes of interest list.

	Differential Expression Statistics				**BLASTx Statistics**				**Genes of Interest Database Search**	
TranscriptID	**Over-disp.**	**logFC**	**logCPM**	**P Value**	**UniProt ID**	**Gene**	**% id**	**e-val**	**Org. name**	**description**
PB.499.11|124cf9|path4:2108-4109(+)|transcript/7418	1.6342	-2.967	4.6335	4.42E-09	NA	NA	NA	NA	NA	NA
PB.3885.3|918572|path1:1406-3716(+)|transcript/9242	3.9624	5.8201	0.8151	5.55E-08	NA	NA	NA	NA	NA	NA
PB.7745.1|transcript/17825:1-1637(+)|transcript/17825	1.0108	-5.0118	2.8031	7.50E-08	Q96KB5	*pbk*	60.58	1.50E-37	NA	NA
PB.3681.1|8a3a64|path3:1-2713(+)|transcript/5882	8.9516	-3.7007	3.9452	9.98E-08	NA	NA	NA	NA	NA	NA
PB.7443.1|transcript/15548:1-1797(+)|transcript/15548	1.0329	-4.4239	2.7361	3.12E-07	P04047	*tk1*	82.06	8.12E-136	*Gallus gallus*	thymidine kinase 1
PB.8897.1|transcript/3462:1-3118(+)|transcript/3462	1.0627	-2.8663	4.1681	3.54E-07	Q5ZMN2	*mcm3*	81.86	0	*Anser cygnoides domesticus*	minichromosome maintenance complex component 3
PB.8300.1|transcript/22013:1-1186(+)|transcript/22013	1	-4.211	2.2007	7.75E-07	Q13257	*mad2l1*	82	2.73E-120	NA	NA
PB.499.1|124cf9|path4:1-2080(+)|transcript/11611	1.4143	-2.5553	5.0358	8.94E-07	Q9I9K9	*gatm*	88.03	0	NA	NA
PB.5332.1|c9b505|path1:1-1445(+)|transcript/19992	3.4956	-3.4674	2.5038	1.66E-06	B0K025	*ostc*	93.75	2.21E-80	NA	NA
PB.2003.2|4ab8e8|path0:1458-4018(+)|transcript/6743	1.5509	-2.4788	4.4606	2.32E-06	NA	NA	NA	NA	NA	NA
PB.2869.1|6b0af0|path2:1-1117(+)|transcript/22387	1.0166	-2.9488	5.0634	2.45E-06	NA	NA	NA	NA	NA	NA
PB.2068.1|4cb9fa|path2:1-2419(+)|transcript/8339	3.4172	-3.356	1.5414	2.96E-06	Q6DEB1	*etnppl*	76.23	0	NA	NA
PB.5596.1|d429b7|path7:37-1550(+)|transcript/19333	3.6864	-2.2876	5.1944	2.98E-06	NA	NA	NA	NA	NA	NA

	Differential Expression Statistics				BLASTx Statistics				Genes of Interest Database Search	
TranscriptID	**Over-disp.**	**logFC**	**logCPM**	**P Value**	**UniProt ID**	**Gene**	**% id**	**e-val**	**Org. name**	**description**
PB.7967.1|transcript/19601:1-1493(+)|transcript/19601	1	-4.9755	2.4905	3.35E-06	Q1LZG6	*ccnb1*	68.08	0	NA	NA
PB.1773.1|42d732|path1:1-1839(+)|transcript/15277	1	-5.7001	1.8053	4.45E-06	Q5H7C0	*cdc20*	77.82	0	NA	NA
PB.7361.1|transcript/14810:1-1862(+)|transcript/14810	1	-3.7332	2.4253	4.87E-06	Q5ZJU5	*nusap1*	42.69	1.56E-46	NA	NA
PB.6779.1|transcript/10281:1-2227(+)|transcript/10281	1	-4.3972	1.979	4.93E-06	Q76I89	*ndc80*	68.22	0	NA	NA
PB.3848.1|906dad|path15:1-1965(+)|transcript/13370	8.5012	-6.1007	4.3178	5.10E-06	NA	NA	NA	NA	NA	NA
PB.8089.1|transcript/20489:1-1427(+)|transcript/20489	1.1594	-4.0435	2.1771	6.18E-06	Q9Y619	*slc25a15*	60.85	3.80E-102	*Anolis carolinensis*	solute carrier family 25 member 15
PB.1636.2|3ea652|path0:21-2775(+)|transcript/5207	1.2619	-3.5307	1.3107	7.66E-06	P79784	*entpd2*	65.16	0	NA	NA
PB.3849.1|906dad|path19:1-2010(+)|transcript/12836	2.7714	-3.5041	4.8059	7.78E-06	NA	NA	NA	NA	NA	NA
PB.7382.1|transcript/14974:1-1845(+)|transcript/14974	1	-6.8615	6.964	1.00E-05	P56594	*cyp2c21*	58.85	0	NA	NA
PB.5022.1|bbf2e4|path1:3-3181(+)|transcript/3405	1.9811	3.6964	5.9112	1.01E-05	Q10126	*f52c9.6*	26.45	8.03E-15	NA	NA
PB.3353.2|7cf053|path5:2504-7730(+)|transcript/274	3.794	-7.5439	4.983	1.02E-05	P07589	*fn1*	71.56	0	NA	NA
PB.4788.1|b34d56|path10:2-1859(+)|transcript/15227	1.3622	-9.844	7.9267	1.09E-05	NA	NA	NA	NA	NA	NA

Table 4-3 and Table 4-4 show the significantly upregulated and downregulated genes (as summarised in Table 4-1) that were assigned a BLASTx hit that matched a gene symbol on the genes of interest list (Table 6-6). All of these genes had significant differential expression between autumn and spring despite not being among the top 25 most differentially expressed based on *p* values. Of the seven genes on the genes of interest list that were downregulated, three are solute carriers: *slc25a15* (solute carrier family 25 member 15) mitochondrial ornithine transporter 1 and two genes that both produced a BLASTx hit for *slc5a1* (solute carrier family 5 member 1) sodium/glucose cotransporter 1. The other four genes were *chrna4* (cholinergic receptor nicotinic alpha 4 subunit), *dpep1* (dipeptidase 1 -renal), *mcm3* and *tk1*, the latter two of which were among the top 25 differentially expressed genes (Table 4-3).The only gene on the genes of interest list with evidence for upregulation was *gja1* (gap junction alpha 1), otherwise known as connexin-43 (or Cx43) (Table 4-4). The full list of significantly upregulated and significantly downregulated genes, and an overall summary of annotations and gene expression data have been uploaded to GitHub.

GO over-representation analysis of the genes upregulated in autumn returned no significantly over-represented GO categories other than the group of genes that were not classified into an ontology. The downregulated genes returned 98 over-represented gene ontologies with significant *p* values, though all these have small fold-enrichment values. The table of significant GO terms is included in Appendix 2 (Table 6-3).

Table 4-3: Seven genes identified as downregulated in the differentially expressed genes comparing March to September, of mRNA sequenced from kidney tissue of six *Tiliqua adelaidensis* individuals (2F 1M per season group) which returned both a BLASTx hit and 'genes of interest' entry

	Differential Expression Statistics				**BLASTx Statistics**				**Genes of Interest Database Search**	
TranscriptID	**Over-disp.**	**logFC**	**logCPM**	**P Value**	**Uniprot ID**	**Gene**	**% id**	**e-val**	**Org name**	**Description**
PB.840.1|1eedc5|path0:1-2091(+)|transcript/11690|m.1514	1	-5.8894	3.5505	0.000198	P09482	*chrna4*	87.961	0	*Gallus gallus*	cholinergic receptor nicotinic alpha 4 subunit
PB.1055.1|295311|path2:1-1979(+)|transcript/12857|m.1877	1	-3.1329	1.6564	0.000285	Q3SZM7	*dpep1*	68.564	0	*Anolis carolinensis*	dipeptidase 1
PB.8897.1|transcript/3462:1-3118(+)|transcript/3462|m.14650	1.0627	-2.8663	4.1681	3.54E-07	Q5ZMN2	*mcm3*	81.863	0	*Anser cygnoides domesticus*	minichromosome maintenance complex component 3
PB.8089.1|transcript/20489:1-1427(+)|transcript/20489|m.13880	1.1594	-4.0435	2.1771	6.18E-06	Q9Y619	*slc25a15*	60.853	3.80E-102	*Anolis carolinensis*	solute carrier family 25 member 15
PB.2610.1|60c22a|path0:1-2155(+)|transcript/10675|m.4798	1.6866	-2.5602	2.5882	0.00082	P53790	*slc5a1*	82.514	0	*Pogona vitticeps*	solute carrier family 5 member 1
PB.2610.2|60c22a|path0:1-2499(+)|transcript/7260|m.4799	2.0509	-3.0467	2.1128	8.78E-05	P53790	*slc5a1*	78.681	0	*Pogona vitticeps*	solute carrier family 5 member 1
PB.7443.1|transcript/15548:1-1797(+)|transcript/15548|m.13231	1.0329	-4.4239	2.7361	3.12E-07	P04047	*tk1*	82.063	8.12E-136	*Gallus gallus*	thymidine kinase 1

Table 4-4: One gene identified as upregulated in the differentially expressed genes comparing March to September, of mRNA sequenced from kidney tissue of six *Tiliqua adelaidensis* individuals (2F 1M per season group) which returned both a BLASTx hit and 'genes of interest' entry

TranscriptID	Differential Expression Statistics				BLASTx Statistics				Genes of Interest Database Search	
	Over-disp.	logFC	logCPM	P Value	Uniprot ID	Gene	% id	e-value	Org name	Description
PB.6136.2|ea3739|path0:15-2762(+)|transcript/5251|m.11463	1.4376	4.7122	0.1626	0.000955	P14154	*gja1*	86.935	1.66E-115	*Gallus gallus*	gap junction protein alpha 1 - Connexin 43

4.5 Discussion

In this chapter, differential gene expression was investigated for putative genes sequenced from kidney tissue of six *T. adelaidensis* individuals during early Austral spring at the end of the wet period/beginning of the dry period and at the beginning of Austral autumn before the end of the dry period. A much larger number of putative genes showed lower expression levels, or downregulation, in autumn compared to genes which showed downregulation in spring. I identified 189 genes that were downregulated, including *slc* genes, *chrna4, mcm3* and *tk1*. A much smaller number (25) of genes were significantly upregulated in autumn, including genes that function in water and ion transport, homeostasis, cell signalling and—in the case of *gja1*—the release of renin, which signals the renin–angiotensin antidiuretic hormone pathway (Figueroa & Duling 2009) and has potential impacts on muscle activity in kidney tubule wall dilation or contraction, thus affecting renal flow and blood pressure (Hanner et al. 2010). These genes are relevant to kidney function and water retention. Understanding what genes are important for potentially surviving dry environments may help in later studies to identify variability or expression of these important genes in different populations of this endangered lizard.

Two of the genes identified as downregulated in autumn were identified as *solute carrier family 5 member 1*, a sodium/glucose co-transporter known to be crucial to sodium transport and glucose reabsorption along the renal tubule (in mice) (Lewis et al. 2021), and *solute carrier family 25 member 15*, a mitochondrial ornithine transporter 1, which was among the top 25 most differentially expressed genes. Active transporters are a crucial component of the functioning of the distal kidney tubules (Dantzler 2003) and are even more crucial to reptiles, which lack a loop of Henle (Doneley et al. 2018; Holz 2020). This section of the kidney tubules in mammals helps control osmoregulation concentration gradients by generating a counter-current exchange between the ascending and descending loop, increasing the concentration of ions withing the medulla, and influencing sodium and water retention or output. This is what allows mammals to create concentrated urine (Singh & Dyce 2018; Sjaastad, Hove & Sand 2016). The downregulation of many genes that serve this purpose may imply that pygmy bluetongues limit renal flow and kidney functions during the drier months to avoid excessive water loss, as they are unable to produce concentrated urine at the kidney.

Solute carrier transporters function to allow substrates to pass through the surface membranes of cells and have been found in all vertebrate lineages (Xiong & Lei 2021). Solute carriers represent

one of the largest superfamilies in the human genome (Schaller & Lauschke 2019) with over 400 membrane-bound proteins classified into 65 subfamilies. A much larger number of *slc* genes was putatively identified in *T. adelaidensis* in Chapter 3 than the number shown to be differentially expressed. All BLAST results exhibited exceptionally low e-values and elevated bit-scores, suggesting a low likelihood of random matches; however *slc25a15* has a percent identity of 60.8, which is lower than other hits and may indicate a poorer match. In comparison, the two genes matching *slc5a1* have percent identity matches of 78.7% and 82.5% to the hit sequence, respectively. These *slc* genes may be different isoforms, or closely related members of *slc* family 5. It is expected that *slc* orthologues would differ extensively between species, but it is also possible that multiple members of a large and closely related gene family such as *slc* may lead to erroneous BLAST hits to closely related gene family representatives based on database availability when comparing across species. A sequence similarity network clusters members of *slc* family 5 tightly together (Schlessinger et al. 2013). Höglund et al. (2011) identified *slc* family 25 (mitochondrial transporters) in every eukaryote genome they explored. More than 30 *slc* families have been found in each of *Homo sapiens*, *Mus musculus*, *Caenorhabditis elegans* and *Drosophila melanogaster* (Höglund et al. 2011) and *slc* genes were shown to be part of the response to dehydration in the cactus mouse (*Peromyscus eremicus*) when placed in artificial conditions and tested for gene expression changes (Kordonowy & MacManes 2017). The fact that this extremely large and variable gene family is highly conserved across vertebrates allows us to infer that the orthologues identified here in *T. adelaidensis* are involved in similar responses to water stress to control the osmotic gradient and prevent water loss.

Solute carriers and aquaporins play a crucial role in ion transport across cell membranes, in cell communication and in maintaining osmotic gradients for a variety of processes. In this respect they are crucial to the functioning of reptile kidneys as the osmotic gradient plays an important role in regulating water and solute excretion or reabsorption throughout the renal tubules. Water retention at the kidney has a direct impact on blood pressure. In most mammals, reptiles and birds, the kidney's nephrons are initially encapsulated by a cluster of capillaries—the glomeruli, which filter blood plasma and then extend into a long tubule with various segments (Singh & Dyce 2018). The proximal tubule of the kidneys largely undertakes passive transfer of water and soluble molecules, while the distal tubule uses active processes for the filtration of remaining water and other soluble molecules. It is in this distal tubule where genes such as *Aqp1* are expressed when activated by the renin–angiotensin system (Takei et al. 1993; Kobayashi & Takei 1996). As reptiles

lack a loop of Henle (Holz 2020), they are incapable of producing hypertonic urine and a larger level of reabsorption must thus be achieved or water loss limited, in other areas of the kidney tubule (Holz 2020). The capacity for these animals to change expression levels and the solute transporter function in this region is therefore directly related to their homeostasis, blood pressure and filtering of waste, and could control the amount of water reabsorbed based on water availability. These genes may, therefore, be crucial to a lizard's ability to conserve water throughout the Australian dry season.

Other genes identified with significant differential gene expression have less explicit roles in the kidney and may not be directly involved in water or solute transport. Both *mcm3* and *tk1* were downregulated in autumn, and are known to play a role in cell proliferation. The former is involved in regulating DNA replication (Shinya et al. 2014) while the latter is an enzyme involved in DNA repair—specifically through phosphorylation of deoxythymidine and deoxyuridine (Konrad et al. 2014)—and by extension, cell repair (Bitter et al. 2020). The final downregulated gene identified in the genes of interest table is *chrna4*, which is commonly associated with synapses but also highly expressed in the human kidney and liver (Zhang et al. 2017) and has a large number of orthologues that are conserved across species (NCBI Resource Coordinators 2016). This gene plays a role in binding acetylcholine and activating acetylcholine-mediated ion channel activity (Zhang et al. 2017). Acetylcholine is the main transmitter of the parasympathetic nervous system (Klein & Cunningham 2021), which, if present (as implied by the presence of its receptor) in the context of the kidney would cause receptor activity and activation of acetylcholine-gated cation selective ion channels, allowing ions to cross the plasma membrane in response to membrane depolarisation response or osmolarity. The downregulation of these genes suggests a lower level of signalling activity in the *T. adelaidensis* kidney at the time of collection in autumn.

Analysis of GO over-representation was carried out separately for upregulated and downregulated genes, similar to the analysis by Griffith et al. (2016) where over-representation was calculated for each tissue group. The upregulated genes did not show any overdispersion, but this was a short list of 19 upregulated genes that were assigned a BLASTx hit. The genes downregulated in autumn had no significantly over-represented GO categories. Included in these are many involved in regulation, metabolic process and specific processes clearly involved in renal function, such as 'positive regulation of nitrogen compound metabolic process' and 'negative regulation of nitrogen compound metabolic process'.

Of all the genes found to be significantly upregulated in autumn, only *gja1* is present on the genes of interest list generated in Chapter 3. This gene, also known as Cx43 is one of a group of connexin proteins that form gap junctions in endothelial cells and vasculature of many organs including kidney renal vasculature, allowing intercellular communication (Saez et al. 2003). Downregulation of this gene in rodents has been studied because of its implications for cancers and diseases such as diabetes. Inhibition of this and related genes results in reduced vasodilation in renal vessels as well as reduced blood flow, increased blood pressure and increased renin secretion (Hanner et al. 2010). The upregulation of this gene signals the release of renin in other vertebrates and triggers the RAAS (Roger, Boutin & Chadjichristos 2022). The released renin metabolises angiotensinogen to angiotensin I, subsequently converted by angiotensin-converting enzyme into angiotensin II, which then acts on receptors causing sodium retention, vasoconstriction, increased sympathetic nervous system activity and the release of aldosterone—this is the RAAS response (Ames et al. 2019). Cx43 has been shown to be differentially expressed in the medullary vasculature and interstitium of rats, and more direct intracellular signalling causes vessel contraction through 'electrical syncytium', propagating action potentials (Xu et al. 2021). It is possible that Cx43 is signalling a similar response in *T. adelaidensis* to reduce renal fluid loss. However, renin release in the first instance may be triggered not only by low blood pressure or dehydration, but through other sympathetic stimulation, and it is possible that capture and handling of these animals during processing contributed to triggering of this response (Ames et al. 2019).

Snake nephrons have six regions starting at the Bowman's capsule in the renal glomerulus: the neck, the proximal convoluted tubule, the distal convoluted tubule, the sexual segment of the kidney, the post terminal region and the collecting ducts (Siegel, Sever & Aldridge 2010). A secondary reproductive function of the kidney has been observed in three vertebrate taxa: squamates, salamanders and stickleback fish (Aldridge et al. 2011). It appears to have evolved independently in these taxa (Siegel, Sever & Aldridge 2010; Aldridge et al. 2011). Well studied in snakes, it most commonly involves the distal convoluted tubules of the kidney, which enlarge in diameter on transition into the sexual segment as well as collecting ducts in some species such as *Leptotyphlops dulcis* (Sever et al. 2012), and the ureter of some lizard taxa (Rheubert et al. 2011). Sections of the renal nephron serve as a secondary sexual organ, which is identifiable through histological analysis of the nephron, involving secretory epithelia that secrete proteins and neutral carbohydrates (Siegel, Sever & Aldridge 2010; Aldridge et al. 2011; Sever et al. 2012). These secretions mix with sperm in a ductus deferens–ureter complex, enter the cloaca together through

urogenital papillae and are transferred to the female during mating (Rheubert, Siegel & Trauth 2015). Thus it has been suggested that the secretions serve a role in sperm health and motility (Rheubert et al. 2011).

The sexual segment of the kidney in squamates is androgen dependent and undergoes hypertrophy and regression on a seasonal and hormonal basis in males (Prasad & Reddy 1972; Aldridge et al. 2011; Rheubert, Siegel & Trauth 2015). A hypertrophied version of this structure is present in females and can be stimulated with hormones to grow (Aldridge et al. 2011). The sexual segment of the kidney may be identified in the inactive season in squamates, though the seasonal hypertrophy recorded for snakes (Sever et al. 2012) has also been observed in lizards and the sexual segment may be easily identified histologically by the tall columnar epithelium with secretory granules at the apices of all epithelial cells (Rheubert et al. 2011). Specific histological analysis of the sexual segment of the kidney of *T. adelaidensis* and closely related species has not been conducted. However a morphological tree for sexual segment of the kidney's location in various reptile species shows skinks possess sexual functions in the collecting ducts and ureter (Rheubert et al. 2011), providing similar secretory functions with similar composition of carbohydrate and protein (Samson et al. 2021). Studies of hypertrophy of the sexual segment of the kidney in the active seasons of other species have shown peak hypertrophy in males during spring (Aldridge et al. 2011). If the Australian pygmy bluetongue skink were to follow this pattern, hypertrophy would be expected in the spring (September) male samples. The sample that shows the greatest variation in gene expression activity (G5; Figure 4-2) is such a sample; however with the data collected here it is not known whether unintentional sequencing of all or part of the sexual segment of the kidney in G5 is the cause of these differences. There are also too few replicates to determine whether the differences between G5 and other samples is unique or would form a pattern involving other males. No attempt was made to avoid particular areas of the kidney during sampling or tissue processing. A potential way to isolate gene expression in the sexual segment of the kidney when it cannot be grossly identified, as is sometimes possible in snakes (Aldridge et al. 2011), may be to conduct single-cell analysis (scRNA-seq) of renal tubular epithelial cells, allowing the identification of gene expression and putative function for particular cell types (Hasegawa 2022). Determining whether specific genes or changes in expression occur only in specific regions of the kidney may be a way to identify the presence and function of the sexual segment of the kidney at a molecular level.

There is a large amount of observable variation in expression in sample G5 (a spring/September male) in both the heatmap and MDS plot. This may suggest that sample G5 is responsible for a large amount of the variation observed in differential gene expression calculated between experimental groups. In the MDS plot, samples from the same season, while not tightly clustered, were closer to each other than to other samples. There are a number of possibilities for gene expression differences at the individual level, including individual variation, variation between years of sampling, or different activity or metabolism requirements of males at this time, and the real reason cannot be narrowed down with the samples and data available. Sample G5 is from the only male collected in spring after removal of sample G6, and there may be basic biological differences between sexes in addition to the effect of the sexual segment of the kidney discussed above. An exploration of differential expression using sex as the comparable factor is presented in the Appendix 2 data analysis document and does show that a number of genes are significantly more upregulated in males. Male pygmy bluetongue lizards are also known to disperse further (Schofield et al. 2012) and emerge in spring to begin looking for a mate (Milne, Bull & Hutchinson 2003) and are therefore subjected to different environmental stressors and exposure where a larger impact on females might be expected due to reproductive stress potentially later in the season. While there are indications that sample G5 may also be the source of some of the overall skew in the data towards upregulated genes, a brief comparison of involving only the two females in each group, while offering little power for interpretation, does show that this skew towards upregulated genes remains when sample G5 is removed from the dataset.

The Australian pygmy bluetongue lizard spends the cooler months (March–September) in burrows in a state of brumation, and during the hotter dry season (September–March) individuals are more active (Ebrahimi & Bull 2014). Sample collection occurred in spring 2014, autumn 2015, spring 2015 and autumn 2016. These years were marked by a strong El Niño climate event driving hotter and drier conditions (Bureau of Meteorology 2022). There is variation in rainfall among these years, with 2015 being the driest overall and drier than the average of all years between 1962 and 2021. The springs of 2015 and 2016 followed a much drier late wet season compared to the monthly mean rainfall for the same years. This may have introduced variation into the results at the individual level due to the small number of samples, and may make them difficult to reproduce, particularly if one were to compare them with results for the extremely wet springs of 2020–22 influenced by La Niña. It is thus important to remember that these data represent an example of gene expression responses of *T. adelaidensis* over a hotter and drier spring and summer.

This study has focused on kidney tissue to measure gene expression in *T. adelaidensis* renal tissue during periods of hydration stress; however, as reptiles, *T. adelaidensis* are uricotelic animals and their liver plays a fundamental role in the production of uric acid to excrete nitrogenous waste (Mora et al. 1965). This prevents excessive water loss via liquid urine, which is not the primary means of secreting waste because of the species' inability to concentrate urine beyond the osmolarity of its blood plasma (Singh & Dyce 2018). Hepatic enzymes like those well studied in mammalian vertebrates have been identified in a number of reptiles. For example, the three most common liver enzymes used in biochemistry panels (Latiwesh et al. 2018) have been identified in the eastern box tortoise (*Terrapene carolina carolina*); they are alanine aminotransferase, aspartate aminotransferase and alkaline phosphatase (Adamovicz et al. 2019). However, tissue-specific presentation of isoenzymes varies greatly and has even been shown to be a result of tissue-specific alternative splicing in the production of glutamine synthetase in *Squalus acanthias,* a ureosmotic elasmobranch (Matthews et al. 2005). The study by Matthews et al. (2005) identified that the different glutamine synthetase mitochondrial enzyme is also present in uricotelic species.

Exploration of gene expression in *T. adelaidensis* livers is a logical next step in exploring the secretion of nitrogenous waste and the role the liver plays in supporting renal function and promoting water retention. This would not only supplement the gene expression analysis conducted here but may provide new insight into alternative splicing in uricotelic squamates. The inability to extract good-quality RNA samples from collected liver tissues prevented their use in this present study and limits the conclusions that can be drawn about overall nitrogenous waste excretion. However, ample tissue samples remain at the South Australian Museum for future analysis.

Many of the genes identified in this chapter directly or indirectly influence cell membrane permeability, ion channel activity or glucose transport, and play a role in maintaining homeostasis. Downregulated genes are just as informative as upregulated genes, and the upregulated *gja1* may prevent excess fluid loss in renal endothelium or vesicles by preventing vasodilation. All of the identified genes play a role in maintenance of homeostasis in these lizards during a time when they are increasing metabolic and physical activity, and when their environment in South Australia becomes hotter and drier.

4.6 References

Aceto, S, Sica, M, De Paolo, S, D'Argenio, V, Cantiello, P, Salvatore, F & Gaudio, L 2014, 'The Analysis of the Inflorescence miRNome of the Orchid Orchis italica Reveals a DEF-Like MADS-Box Gene as a New miRNA Target', Plos One, vol. 9, no. 5, p. 12.

Adamovicz, L, Griffioen, J, Cerreta, A, Lewbart, GA & Allender, MC 2019, 'Tissue Enzyme Activities In Free-Living Eastern Box Turtles (*Terrapene carolina Carolina*)', *Journal of Zoo and Wildlife Medicine*, vol. 50, no. 1, pp. 45-54.

Allaire, J., Horner, J., Xie, Y., Marti, V. & Porte, N. 2019, 'markdown: Render Markdown with the C Library 'Sundown'. R package version 1.1.', https://CRAN.R-project.org/package=markdown

Ames, MK, Atkins, CE & Pitt, B 2019, 'The renin-angiotensin-aldosterone system and its suppression', *J Vet Intern Med*, vol. 33, no. 2, pp. 363-82.

Babonis, LS, Miller, SN & Evans, DH 2011, 'Renal responses to salinity change in snakes with and without salt glands', *Journal of Experimental Biology*, vol. 214, no. 13, pp. 2140-56.

Barnier, J. 2021, 'rmdformats: HTML Output Formats and Templates for 'rmarkdown' Documents. R package version 1.0.3.', https://CRAN.R-project.org/package=rmdformats

Bentley, BP, Haas, BJ, Tedeschi, JN & Berry, O 2017, 'Loggerhead sea turtle embryos (*Caretta caretta*) regulate expression of stress response and developmental genes when exposed to a biologically realistic heat stress', *Mol Ecol*, vol. 26, no. 11, pp. 2978-92.

Bitter, EE, Townsend, MH, Erickson, R, Allen, C & O'Neill, KL 2020, 'Thymidine kinase 1 through the ages: a comprehensive review', *Cell & Bioscience*, vol. 10, no. 1, p. 138.

Brandley, MC, Young, RL, Warren, DL, Thompson, MB & Wagner, GP 2012, 'Uterine Gene Expression in the Live-Bearing Lizard, *Chalcides ocellatus*, Reveals Convergence of Squamate Reptile and Mammalian Pregnancy Mechanisms', *Genome Biology and Evolution*, vol. 4, no. 3, pp. 394-411.

Bray, NL, Pimentel, H, Melsted, P & Pachter, L 2016, 'Near-optimal probabilistic RNA-seq quantification (vol 34, pg 525, 2016)', *Nature Biotechnology*, vol. 34, no. 8, pp. 888-.

Bureau of Meteorology 2022, 'ENSO Outlook', Australian Government, Bureau of Meteorology, http://www.bom.gov.au/climate/enso/outlook/ [Accessed 21 Mar. 2022].

Chen, Y, Lun, A & Smyth, G 2016, 'From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline [version 2; peer review: 5 approved]', *F1000Research*, vol. 5, no. 1438.

Chen, Y, Lun, AT & Smyth, GK 2016, 'From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline', *F1000Res*, vol. 5, p. 1438.

Chow, JC, Kyritsis, N, Mills, M, Godfrey, MH, Harms, CA, Anderson, PE & Shedlock, AM 2021, 'Tissue and Temperature-Specific RNA-Seq Analysis Reveals Genomic Versatility and Adaptive Potential in Wild Sea Turtle Hatchlings (*Caretta caretta*)', *Animals*, vol. 11, no. 11.

Clusella-Trullas, S, Blackburn, TM & Chown, SL 2011, 'Climatic predictors of temperature performance curve parameters in ectotherms imply complex responses to climate change', *Am Nat*, vol. 177, no. 6, pp. 738-51.

Dantzler, WH 2003, 'Regulation of renal proximal and distal tubule transport: sodium, chloride and organic anions', *Comp Biochem Physiol A Mol Integr Physiol*, vol. 136, no. 3, pp. 453-78.

Das, D, Singh, SK, Bierstedt, J, Erickson, A, Galli, GLJ, Crossley, DA & Rhen, T 2020, 'Draft Genome of the Common Snapping Turtle, *Chelydra serpentina*, a Model for Phenotypic Plasticity in Reptiles', *G3-Genes Genomes Genetics*, vol. 10, no. 12, pp. 4299-314.

Doneley, B, Monks, D, Johnson, R & Carmel, B 2018, *Reptile Medicine and Surgery in Clinical Practice*, Wiley Blackwell, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA.

Ebrahimi, M & Bull, CM 2014, 'Short-Term Dispersal Response of an Endangered Australian Lizard Varies with Time of Year', *Plos One*, vol. 9, no. 8, p. 7.

Fassbinder-Orth, CA 2014, 'Methods for Quantifying Gene Expression in Ecoimmunology: From qPCR to RNA-Seq', *Integrative and Comparative Biology*, vol. 54, no. 3, pp. 396-406.

Feiner, N, Rago, A, While, GM & Uller, T 2018, 'Developmental plasticity in reptiles: Insights from temperature-dependent gene expression in wall lizard embryos', *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, vol. 329, no. 6-7, pp. 351-61.

Figueroa, XF & Duling, BR 2009, 'Gap junctions in the control of vascular function', *Antioxid Redox Signal*, vol. 11, no. 2, pp. 251-66.

Fleming, A, Wibbels, T, Skipper, JK & Crews, D 1999, 'Developmental expression of steroidogenic factor 1 in a turtle with temperature-dependent sex determination', *General and Comparative Endocrinology*, vol. 116, no. 3, pp. 336-46.

Fulgione, D, Trapanese, M, Maselli, V, Rippa, D, Itri, F, Avallone, B, Van Damme, R, Monti, DM & Raia, P 2014, 'Seeing through the skin: dermal light sensitivity provides cryptism in moorish gecko', *Journal of Zoology*, vol. 294, no. 2, pp. 122-8.

Gagolewski, M. 2021, 'stringi: Fast and portable character string processing in R_. R package version 1.6.1.', https://stringi.gagolewski.com/

Geyle, HM, Tingley, R, Amey, AP, Cogger, H, Couper, PJ, Cowan, M, Craig, MD, Doughty, P, Driscoll, DA & Ellis, RJ 2020, 'Reptiles on the brink: identifying the Australian terrestrial snake and lizard species most at risk of extinction', *Pacific Conservation Biology*, vol. 27, no. 1, pp. 3-12.

Griffith, OW, Brandley, MC, Belov, K & Thompson, MB 2016, 'Reptile Pregnancy Is Underpinned by Complex Changes in Uterine Gene Expression: A Comparative Analysis of the Uterine Transcriptome in Viviparous and Oviparous Lizards', *Genome Biology and Evolution*, vol. 8, no. 10, pp. 3226-39.

Hanner, F, Sorensen, CM, Holstein-Rathlou, N-H & Peti-Peterdi, J 2010, 'Connexins and the kidney', *American journal of physiology. Regulatory, integrative and comparative physiology*, vol. 298, no. 5, pp. R1143-R55.

Harry, JL, Williams, KL & Briscoe, DA 1990, 'Sex determination in loggerhead turtles: Differential expression of two hnRNP proteins', *Development*, vol. 109, no. 2, pp. 305-12.

Hofmann, S, Kuhl, H, Baniya, CB & Stöck, M 2019, 'Multi-Tissue Transcriptomes Yield Information on High-Altitude Adaptation and Sex-Determination in *Scutiger cf. sikimmensis*', *Genes*, vol. 10, no. 11, p. 13.

Höglund, PJ, Nordström, KJV, Schiöth, HB & Fredriksson, R 2011, 'The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species', *Molecular biology and evolution*, vol. 28, no. 4, pp. 1531-41.

Holz, P 2020, 'Anatomy and Physiology of the Reptile Renal System', *The veterinary clinics of North America. Exotic animal practice*, vol. 23, pp. 103-14.

IPCC 2022, 'Climate Change 2022: Impacts, Adaptation, and Vulnerability' Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig,

S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)], *Cambridge University Press*, Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844

Jin, YT, Brandt, DYC, Li, JS, Wo, YB, Tong, HJ & Shchur, V 2021, 'Elevation as a selective force on mitochondrial respiratory chain complexes of the Phrynocephalus lizards in the Tibetan plateau', *Current Zoology*, vol. 67, no. 2, pp. 191-9.

Junqueira-de-Azevedo, IL, Campos, PF, Ching, AT & Mackessy, SP 2016, 'Colubrid Venom Composition: An -Omics Perspective', *Toxins*, vol. 8, no. 8.

Klein, BG & Cunningham, JG 2021, *Cunningham's Textbook of Veterinary Physiology*, 6th ed. edn, W.B. Saunders.

Konrad, A, Lai, J, Mutahir, Z, Piškur, J & Liberles, DA 2014, 'The Phylogenetic Distribution and Evolution of Enzymes Within the Thymidine Kinase 2-like Gene Family in Metazoa', *Journal of Molecular Evolution*, vol. 78, no. 3, pp. 202-16.

Kordonowy, L & MacManes, M 2017, 'Characterizing the reproductive transcriptomic correlates of acute dehydration in males in the desert-adapted rodent, *Peromyscus eremicus*', *Bmc Genomics*, vol. 18, no. 1, p. 473.

Latiwesh, OB, Younis, MYG, Shakila, S, Abdulmalik, F, Alammari, JA, Min, Y, Hussain, A & Peela, JR 2018, 'Hepatic Enzymes Changes In Chronic Kidney Disease Patients- A Need For Modified Reference Values', *Journal of Evolution of Medical and Dental Sciences-Jemds*, vol. 7, no. 16, pp. 1949-54.

Lewis, S, Chen, L, Raghuram, V, Khundmiri, SJ, Chou, CL, Yang, CR & Knepper, MA 2021, '"SLC-omics" of the kidney: solute transporters along the nephron', *Am J Physiol Cell Physiol*, vol. 321, no. 3, pp. C507-c18.

Matthews, GD, Gould, RM & Vardimon, L 2005, 'A single glutamine synthetase gene produces tissue-specific subcellular localization by alternative splicing', *Febs Letters*, vol. 579, no. 25, pp. 5527-34.

McCarthy, DJ, Chen, Y & Smyth, GK 2012, 'Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation', *Nucleic Acids Research*, vol. 40, no. 10, pp. 4288-97.

McLean, CA, Lutz, A, Rankin, KJ, Elliott, A, Moussalli, A & Stuart-Fox, D 2019, 'Red carotenoids and associated gene expression explain colour variation in frillneck lizards', *Proceedings of the Royal Society B-Biological Sciences*, vol. 286, no. 1907, p. 9.

Milne, T, Bull, CM & Hutchinson, MN 2003, 'Use of burrows by the endangered pygmy blue-tongue lizard, *Tiliqua adelaidensis* (Scincidae)', *Wildlife Research*, vol. 30, no. 5, pp. 523-8.

Mora, J, Martuscelli, J, Ortiz Pineda, J & Soberon, G 1965, 'The Regulation Of Urea-Biosynthesis Enzymes In Vertebrates', *Biochem J*, vol. 96, no. 1, pp. 28-35.

NCBI Resource Coordinators, 2016, 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Research*, vol. 44, no. D1, pp. D7-D19.

Neuwirth, E. 2014, 'RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.', https://CRAN.R-project.org/package=RColorBrewer

Nielsen, S, Chou, CL, Marples, D, Christensen, EI, Kishore, BK & Knepper, MA 1995, 'Vasopressin increases water permeability of kidney collecting duct by inducing translocation of aquaporin-CD water channels to plasma membrane', *Proc Natl Acad Sci U S A*, vol. 92, no. 4, pp. 1013-7.

Phipson, B., Trigos, A., Ritchie, M., Su, S., Doyle, M., Dashnow, H. & Law, C. n.d., 'RNA-seq analysis in R - Differential expression analysis', *Combine Australia Workshop*, https://combine-australia.github.io/RNAseq-R/06-rnaseq-day1.html#Data_files_and_Resources

R Core Team 2019, 'R: A language and environment for statistical computing. R Foundation for Statistical Computing', *Vienna, Austria,* https://www.R-project.org/

R Studio Team 2020, 'RStudio: Integrated Development for R. RStudio', *PBC, Boston, MA.,* http://www.rstudio.com/

Raghavan, V, Kraft, L, Mesny, F & Rigerte, L 2022, 'A simple guide to de novo transcriptome assembly and annotation', Briefings in Bioinformatics, vol. 23, no. 2, p. bbab563.

Razgour, O, Forester, B, Taggart, JB, Bekaert, M, Juste, J, Ibanez, C, Puechmaille, SJ, Novella-Fernandez, R, Alberdi, A & Manel, S 2019, 'Considering adaptive genetic variation in climate change vulnerability assessment reduces species range loss projections', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 21, pp. 10418-23.

Ren, Y, Tseng, E, Smith, TPL, Hiendleder, S, Williams, JL & Low, WY 2023, 'Long read isoform sequencing reveals hidden transcriptional complexity between cattle subspecies', Bmc Genomics, vol. 24, no. 1, p. 108.

Robinson, MD, McCarthy, DJ & Smyth, GK 2010, 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, vol. 26, no. 1, pp. 139-40.

Roger, E, Boutin, L & Chadjichristos, CE 2022, 'The Role of Connexin 43 in Renal Disease: Insights from In Vivo Models of Experimental Nephropathy', *International Journal of Molecular Sciences*, vol. 23, no. 21, p. 13090.

Saez, JC, Berthoud, VM, Branes, MC, Martinez, AD & Beyer, EC 2003, 'Plasma membrane channels formed by connexins: their regulation and functions', *Physiol Rev*, vol. 83, no. 4, pp. 1359-400.

Schaller, L & Lauschke, VM 2019, 'The genetic landscape of the human solute carrier (SLC) transporter superfamily', *Human genetics*, vol. 138, no. 11-12, pp. 1359-77.

Schlessinger, A, Yee, SW, Sali, A & Giacomini, KM 2013, 'SLC classification: an update', *Clinical pharmacology and therapeutics*, vol. 94, no. 1, pp. 19-23.

Schofield, JA, Fenner, AL, Pelgrim, K & Bull, CM 2012, 'Male-biased movement in pygmy bluetongue lizards: implications for conservation', *Wildlife Research*, vol. 39, no. 8, pp. 677-84.

Shinya, M, Machiki, D, Henrich, T, Kubota, Y, Takisawa, H & Mimura, S 2014, 'Evolutionary diversification of MCM3 genes in *Xenopus laevis* and *Danio rerio*', *Cell Cycle*, vol. 13, no. 20, pp. 3271-81.

Singh, B & Dyce, KM 2018, *Dyce, Sack, and Wensing's textbook of veterinary anatomy*, Fifth edition. edn, Textbook of veterinary anatomy, Elsevier, St. Louis, Missouri.

Sjaastad, OV, Hove, K & Sand, O 2016, *Physiology of domestic animals*, Third edition. edn, Scandinavian Veterinary Press, Oslo, Norway.

Soneson, C, Matthes, KL, Nowicka, M, Law, CW & Robinson, MD 2016, 'Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage', Genome Biology, vol. 17, no. 1, p. 12.

Tan, CH & Tan, KY 2021, 'De novo venom-gland transcriptomics of spine-bellied sea snake (*Hydrophis curtus*) from penang, malaysia—next-generation sequencing, functional annotation and toxinological correlation', *Toxins*, vol. 13, no. 2.

Tedeschi, JN, Kennington, WJ, Berry, O, Whiting, S, Meekan, M & Mitchell, NJ 2015, 'Increased expression of Hsp70 and Hsp90 mRNA as biomarkers of thermal stress in loggerhead turtle embryos (*Caretta Caretta*)', *J Therm Biol*, vol. 47, pp. 42-50.

Tedeschi, JN, Kennington, WJ, Tomkins, JL, Berry, O, Whiting, S, Meekan, MG & Mitchell, NJ 2016, 'Heritable variation in heat shock gene expression: a potential mechanism for adaptation to thermal stress in embryos of sea turtles', *Proceedings of the Royal Society B: Biological Sciences*, vol. 283, no. 1822, p. 20152320.

Todd, EV, Black, MA & Gemmell, NJ 2016, 'The power and promise of RNA-seq in ecology and evolution', *Mol Ecol*, vol. 25, no. 6, pp. 1224-41.

Torres-Maldonado, L, Moreno-Mendoza, N, Landa, A & Merchant-Larios, H 2001, 'Timing of SOX9 downregulation and female sex determination in gonads of the sea turtle *Lepidochelys olivacea*', *Journal of Experimental Zoology*, vol. 290, no. 5, pp. 498-503.

The UniProt Consortium 2018, 'UniProt: the universal protein knowledgebase', Nucleic Acids Research, vol. 46, no. 5, p. 2699.

The UniProt Consortium 2021, 'UniProt: the universal protein knowledgebase in 2021', Nucleic Acids Research, vol. 49, no. D1, pp. D480-D489.

The UniProt Consortium 2023, 'UniProt: the Universal Protein Knowledgebase in 2023', Nucleic Acids Research, vol. 51, no. D1, pp. D523-D531.

Van den Berge, K, Hembach, KM, Soneson, C, Tiberi, S, Clement, L, Love, MI, Patro, R & Robinson, MD 2019, 'RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis', *Annual Review of Biomedical Data Science*, vol. 2, no. 1, pp. 139-73.

Viala, VL, Hildebrand, D, Trusch, M, Fucase, TM, Sciani, JM, Pimenta, DC, Arni, RK, Schlüter, H, Betzel, C, Mirtschin, P, Dunstan, N & Spencer, PJ 2015, 'Venomics of the Australian eastern brown snake (*Pseudonaja textilis*): Detection of new venom proteins and splicing variants', *Toxicon*, vol. 107, pp. 252-65.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M. & Venables, B. 2020, 'gplots: Various R

Programming Tools for Plotting Data. R package version 3.1.1.', https://CRAN.R-project.org/package=gplots

Wickham, H. 2019, 'stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0.', https://CRAN.R-project.org/package=stringr

Wickham, H., François, R., Henry, L. & Müller, K. 2021, 'dplyr: A Grammar of Data Manipulation. R package version 1.0.6.', https://CRAN.R-project.org/package=dplyr

Williams, TD, Gensberg, K, Minchin, SD & Chipman, JK 2003, 'A DNA expression array to detect toxic stress response in European flounder (*Platichthys flesus*)', *Aquatic Toxicology*, vol. 65, no. 2, pp. 141-57.

Xu, C, Dolby, GA, Drake, KK, Esque, TC & Kusumi, K 2020, 'Immune and sex-biased gene expression in the threatened Mojave desert tortoise, *Gopherus agassizii*', *Plos One*, vol. 15, no. 8.

Xu, Y, Hu, J, Yilmaz, DE & Bachmann, S 2021, 'Connexin43 is differentially distributed within renal vasculature and mediates profibrotic differentiation in medullary fibroblasts', *American Journal of Physiology-Renal Physiology*, vol. 320, no. 1, pp. F17-F30.

Xie, Y. 2021, 'formatR: Format R Code Automatically. R package version 1.11.', https://CRAN.R-project.org/package=formatR

Xiong, Y & Lei, FM 2021, '*SLC2A12* of SLC2 Gene Family in Bird Provides Functional Compensation for the Loss of *SLC2A4* Gene in Other Vertebrates', Molecular biology and evolution, vol. 38, no. 4, pp. 1276-91.

Zhang, B, Madden, P, Gu, J, Xing, X, Sankar, S, Flynn, J, Kroll, K & Wang, T 2017, 'Uncovering the transcriptomic and epigenomic landscape of nicotinic receptor genes in non-neuronal tissues', *Bmc Genomics*, vol. 18, no. 1, p. 439.

5. Thesis Discussion

Tiliqua adelaidensis returning to natural spider burrow - C. Maher

5.1 Discussion

This thesis describes the generation of annotated kidney mRNA reference transcripts using full-length transcript sequencing, and an analysis of gene expression of these genes using short-read sequencing technologies, to identify transcripts expressed in renal tissue of *T. adelaidensis* and identify genes under differential expression in different seasons. The overarching aim was to identify genes that may play a role in the pygmy bluetongue's ability to survive in a semi-arid habitat, providing a valuable genomic reference in a threatened skink. The primary objectives achieved in this thesis and the chapters in which they are addressed are:

1. Sequence and assemble transcripts from RNA isolated from *T. adelaidensis* kidney tissues (Chapter 2 & 3).
2. Compare the transcripts generated through *de novo* assembly of short-read RNA-seq and long-read Iso-Seq techniques and discuss these methods in the context of *T. adelaidensis* gene annotation (Chapter 3).
3. Conduct a literature and gene database search to establish a reference list of genes that have been identified in other reptiles and may be of interest regarding kidney function (Chapter 3; Appendix 4).
4. Annotate assembled transcripts from RNA isolated from *T. adelaidensis* kidney tissues to identify putative genes present that may play a role in kidney function, including homeostasis and water retention (Chapter 3).
5. Conduct a differential gene expression analysis to compare genes expressed in different seasons (early spring [September] and autumn [March/April]) using kidney tissue of *T. adelaidensis* (Chapter 4).

These objectives were achieved through comparison of RNA sequenced from six Australian pygmy bluetongue lizard kidney tissue samples selected for poly(A) mRNA and sequenced using both short-read (Illumina HiSeq) and long-read (Pacific Biosciences Iso-Seq) technologies to generate reference transcripts and perform an expression analysis. I generated a set of reference transcript sequences that may be used as references for further functional analyses of *T. adelaidensis* RNA (Chapter 3). Expression was compared between two sampling periods: early Austral spring (September) following the wet part of the year (two males and two females) and the beginning of the Austral autumn (March and April) at the end of the dry part of the year (one male and three females). The summer period between these two collection periods, as well as being the dry season, also coincides with this species' breeding season and is when they are most active. The integration of literature and gene database searches with annotated genes and genes with significant differential expression allowed identification of transcripts and putative genes that have

been implicated in renal processes in other species and may provide insight into a species' ability to respond to climate change, which may be of interest in future conservation efforts.

Rather than relying on *de novo* assembly methods, the long-read clustered putative gene set from Chapter 3 was annotated, generating the reference for expression counts in Chapter 4 so that expression counts could be identified to putative genes, enabling better inferences of biological function to be made. This is not the first instance of combining methods to ensure accuracy of results (Roscito et al. 2018) and has facilitated development of new transcript tools for *T. adelaidensis*. The individuals and tissues collected in this study provide invaluable genomic resources for future studies of this threatened species.

Both long- and short-read sequencing were completed on these kidney tissues. The long reads were used to identify and annotate putative genes from the canonical transcripts generated, and these were used as a reference to estimate gene counts and quantify seasonal differential gene expression using short-read data. *De novo* assembly of transcripts is an ongoing challenge for biologists, with new solutions and tools frequently released. This compounds an already complicated process with enumerable options and approaches that must be tailored to the goal of a study. A recent review by Raghavan et al. (2022) also outlines many of the available tools and their purpose for RNA-seq analysis when *de novo* assembly is required. This thesis outlines one possible pipeline for an Australian skink to analyse potential genes that may impact the species' ability to survive dry periods in an arid habitat and to make inferences about genes of importance when assessing the impact of climate change or translocations on populations.

5.1.1 Transcript Assembly Comparison

A comparison of completeness for short-read *de novo* assembled transcripts and full-length long-read transcripts was done using BUSCO scores (Manni et al. 2021) of highly conserved genes present in the assemblies. This showed that slightly fewer, but a very similar proportion of complete BUSCOs were identified in G1, G2, G3, G4, G5 and G8 (the first short-read sequencing run) compared to the long-read sequences. All six of these samples each had the highest number of fragmented BUSCOs compared to other transcriptomes, not surprising for a *de novo* assembly using Trinity (Rana et al. 2016). Interestingly G3, G4 and G5, the short-reads with the lowest BUSCO scores, had the highest RIN scores among these six samples prior to sequencing (Appendix 1, Table 6-1). As expected for deeper read depths, despite the shorter initial insert size (Lighten, van

Oosterhout & Bentzen 2014), samples G6 and G7 had much higher BUSCO completeness scores than any other samples, and a lower proportion of fragmented BUSCOs than the other short-read samples, and it is their inclusion that has increased the combined Trinity assembly's completeness score. The depth of this sequencing may have allowed shorter or rare transcripts to become more amplified in these samples.

The long-read samples, while detecting marginally higher BUSCOs than the initial short-read sequencing run, were outperformed by the deeper Trinity assembly on completeness. However, nearly all BUSCOs detected in these samples were complete, and comparison of the stages of analysis—'(LR) High Quality Iso-Seq Transcripts'; the predicted open read frames '(LR) Unique Isoform CDS'; clustering; and pulling out the full-length transcript sequence for each cluster of predicted coding region '(LR) Clustered Reference Transcripts'—offers a unique way to observe the effects of these steps. The initial sample '(LR) High Quality Iso-Seq Transcripts' has a high level of BUSCO duplication; this file still contains isoforms of the same transcripts, which is to be expected. The next step '(LR) Unique Isoform CDS' shows a decrease in complete BUSCOs and an increase in fragmented hits. As this file only contained predicted coding regions with the untranslated flanking regions removed and kept in a separate file, this is also as expected. The next step is where final clustering has been completed. The file '(LR) Clustered Reference Transcripts' contains a full-length transcript for each of retained coding sequences, which is reflected in the BUSCO result by less fragmented matches in this transcript set, but also a lower level of duplicated BUSCOs—also expected now that the putative genes have been clustered.

In the absence of a genome, using long-read sequencing as a supplement to short-read sequencing methods provides more guidance in assemblies. For example, a combination of 74X Illumina short-read and Pacific Biosciences long-read technologies was used to generate a highly complete genome for *S. merianae* (Roscito et al. 2018). When generating transcripts, long-read data have similar benefits and increase completeness of full-length transcripts, as well as providing greater potential for gene and isoform identification (Grabherr et al. 2011; Haas et al. 2013). While the challenges of differentiating similar transcript isoforms is most evident in the study of many cereal crops, with the added complication of polyploidy (Gruenheit et al. 2012), it is still an important consideration here. Full-length transcripts generated using long-read data may then be used to align short-read sequencing to generate gene expression counts. Using both these

technologies provides each with a secondary data source that can be used for comparison and validation.

In the absence of long-read data the assembly of *de novo* assembled short-read RNA-seq containing all sample data may be used as a 'pseudo reference', for example. The duplication evident through the mapping of duplicated BUSCOs and the potential for transcript splice isoforms must be accounted for similarly to how the long-read set was clustered into putative genes in this study. A number of tools have been developed that compute isoform quantification (Zhang et al. 2017) and count expression levels for similar or overlapping reference transcript sequences and splice isoforms. These include Salmon's equivalence classes (Patro et al. 2017); multi-mapping of reads in bowtie (Langmead & Salzberg 2012), which can then be clustered and counted by programs such as Corset (Davidson & Oshlack 2014); and the construction of 'super transcripts' as the longest reference transcript for a set of exons on a gene (Davidson, Hawkins & Oshlack 2017).

5.1.2 Annotated Genes

I produced a dataset of 13,882 unique, non-redundant transcripts, which were clustered into a list of 9,813 representative putative loci or genes. A number of these genes were identified as having functions relevant to the research aims, with 393 on the genes of interest list represented by 955 *T. adelaidensis* transcripts (Appendix 3, Table 6-6). Genes with the most transcript annotations included a number of aquaporins, and an extremely large number of solute carriers, notably *slc25a45*, *slc2a11*, *slc34a1* and *slc7a3*, that had ten or more transcripts that annotated to that gene (Appendix 3).

Other genes with a large number (>10) of transcripts that produced a BLAST hit for the corresponding protein ID include aldehyde dehydrogenase, a catalyst for the oxidation of aldehydes; aldolase fructose biphosphate B and C, involved in several processes including response to insulin; argininosuccinate synthase-1 (*ass1*); and calbindin-1 (*calb1*), which enables calcium-binding activity but also regulates calcium ion concentration and absorption, and has been shown to have expression changes in cattle under endoplasmic reticulum stress (Chiappisi et al. 2017); collectrin (*cltrn*), another transmembrane amino acid transporter; HSPs (notably *hsp90ab1* had the most matching transcripts); and prolyl 4-hydroxylase subunit beta (*p4hb*), which is involved in immune response and activity of interleukin-7 but has ubiquitous expression in a number of tissues and has been shown to be a biomarker of poor prognosis in the kidneys of cancer patients due to

its much higher presence in unhealthy kidneys (Xie et al. 2020). The differentially expressed genes identified in Chapter 4 form important references confirming the presence of these gene orthologues in *T. adelaidensis* and any variation in sequence and isoforms found.

5.1.3 Differentially Expressed Genes

The common dispersion factor calculated for all genes across seasonal period was not unusual, with an expected value of 0.05–0.20 for cloned cell lines or inbred mice and up to 0.30 in humans (Chen, Lun & Smyth 2016). Individual differentially expressed genes in the dataset do show a much larger range of tagwise dispersions, but this indicates that a not unreasonable skew in the data despite the appearance of a bias towards upregulated genes in spring.

Three genes identified as solute carriers were identified as significantly downregulated in Austral autumn, and these were discussed in detail in Chapter 4. Two other genes were identified in the genes of interest list and not only showed as significantly upregulated in spring, but were among the top 25 significantly upregulated genes in spring from among 157 upregulated genes. These were *mcm3* and *tk1*. The only upregulated gene also identified on the genes of interest list was *gja1*, which plays a structural role in protein gap junctions and the renin–angiotensin response (Figueroa & Duling 2009). Upregulation of this gene in the context of renal tissue would cause a release of renin signalment and affect the interstitial pressure at the glomerulus of the kidney, affecting the exchange of plasma from blood vessels into the renal tubules (Roger, Boutin & Chadjichristos 2022). Differential expression in both upwards and downwards directions is equally important when considering biological consequences. It is possible that regulation of this gene implies a similar function to that seen in birds. Birds have been shown to use the circulatory portal system to control differences in hydrostatic pressure or bypass the glomerulus in the kidney; for example, broad-tailed hummingbirds cease renal filtration while resting so do not lose water (Bakken et al. 2004). It is possible that pygmy bluetongues are also able to cease renal function. This may also explain the lower expression of many other genes identified as having roles in cellular and renal activity in favour of uric acid production in the liver (Sjaastad, Hove & Sand 2016; Vigetti et al. 2002) or water reabsorption in the rectal cloaca. This has been observed in other Australian reptiles (Braysher 1976) when water becomes scarce, as would have been the case in the current study for the autumn (March) collection period at the end of the dry season.

In this thesis, the focus has been on genes that showed a significant difference in gene expression or were identified in *T. adelaidensis*, and on the genes of interest list, flagging them as genes that likely play a role in kidney function and water homeostasis. Genes that met either category for assessment were analysed and, similar to results produced by Marra et al. (2012), combining the approaches to summarising the data (using a list of expected genes of interest, and basing analysis on the most differentially expressed genes) provided successful for identifying genes relevant to kidney function and water retention, including solute carrier genes. This guided the analysis to focus on important genes whose functions were then researched in more detail.

5.1.4 Limitations

First and foremost, this study is limited by the nature of the study species, *T. adelaidensis*. As this is a threatened and protected native species, permits and ethics were required for collection of a minimal number of samples; two males and two females from each proposed experimental group (spring and autumn). This cryptic and solitary species is not only difficult to find (Milne, Bull & Hutchinson 2003), but also does not have significant sexual dimorphism. Sex was not confirmed until dissection while tissue sampling; therefore the groups were not balanced as planned, with two males and two females in spring, and one male and three females in autumn. This slightly lower representation of the males must be considered when interpreting results, as sex-biased gene expression—particularly when driven by hormone transmitters—has been demonstrated in *Anolis sagrei* (Cox et al. 2017) and *Megalobrama amblycephal* (Lv et al. 2021), and shown to result in differences in immune gene expression responses in the latter. While eight individuals were originally collected, sequencing batch effects resulted in the final gene expression analysis in Chapter 4 being conducted on n = 6 individuals, in experimental groups of n = 3. When analysed in EdgeR, this amount of data has been shown to be sufficient (as a minimum) to identify differentially expressed genes (Schurch et al. 2016).

The main limitation to this project design results from the length of the project and timing of original planning. The kits purchased in 2015 and 2016 for short-read mRNA sample preparation were not designed to amplify stranded RNA and cDNA libraries. Therefore G1, G2, G3, G4, G5 and G8 were all prepared as non-stranded libraries, while G6 and G7—prepared and sequenced at a later date—were stranded. This is addressed in Chapter 3 by treating all pooled samples as non-stranded, as per Trinity's user manual (Haas et al. 2013). In the final analysis and gene expression conducted in Chapter 4, all short reads were aligned by Kallisto directly after cleaning to full

transcripts of putative genes generated using long-read technology; strand-specific information was not supplied as the analysis was run in paired-end mode. Note that G1, G2, G3, G4, G5 and G8 were all sequenced together on an Illumina HiSeq and library prep was done with the same kit. Due to time constraints and logistical factors in the lab, samples G6 and G7 were prepared in a separate batch using a different kit and sequenced on an Illumina HiSeq with fewer samples (Chapter 3); they had a much higher read depth compared to other short-read samples. These two samples are from a spring male and an autumn female. Further exploration of the implications of the two resulting sequencing batches is presented in Appendix 2. Unfortunately, the effects of these two samples could not be normalised and they were ultimately excluded, though data summaries are provided in Appendix 2 for comparison. The removal of G6 and G7 from the dataset resulted in two balanced groups of three per season, each with two females and one male. The BUSCO comparison in Chapter 3 shows that greater completeness was gained from short- than from long-read sequencing and that a hypothetical read depth of 15,000 150PE reads per sample performed as well as the Pacific Biosciences Iso-Seq for expected BUSCO sequence discovery. Unless there is a particular interest in transcript discovery or rare transcripts, this level of sequencing depth will likely be sufficient for future studies and—particularly with the increasing capacities of newer sequencing machines—will allow all samples in a study to be sequenced at the same time and thus avoid potential batch effects.

5.1.5 Implications and Future Directions

The focus on kidney tissue in this study was tailored to the research question in relation to renal function and water homeostasis during different seasons in this species. Other tissue samples that could have supplemented the analysis by potentially generating a larger pool of annotated genes through capturing sequences expressed in a variety of tissues could not be included because of repeated extraction failures (Appendix 1 Table 6-1) and ultimately, time and budget limitations. Liver samples were prioritised for this because of the liver's role in the production of uric acid (Sjaastad, Hove & Sand 2016) and likely role in allowing solid waste to be excreted separately to the kidneys, avoiding water loss. The other tissue samples collected remain available for this possibility to be explored in the future. While not explored here due to various financial and time constraints, there is potential for future gene annotations or gene expression analyses to be completed on genes expressed in other tissues in *T. adelaidensis*, and for this purpose liver tissues should be a priority.

At sample collection, as many tissue samples as possible were taken and remain in storage at –80°C at the SARFMEE and South Australian Museum. These tissues include lung, stomach, tongue, heart, hindgut, liver, epididymis, eye, vomeronasal, brain, muscle, whole frozen blood and testes/ovary as relevant, as well as the remaining kidney tissue. There is also blood on FTA paper samples, scat and cloacal swab samples and blood smears. Although collection was completed in 2015–17, it is expected that these tissues will be viable for some time if they remain frozen. Andreasson et al. (2013) has shown that extracted RNA from tissues stored long term at –80°C does not show a significant drop in quality. This study focused on water retention, but an exploration of liver tissue would be an important next step in gaining a greater understanding of how *T. adelaidensis* removes urates and other waste from the system, given that reptiles produce uric acid in the liver for excretion via the cloaca in solid form to compensate for not having the ability to concentrate urine like mammals (Singh & Dyce 2018). Sequencing RNA from brain, testes and ovaries as well as brain tissue may also provide insights into endocrine regulatory factors for a number of these biological processes.

A large portion of this research was focused on gene discovery and building the tools to perform the experimental gene expression comparison. A more curated approach to validating individual gene annotations could be undertaken on these sequences in the future, particularly once more Australian reptile genomes become available for comparison through the AusARG (2020) targeting Australian species of conservation interest. Genes that were not able to be annotated here could be further explored, along with teasing out specific isoforms that annotated to the same putative gene, and their specific functions (Amin et al. 2014; Livernois et al. 2016). Full-length reads have also been used previously to improve genome assemblies (Ali, Thorgaard & Salem 2021).

The genes identified in this study add another potential source from which to identify orthologues. Primers from closely related species have already been proven to successfully amplify across species in *Egernia* and *Tiliqua* genera (Ansari, Talat Hojat et al. 2014). Of the 33 papers investigated in greater detail in the literature review in Appendix 4.2 based on their specific focus on differential gene expression in a squamate species, only one article returned in the search involved a specific focus on the effect of an environmental stimulus on gene expression in the kidney of a squamate (Trinchella et al. 2006). The variable measured in that study was cadmium exposure to developing embryos of the Italian wall lizard, *Podarcis sicula*. Further data mining and analysis of these results has the potential to contribute to population-level analyses and genetic studies on this species

long into the future. The data obtained in this study may also allow the development of primers that can be used to detect identified genes in other members of the population without requiring whole tissue samples from further animals.

Genetic diversity must be considered in conservation, especially when assessing the specific impact climate change will have on a threatened species (Pauls et al. 2013). Lizards have been shown to have a 'thermal type' and preference for the temperature ranges and activity level to which they are adapted (Goulet, Thompson & Chapple 2017). Genomic analysis in relation to these traits is crucial for understanding a species' ability to adapt to changing conditions. Functional genes, and whether changes in expression can play a role in adaptation in the shorter term, may become the key to survival for some species as the environmental conditions change faster than species can evolve or disperse in response to conditions (Bernatchez 2016). More information about a species' biology and responses to increasing pressures being placed on individual survival is required to fully assess its vulnerability to climate change, and ultimately its conservation status (Foden et al. 2019). Including an assessment of population genomic health and diversity (Ralls et al. 2020), as well as predicted capacity for adaptation to changing conditions in the short or long term, is becoming increasingly important to consider for conservation efforts that aim to mitigate the impact of climate change on species (Pauls et al. 2013). There is strong evidence that considering adaptive genetic variation alongside other ecological models has considerable effects on future range model predictions (Razgour et al. 2019) and thus has the capacity to influence conservation decisions, highlighting the importance of increasing genomic resources for threatened species to supplement our understanding of their biology and better inform conservation decisions.

Transcriptomic information is a valuable resource even in species with a draft genome published, such as the fathead minnow (*Pimephales promelas*) for which transcriptome and proteome data were used to supplement genomic information, build tissue-specific transcript profiles and identify cellular pathways enriched during homeostasis for reference in future studies (Lavelle et al. 2018). With the sexual segment of the kidney of many squamates (and potentially that of *T. adelaidensis,* if it is present) not as grossly evident as in some snake species during seasonal heteroplasia (Aldridge et al. 2011), scRNA-seq could provide the potential for a deeper analysis of specific cell types within the renal tissue and correlation of gene expression data to spatial coordinate data within tissue (Chen et al. 2021). Although cost and tissue availability remain limiting factors, identification of the sexual segment of the kidney and different sections of nephrons through spatial transcriptomic

techniques and inference based on their expression profiles would provide a more detailed picture of cellular structure (Shen et al. 2022; Williams et al. 2022).

Although *T. adelaidensis* is one of the most studied reptiles in South Australia, few comprehensive studies have examined the role of functional genes in the species (Appendix 4.1, Figure 6-10 & Figure 6-11). Genetic studies on neutral markers have been used in trial translocation and population augmentation studies (Clive et al. 2020); to determine paternity of offspring during studies; and to identify promiscuous mating and multiple paternity (Schofield et al. 2014). However, the sole study on functional genes conducted on *T. adelaidensis* was an exploration of the applicability of major histocompatibility complex immune gene primers developed for *T. rugosa* (Ansari, T. H. et al. 2015). Understanding the role of functional genes in a species, the polymorphisms present in different populations, and whether expression of these genes is affected by environmental conditions, as I have done here, is a crucial addition to our knowledge when developing translocation proposals and planning conservation efforts (Allendorf, Hohenlohe & Luikart 2010; Hogg et al. 2022; Kohn et al. 2006; Ouborg et al. 2010).

5.2 Summary and Conclusion

Here I have generated a reference list of canonical full-length transcripts (Chapter 3) annotated to one of 6,346 identified unique genes based on the protein ID and GO analysis (Chapter 3). This list of reference genes guided analysis of differential expression (Chapter 4), which may play a role in the pygmy bluetongue's ability to survive in a semi-arid habitat. The genes identified in this thesis—particularly those shown to have differential expression during the period of sampling—may also play a broader role in short-term adaptation and plasticity as this species is forced to adapt to increasing temperature and rainfall/drought extremes. The bioinformatic pipeline developed in this thesis (Appendix 2) has already contributed to guiding and streamlining the analysis protocol for gene expression studies in the related sleepy lizard (*T. rugosa*; O'Reilly, in preparation). This thesis contributes new scientific knowledge of genomics in an endangered Australian skink to assist conservation efforts in response to climate change as a key threatening process and provides a guide for RNA-seq studies in a novel species.

5.3 References

Ali, A, Thorgaard, GHH & Salem, M 2021, 'PacBio Iso-Seq Improves the Rainbow Trout Genome Annotation and Identifies Alternative Splicing Associated With Economically Important Phenotypes', *Frontiers in Genetics*, vol. 12, pp. 683408-.

Allendorf, FW, Hohenlohe, PA & Luikart, G 2010, 'Genomics and the future of conservation genetics', *Nat Rev Genet*, vol. 11, no. 10, pp. 697-709.

Amin, NM, Tandon, P, Nishimura, EO & Conlon, FL 2014, 'RNA-seq in the tetraploid *Xenopus laevis* enables genome-wide insight in a classic developmental biology model organism', *Methods*, vol. 66, no. 3, pp. 398-409.

Andreasson, A, Kiss, NB, Juhlin, CC & Hoog, A 2013, 'Long-Term Storage of Endocrine Tissues at-80 degrees C Does Not Adversely Affect RNA Quality or Overall Histomorphology', *Biopreservation and Biobanking*, vol. 11, no. 6, pp. 366-70.

Ansari, TH, Bertozzi, Miller, RD & Gardner, MG 2015, 'MHC in a monogamous lizard - Characterization of class I MHC genes in the Australian skink *Tiliqua rugosa*', *Developmental and Comparative Immunology*, vol. 53, no. 2, pp. 320-7.

Ansari, TH, Bertozzi, T, Hacking, J, Cooper, SJB & Gardner, MG 2014, 'Random non-coding fragments of lizard DNA: anonymous nuclear loci for the Australian skink, *Tiliqua rugosa*, and their utility in other Egernia-group species', *Australian Journal of Zoology*, vol. 62, no. 6.

Bakken, BH, McWhorter, TJ, Tsahar, E & Del Rio, CM 2004, 'Hummingbirds arrest their kidneys at night: diel variation in glomerular filtration rate in *Selasphorus platycercus*', *J Exp Biol*, vol. 207, no. Pt 25, pp. 4383-91.

Bernatchez, L 2016, 'On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishesa', *Journal of Fish Biology*, vol. 89, no. 6, pp. 2519-56.

Braysher, ML 1976, 'The excretion of hyperosmotic urine and other aspects of the electrolyte balance of the lizard Amphibolurus maculosus', *Comparative Biochemistry and Physiology Part A: Physiology*, vol. 54, no. 3, pp. 341-5.

Chen, YW, Qian, WZ, Li, L, Cai, LF, Yin, K, Jiang, SW, Song, J, Han, RPS & Yang, CY 2021, 'Mapping Gene Expression in the Spatial Dimension', Small Methods, vol. 5, no. 11, p. 21.

Chen, Y, Lun, AT & Smyth, GK 2016, 'From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline', *F1000Res*, vol. 5, p. 1438.

Chiappisi, E, Ringseis, R, Eder, K & Gessner, DK 2017, 'Effect of endoplasmic reticulum stress on metabolic and stress signaling and kidney-specific functions in Madin-Darby bovine kidney cells', *Journal of Dairy Science*, vol. 100, no. 8, pp. 6689-706.

Clive, LFR, Gardner, MG, Clayton, J, Baring, R, Hutchinson, MN, Fenner, A & Bull, CM 2020, 'Population augmentation had no effect on the abundance or body condition of conspecifics and co-occurring lizard species in a native grassland community', *Austral Ecology*, vol. 45, no. 4, pp. 478-87.

Cox, RM, Cox, CL, McGlothlin, JW, Card, DC, Andrew, AL & Castoe, TA 2017, 'Hormonally Mediated Increases in Sex-Biased Gene Expression Accompany the Breakdown of Between-Sex Genetic Correlations in a Sexually Dimorphic Lizard', *Am Nat*, vol. 189, no. 3, pp. 315-32.

Davidson, NM, Hawkins, ADK & Oshlack, A 2017, 'SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes', *Genome Biology*, vol. 18, p. 10.

Davidson, NM & Oshlack, A 2014, 'Corset: enabling differential gene expression analysis for de novo assembled transcriptomes', *Genome Biol*, vol. 15, no. 7, p. 410.

Figueroa, XF & Duling, BR 2009, 'Gap junctions in the control of vascular function', *Antioxid Redox Signal*, vol. 11, no. 2, pp. 251-66.

Goulet, CT, Thompson, MB & Chapple, DG 2017, 'Repeatability and correlation of physiological traits: Do ectotherms have a "thermal type"?', *Ecol Evol*, vol. 7, no. 2, pp. 710-9.

Grabherr, MG, Haas, BJ, Yassour, M, Levin, JZ, Thompson, DA, Amit, I, Adiconis, X, Fan, L, Raychowdhury, R, Zeng, Q, Chen, Z, Mauceli, E, Hacohen, N, Gnirke, A, Rhind, N, di Palma, F, Birren, BW, Nusbaum, C, Lindblad-Toh, K, Friedman, N & Regev, A 2011, 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nat Biotechnol*, vol. 29, no. 7, pp. 644-52.

Gruenheit, N, Deusch, O, Esser, C, Becker, M, Voelckel, C & Lockhart, P 2012, 'Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants', *Bmc Genomics*, vol. 13, p. 19.

Haas, BJ, Papanicolaou, A, Yassour, M, Grabherr, M, Blood, PD, Bowden, J, Couger, MB, Eccles, D, Li, B, Lieber, M, MacManes, MD, Ott, M, Orvis, J, Pochet, N, Strozzi, F, Weeks, N, Westerman, R, William, T, Dewey, CN, Henschel, R, Leduc, RD, Friedman, N & Regev, A 2013, 'De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis', *Nature Protocols*, vol. 8, no. 8, pp. 1494-512.

Hogg, C, J., Ottewell, K, Latch, P, Rossetto, M, Biggs, J, Gilbert, A, Richmond, S & Belov, K 2022, 'Threatened Species Initiative: Empowering conservation action using genomic resources', *Proceedings of the National Academy of Sciences*, vol. 119, no. 4, p. e2115643118.

Kohn, MH, Murphy, WJ, Ostrander, EA & Wayne, RK 2006, 'Genomics and conservation genetics', *Trends Ecol Evol*, vol. 21, no. 11, pp. 629-37.

Langmead, B & Salzberg, SL 2012, 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, vol. 9, no. 4, pp. 357-9.

Lavelle, C, Smith, LC, Bisesi, JH, Vu, FH, Silva-Sanchez, C, Moraga-Amador, D, Buerger, AN, Garcia-Reyero, N, Sabo-Attwood, T & Denslow, ND 2018, 'Tissue-Based Mapping of the Fathead Minnow (Pimephales promelas) Transcriptome and Proteome', Frontiers in Endocrinology, vol. 9, p. 18.

Lighten, J, van Oosterhout, C & Bentzen, P 2014, 'Critical review of NGS analyses for de novo genotyping multigene families', *Mol Ecol*, vol. 23, no. 16, pp. 3957-72.

Livernois, A, Hardy, K, Domaschenz, R, Papanicolaou, A, Georges, A, Sarre, SD, Rao, S, Ezaz, T & Deakin, JE 2016, 'Identification of interleukin genes in *Pogona vitticeps* using a de novo transcriptome assembly from RNA-seq data', *Immunogenetics*, vol. 68, no. 9, pp. 719-31.

Lv, M, Chen, X, Huang, X, Liu, N, Wang, W & Liu, H 2021, 'Transcriptome Analysis Reveals Sexual Disparities between Olfactory and Immune Gene Expression in the Olfactory Epithelium of *Megalobrama amblycephala*', *International Journal of Molecular Sciences*, vol. 22, no. 23, p. 13017.

Marra, NJ, Eo, SH, Hale, MC, Waser, PM & DeWoody, JA 2012, 'A priori and a posteriori approaches for finding genes of evolutionary interest in non-model species: Osmoregulatory genes in the kidney transcriptome of the desert rodent *Dipodomys spectabilis* (banner-tailed kangaroo rat)', *Comparative Biochemistry and Physiology D-Genomics & Proteomics*, vol. 7, no. 4, pp. 328-39.

Milne, T, Bull, CM & Hutchinson, MN 2003, 'Use of burrows by the endangered pygmy blue-tongue lizard, *Tiliqua adelaidensis* (Scincidae)', *Wildlife Research*, vol. 30, no. 5, pp. 523-8.

O'Reilly, R., Maher, C., Bertozzi, T. & Gardner, M. G. 2022, 'Differential gene expression analysis of a non-model organism (*Tiliqua rugosa*) suggests upregulation of adaptive immune genes with the bobtail flu', Manuscript in preparation.

Ouborg, NJ, Pertoldi, C, Loeschcke, V, Bijlsma, RK & Hedrick, PW 2010, 'Conservation genetics in transition to conservation genomics', *Trends Genet*, vol. 26, no. 4, pp. 177-87.

Patro, R, Duggal, G, Love, MI, Irizarry, RA & Kingsford, C 2017, 'Salmon provides fast and bias-aware quantification of transcript expression', *Nature Methods*, vol. 14, no. 4, pp. 417-9.

Pauls, SU, Nowak, C, Balint, M & Pfenninger, M 2013, 'The impact of global climate change on genetic diversity within populations and species', *Mol Ecol*, vol. 22, no. 4, pp. 925-46.

Raghavan, V, Kraft, L, Mesny, F & Rigerte, L 2022, 'A simple guide to de novo transcriptome assembly and annotation', *Briefings in Bioinformatics*, vol. 23, no. 2, p. bbab563.

Rana, SB, Zadlock, FJ, Zhang, ZP, Murphy, WR & Bentivegna, CS 2016, 'Comparison of De Novo Transcriptome Assemblers and k-mer Strategies Using the Killifish, *Fundulus heteroclitus*', *Plos One*, vol. 11, no. 4, p. 16.

Roger, E, Boutin, L & Chadjichristos, CE 2022, 'The Role of Connexin 43 in Renal Disease: Insights from In Vivo Models of Experimental Nephropathy', *International Journal of Molecular Sciences*, vol. 23, no. 21, p. 13090.

Roscito, JG, Sameith, K, Pippel, M, Francoijs, KJ, Winkler, S, Dahl, A, Papoutsoglou, G, Myers, G & Hiller, M 2018, 'The genome of the tegu lizard *Salvator merianae*: combining Illumina, PacBio, and optical mapping data to generate a highly contiguous assembly', *Gigascience*, vol. 7, no. 12, p. 13.

Schofield, JA, Gardner, MG, Fenner, AL & Bull, CM 2014, 'Promiscuous mating in the endangered Australian lizard *Tiliqua adelaidensis*: a potential windfall for its conservation', *Conservation Genetics*, vol. 15, no. 1, pp. 177-85.

Schurch, NJ, Schofield, P, Gierliński, M, Cole, C, Sherstnev, A, Singh, V, Wrobel, N, Gharbi, K, Simpson, GG, Owen-Hughes, T, Blaxter, M & Barton, GJ 2016, 'How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?', *RNA*, vol. 22, no. 6, pp. 839-51.

Shen, XH, Zhao, YC, Wang, Z & Shi, QH 2022, 'Recent advances in high-throughput single-cell transcriptomics and spatial transcriptomics', Lab on a Chip, vol. 22, no. 24, pp. 4774-91.

Singh, B & Dyce, KM 2018, *Dyce, Sack, and Wensing's textbook of veterinary anatomy*, Fifth edition. edn, Textbook of veterinary anatomy, Elsevier, St. Louis, Missouri.

Sjaastad, OV, Hove, K & Sand, O 2016, *Physiology of domestic animals*, Third edition. edn, Scandinavian Veterinary Press, Oslo, Norway.

The Australian Amphibians and Reptile Genomics initiative 2020, https://doi.org/10.25953/xer9-2e12

Trinchella, F, Riggio, M, Filosa, S, Volpe, MG, Parisi, E & Scudiero, R 2006, 'Cadmium distribution and metallothionein expression in lizard tissues following acute and chronic cadmium intoxication', *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology*, vol. 144, no. 3, pp. 272-8.

Vigetti, D, Pollegioni, L, Monetti, C, Prati, M, Bernardini, G & Gornati, R 2002, 'Property comparison of recombinant amphibian and mammalian allantoicases', *Febs Letters*, vol. 512, no. 1-3, pp. 323-8.

Williams, CG, Lee, HJ, Asatsuma, T, Vento-Tormo, R & Haque, A 2022, 'An introduction to spatial transcriptomics for biomedical research', Genome Medicine, vol. 14, no. 1, p. 68.

Xie, L, Li, H, Zhang, L, Ma, X, Dang, Y, Guo, J, Liu, J, Ge, L, Nan, F, Dong, H, Yan, Z & Guo, X 2020, 'Autophagy-related gene P4HB: a novel diagnosis and prognosis marker for kidney renal clear cell carcinoma', *Aging (Albany NY)*, vol. 12, no. 2, pp. 1828-42.

Zhang, C, Zhang, B, Lin, L-L & Zhao, S 2017, 'Evaluation and comparison of computational tools for RNA-seq isoform quantification', *Bmc Genomics*, vol. 18, no. 1, p. 583.

6. Appendices

Pseudonaja textilis attempting predation *Tiliqua adelaidensis* in burrow – C. Maher

6.1 Appendix 1: General

6.1.1 Quality Assessment Summary

Table 6-1: Summary Table of quality scoring using an Agilent Bioanalyser for total RNA extractions before poly(A) selection using the Qiagen RNeasy plus kit, from various *Tiliqua adelaidensis* tissue samples. Sample IDs highlighted in yellow are the final sequenced samples.

Sample ID #	Tissue	Date extracted	Date on report	Assay Class	Sample tube name	RIN	RNA Area	Concentration (ng/µL)	Ratio 28s/18s	µg RNA eluted
G1	Kidney	7/11/2016	23/11/2018	Eucaryote Total RNA Nano	G1-Kidney	5.9	78.2	52	0.6	1.56
G2	Kidney	14/11/2016	23/11/2018	Eucaryote Total RNA Nano	G2-Kidney	6.7	174.3	116	0.5	3.48
G3	Kidney	14/11/2016	23/11/2018	Eucaryote Total RNA Nano	G3-Kidney	7	675.3	451	0.6	13.53
G4	Kidney	14/11/2016	23/11/2018	Eucaryote Total RNA Nano	G4-Kidney	7.1	192.4	128	3.4	3.84
G5	Kidney	7/11/2016	23/11/2018	Eucaryote Total RNA Nano	G5-Kidney	7	299.6	200	0.7	6
G8	Kidney	14/11/2016	23/11/2018	Eucaryote Total RNA Nano	G8-Kidney	6.8	347.2	232	0.6	6.96
G4	Spleen	6/03/2018	18/04/2018	Eucaryote Total RNA Pico	G4-Spleen	n/a	5.8	0.11	0	0.0033
G6	Kidney	6/03/2018	18/04/2018	Eucaryote Total RNA Pico	G6-Kidney	4.5	3515.7	66.654	0.3	1.99962
G7	Kidney	6/03/2018	18/04/2018	Eucaryote Total RNA Pico	G7-Kidney	6.9	64.8	1.229	0.3	0.03687
G5	Liver	1/05/2018	7/05/2018	Eucaryote Total RNA Nano	G5-Liver-a	7.1	480.5	270	0.8	10.8
G5	Liver	1/05/2018	7/05/2018	Eucaryote Total RNA Nano	G5-Liver-b	n/a	1.8	1	0	0.04
G8	Liver	1/05/2018	7/05/2018	Eucaryote Total RNA Nano	G8-Liver-a	7.7	25.3	14	0.9	0.56
G8	Liver	1/05/2018	7/05/2018	Eucaryote Total RNA Nano	G8-Liver-b	7.6	119.3	67	0.7	2.68
G2	Liver	5/06/2018	14/06/2018	Eucaryote Total RNA Nano	G2-Liver	7.8	2006	543	0.9	21.72
G4	Liver	5/06/2018	14/06/2018	Eucaryote Total RNA Nano	G4-Liver	n/a	3	1	1.2	0.04
G6	Kidney	5/06/2018	14/06/2018	Eucaryote Total RNA Nano	G6-Kidney	8.6	1958.4	530	1.1	15.9
G7	Kidney	5/06/2018	14/06/2018	Eucaryote Total RNA Nano	G7-Kidney	7.9	566.7	153	0.9	4.59
G4	Liver	24/07/2018	27/07/2018	Eucaryote Total RNA Nano	G4-Liver	1	10.6	17	0	0.68
G8	Liver	24/07/2018	27/07/2018	Eucaryote Total RNA Nano	G8-Liver	n/a	4.2	7	1.4	0.28
G4	Liver	30/07/2018	10/08/2018	Eucaryote Total RNA Nano	G4-Liver-a	n/a	4.7	3	0	0.12
G4	Liver	30/07/2018	10/08/2018	Eucaryote Total RNA Nano	G4-Liver-b	n/a	5.8	4	0	0.16
G5	Liver	30/07/2018	10/08/2018	Eucaryote Total RNA Nano	G5-Liver-a	n/a	11.2	7	0	0.28
G5	Liver	30/07/2018	10/08/2018	Eucaryote Total RNA Nano	G5-Liver-b	n/a	5.1	3	0	0.12
G8	Liver	30/07/2018	10/08/2018	Eucaryote Total RNA Nano	G8-Liver-a	7.8	127.8	80	0.7	3.2
G8	Liver	30/07/2018	10/08/2018	Eucaryote Total RNA Nano	G8-Liver-b	8	96.5	60	0.7	2.4
G4	Liver	21/08/2018	23/08/2018	Eucaryote Total RNA Nano	G4-Liver-a	n/a	1.4	1	0	0.04
G4	Liver	21/08/2018	23/08/2018	Eucaryote Total RNA Nano	G4-Liver-b	n/a	5.5	4	0	0.16
G3	Liver	21/08/2018	23/08/2018	Eucaryote Total RNA Nano	G3-Liver	n/a	5	4	0	0.16
G4	Liver	29/08/2018	14/09/2018	Eucaryote Total RNA Nano	G4-Liver-a	n/a	14.3	6	0	0.24
G4	Liver	29/08/2018	14/09/2018	Eucaryote Total RNA Nano	G4-Liver-b	n/a	20.6	9	0	0.36
G3	Liver	24/09/2018	9/10/2018	Eucaryote Total RNA Nano	G3-Liver	n/a	3.4	2	0	0.08
G4	Liver	24/09/2018	9/10/2018	Eucaryote Total RNA Nano	G4-Liver	n/a	7.7	5	0	0.2
G6	Liver	24/09/2018	9/10/2018	Eucaryote Total RNA Nano	G6-Liver	n/a	1.1	1	0	0.04
G7	Liver	24/09/2018	9/10/2018	Eucaryote Total RNA Nano	G7-Liver	n/a	13	9	0.3	0.36
G6	Testes	21/01/2019	24/01/2019	Eucaryote Total RNA Nano	G6-Testes-a	9.1	604.3	220	1.2	6.6
G6	Testes	21/01/2019	24/01/2019	Eucaryote Total RNA Nano	G6-Testes-b	9.3	281	102	1.2	3.06

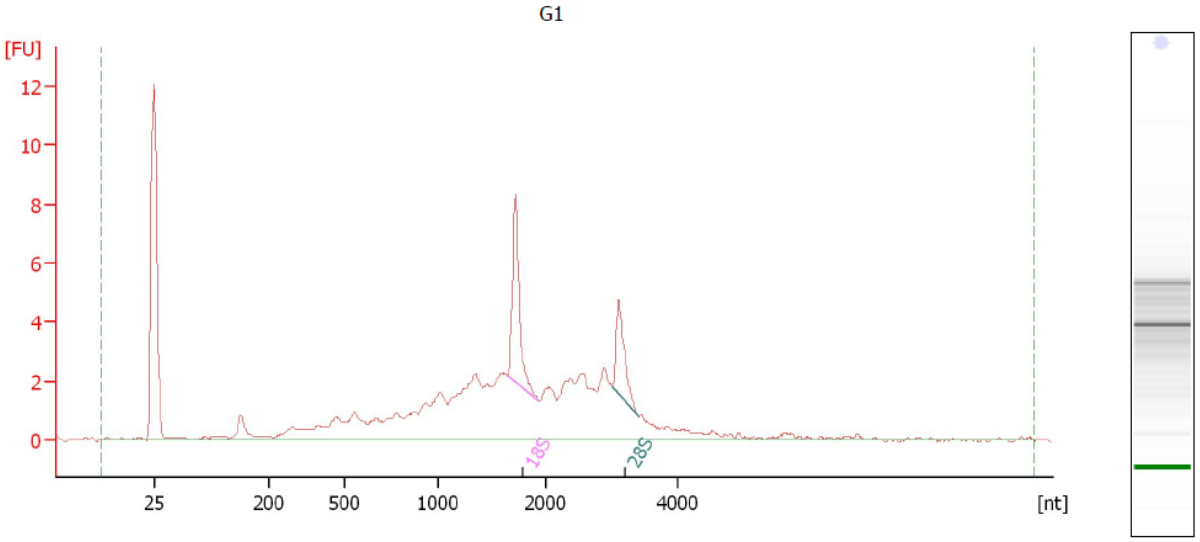
6.1.2 Bioanalyser Reports

All reports below (Figure 6-1; Figure 6-2; Figure 6-3; Figure 6-4; Figure 6-5; Figure 6-6; Figure 6-7; Figure 6-8) were run by SA Pathology on an Agilent Bioanalyser, Eukaryote Total RNA Nano assay class 2100 expert, version 2.6 with ng sensitivity (Copyright 2003–2009 Agilent Technologies, Inc.)

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE54704589_2018-11-19_17-20-05.xad

Created: 11/19/2018 5:20:05 PM
Modified: 11/23/2018 11:19:44 AM

Electropherogram Summary Continued ...



Overall Results for sample 1 : G1

RNA Area:	78.2	RNA Integrity Number (RIN):	5.9 (B.02.08)
RNA Concentration:	52 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.6	Result Flagging Label:	RIN: 5.90

Fragment table for sample 1 : G1

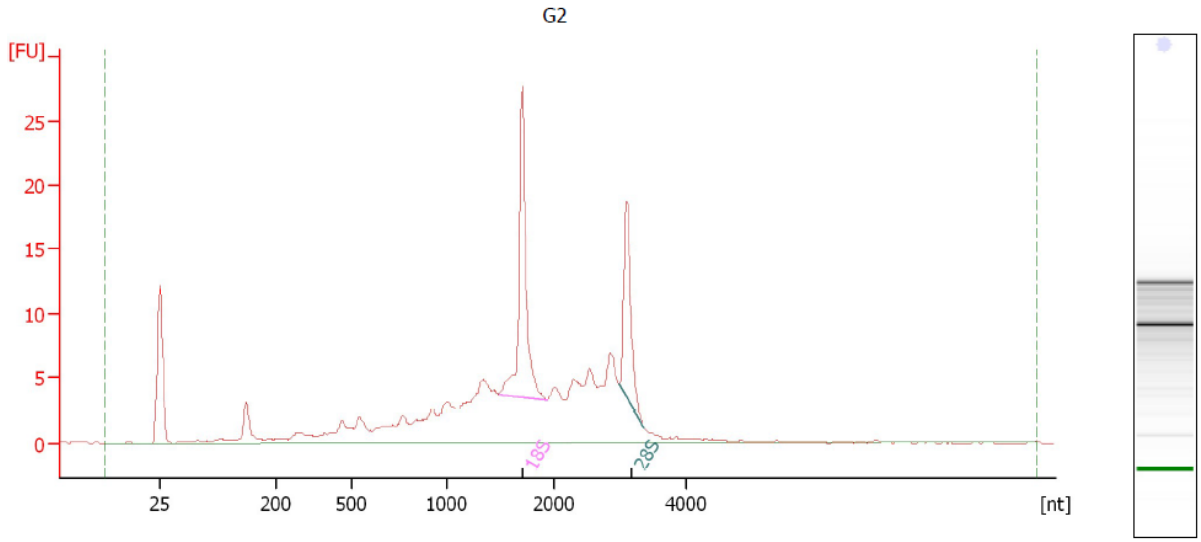
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,651	1,931	6.2	8.0
28S	3,004	3,405	3.4	4.4

Figure 6-1: Bioanalyser report for total RNA extraction of sample G1 *Tiliqua adelaidensis* kidney tissue.

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE54704589_2018-11-19_17-20-05.xad

Created: 11/19/2018 5:20:05 PM
Modified: 11/23/2018 11:19:44 AM

Electropherogram Summary Continued ...



Overall Results for sample 2 : G2

RNA Area:	174.3	RNA Integrity Number (RIN):	6.7 (B.02.08)
RNA Concentration:	116 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.5	Result Flagging Label:	RIN: 6.70

Fragment table for sample 2 : G2

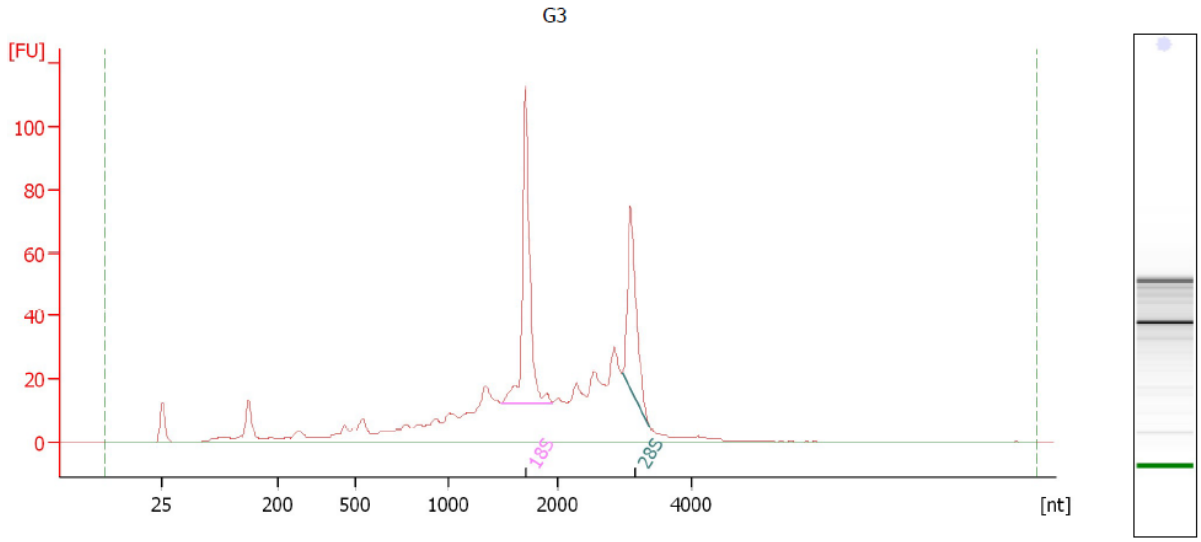
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,490	1,924	26.3	15.1
28S	2,971	3,356	13.6	7.8

Figure 6-2: Bioanalyser report for total RNA extraction of sample G2 *Tiliqua adelaidensis* kidney tissue.

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE54704589_2018-11-19_17-20-05.xad

Created: 11/19/2018 5:20:05 PM
Modified: 11/23/2018 11:19:44 AM

Electropherogram Summary Continued ...



Overall Results for sample 3 : G3

RNA Area:	675.3	RNA Integrity Number (RIN):	7 (B.02.08)
RNA Concentration:	451 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.6	Result Flagging Label:	RIN:7

Fragment table for sample 3 : G3

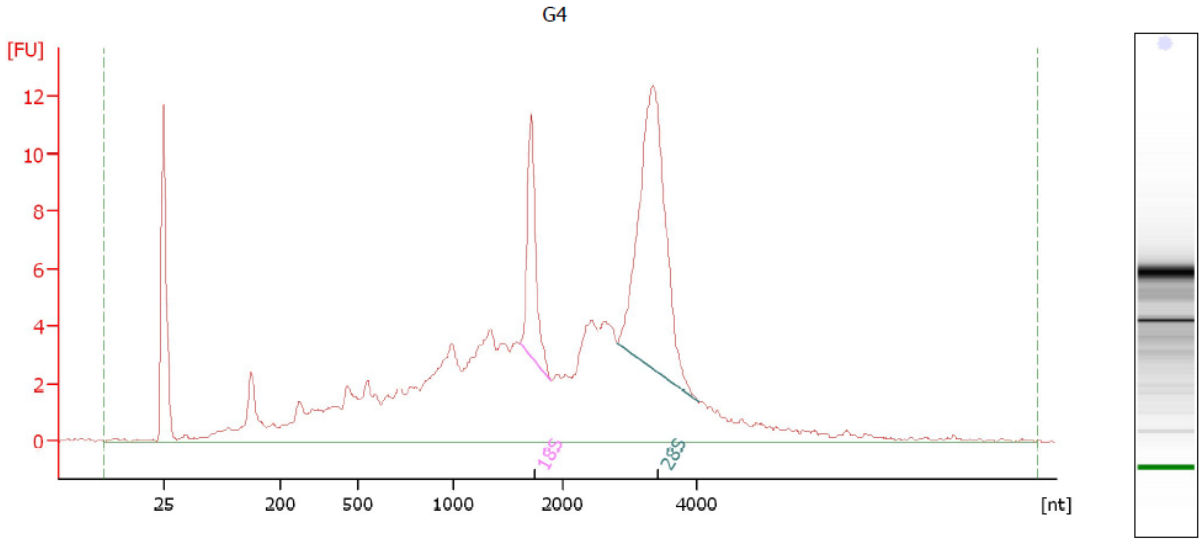
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,477	1,952	109.2	16.2
28S	2,969	3,380	66.2	9.8

Figure 6-3: Bioanalyser report for total RNA extraction of sample G3 *Tiliqua adelaidensis* kidney tissue.

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE54704589_2018-11-19_17-20-05.xad

Created: 11/19/2018 5:20:05 PM
Modified: 11/23/2018 11:19:44 AM

Electropherogram Summary Continued ...



Overall Results for sample 4 : G4

RNA Area:	192.4	RNA Integrity Number (RIN):	7.1 (B.02.08)
RNA Concentration:	128 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	3.4	Result Flagging Label:	RIN: 7.10

Fragment table for sample 4 : G4

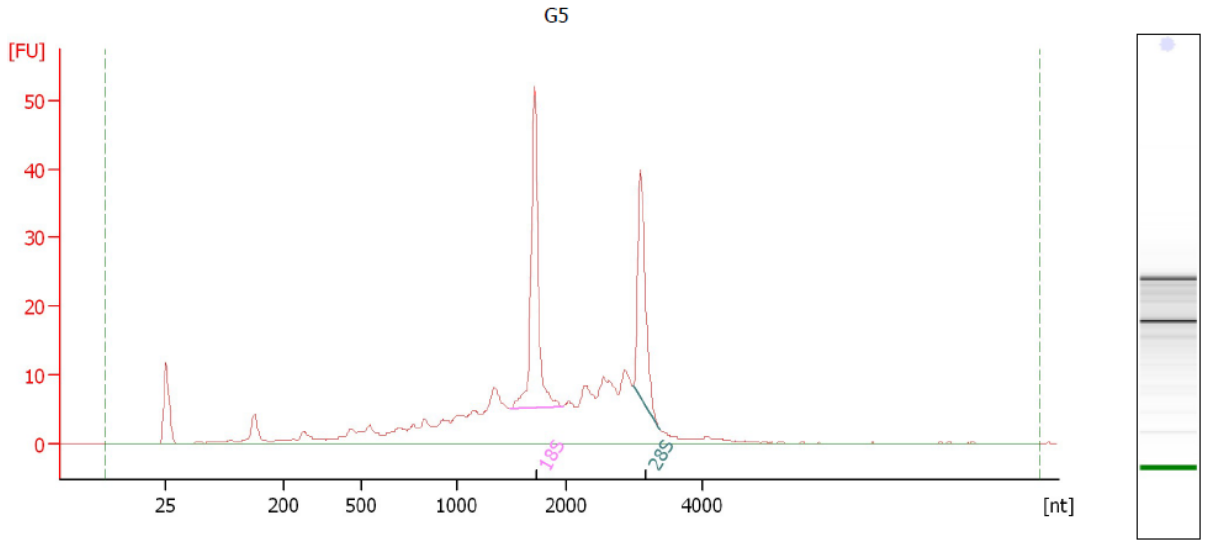
Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,602	1,885	9.9	5.2
28S	2,810	4,006	33.3	17.3

Figure 6-4: Bioanalyser report for total RNA extraction of sample G4 *Tiliqua adelaidensis* kidney tissue.

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE54704589_2018-11-19_17-20-05.xad

Created: 11/19/2018 5:20:05 PM
Modified: 11/23/2018 11:19:44 AM

Electropherogram Summary Continued ...



Overall Results for sample 5 : G5

RNA Area:	299.6	RNA Integrity Number (RIN):	7 (B.02.08)
RNA Concentration:	200 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.7	Result Flagging Label:	RIN:7

Fragment table for sample 5 : G5

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,483	1,954	49.1	16.4
28S	2,978	3,385	31.9	10.6

Figure 6-5: Bioanalyser report for total RNA extraction of sample G5 *Tiliqua adelaidensis* kidney tissue.

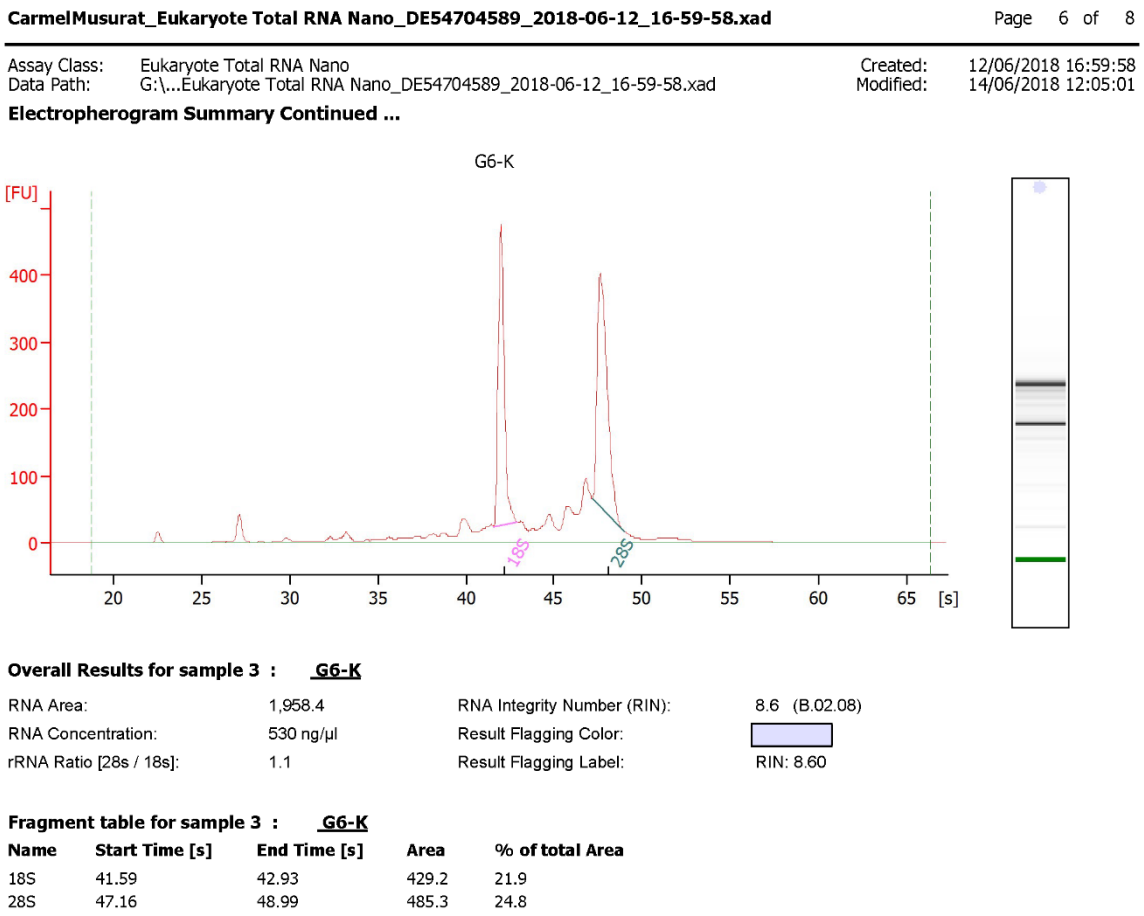
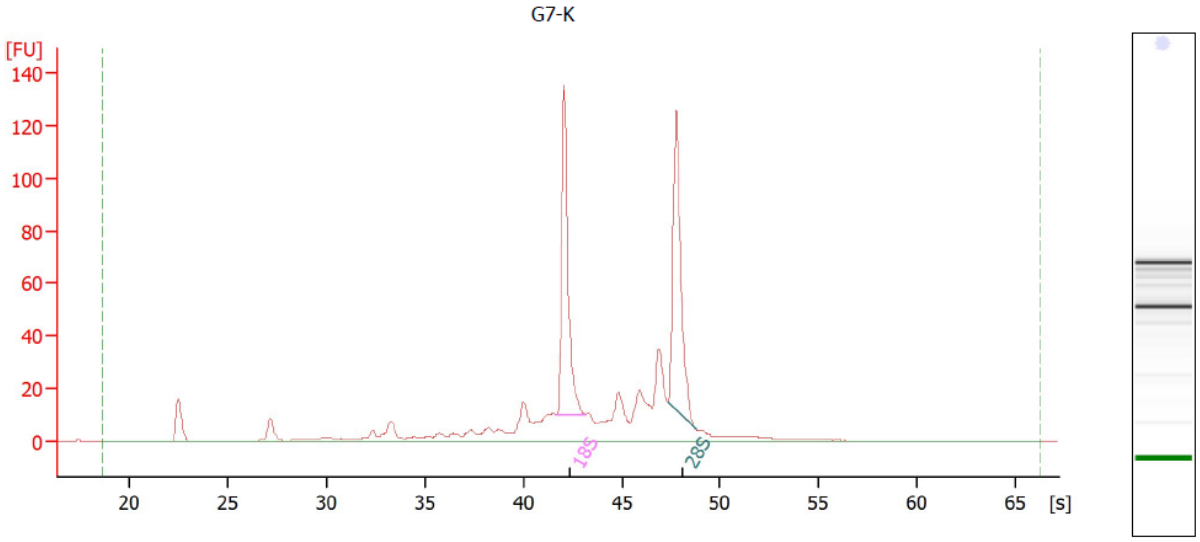


Figure 6-6: Bioanalyser report for total RNA extraction of sample G6 *Tiliqua adelaidensis* kidney tissue. This sample G6 had the highest RIN score, low amounts of short-length RNA to indicate low degradation, and a usable concentration, and was therefore the sample used for long-read sequencing.

Assay Class: Eukaryote Total RNA Nano
Data Path: G:\...Eukaryote Total RNA Nano_DE54704589_2018-06-12_16-59-58.xad

Created: 12/06/2018 16:59:58
Modified: 14/06/2018 12:05:01

Electropherogram Summary Continued ...



Overall Results for sample 4 : G7-K

RNA Area:	566.7	RNA Integrity Number (RIN):	7.9 (B.02.08)
RNA Concentration:	153 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.9	Result Flagging Label:	RIN: 7.90

Fragment table for sample 4 : G7-K

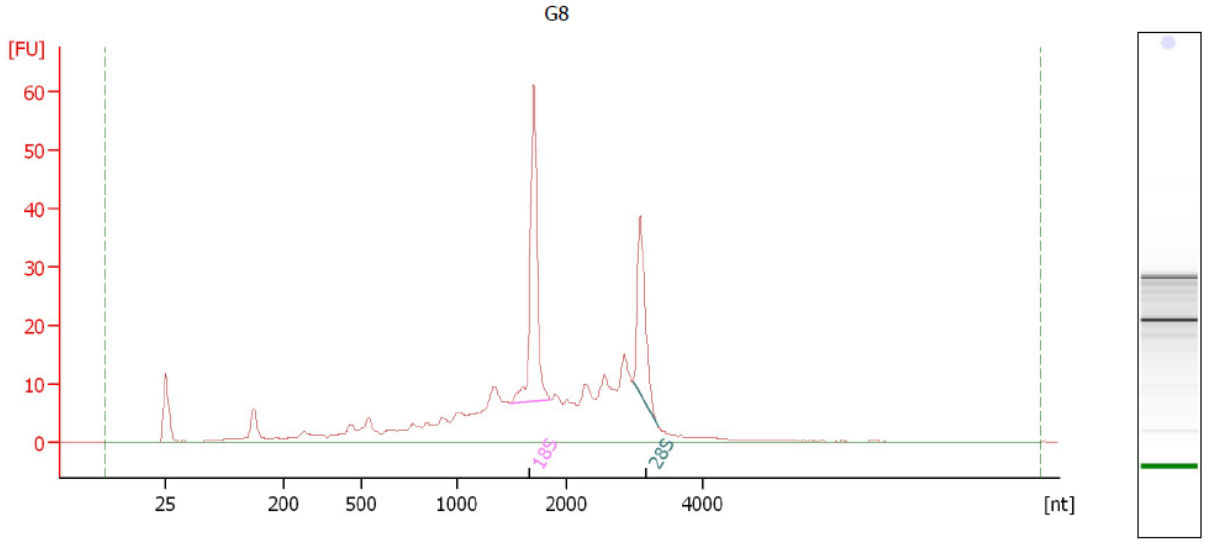
Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.69	43.13	115.7	20.4
28S	47.35	48.79	109.0	19.2

Figure 6-7: Bioanalyser report for total RNA extraction of sample G7 *Tiliqua adelaidensis* kidney tissue.

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE54704589_2018-11-19_17-20-05.xad

Created: 11/19/2018 5:20:05 PM
Modified: 11/23/2018 11:19:44 AM

Electropherogram Summary Continued ...



Overall Results for sample 6 : G8

RNA Area:	347.2	RNA Integrity Number (RIN):	6.8 (B.02.08)
RNA Concentration:	232 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.6	Result Flagging Label:	RIN: 6.80

Fragment table for sample 6 : G8

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,483	1,851	53.2	15.3
28S	2,978	3,343	30.8	8.9

Figure 6-8: Bioanalyser report for total RNA extraction of sample G8 *Tiliqua adelaidensis* kidney tissue.

6.1.3 A Note on Other Tissues

Liver samples were initially a secondary focus because of the role of this tissue in producing uric acid; however, total RNA extractions repeatedly did not have a suitable RIN score or concentration (Table 6-1), and this portion of the project was abandoned because of limits in kit reagents and budget, and an insufficient number of working samples for a robust analysis to justify sequencing.

After repeated extractions with low concentrations from the only collected spleen and other tissues, these were also not included. A single testes RNA sample was sequenced alongside the second sequencing run including kidney samples G6 and G7 but was not included in further analysis.

6.1.4 FastQC/NGS reports

Full quality reports (as .html) are available on GitHub for download, including quality scoring before and after adapter and quality trimming of short-read sequencing.

6.1.5 List of Figures

[Note Figures are numbered by Chapter 2 to Chapter 5. Figures labelled Section 6 are in an Appendix]

6.1.6 List Of Tables

[Note Tables are numbered by Chapter 2 to Chapter 5. Tables labelled Section 6 are in an Appendix]

6.2 Appendix 2: Gene Expression Analysis

6.2.1 Annotated Genes which were Differentially Expressed

Table 6-2: List of all annotated genes found to be differentially expressed in autumn (March/ April) compared to spring (September) based on P value, from poly(A) selected mRNA extracted from kidney tissue of six *Tiliqua adelaidensis* individuals (2F, 1M per group) and sequenced on an Illumina Hi-Seq platform. Annotation was conducted using BLASTx against the UniProt Swiss Prot database, genes, full names, and descriptions correspond to the UniProtKB entry for each UniProt ID (The UniProt Consortium 2021; The UniProt Consortium 2023). Other publication references for function were kept where they were explicitly referenced within the description on UniProtKB database. Descriptions marked ^ are transcripts which was assigned a BLASTx hit for the same gene ID as the previous entry.

Genes Upregulated in Autumn								
	Differential Expression Statistics				**BLASTx Statistics**			
TranscriptID	**Over-disp.**	**logFC**	**logCPM**	**P Value**	**UniProtID**	**Gene**	**% id**	**Name and Function** - Directly quoted from the UniProtKB database based on UniProtID
PB.3807.1|8f2061|path47:7-1741(+)|transcript/14413|m.6997	7.07761	3.8376	4.5013	1.59E-05	Q3SZB4	*acadm*	84.135	Medium-chain specific acyl-CoA dehydrogenase, one of the acyl-CoA dehydrogenases that catalyse the first step of mitochondrial fatty acid beta-oxidation, an aerobic process breaking down fatty acids into acetyl-CoA and allowing the production of energy from fats.
PB.4931.3|b8571c|path20:38-3608(+)|transcript/1966|m.9059	7.015262	6.6718	0.9626	0.00012	Q95241	*app*	92.14	Amyloid-beta precursor protein, Functions as a cell surface receptor and performs physiological functions on the surface of neurons relevant to neurite growth, neuronal adhesion and axonogenesis. Interaction between APP molecules on neighbouring cells promotes synaptogenesis. Involved in cell mobility and transcription regulation through protein-protein interactions (By similarity).
PB.3445.2|8111f0|path13:1-2049(+)|transcript/12812|m.6304	5.171535	2.3049	6.6454	7.49E-05	P31335	*atic*	87.291	Bifunctional purine biosynthesis protein ATIC, Bifunctional enzyme that catalyses the last two steps of purine biosynthesis (Wolan et al. 2002).

PB.3445.6|8111f0|path13:4-2048(+)|transcript/17986|m.6308	3.134023	2.75	6.0046	0.00018	P31335	*atic*	86.957	^
PB.3446.2|8111f0|path4:1-2608(+)|transcript/5871|m.6310	11.65428	2.1394	3.6657	0.00025	P31335	*atic*	87.521	^
PB.9128.1|transcript/4808:1-2827(+)|transcript/4808|m.14870	1	1.7645	5.722	0.00108	P38936	*cdkn1a*	38.75	Cyclin-dependent kinase inhibitor 1, May be involved in p53/TP53 mediated inhibition of cellular proliferation in response to DNA damage. Binds to and inhibits cyclin-dependent kinase activity, preventing phosphorylation of critical cyclin-dependent kinase substrates and blocking cell cycle progression (Ducoux et al. 2001). Plays an important role in controlling cell cycle progression and DNA damage-induced G2 arrest (LaBaer et al. 1997).
PB.5276.2|c71572|path0:32-2931(+)|transcript/4500|m.9727	1.084631	2.0844	6.6256	0.00075	Q7Z692	*ceacam19*	38.168	Carcinoembryonic antigen-related cell adhesion molecule 19
PB.876.2|206df6|path10:20-1995(+)|transcript/12799|m.1572	6.58351	3.5189	3.2567	1.89E-05	P24461	*cyp2g1*	62.146	Cytochrome P450 2G1, Cytochromes P450 are a group of heme-thiolate monooxygenases. This isozyme seems to be implicated in olfaction
PB.876.5|206df6|path10:31-1974(+)|transcript/14832|m.1575	4.693415	3.9117	1.6344	0.00016	P24461	*cyp2g1*	61.765	^
PB.879.1|206df6|path7:41-2171(+)|transcript/12507|m.1583	16.00068	8.8151	1.5232	1.40E-05	P24461	*cyp2g1*	56.41	^
PB.5022.1|bbf2e4|path1:3-3181(+)|transcript/3405|m.9303	1.98107	3.6964	5.9112	1.01E-05	Q10126	*f52c9.6*	26.446	Putative uncharacterized transposon-derived protein F52C9.6 (could be the product of a pseudogene)
PB.6136.2|ea3739|path0:15-2762(+)|transcript/5251|m.11463	1.437629	4.7122	0.1626	0.00095	P14154	*gja1*	86.935	Gap junction alpha-1 protein, one gap junction consists of a cluster of closely packed pairs of transmembrane channels, the connexons, through which materials of low MW diffuse from one cell to a

								neighbouring cell. Plays an essential role in gap junction communication in the ventricles.
PB.3917.1|92fcf5|path2:1-2087(+)|transcript/12348|m.7251	1.838971	1.864	5.7922	0.00035	Q99ME9	*gtpbp4*	87.154	GTP-binding protein 4, Involved in the biogenesis of the 60S ribosomal subunit. Acts as TP53 repressor, preventing TP53 stabilization and cell cycle arrest.
PB.1305.2|31fdc7|path1:93-2420(+)|transcript/10029|m.2322	4.775751	1.8244	5.419	0.00095	P52191	*kcnj16*	66.667	Inward rectifier potassium channel 16, Inward rectifier potassium channels are characterized by a greater tendency to allow potassium to flow into the cell rather than out of it. Their voltage dependence is regulated by the concentration of extracellular potassium; as external potassium is raised, the voltage range of the channel opening shifts to more positive voltages. The inward rectification is mainly due to the blockage of outward current by internal magnesium... In the kidney, together with KCNJ10, mediates basolateral K+ recycling in distal tubules; this process is critical for Na+ reabsorption at the tubules.
PB.3538.2|8589d3|path1:3-4422(+)|transcript/649|m.6552	1.220683	1.7838	5.2739	0.00044	Q9BT23	*limd2*	47.887	LIM domain-containing protein 2, Acts as an activator of the protein kinase ILK, thereby regulating cell motility (Peng et al. 2014)
PB.3413.1|7f6d24|path1:1-1736(+)|transcript/16515|m.6266	5.233853	6.0796	-0.766	7.78E-05	P17694	*ndufs2*	84.718	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial, Core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (Complex I) which catalyses electron transfer from NADH through the respiratory chain, using ubiquinone as an electron acceptor (Sazanov et al. 2000; Lemma-Gray et al. 2008)
PB.7021.1|transcript/12099:1-2056(+)|transcript/12099|m.12792	1.022081	3.3777	4.545	1.88E-05	Q98SV0	*sepp1b*	51.969	Selenoprotein Pb, might be responsible for some of the extracellular antioxidant defence properties of selenium.

PB.8033.1|transcript/20076:1-1439(+)|transcript/20076|m.13828	1	1.973	5.1985	0.00022	P34955	*serpina1*	45.73	Alpha-1-antiproteinase, Inhibitor of serine proteases. Its primary target is elastase, but it also has a moderate affinity for plasmin and thrombin.
PB.5202.1|c35005|path18:1-2621(+)|transcript/6504|m.9616	2.620448	2.0001	5.4238	0.00069	P22309	*ugt1a1*	63.362	UDP-glucuronosyltransferase 1A1, UDP-glucuronosyltransferase (UGT) that catalyses phase II biotransformation reactions in which lipophilic substrates are conjugated with glucuronic acid to increase the metabolite's water solubility, thereby facilitating excretion into either the urine or bile (Gagné et al. 2002; Lépine et al. 2004; Girard et al. 2007; Udomuksorn et al. 2007; Itäaho et al. 2008; Sneitz et al. 2010; Sneitz et al. 2013).
Genes Downregulated in Autumn								
	Differential Expression Statistics				**BLASTx Statistics**			
TranscriptID	**Over-disp.**	**logFC**	**logCPM**	**P Value**	**UniProtID**	**Gene**	**% id**	**Name and Description** - Directly quoted from the UniProtKB database based on UniProtID (The UniProt Consortium 2021; The UniProt Consortium 2023)
PB.966.1|257403|path2:1-1611(+)|transcript/18188|m.1746	1.125528	-7.976	7.0879	7.08E-05	Q5NVF6	*acp2*	46.686	Lysosomal acid phosphatase
PB.966.3|257403|path2:1-1480(+)|transcript/20803|m.1748	2.668011	-8.715	1.4271	0.00069	Q5NVF6	*acp2*	44.092	^
PB.6698.1|ff02c7|path2:180-1998(+)|transcript/19664|m.12475	1	-7.294	8.7518	0.00012	P15309	*acp3*	48.563	Prostatic acid phosphatase (*Homo sapiens*), A non-specific tyrosine phosphatase that dephosphorylates a diverse number of substrates under acidic conditions (pH 4-6) (Zhang, Ostanin & Van Etten 1997; Hiroyama & Takenawa 1999; Tanaka et al. 2004; Chuang et al. 2010).
PB.6631.3|fc790c|path5:90-1513(+)|transcript/21158|m.12302	3.14774	-7.399	5.9701	0.00012	Q3KQG9	*acp4*	44.906	Testicular acid phosphatase homolog

PB.6119.1|e99bf7|path16:1-1504(+)|transcript/19547|m.11431	1	-8.534	4.2389	0.00101	A6H730	*acpp*	48.718	Prostatic acid phosphatase (*Bos taurus*)
PB.6629.1|fc790c|path11:1-1272(+)|transcript/20911|m.12297	1.416825	-7.826	9.8174	6.49E-05	A6H730	*acpp*	47.977	^
PB.6629.2|fc790c|path11:813-1508(+)|transcript/23958|m.12298	1.681319	-7.24	5.6571	3.92E-05	A6H730	*acpp*	47.115	^
PB.6630.1|fc790c|path12:93-3778(+)|transcript/1704|m.12299	1.341913	-7.342	6.7004	3.66E-05	A6H730	*acpp*	47.879	^
PB.6699.2|ff02c7|path8:216-1992(+)|transcript/15774|m.12478	2.566849	-7.922	2.5777	0.00046	A6H730	*acpp*	47.44	^
PB.3285.1|7a2645|path2:1-3617(+)|transcript/1897|m.6054	4.181166	-2.637	2.8725	0.00048	P33121	*acsl1*	78.561	Long-chain-fatty-acid--CoA ligase 1, Catalyses the conversion of long-chain fatty acids to their active form acyl-CoAs for both synthesis of cellular lipids, and degradation via beta-oxidation (Golej et al. 2011; Nakahara et al. 2012; Ohkuni, Ohno & Kihara 2013).
PB.2726.1|660a4b|path2:1-1412(+)|transcript/20515|m.4978	1.022723	-2.603	5.2801	0.00071	P63269	*actg2*	100	Actin, gamma-enteric smooth muscle, Actins are highly conserved proteins that are involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cells.
PB.5398.2|cbf34d|path5:582-2207(+)|transcript/17939|m.10003	2.311738	-3.876	4.4353	0.00028	P11884	*aldh2*	86.072	Aldehyde dehydrogenase, mitochondrial, required for clearance of cellular formaldehyde, a cytotoxic and carcinogenic metabolite that induces DNA damage.
PB.3464.2|81f3ff|path2:2-2336(+)|transcript/8627|m.6352	1.338924	-4.743	3.0846	0.00023	Q9JHZ2	*ankh*	94.32	Progressive ankylosis protein, regulates intra- and extracellular levels of inorganic pyrophosphate (PPi), probably functioning as PPi transporter.

PB.8053.1|transcript/20242:1-1425(+)|transcript/20242|m.13847	1	-3.261	0.9045	0.00032	Q9NQW6	*anln*	85.577	Anillin, Required for cytokinesis (Zhao & Fang 2005). Essential for the structural integrity of the cleavage furrow and for completion of cleavage furrow ingression. Plays a role in bleb assembly during metaphase and anaphase of mitosis (Gbadegesin et al. 2014).
PB.4486.2|a7cd86|path1:70-1974(+)|transcript/14382|m.8265	1.267642	-3.18	4.4573	0.00096	P23614	*basp1*	74.648	Brain acid soluble protein 1 homolog, may play a specific role in the development of tissues.
PB.8832.1|transcript/3117:1-3189(+)|transcript/3117|m.14587	1	-4.524	1.332	1.40E-05	O43683	*bub1*	62.606	Mitotic checkpoint serine/threonine-protein kinase BUB1, Serine/threonine-protein kinase that performs 2 crucial functions during mitosis: it is essential for spindle-assembly checkpoint signalling and for correct chromosome alignment.
PB.7313.1|transcript/1444:1-3746(+)|transcript/1444|m.13115	1	-4.151	3.7258	1.82E-05	O60566	*bub1b*	50	Mitotic checkpoint serine/threonine-protein kinase BUB1 beta, Essential component of the mitotic checkpoint. Required for normal mitosis progression. The mitotic checkpoint delays anaphase until all chromosomes are properly attached to the mitotic spindle.
PB.1839.1|45b190|path2:1-1051(+)|transcript/22627|m.3370	1.063444	-6.602	7.6498	0.00022	P07660	*calca*	81.159	Calcitonin, causes a rapid but short-lived drop in the level of calcium and phosphate in blood by promoting the incorporation of those ions in the bones.
PB.5964.3|e3291f|path2:9-3734(+)|transcript/12644|m.11062	5.302731	-10.42	3.0854	0.00019	P07660	*calca*	48.78	^
PB.5964.5|e3291f|path2:9-3697(+)|transcript/7077|m.11064	2.285255	-5.931	5.4205	4.39E-05	P07660	*calca*	49.398	^
PB.6088.5|e88f16|path1:162-2531(+)|transcript/18623|m.11356	1.990285	-6.579	7.9945	8.98E-05	P10286	*calca*	48.387	Calcitonin gene-related peptide, CGRP induces vasodilation. It dilates a variety of vessels including the coronary, cerebral and systemic vasculature.

PB.6088.11|e88f16|path1:164-2792(+)|transcript/22345|m.11362	1.838208	-6.726	9.184	0.00013	P10286	*calca*	48.387	^
PB.9705.1|transcript/8517:1-2388(+)|transcript/8517|m.15476	1.056813	-8.457	3.0161	0.00014	O14815	*capn9*	78.403	Calpain-9, Calcium-regulated non-lysosomal thiol-protease.
PB.5901.2|e0b493|path0:13-2360(+)|transcript/8664|m.10952	1.305984	-5.237	3.5189	8.88E-05	Q91WT9	*cbs*	86.093	Cystathionine beta-synthase (*Mus musculus*), Hydro-lyase catalysing the first step of the transsulfuration pathway
PB.5901.3|e0b493|path0:14-2006(+)|transcript/7052|m.10953	2.999079	-7.735	5.3913	5.70E-05	Q58H57	*cbs*	70.038	Cystathionine beta-synthase (*Macaca fascicularis*), Hydro-lyase catalysing the first step of the transsulfuration pathway
PB.7967.1|transcript/19601:1-1493(+)|transcript/19601|m.13768	1	-4.975	2.4905	3.35E-06	Q1LZG6	*ccnb1*	68.075	G2/mitotic-specific cyclin-B1, Essential for the control of the cell cycle at the G2/M (mitosis) transition.
PB.6863.1|transcript/1087:1-4085(+)|transcript/1087|m.12652	1.001862	-3.918	1.3506	0.00012	P41002	*ccnf*	62.527	Cyclin-F, Substrate recognition component of a SCF (SKP1-CUL1-F-box protein) E3 ubiquitin-protein ligase complex which mediates the ubiquitination and subsequent proteasomal degradation of target proteins (D'Angiolella et al. 2010; D'Angiolella et al. 2012; Choudhury et al. 2016; Walter et al. 2016; Williams et al. 2016; Lee et al. 2018).
PB.1773.1|42d732|path1:1-1839(+)|transcript/15277|m.3183	1	-5.7	1.8053	4.45E-06	Q5H7C0	*cdc20*	77.822	Cell division cycle protein 20 homolog, required for full ubiquitin ligase activity of the anaphase-promoting complex/cyclosome (APC/C) and may confer substrate specificity upon the complex.
PB.9131.1|transcript/4816:1-2838(+)|transcript/4816|m.14872	1	-2.381	1.9817	0.00046	Q99741	*cdc6*	69.35	Cell division control protein 6 homolog, Involved in the initiation of DNA replication. Also participates in checkpoint controls that ensure DNA replication is completed before mitosis is initiated.

PB.8114.1|transcript/20645:1-1387(+)|transcript/20645|m.13903	1	-3.789	1.4206	6.07E-05	P13863	*cdk1*	92.384	Cyclin-dependent kinase 1, Plays a key role in the control of the eukaryotic cell cycle by modulating the centrosome cycle as well as mitotic onset; promotes G2-M transition, and regulates G1 progress and G1-S transition via association with multiple interphase cyclins. Required in higher cells for entry into S-phase and mitosis. May play a role in regulating the amplitude of the cyclic expression of circadian clock genes.
PB.840.1|1eedc5|path0:1-2091(+)|transcript/11690|m.1514	1	-5.889	3.5505	0.0002	P09482	*chrna4*	87.961	Neuronal acetylcholine receptor subunit alpha-4, After binding acetylcholine, the AChR responds by an extensive change in conformation that affects all subunits and leads to opening of an ion-conducting channel across the plasma membrane.
PB.4480.1|a796d3|path2:1-4660(+)|transcript/541|m.8250	1.329697	-2.833	2.7053	2.19E-05	P02457	*col1a1*	92.193	Collagen alpha-1(I) chain, Type I collagen is a member of group I collagen (fibrillar forming collagen).
PB.5695.2|d72911|path1:1745-4544(+)|transcript/5044|m.10570	1.792256	-2.168	3.2302	0.00078	P02467	*col1a2*	83.516	Collagen alpha-2(I) chain, Type I collagen is a member of group I collagen (fibrillar forming collagen).
PB.1490.1|38bf54|path13:1-5562(+)|transcript/179|m.2719	2.72325	-3.762	1.7506	3.78E-05	P12105	*col3a1*	75.676	Collagen alpha-1(III) chain, Collagen type III occurs in most soft connective tissues along with type I collagen.
PB.1490.2|38bf54|path13:1498-5563(+)|transcript/9326|m.2720	1.15342	-2.346	5.6532	0.00066	P12105	*col3a1*	75.676	^
PB.7382.1|transcript/14974:1-1845(+)|transcript/14974|m.13176	1	-6.862	6.964	1.00E-05	P56594	*cyp2c21*	58.848	Cytochrome P450 2C21, Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics. Showed testosterone hydrolase activity.

PB.3153.1|74febd|path1:214-2038(+)|transcript/15385|m.5765	1.186255	-7.875	2.4917	0.00087	P15129	*cyp4b1*	57.267	Cytochrome P450 4B1, Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics.
PB.3935.1|93b55a|path0:1-2494(+)|transcript/7332|m.7285	1.353726	-4.156	2.5244	8.84E-05	P43346	*dck*	82.51	Deoxycytidine kinase, Phosphorylates the deoxyribonucleosides deoxycytidine, deoxyguanosine and deoxyadenosine.
PB.1567.2|3bbed1|path1:1-2767(+)|transcript/5340|m.2848	1	-2.913	4.6392	0.00099	P02542	*des*	85.969	Desmin, Muscle-specific type III intermediate filament essential for proper muscular structure and function. Plays a crucial role in maintaining the structure of sarcomeres
PB.1405.4|35531f|path14:3-2255(+)|transcript/10599|m.2470	1.823119	-1.918	4.48	0.00105	Q9Y394	*dhrs7*	67.687	Dehydrogenase/reductase SDR family member 7, NADPH-dependent oxidoreductase which catalyses the reduction of a variety of compounds bearing carbonyl groups including steroids, retinoids and xenobiotics (Štambergová et al. 2014; Štambergová et al. 2016; Araya et al. 2017; Zemanová et al. 2017).
PB.8861.1|transcript/326:1-5053(+)|transcript/326|m.14615	1	-3.1	0.6078	0.00045	Q9NSV4	*diaph3*	69.899	Protein diaphanous homolog 3, Actin nucleation and elongation factor required for the assembly of F-actin structures, such as actin cables and stress fibers.
PB.1055.1|295311|path2:1-1979(+)|transcript/12857|m.1877	1	-3.133	1.6564	0.00029	Q3SZM7	*dpep1*	68.564	Dipeptidase 1, Hydrolyses a wide range of dipeptides including the conversion of leukotriene D4 to leukotriene E4. Hydrolyses cystinyl-bis-glycine (cys-bis-gly) formed during glutathione degradation.
PB.1636.2|3ea652|path0:21-2775(+)|transcript/5207|m.2967	1.261881	-3.531	1.3107	7.66E-06	P79784	*entpd2*	65.161	Ectonucleoside triphosphate diphosphohydrolase 2, In the nervous system, could hydrolyse ATP and other nucleotides to regulate purinergic neurotransmission. Hydrolyses ADP only to a marginal extent (By similarity).
PB.3631.1|88a4ef|path3:1-1773(+)|transcript/14984|m.6705	1	-4.035	6.0338	0.00033	Q86YB8	*ero1b*	89.245	ERO1-like protein beta, Oxidoreductase involved in disulfide bond formation in the endoplasmic reticulum.

PB.6093.1|e8a4b2|path0:1-2765(+)|transcript/5519|m.11383	1.004373	-3.461	1.8162	5.74E-05	Q56NI9	*esco2*	45.789	N-acetyltransferase ESCO2, Acetyltransferase required for the establishment of sister chromatid cohesion (Vega et al. 2005; Hou & Zou 2005).
PB.2068.1|4cb9fa|path2:1-2419(+)|transcript/8339|m.3709	3.41723	-3.356	1.5414	2.96E-06	Q6DEB1	*etnppl*	76.228	Ethanolamine-phosphate phospho-lyase, Catalyses the pyridoxal-phosphate-dependent breakdown of phosphoethanolamine, converting it to ammonia, inorganic phosphate and acetaldehyde (does not seem to possess aminotransferare activity) (By similarity)
PB.2068.2|4cb9fa|path2:83-2407(+)|transcript/8860|m.3710	6.445415	-3.559	0.1573	0.00016	Q6DEB1	*etnppl*	77.393	^
PB.7942.1|transcript/19401:1-1535(+)|transcript/19401|m.13662	1.049059	-2.994	1.4689	0.00017	Q8BHL6	*faap24*	62.376	Fanconi anemia core complex-associated protein 24, Plays a role in DNA repair through recruitment of the FA core complex to damaged DNA. Regulates FANCD2 monoubiquitination upon DNA damage. Induces chromosomal instability as well as hypersensitivity to DNA cross-linking agents, when repressed. Targets FANCM/FAAP24 complex to the DNA, preferentially to single strand DNA (By similarity).
PB.9768.1|transcript/9048:1-2313(+)|transcript/9048|m.15534	1	-4.171	1.0688	4.89E-05	Q00597	*fancc*	50.356	Fanconi anemia group C protein, DNA repair protein that may operate in a post replication repair or a cell cycle checkpoint function. May be implicated in interstrand DNA cross-link repair and in the maintenance of normal chromosome stability.
PB.7086.1|transcript/12535:1-2021(+)|transcript/12535|m.12846	1	-2.193	2.4442	0.00066	Q9UKT4	*fbxo5*	49.878	F-box only protein 5, Regulator of APC activity during mitotic and meiotic cell cycle (Miller et al. 2006; Di Fiore & Pines 2007; Machida & Dutta 2007; Verschuren et al. 2007; Frye et al. 2013; Wang & Kirschner 2013) (as both substrate and inhibitor)
PB.3353.1|7cf053|path5:1-7730(+)|transcript/10|m.6157	2.638708	-6.67	5.325	2.12E-05	P02751	*fn1*	75.212	Fibronectin (*Homo sapiens*), Fibronectins bind cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin. Fibronectins are

								involved in cell adhesion, cell motility, opsonization, wound healing, and maintenance of cell shape
PB.3353.2|7cf053|path5:2504-7730(+)|transcript/274|m.6158	3.793988	-7.544	4.983	1.02E-05	P07589	*fn1*	71.557	Fibronectin (*Bos taurus*)
PB.3353.4|7cf053|path5:5093-7737(+)|transcript/6911|m.6160	4.073233	-7.305	4.0884	6.31E-05	P04937	*fn1*	68.932	Fibronectin (*Rattus norvegicus*)
PB.1658.1|3f1351|path1:1-2676(+)|transcript/6659|m.3008	1.038635	-3.958	4.2846	0.00049	Q9UFP1	*gask1a*	60.267	Golgi-associated kinase 1A
PB.499.1|124cf9|path4:1-2080(+)|transcript/11611|m.837	1.414303	-2.555	5.0358	8.94E-07	Q9I9K9	*gatm*	88.028	Glycine amidinotransferase, mitochondrial (*Gallus gallus*), Catalyses the biosynthesis of guanidinoacetate, the immediate precursor of creatine. Creatine plays a vital role in energy metabolism in muscle tissues. May play a role in embryonic and central nervous system development (By similarity).
PB.499.2|124cf9|path4:1-4611(+)|transcript/19898|m.838	2.386387	-2.347	2.2057	0.00042	P50441	*gatm*	78.621	Glycine amidinotransferase, mitochondrial (*Sus scrofa*), Transamidinase that catalyses the transfer of the amidino group of L-arginine onto the amino moiety of acceptor metabolites such as glycine, beta-alanine, gamma-aminobutyric acid (GABA) and taurine yielding the corresponding guanidine derivatives (By similarity).
PB.4939.1|b87d61|path1:1-1176(+)|transcript/22771|m.9073	1.745864	-7.577	6.1362	0.00011	P39905	*gdnf*	38.208	Glial cell line-derived neurotrophic factor, Neurotrophic factor that enhances survival and morphological differentiation of dopaminergic neurons and increases their high-affinity dopamine uptake.

PB.1195.2|2e529e|path0:676-3190(+)|transcript/7266|m.2146	1	-5.951	11.555	0.00062	O42241	*gnrh2*	46.939	Progonadoliberin-2, Stimulates the secretion of gonadotropins.
PB.2953.1|6d2bbb|path2:1-1810(+)|transcript/15503|m.5409	1.156076	-2.542	2.1272	0.00035	Q92522	*h1-10*	91.667	Histone H1.10, Histones H1 are necessary for the condensation of nucleosome chains into higher-order structures.
PB.6627.1|fc6596|path2:87-6001(+)|transcript/115|m.12291	1.957726	-1.699	4.4342	0.00088	Q9Y6Y0	*ivns1abp*	87.611	Influenza virus NS1A-binding protein, involved in many cell functions, including pre-mRNA splicing, the aryl hydrocarbon receptor (AHR) pathway, F-actin organization and protein ubiquitination. Plays a role in the dynamic organization of the actin skeleton as a stabilizer of actin filaments by association with F-actin through Kelch repeats. Protects cells from cell death induced by actin destabilization (By similarity).
PB.8891.1|transcript/3421:1-3129(+)|transcript/3421|m.14644	1.008132	-3.237	1.8718	0.00032	Q29RT6	*kif20a*	61.252	Kinesin-like protein KIF20A, Mitotic kinesin required for chromosome passenger complex (CPC)-mediated cytokinesis.
PB.1432.1|3646e0|path0:1-2286(+)|transcript/9645|m.2636	1	-3.073	2.9565	0.00012	Q14807	*kif22*	58.717	Kinesin-like protein KIF22, Kinesin family member that is involved in spindle formation and the movements of chromosomes during mitosis and meiosis. Binds to microtubules and to DNA (By similarity).
PB.8148.1|transcript/20926:1-1346(+)|transcript/20926|m.13933	1.041722	-3.707	0.5341	0.00033	Q02241	*kif23*	71.984	Kinesin-like protein KIF23, Component of the centralspindlin complex that serves as a microtubule-dependent and Rho-mediated signalling required for the myosin contractile ring formation during the cell cycle cytokinesis.
PB.7158.1|transcript/1314:1-3906(+)|transcript/1314|m.12913	1	-2.746	1.9904	0.00021	Q90640	*kif4*	55.153	Chromosome-associated kinesin KIF4, Iron-sulfur (Fe-S) cluster binding motor protein that has a role in chromosome segregation during mitosis (By similarity). Required for mitotic chromosomal positioning and bipolar spindle stabilization

PB.4927.1|b80aac|path0:1-2366(+)|transcript/8622|m.9052	1	-4.003	3.5359	2.94E-05	Q5ZM46	*lbh*	59.42	Protein LBH, Transcriptional activator.
PB.8300.1|transcript/22013:1-1186(+)|transcript/22013|m.14070	1	-4.211	2.2007	7.75E-07	Q13257	*mad2l1*	82	Mitotic spindle assembly checkpoint protein MAD2A, Component of the spindle-assembly checkpoint that prevents the onset of anaphase until all chromosomes are properly aligned at the metaphase plate (Ji et al. 2018; Luo et al. 2004).
PB.8897.1|transcript/3462:1-3118(+)|transcript/3462|m.14650	1.062696	-2.866	4.1681	3.54E-07	Q5ZMN2	*mcm3*	81.863	DNA replication licensing factor MCM3, Acts as component of the MCM2-7 complex (MCM complex) which is the replicative helicase essential for 'once per cell cycle' DNA replication initiation and elongation in eukaryotic cells (By similarity).
PB.8722.1|transcript/2571:1-3363(+)|transcript/2571|m.14482	1	-2.567	2.8401	1.16E-05	P30664	*mcm4-b*	91.02	DNA replication licensing factor mcm4-B, Acts as component of the MCM2-7 complex (MCM complex) which is the replicative helicase essential for 'once per cell cycle' DNA replication initiation and elongation in eukaryotic cells (By similarity).
PB.9414.1|transcript/6535:1-2598(+)|transcript/6535|m.15211	1.037284	-2.33	2.9514	0.00019	P33992	*mcm5*	89.373	DNA replication licensing factor MCM5, Acts as component of the MCM2-7 complex (MCM complex) which is the replicative helicase essential for 'once per cell cycle' DNA replication initiation and elongation in eukaryotic cells (Rzechorzek et al. 2020; Joneset al. 2021; Jenkyn-Bedford et al. 2021; Baris et al. 2022; Tsuji, Ficarro & Jiang 2006).
PB.7922.1|transcript/1924:1-3613(+)|transcript/1924|m.13646	1	-3.646	2.1521	1.46E-05	Q9BPX3	*ncapg*	58.92	Condensin complex subunit 3, Regulatory subunit of the condensin complex, a complex required for conversion of interphase chromatin into mitotic-like condense chromosomes.
PB.6779.1|transcript/10281:1-2227(+)|transcript/10281|m.12578	1	-4.397	1.979	4.93E-06	Q76I89	*ndc80*	68.218	Kinetochore protein NDC80 homolog, Acts as a component of the essential kinetochore-associated NDC80 complex, which is required for chromosome segregation and spindle checkpoint activity (Hori et al. 2003)

PB.7285.1|transcript/14212:1-1913(+)|transcript/14212|m.13021	1.091503	-3.896	3.6839	0.00035	O42414	*nfasc*	76.658	Neurofascin, Cell adhesion, ankyrin-binding protein which may be involved in neurite extension, axonal guidance, synaptogenesis, myelination and neuron-glial cell interactions (By similarity).
PB.9674.1|transcript/8316:1-2368(+)|transcript/8316|m.15446	1	-5.884	2.647	0.00053	Q8IY84	*nim1k*	74.771	Serine/threonine-protein kinase NIM1
PB.9559.1|transcript/7471:1-2422(+)|transcript/7471|m.15343	1.027035	-2.573	3.7677	0.00076	B3SV56	*nr1d1*	65.306	Nuclear receptor subfamily 1 group D member 1, Transcriptional repressor which coordinates circadian rhythm and metabolic pathways in a heme-dependent manner... Also regulates genes involved in metabolic functions, including lipid and bile acid metabolism, adipogenesis, gluconeogenesis and the macrophage inflammatory response (By similarity).
PB.6082.1|e84280|path0:1-1881(+)|transcript/14136|m.11342	1.000032	-2.578	2.7922	0.00027	Q76I90	*nuf2*	62.582	Kinetochore protein Nuf2, Acts as a component of the essential kinetochore-associated NDC80 complex, which is required for chromosome segregation and spindle checkpoint activity (Hori et al. 2003)
PB.7361.1|transcript/14810:1-1862(+)|transcript/14810|m.13158	1	-3.733	2.4253	4.87E-06	Q5ZJU5	*nusap1*	42.693	Nucleolar and spindle-associated protein 1, Microtubule-associated protein with the capacity to bundle and stabilize microtubules. May associate with chromosomes and promote the organization of mitotic spindle microtubules around them (By similarity).

PB.5332.1|c9b505|path1:1-1445(+)|transcript/19992|m.9895	3.495611	-3.467	2.5038	1.66E-06	B0K025	*ostc*	93.75	Oligosaccharyltransferase complex subunit OSTC, Specific component of the STT3A-containing form of the oligosaccharyl transferase (OST) complex that catalyses the initial transfer of a defined glycan (Glc3Man9GlcNAc2 in eukaryotes) from the lipid carrier dolichol-pyrophosphate to an asparagine residue within an Asn-X-Ser/Thr consensus motif in nascent polypeptide chains, the first step in protein N-glycosylation. N-glycosylation occurs cotranslationally and the complex associates with the Sec61 complex at the channel-forming translocon complex that mediates protein translocation across the endoplasmic reticulum (ER). All subunits are required for a maximal enzyme activity. May be involved in N-glycosylation of APP (amyloid-beta precursor protein). Can modulate gamma-secretase cleavage of APP by enhancing endoprotelysis of PSEN1.
PB.7745.1|transcript/17825:1-1637(+)|transcript/17825|m.13496	1.010844	-5.012	2.8031	7.50E-08	Q96KB5	*pbk*	60.577	Lymphokine-activated killer T-cell-originated protein kinase, Phosphorylates MAP kinase p38. Seems to be active only in mitosis. May also play a role in the activation of lymphoid cells.
PB.8500.1|transcript/23403:1-818(+)|transcript/23403|m.14254	1.053736	-3.824	2.652	3.22E-05	Q5HZL4	*pclaf*	61.345	PCNA-associated factor, PCNA-binding protein that acts as a regulator of DNA repair during DNA replication... Also acts as a regulator of centrosome number (By similarity).
PB.5229.1|c496d1|path0:1-3179(+)|transcript/3063|m.9657	1	-2.212	4.7525	0.00034	Q2KIL5	*pdia5*	70.428	Protein disulfide isomerase A5

PB.2798.1|680426|path0:1-3624(+)|transcript/1864|m.5088	1	-5.616	2.8942	0.00081	P11369	*pol*	29.054	LINE-1 retrotransposable element ORF2 protein, has reverse transcriptase activity required for target-primed reverse transcription of the LINE-1 element mRNA, a crucial step in LINE-1 retrotransposition. Also has endonuclease activity that allows the introduction of nicks in the chromosomal target DNA. Cleaves DNA in AT-rich regions between a 5' stretch of purines and a 3' stretch of pyrimidines, corresponding to sites of LINE-1 integration in the genome. Conformational properties of the target DNA sequence rather than specific nucleotides are key determinants of the ORF2p capacity for sequence-specific DNA recognition. Does not bend the DNA helix unlike related endonucleases but causes compression near the cleavage site.
PB.7552.1|transcript/1637:1-3730(+)|transcript/1637|m.13323	1	-4.098	-0.005	0.00108	Q8CJ52	*prom2*	46.326	Prominin-2 (testosterone regulated, membrane glycoprotein, cholesterol binding)
PB.3213.3|772cba|path4:80-2024(+)|transcript/11530|m.5935	1	-4.974	8.732	0.00041	Q05186	*rcn1*	67.881	Reticulocalbin-1, May regulate calcium-dependent activities in the endoplasmic reticulum lumen or post-ER compartment.
PB.8421.1|transcript/22816:1-1007(+)|transcript/22816|m.14185	1	-4.815	3.0596	0.00024	Q9H1X1	*rsph9*	66.667	Radial spoke head protein 9 homolog, Functions as part of axonemal radial spoke complexes that play an important part in the motility of sperm and cilia (Castleman et al. 2009)... Required for motility of olfactory and neural cilia and for the structural integrity of ciliary axonemes in both 9+0 and 9+2 motile cilia (By similarity).

PB.2144.1|501a88|path2:1-2773(+)|transcript/5134|m.3933	4.122911	-2.694	0.1598	0.00106	Q5R7A7	*sgk3*	92.638	Serine/threonine-protein kinase Sgk3, Serine/threonine-protein kinase which is involved in the regulation of a wide variety of ion channels, membrane transporters, cell growth, proliferation, survival and migration. Up-regulates Na+ channels: SCNN1A/ENAC and SCN5A, K+ channels: KCNA3/KV1.3, KCNE1, KCNQ1 and KCNH2/HERG, epithelial Ca2+ channels: TRPV5 and TRPV6, chloride channel: BSND, creatine transporter: SLC6A8, Na+/dicarboxylate cotransporter: SLC13A2/NADC1, Na+-dependent phosphate cotransporter: SLC34A2/NAPI-2B, amino acid transporters: SLC1A5/ASCT2 and SLC6A19, glutamate transporters: SLC1A3/EAAT1, SLC1A6/EAAT4 and SLC1A7/EAAT5, glutamate receptors: GRIA1/GLUR1 and GRIK2/GLUR6, Na+/H+ exchanger: SLC9A3/NHE3, and the Na+/K+ ATPase. Plays a role in the regulation of renal tubular phosphate transport and bone density. Phosphorylates NEDD4L and GSK3B. Positively regulates ER transcription activity through phosphorylation of FLII. Negatively regulates the function of ITCH/AIP4 via its phosphorylation and thereby prevents CXCR4 from being efficiently sorted to lysosomes (By similarity).
PB.8089.1|transcript/20489:1-1427(+)|transcript/20489|m.13880	1.159437	-4.044	2.1771	6.18E-06	Q9Y619	*slc25a15*	60.853	Mitochondrial ornithine transporter 1, Mitochondrial ornithine-citrulline antiporter (Fiermonte et al. 2003; Monné et al 2012) (Probable). Catalyses the exchange between cytosolic ornithine and mitochondrial citrulline plus an H+, the proton compensates the positive charge of ornithine thus leading to an electroneutral transport. Plays a crucial role in the urea cycle, by connecting the cytosolic and the intramitochondrial reactions of the urea cycle (Fiermonte et al. 2003; Monné et al 2012) (Probable). Lysine and arginine are also transported by the antiport mechanism (Fiermonte et al. 2003)

								(Probable). In addition, catalyses an electroneutral exchange of ornithine or lysine for H+, a reaction driven by the pH gradient across the inner membrane (By similarity).
PB.2610.1|60c22a|path0:1-2155(+)|transcript/10675|m.4798	1.686574	-2.56	2.5882	0.00082	P53790	*slc5a1*	82.514	Sodium/glucose cotransporter 1, Electrogenic Na+-coupled sugar simporter that actively transports D-glucose or D-galactose at the plasma membrane, with a Na+ to sugar coupling ratio of 2:1. Transporter activity is driven by a transmembrane Na+ electrochemical gradient set by the Na+/K+ pump. Has a primary role in the transport of dietary monosaccharides from enterocytes to blood. Responsible for the absorption of D-glucose or D-galactose across the apical brush-border membrane of enterocytes, whereas basolateral exit is provided by GLUT2. Additionally, functions as a D-glucose sensor in enteroendocrine cells, triggering the secretion of the incretins GCG and GIP that control food intake and energy homeostasis. Together with SGLT2, functions in reabsorption of D-glucose from glomerular filtrate, playing a nonredundant role in the S3 segment of the proximal tubules. Transports D-glucose into endometrial epithelial cells, controlling glycogen synthesis and nutritional support for the embryo as well as the decidual transformation of endometrium prior to conception. Acts as a water channel enabling passive water transport in response to the osmotic gradient

								created upon sugar and Na+ uptake. Has high water conductivity comparable to aquaporins and therefore is expected to play an important role in transepithelial water permeability, especially in the small intestine (By similarity).
PB.2610.2|60c22a|path0:1-2499(+)|transcript/7260|m.4799	2.0509	-3.047	2.1128	8.78E-05	P53790	*slc5a1*	78.681	^
PB.9307.1|transcript/5854:1-2788(+)|transcript/5854|m.15112	1.058944	-2.463	1.6343	0.00067	O35400	*sult2b1*	48.921	Sulfotransferase 2B1, Sulfotransferase that utilizes 3'-phospho-5'-adenylyl sulfate (PAPS) as sulfonate donor to catalyse the sulfate conjugation. Preferentially sulfonates cholesterol (Shimizu et al 2003).
PB.1854.1|466cde|path0:1-1088(+)|transcript/22568|m.3394	1	-7.182	6.7279	5.27E-05	Q7YRQ8	*tfpi2*	35.16	Tissue factor pathway inhibitor 2, May play a role in the regulation of plasmin-mediated matrix remodelling. Inhibits trypsin, plasmin, factor VIIa/tissue factor and weakly factor Xa. Has no effect on thrombin.

PB.7443.1|transcript/15548:1-1797(+)|transcript/15548|m.13231	1.032928	-4.424	2.7361	3.12E-07	P04047	*tk1*	82.063	Thymidine kinase, cytosolic, Cell-cycle-regulated enzyme of importance in nucleotide metabolism. Catalyses the first enzymatic step in the salvage pathway converting thymidine into thymidine monophosphate. (Two forms have been identified in animal cells, one in cytosol and one in mitochondria).
PB.8961.1|transcript/3826:1-3036(+)|transcript/3826|m.14712	1	-3.035	1.7323	9.61E-05	P10039	*tnc*	72.269	Tenascin, Extracellular matrix protein implicated in guidance of migrating neurons as well as axons during development, synaptic plasticity as well as neuronal regeneration. Ligand for integrins alpha-8/beta-1, alpha-9/beta-1, alpha-V/beta-3 and alpha-V/beta-6.
PB.8375.1|transcript/22523:1-1083(+)|transcript/22523|m.14140	1.07585	-2.315	5.7375	0.00015	Q8IZW8	*tns4*	63.158	Tensin-4, Promotes EGF-induced cell migration by displacing tensin TNS3 from the cytoplasmic tail of integrin ITGB1 which results in dissociation of TNS3 from focal adhesions, disassembly of actin stress fibers and initiation of cell migration (Katz et al. 2007).
PB.3293.1|7a863a|path0:1-2842(+)|transcript/5562|m.6070	2.67454	-3.912	0.2536	0.00029	Q9ULW0	*tpx2*	52.45	Targeting protein for Xklp2, Spindle assembly factor required for normal assembly of mitotic spindles. Required for normal assembly of microtubules during apoptosis. Required for chromatin and/or kinetochore dependent microtubule nucleation. Mediates AURKA localization to spindle microtubules (Bird & Hyman 2008; Moss, Wilde & Lane 2009).
PB.5200.2|c35005|path13:14-2538(+)|transcript/17737|m.9612	17.85339	-2.154	1.6112	0.00106	P19224	*ugt1a6*	82	UDP-glucuronosyltransferase 1-6, UDPGT is of major importance in the conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds. This isoform has specificity for phenols. Isoform 3 lacks transferase activity but acts as a negative regulator of isoform 1 (By similarity).
PB.2459.1|5bf7a7|path0:1-4382(+)|transcript/736|m.4458	1.104758	-4.323	3.7225	0.00019	Q4I1B1	*vac8*	37.548	Vacuolar protein 8, Functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole.

PB.5995.1|e4623f|path4:1-2089(+)|transcript/12938|m.11182	1.300352	-5.217	3.3217	0.00105	Q9TT15	*vdac1*	95.238	Voltage-dependent anion-selective channel protein 1, Forms a channel through the mitochondrial outer membrane and also the plasma membrane. The channel at the outer mitochondrial membrane allows diffusion of small hydrophilic molecules; in the plasma membrane it is involved in cell volume regulation and apoptosis. It adopts an open conformation at low or zero membrane potential and a closed conformation at potentials above 30-40 mV. The open state has a weak anion selectivity whereas the closed state is cation selective. Binds various signalling molecules, including the sphingolipid ceramide, the phospholipid phosphatidylcholine, and the sterol cholesterol. In depolarized mitochondria, acts downstream of PRKN and PINK1 to promote mitophagy or prevent apoptosis; polyubiquitination by PRKN promotes mitophagy, while monoubiquitination by PRKN decreases mitochondrial calcium influx which ultimately inhibits apoptosis. May participate in the formation of the permeability transition pore complex (PTPC) responsible for the release of mitochondrial products that triggers apoptosis. May mediate ATP export from cells. (Inhibited by nitric oxide). (By similarity).
PB.3026.1|6fad6f|path6:1-506(+)|transcript/24748|m.5523	1.502686	-6.535	8.8699	0.00011	P81401	*vip*	40.625	VIP peptides, VIP causes vasodilation, lowers arterial blood pressure, stimulates myocardial contractility, increases glycogenolysis and relaxes the smooth muscle of trachea, stomach and gall bladder. PHI also causes vasodilation.
PB.3027.1|6fad6f|path8:1-772(+)|transcript/23677|m.5525	1.927018	-8.553	8.6697	4.73E-05	P81401	*vip*	40.625	^

6.2.1.a References

Araya, S, Kratschmar, DV, Tsachaki, M, Stücheli, S, Beck, KR & Odermatt, A 2017, 'DHRS7 (SDR34C1) - A new player in the regulation of androgen receptor function by inactivation of 5α-dihydrotestosterone?', J Steroid Biochem Mol Biol, vol. 171, pp. 288-95.

Baris, Y, Taylor, MRG, Aria, V & Yeeles, JTP 2022, 'Fast and efficient DNA replication with purified human proteins', Nature, vol. 606, no. 7912, pp. 204-10.

Bird, AW & Hyman, AA 2008, 'Building a spindle of the correct length in human cells requires the interaction between TPX2 and Aurora A', J Cell Biol, vol. 182, no. 2, pp. 289-300.

Castleman, VH, Romio, L, Chodhari, R, Hirst, RA, de Castro, SC, Parker, KA, Ybot-Gonzalez, P, Emes, RD, Wilson, SW, Wallis, C, Johnson, CA, Herrera, RJ, Rutman, A, Dixon, M, Shoemark, A, Bush, A, Hogg, C, Gardiner, RM, Reish, O, Greene, ND, O'Callaghan, C, Purton, S, Chung, EM & Mitchison, HM 2009, 'Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities', Am J Hum Genet, vol. 84, no. 2, pp. 197-209.

Choudhury, R, Bonacci, T, Arceci, A, Lahiri, D, Mills, CA, Kernan, JL, Branigan, TB, DeCaprio, JA, Burke, DJ & Emanuele, MJ 2016, 'APC/C and SCF(cyclin F) Constitute a Reciprocal Feedback Circuit Controlling S-Phase Entry', Cell Rep, vol. 16, no. 12, pp. 3359-72.

Chuang, TD, Chen, SJ, Lin, FF, Veeramani, S, Kumar, S, Batra, SK, Tu, Y & Lin, MF 2010, 'Human prostatic acid phosphatase, an authentic tyrosine phosphatase, dephosphorylates ErbB-2 and regulates prostate cancer cell growth', J Biol Chem, vol. 285, no. 31, pp. 23598-606.

D'Angiolella, V, Donato, V, Forrester, FM, Jeong, YT, Pellacani, C, Kudo, Y, Saraf, A, Florens, L, Washburn, MP & Pagano, M 2012, 'Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair', CELL, vol. 149, no. 5, pp. 1023-34.

D'Angiolella, V, Donato, V, Vijayakumar, S, Saraf, A, Florens, L, Washburn, MP, Dynlacht, B & Pagano, M 2010, 'SCF(Cyclin F) controls centrosome homeostasis and mitotic fidelity through CP110 degradation', Nature, vol. 466, no. 7302, pp. 138-42.

Di Fiore, B & Pines, J 2007, 'Emi1 is needed to couple DNA replication with mitosis but does not regulate activation of the mitotic APC/C', J Cell Biol, vol. 177, no. 3, pp. 425-37.

Ducoux, M, Urbach, S, Baldacci, G, Hübscher, U, Koundrioukoff, S, Christensen, J & Hughes, P 2001, 'Mediation of proliferating cell nuclear antigen (PCNA)-dependent DNA replication through a conserved p21(Cip1)-like PCNA-binding motif present in the third subunit of human DNA polymerase delta', J Biol Chem, vol. 276, no. 52, pp. 49258-66.

Fiermonte, G, Dolce, V, David, L, Santorelli, FM, Dionisi-Vici, C, Palmieri, F & Walker, JE 2003, 'The mitochondrial ornithine transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution of two human isoforms', J Biol Chem, vol. 278, no. 35, pp. 32778-83.

Frye, JJ, Brown, NG, Petzold, G, Watson, ER, Grace, CR, Nourse, A, Jarvis, MA, Kriwacki, RW, Peters, JM, Stark, H & Schulman, BA 2013, 'Electron microscopy structure of human APC/C(CDH1)-EMI1 reveals multimodal mechanism of E3 ligase shutdown', Nat Struct Mol Biol, vol. 20, no. 7, pp. 827-35.

Gagné, JF, Montminy, V, Belanger, P, Journault, K, Gaucher, G & Guillemette, C 2002, 'Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38)', Mol Pharmacol, vol. 62, no. 3, pp. 608-17.

Gbadegesin, RA, Hall, G, Adeyemo, A, Hanke, N, Tossidou, I, Burchette, J, Wu, G, Homstad, A, Sparks, MA, Gomez, J, Jiang, R, Alonso, A, Lavin, P, Conlon, P, Korstanje, R, Stander, MC, Shamsan, G, Barua, M, Spurney, R, Singhal, PC, Kopp, JB, Haller, H, Howell, D, Pollak, MR, Shaw, AS, Schiffer, M & Winn, MP 2014, 'Mutations in the gene that encodes the F-actin binding protein anillin cause FSGS', J Am Soc Nephrol, vol. 25, no. 9, pp. 1991-2002.

Girard, H, Lévesque, E, Bellemare, J, Journault, K, Caillier, B & Guillemette, C 2007, 'Genetic diversity at the UGT1 locus is amplified by a novel 3' alternative splicing mechanism leading to nine additional UGT1A proteins that act as regulators of glucuronidation activity', Pharmacogenet Genomics, vol. 17, no. 12, pp. 1077-89.

Golej, DL, Askari, B, Kramer, F, Barnhart, S, Vivekanandan-Giri, A, Pennathur, S & Bornfeldt, KE 2011, 'Long-chain acyl-CoA synthetase 4 modulates prostaglandin E_2 release from human arterial smooth muscle cells', J Lipid Res, vol. 52, no. 4, pp. 782-93.

Hiroyama, M & Takenawa, T 1999, 'Isolation of a cDNA encoding human lysophosphatidic acid phosphatase that is involved in the regulation of mitochondrial lipid biosynthesis', J Biol Chem, vol. 274, no. 41, pp. 29172-80.

Hori, T, Haraguchi, T, Hiraoka, Y, Kimura, H & Fukagawa, T 2003, 'Dynamic behavior of Nuf2-Hec1 complex that localizes to the centrosome and centromere and is essential for mitotic progression in vertebrate cells', J Cell Sci, vol. 116, no. Pt 16, pp. 3347-62.

Hou, F & Zou, H 2005, 'Two human orthologues of Eco1/Ctf7 acetyltransferases are both required for proper sister-chromatid cohesion', Mol Biol Cell, vol. 16, no. 8, pp. 3908-18.

Itäaho, K, Mackenzie, PI, Ikushiro, S, Miners, JO & Finel, M 2008, 'The configuration of the 17-hydroxy group variably influences the glucuronidation of beta-estradiol and epiestradiol by human UDP-glucuronosyltransferases', Drug Metab Dispos, vol. 36, no. 11, pp. 2307-15.

Jenkyn-Bedford, M, Jones, ML, Baris, Y, Labib, KPM, Cannone, G, Yeeles, JTP & Deegan, TD 2021, 'A conserved mechanism for regulating replisome disassembly in eukaryotes', Nature, vol. 600, no. 7890, pp. 743-7.

Ji, W, Luo, Y, Ahmad, E & Liu, ST 2018, 'Direct interactions of mitotic arrest deficient 1 (MAD1) domains with each other and MAD2 conformers are required for mitotic checkpoint signaling', J Biol Chem, vol. 293, no. 2, pp. 484-96.

Jones, ML, Baris, Y, Taylor, MRG & Yeeles, JTP 2021, 'Structure of a human replisome shows the organisation and interactions of a DNA replication machine', Embo j, vol. 40, no. 23, p. e108819.

Katz, M, Amit, I, Citri, A, Shay, T, Carvalho, S, Lavi, S, Milanezi, F, Lyass, L, Amariglio, N, Jacob-Hirsch, J, Ben-Chetrit, N, Tarcic, G, Lindzen, M, Avraham, R, Liao, YC, Trusk, P, Lyass, A, Rechavi, G, Spector, NL, Lo, SH, Schmitt, F, Bacus, SS & Yarden, Y 2007, 'A reciprocal tensin-3-cten switch mediates EGF-driven mammary cell migration', Nat Cell Biol, vol. 9, no. 8, pp. 961-9.

LaBaer, J, Garrett, MD, Stevenson, LF, Slingerland, JM, Sandhu, C, Chou, HS, Fattaey, A & Harlow, E 1997, 'New functional activities for the p21 family of CDK inhibitors', Genes Dev, vol. 11, no. 7, pp. 847-62.

Lee, A, Rayner, SL, Gwee, SSL, De Luca, A, Shahheydari, H, Sundaramoorthy, V, Ragagnin, A, Morsch, M, Radford, R, Galper, J, Freckleton, S, Shi, B, Walker, AK, Don, EK, Cole, NJ, Yang, S, Williams, KL, Yerbury, JJ, Blair, IP, Atkin, JD, Molloy, MP & Chung, RS 2018, 'Pathogenic mutation in the ALS/FTD gene, CCNF, causes elevated Lys48-linked ubiquitylation and defective autophagy', Cell Mol Life Sci, vol. 75, no. 2, pp. 335-54.

Lemma-Gray, P, Valusová, E, Carroll, CA, Weintraub, ST, Musatov, A & Robinson, NC 2008, 'Subunit analysis of bovine heart complex I by reversed-phase high-performance liquid chromatography, electrospray ionization-tandem mass spectrometry, and matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry', Anal Biochem, vol. 382, no. 2, pp. 116-21.

Lépine, J, Bernard, O, Plante, M, Têtu, B, Pelletier, G, Labrie, F, Bélanger, A & Guillemette, C 2004, 'Specificity and regioselectivity of the conjugation of estradiol, estrone, and their catecholestrogen and methoxyestrogen metabolites by human uridine diphospho-glucuronosyltransferases expressed in endometrium', J Clin Endocrinol Metab, vol. 89, no. 10, pp. 5222-32.

Luo, X, Tang, Z, Xia, G, Wassmann, K, Matsumoto, T, Rizo, J & Yu, H 2004, 'The Mad2 spindle checkpoint protein has two distinct natively folded states', Nat Struct Mol Biol, vol. 11, no. 4, pp. 338-45.

Machida, YJ & Dutta, A 2007, 'The APC/C inhibitor, Emi1, is essential for prevention of rereplication', Genes Dev, vol. 21, no. 2, pp. 184-94.

Miller, JJ, Summers, MK, Hansen, DV, Nachury, MV, Lehman, NL, Loktev, A & Jackson, PK 2006, 'Emi1 stably binds and inhibits the anaphase-promoting complex/cyclosome as a pseudosubstrate inhibitor', Genes Dev, vol. 20, no. 17, pp. 2410-20.

Monné, M, Miniero, DV, Daddabbo, L, Robinson, AJ, Kunji, ER & Palmieri, F 2012, 'Substrate specificity of the two mitochondrial ornithine carriers can be swapped by single mutation in substrate binding site', J Biol Chem, vol. 287, no. 11, pp. 7925-34.

Moss, DK, Wilde, A & Lane, JD 2009, 'Dynamic release of nuclear RanGTP triggers TPX2-dependent microtubule assembly during the apoptotic execution phase', J Cell Sci, vol. 122, no. Pt 5, pp. 644-55.

Nakahara, K, Ohkuni, A, Kitamura, T, Abe, K, Naganuma, T, Ohno, Y, Zoeller, RA & Kihara, A 2012, 'The Sjögren-Larsson syndrome gene encodes a hexadecenal dehydrogenase of the sphingosine 1-phosphate degradation pathway', Mol Cell, vol. 46, no. 4, pp. 461-71.

Ohkuni, A, Ohno, Y & Kihara, A 2013, 'Identification of acyl-CoA synthetases involved in the mammalian sphingosine 1-phosphate metabolic pathway', Biochem Biophys Res Commun, vol. 442, no. 3-4, pp. 195-201.

Peng, H, Talebzadeh-Farrooji, M, Osborne, MJ, Prokop, JW, McDonald, PC, Karar, J, Hou, Z, He, M, Kebebew, E, Orntoft, T, Herlyn, M, Caton, AJ, Fredericks, W, Malkowicz, B, Paterno, CS, Carolin, AS, Speicher, DW, Skordalakes, E, Huang, Q, Dedhar, S, Borden, KLB & Rauscher, FJ, 3rd 2014, 'LIMD2 is a small LIM-only protein overexpressed in metastatic lesions that regulates cell motility and tumor progression by directly binding to and activating the integrin-linked kinase', Cancer Res, vol. 74, no. 5, pp. 1390-403.

Rzechorzek, NJ, Hardwick, SW, Jatikusumo, VA, Chirgadze, DY & Pellegrini, L 2020, 'CryoEM structures of human CMG-ATPγS-DNA and CMG-AND-1 complexes', Nucleic Acids Res, vol. 48, no. 12, pp. 6980-95.

Sazanov, LA, Peak-Chew, SY, Fearnley, IM & Walker, JE 2000, 'Resolution of the membrane domain of bovine complex I into subcomplexes: implications for the structural organization of the enzyme', Biochemistry, vol. 39, no. 24, pp. 7229-35.

Shimizu, C, Fuda, H, Yanai, H & Strott, CA 2003, 'Conservation of the hydroxysteroid sulfotransferase SULT2B1 gene structure in the mouse: pre- and postnatal expression, kinetic analysis of isoforms, and comparison with prototypical SULT2A1', Endocrinology, vol. 144, no. 4, pp. 1186-93.

Sneitz, N, Bakker, CT, de Knegt, RJ, Halley, DJ, Finel, M & Bosma, PJ 2010, 'Crigler-Najjar syndrome in The Netherlands: identification of four novel UGT1A1 alleles, genotype-phenotype correlation, and functional analysis of 10 missense mutants', Hum Mutat, vol. 31, no. 1, pp. 52-9.

Sneitz, N, Vahermo, M, Mosorin, J, Laakkonen, L, Poirier, D & Finel, M 2013, 'Regiospecificity and stereospecificity of human UDP-glucuronosyltransferases in the glucuronidation of estriol, 16-epiestriol, 17-epiestriol, and 13-epiestradiol', Drug Metab Dispos, vol. 41, no. 3, pp. 582-91.

Stambergova, H, Skarydova, L, Dunford, JE & Wsol, V 2014, 'Biochemical properties of human dehydrogenase/reductase (SDR family) member 7', Chem Biol Interact, vol. 207, pp. 52-7.

Štambergová, H, Zemanová, L, Lundová, T, Malčeková, B, Skarka, A, Šafr, M & Wsól, V 2016, 'Human DHRS7, promising enzyme in metabolism of steroids and retinoids?', J Steroid Biochem Mol Biol, vol. 155, no. Pt A, pp. 112-9.

Tanaka, M, Kishi, Y, Takanezawa, Y, Kakehi, Y, Aoki, J & Arai, H 2004, 'Prostatic acid phosphatase degrades lysophosphatidic acid in seminal plasma', FEBS Lett, vol. 571, no. 1-3, pp. 197-204.

The UniProt Consortium 2021, 'UniProt: the universal protein knowledgebase in 2021', Nucleic Acids Research, vol. 49, no. D1, pp. D480-D489.

The UniProt Consortium 2023, 'UniProt: the Universal Protein Knowledgebase in 2023', Nucleic Acids Research, vol. 51, no. D1, pp. D523-D31.

Tsuji, T, Ficarro, SB & Jiang, W 2006, 'Essential role of phosphorylation of MCM2 by Cdc7/Dbf4 in the initiation of DNA replication in mammalian cells', Mol Biol Cell, vol. 17, no. 10, pp. 4459-72.

Udomuksorn, W, Elliot, DJ, Lewis, BC, Mackenzie, PI, Yoovathaworn, K & Miners, JO 2007, 'Influence of mutations associated with Gilbert and Crigler-Najjar type II syndromes on the glucuronidation kinetics of bilirubin and other UDP-glucuronosyltransferase 1A substrates', Pharmacogenet Genomics, vol. 17, no. 12, pp. 1017-29.

Vega, H, Waisfisz, Q, Gordillo, M, Sakai, N, Yanagihara, I, Yamada, M, van Gosliga, D, Kayserili, H, Xu, C, Ozono, K, Jabs, EW, Inui, K & Joenje, H 2005, 'Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion', Nat Genet, vol. 37, no. 5, pp. 468-70.

Verschuren, EW, Ban, KH, Masek, MA, Lehman, NL & Jackson, PK 2007, 'Loss of Emi1-dependent anaphase-promoting complex/cyclosome inhibition deregulates E2F target expression and elicits DNA damage-induced senescence', Mol Cell Biol, vol. 27, no. 22, pp. 7955-65.

Walter, D, Hoffmann, S, Komseli, ES, Rappsilber, J, Gorgoulis, V & Sørensen, CS 2016, 'SCF(Cyclin F)-dependent degradation of CDC6 suppresses DNA re-replication', Nat Commun, vol. 7, p. 10530.

Wang, W & Kirschner, MW 2013, 'Emi1 preferentially inhibits ubiquitin chain elongation by the anaphase-promoting complex', Nat Cell Biol, vol. 15, no. 7, pp. 797-806.

Williams, KL, Topp, S, Yang, S, Smith, B, Fifita, JA, Warraich, ST, Zhang, KY, Farrawell, N, Vance, C, Hu, X, Chesi, A, Leblond, CS, Lee, A, Rayner, SL, Sundaramoorthy, V, Dobson-Stone, C, Molloy, MP, van Blitterswijk, M, Dickson, DW, Petersen, RC, Graff-Radford, NR, Boeve, BF, Murray, ME, Pottier, C, Don, E, Winnick, C, McCann, EP, Hogan, A, Daoud, H, Levert, A, Dion, PA, Mitsui, J, Ishiura, H, Takahashi, Y, Goto, J, Kost, J, Gellera, C, Gkazi, AS, Miller, J, Stockton, J, Brooks, WS, Boundy, K, Polak, M, Muñoz-Blanco, JL, Esteban-Pérez, J, Rábano, A, Hardiman, O, Morrison, KE, Ticozzi, N, Silani, V, de Belleroche, J, Glass, JD, Kwok, JB, Guillemin, GJ, Chung, RS, Tsuji, S, Brown, RH, Jr., García-Redondo, A, Rademakers, R, Landers, JE, Gitler, AD, Rouleau, GA, Cole, NJ, Yerbury, JJ, Atkin, JD, Shaw, CE, Nicholson, GA & Blair, IP 2016, 'CCNF mutations in amyotrophic lateral sclerosis and frontotemporal dementia', Nat Commun, vol. 7, p. 11253.

Wolan, DW, Greasley, SE, Beardsley, GP & Wilson, IA 2002, 'Structural insights into the avian AICAR transformylase mechanism', Biochemistry, vol. 41, no. 52, pp. 15505-13.

Zemanová, L, Kirubakaran, P, Pato, IH, Štambergová, H & Vondrášek, J 2017, 'The identification of new substrates of human DHRS7 by molecular modeling and in vitro testing', Int J Biol Macromol, vol. 105, no. Pt 1, pp. 171-82.

Zhang, Z, Ostanin, K & Van Etten, RL 1997, 'Covalent modification and site-directed mutagenesis of an active site tryptophan of human prostatic acid phosphatase', Acta Biochim Pol, vol. 44, no. 4, pp. 659-72.

Zhao, WM & Fang, G 2005, 'Anillin is a substrate of anaphase-promoting complex/cyclosome (APC/C) that controls spatial contractility of myosin during late cytokinesis', J Biol Chem, vol. 280, no. 39, pp. 33516-24.

6.2.2 Workflow - Bioinformatic Pipeline

6.2.2.a Online Repositories

Finalised bioinformatic pipeline methods used in each chapter are available, with brief explanations of their purpose, on GitHub.: https://github.com/Carmel-src/T.adelaidensis_SuppInfo

Individual scripts and datasets have been uploaded to this repository and are linked to in order, where relevant in the full workflow document (Appendix 2), which is also provided as a separate pdf. These methods include hard-coded variables and directories (across multiple machines), and often assume specific filename conventions, or that a specific structure of directories, program modules and environments are in place.

6.2.2.b Additional Uploaded Data Files

The following files on GitHub contain the list of all upregulated and downregulated genes with an annotated gene symbol based on BLASTx hit, in the same format as provided in Table 4-3 and Table 4-4 (which only list genes *also* found on the genes of interest list).

Downregulated genes (regardless of annotation status)

(N = 19 annotated out of 25 downregulated)

Upregulated genes (regardless of annotation status)

(N = 99 annotated out of 189 upregulated)

Full table of all annotated transcripts with expression data where relevant

(Transcripts: N = 12,532 out of 13,882 transcripts with a BLASTx hit)

(Genes used for expression analysis: N = 8,906 out of 9,813 genes successfully annotated)

6.2.3 Gene Ontology Overrepresentation Analysis - Significantly downregulated Genes

Table 6-3: Gene Ontology Overdispersion analysis of downregulated genes in *Tiliqua adelaidensis* kidney tissue in autumn compared to spring. Poly(A) selected mRNA was extracted and analysed using Illumina Hi-seq PE reads and quantified against full-length transcripts sequenced from G6, one of the six included individuals and used as a reference. BLASTx hits to gene ID were used as the input IDs in PANTHER 18.0 online portal so that enrichment could be performed on differentially expressed genes using genes with a BLASTx hit as the custom reference list. Panther run using Fishers exact test and calculate false discovery rate as the correction. Only gene IDs with a P value <0.05 are shown. Shaded areas of table indicate related groups of GO terms.

GO biological process complete	**expected**	**raw P value**	**FDR**
Unclassified	3.28	5.32E-76	6.68E-72
organelle organization	13.17	7.25E-07	2.02E-04
cellular component organization	21.7	3.19E-12	2.36E-09
cellular component organization or biogenesis	23.13	4.24E-13	3.80E-10
cellular process	51.41	4.74E-49	1.99E-45
response to stress	12.44	1.33E-06	3.10E-04
response to stimulus	24.34	5.83E-14	5.63E-11
response to organic substance	9.7	5.16E-05	8.00E-03
response to chemical	12.44	1.33E-06	3.04E-04
phosphate-containing compound metabolic process	7.85	3.61E-04	4.67E-02
phosphorus metabolic process	8.03	3.63E-04	4.65E-02
cellular metabolic process	27.39	2.26E-16	2.58E-13
metabolic process	35.3	1.90E-23	3.98E-20
nitrogen compound transport	8.36	2.23E-04	2.98E-02
transport	15.45	3.68E-08	1.28E-05
establishment of localization	16.12	1.10E-08	4.31E-06
localization	18.36	6.67E-10	3.35E-07
organic substance transport	9.98	3.10E-05	5.19E-03
regulation of molecular function	8.77	1.37E-04	2.01E-02
biological regulation	37.18	1.94E-25	6.08E-22
regulation of biological quality	10.25	3.24E-05	5.22E-03
signal transduction	12.53	1.29E-06	3.12E-04
cellular response to stimulus	19.32	1.54E-10	9.18E-08
regulation of cellular process	33.12	2.56E-21	4.01E-18
regulation of biological process	35.55	1.05E-23	2.62E-20
signalling	13.2	7.29E-07	1.95E-04
cell communication	13.85	2.34E-07	7.35E-05
regulation of DNA-templated transcription	9.98	3.10E-05	5.12E-03
regulation of RNA biosynthetic process	10.05	3.09E-05	5.32E-03
regulation of macromolecule biosynthetic process	12.56	1.28E-06	3.16E-04
regulation of macromolecule metabolic process	20.46	3.23E-11	2.14E-08
regulation of metabolic process	22.89	5.02E-13	4.20E-10
regulation of biosynthetic process	13.53	4.07E-07	1.25E-04
regulation of RNA metabolic process	11.48	6.66E-06	1.37E-03
regulation of nucleobase-containing compound metabolic process	12.73	1.28E-06	3.27E-04
regulation of nitrogen compound metabolic process	19.02	1.79E-10	9.74E-08
regulation of cellular metabolic process	18.54	3.64E-10	1.90E-07

regulation of primary metabolic process	19.73	7.38E-11	4.63E-08
regulation of cellular biosynthetic process	13.1	7.22E-07	2.06E-04
regulation of gene expression	15.63	2.10E-08	7.54E-06
cellular response to chemical stimulus	8.62	1.40E-04	1.93E-02
establishment of localization in cell	8.73	1.38E-04	1.96E-02
cellular localization	12.57	1.28E-06	3.21E-04
organic substance biosynthetic process	12.9	7.55E-07	1.98E-04
biosynthetic process	13.17	7.25E-07	1.98E-04
organic substance metabolic process	33.41	1.38E-21	2.48E-18
organic substance catabolic process	8.56	2.32E-04	3.04E-02
catabolic process	9.79	5.28E-05	8.08E-03
protein-containing complex organization	8.78	1.37E-04	1.98E-02
system development	10.7	1.96E-05	3.42E-03
anatomical structure development	15.85	1.96E-08	7.44E-06
developmental process	17.64	1.42E-09	6.36E-07
multicellular organism development	12.27	2.22E-06	4.97E-04
multicellular organismal process	18.22	6.80E-10	3.28E-07
regulation of signal transduction	10.31	1.89E-05	3.40E-03
regulation of response to stimulus	13.44	4.13E-07	1.23E-04
regulation of cell communication	11.46	6.62E-06	1.39E-03
regulation of signalling	11.52	6.75E-06	1.37E-03
small molecule metabolic process	8.86	1.39E-04	1.93E-02
cellular nitrogen compound biosynthetic process	8.68	1.38E-04	1.95E-02
cellular nitrogen compound metabolic process	16.88	5.30E-09	2.30E-06
nitrogen compound metabolic process	29.58	5.36E-18	6.73E-15
cellular biosynthetic process	11.27	6.53E-06	1.41E-03
negative regulation of cellular process	16.25	1.04E-08	4.36E-06
negative regulation of biological process	18.17	6.93E-10	3.22E-07
animal organ development	8.98	1.43E-04	1.96E-02
regulation of protein metabolic process	10.09	3.10E-05	5.05E-03
regulation of multicellular organismal process	9	1.45E-04	1.95E-02
positive regulation of nitrogen compound metabolic process	11.46	6.62E-06	1.41E-03
positive regulation of metabolic process	13.86	2.33E-07	7.49E-05
positive regulation of biological process	21.7	3.19E-12	2.50E-09
negative regulation of nitrogen compound metabolic process	8.13	3.73E-04	4.73E-02
negative regulation of metabolic process	10.38	1.87E-05	3.39E-03
regulation of cellular component organization	9.12	8.45E-05	1.28E-02
protein modification process	10	3.09E-05	5.25E-03
macromolecule modification	11.01	1.12E-05	2.19E-03
macromolecule metabolic process	24.92	2.08E-14	2.18E-11
protein metabolic process	16.17	1.07E-08	4.34E-06
organonitrogen compound metabolic process	21.37	6.42E-12	4.47E-09
primary metabolic process	32.03	2.61E-20	3.63E-17
positive regulation of cellular metabolic process	10.79	1.11E-05	2.22E-03
positive regulation of cellular process	19.32	1.54E-10	8.76E-08
RNA metabolic process	8.26	2.24E-04	2.96E-02
nucleic acid metabolic process	10.77	1.12E-05	2.16E-03
nucleobase-containing compound metabolic process	13.76	4.25E-07	1.24E-04
organic cyclic compound metabolic process	15.77	1.99E-08	7.33E-06

heterocycle metabolic process	14.59	1.27E-07	4.31E-05
cellular aromatic compound metabolic process	14.65	1.29E-07	4.27E-05
cellular component assembly	10.45	1.85E-05	3.47E-03
cellular component biogenesis	12.07	2.25E-06	4.97E-04
negative regulation of macromolecule metabolic process	9.37	8.69E-05	1.28E-02
positive regulation of macromolecule metabolic process	12.51	1.30E-06	3.08E-04
cell differentiation	10.52	1.86E-05	3.43E-03
cellular developmental process	10.68	1.94E-05	3.44E-03
gene expression	9.32	8.56E-05	1.28E-02
protein localization	9.59	5.12E-05	8.04E-03
cellular macromolecule localization	9.61	5.12E-05	8.13E-03
macromolecule localization	11.11	1.16E-05	2.20E-03

6.3 Appendix 3: Selection of "Genes of Interest" from the NCBI Gene Database

Initially a selection of terms was searched within Squamata to collect representatives of genes from species more closely related to *T. adelaidensis*. This search was extended to all Sauria with the same functional terms so that any genes not annotated in scaled reptiles would not be excluded (NCBI Gene database accessed 18 May 2020) (Table 6-4). Using these initial test searches, it was confirmed that Squamata search returns are a complete sub-set of Sauria returns with no loss of data; that is, there are no Squamata results that are not also assigned to Sauria in NCBI. Therefore, the final refined search query as displayed in Chapter 3.3.4 was conducted for all Sauria. Squamate organisms were then favoured when duplicates were removed.

Table 6-4: Total results for an open search with no other query terms for taxa groups of interest. NCBI Gene database accessed 14 June 2020.

Organism	# Total results
Searched June 14, 2020	
(Sauria[Organism])	1931889
(Reptil*[Organism])	542802
(Squamata[Organism])	278726

The NCBI taxonomy tree applies the term 'Sauria' as a clade including Archelosauria (Archosauria and Testudines) and Lepidosauria (Sphenodontia and Squamata). This term was used to maximise results as above. Many genes are represented by *Gallus gallus* in the final list (favoured over other avian species because of a well-annotated genome), and the chicken could be considered the most 'mainstream' model organism represented here.

A list of all Sauria organisms returned by NCBI search and total number of genes represented in each, before and after duplicate removal, is available on GitHub. Duplicates were removed based on gene name. Order was determined first by taxonomic distance and availability of genomic data; remaining taxa are sorted by default (alphabetical). The order in this table represents the (descending) order in which gene duplicates were favoured; the first occurrence of each gene name was retained. Search conducted 15 June 2020 (Table 6-5).

Table 6-5: Summary totals for NCBI search totals (metadata for Table 6-6)

Summary NCBI search totals	
Genes Returned for 104 Taxa	40995
Duplicate Gene Symbols	32,993
Unique Gene symbols remaining for 102 Taxa	8,002
Removed LOC# (loci assigned a unique number and with putative genes in the description but not assigned a gene symbol)	7,009
Unique Gene Symbols Retained for 102 taxa	993

6.3.1 Annotation: Full Table of Potential Genes of Interest with Count of Transcripts Identified in *Tiliqua adelaidensis*

List of candidate genes of interest filtered from NCBI search for genes represented by reptile species (including birds) with roles in water homeostasis and suspected renal functions, which were also identified in *Tiliqua adelaidensis* transcript dataset. Table shows 393 out of the 993 genes identified in NCBI database search which were identified in *T. adelaidensis* based on comparison of Gene name IDs from a BLASTx search against the UniProt Swiss-Prot database. BLASTx parameters *-outfmt 6, -e-value=1x10^{-5}, -max target seqs=1, -max hsps=1*. Comparison of these two lists is based on gene symbol only (BLASTx may have matched to any representative organism available to assign a gene name), organism names are retained from the NCBI gene search here to illustrate the proportion of results which was not represented in squamates. Presence column indicates the number of *T. adelaidensis* transcript isoforms matched to a particular gene symbol only. Genes are listed alphabetically so that gene families are grouped together, and then by number of isoform matches. Org. name identifies the most relevant organism NCBI search returned a result for the gene name as per database search query after duplicates were removed.
Shading of transcript presence: >10 matches = green, >5-10 matches = yellow, 1-5 matches = red.

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
abcc4	ATP binding cassette subfamily C member 4	*Gallus gallus*	418791	2
aco1	aconitase 1, soluble	*Gallus gallus*	373916	4
acsbg2	acyl-CoA synthetase bubblegum family member 2	*Gallus gallus*	420090	3
acta1	actin alpha 1, skeletal muscle	*Gallus gallus*	421534	2
actn1	actinin, alpha 1	*Gallus gallus*	373918	2
ada	adenosine deaminase	*Gallus gallus*	419194	1
adck5	aarF domain-containing kinase 5	*Alligator mississippiensis*	102567875	1
adi1	acireductone dioxygenase 1	*Gallus gallus*	421918	3
adora1	adenosine A1 receptor	*Gallus gallus*	374212	1
adss2	adenylosuccinate synthase 2	*Gallus gallus*	428579	3
agt	angiotensinogen	*Gallus gallus*	421543	1
ak1	adenylate kinase 1	*Gallus gallus*	396002	2
alcam	activated leukocyte cell adhesion molecule	*Gallus gallus*	396092	1
aldh1a1	aldehyde dehydrogenase 1 family member A1	*Gallus gallus*	395264	13
aldob	aldolase, fructose-bisphosphate B	*Gallus gallus*	427308	12
aldoc	aldolase, fructose-bisphosphate C	*Gallus gallus*	395492	1
apoa1	apolipoprotein AI	*Gallus gallus*	396536	10
apob	apolipoprotein B	*Gallus gallus*	396535	3
apobec1	apolipoprotein B mRNA editing enzyme catalytic subunit 1	*Anas platyrhynchos*	101792029	4
aqp1	aquaporin 1 (Colton blood group)	*Protobothrops mucrosquamatus*	107284354	2

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
aqp11	aquaporin 11	*Anolis carolinensis*	100566247	1
aqp2	aquaporin 2	*Merops nubicus*	103773696	6
aqp9	aquaporin 9	*Pogona vitticeps*	110074546	1
arhgef16	Rho guanine nucleotide exchange factor 16	*Notechis scutatus*	113428811	1
arntl	aryl hydrocarbon receptor nuclear translocator like	*Gallus gallus*	374115	1
asl1	argininosuccinate lyase 1	*Gallus gallus*	396498	3
asl2	argininosuccinate lyase 2	*Gallus gallus*	417545	3
asns	asparagine synthetase (glutamine-hydrolyzing)	*Gallus gallus*	420574	1
asnsd1	asparagine synthetase domain containing 1	*Alligator mississippiensis*	102565117	2
ass1	argininosuccinate synthase 1	*Pogona vitticeps*	110079638	17
atp6v1a	ATPase H+ transporting V1 subunit A	*Gallus gallus*	395821	4
atp8b1	ATPase phospholipid transporting 8B1	*Cyanistes caeruleus*	111940982	1
b9d2	B9 domain containing 2	*Cuculus canorus*	104067755	1
baz1a	bromodomain adjacent to zinc finger domain 1A	*Chrysemys picta*	101947548	2
bckdha	branched chain keto acid dehydrogenase E1 subunit alpha	*Notechis scutatus*	113426026	1
bcl2l1	BCL2 like 1	*Gallus gallus*	373954	1
bet1l	Bet1 golgi vesicular membrane trafficking protein like	*Geospiza fortis*	102034857	1
bpi	bactericidal permeability increasing protein	*Trachemys scripta elegans*	117886185	3
btg2	BTG anti-proliferation factor 2	*Gallus gallus*	419932	2
ca2	carbonic anhydrase 2	*Gallus gallus*	396257	4
cactin	cactin, spliceosome C complex subunit	*Gekko japonicus*	107123476	1
calb1	calbindin 1	*Gallus gallus*	396519	6
cdh5	cadherin 5	*Gallus gallus*	374068	1
cdk5rap1	CDK5 regulatory subunit associated protein 1	*Opisthocomus hoazin*	104330570	2
cebpz	CCAAT enhancer binding protein zeta	*Apteryx mantelli mantelli*	106485420	1
chrna10	cholinergic receptor, nicotinic, alpha 10 (neuronal)	*Gallus gallus*	430628	1
chrna4	cholinergic receptor nicotinic alpha 4 subunit	*Gallus gallus*	395606	1
chst15	carbohydrate sulfotransferase 15	*Buceros rhinoceros silvestris*	104501097	3
clcn6	chloride voltage-gated channel 6	*Podarcis muralis*	114602802	1
clcn7	chloride voltage-gated channel 7	*Colius striatus*	104560943	1
clip4	CAP-Gly domain-containing linker protein family member 4	*Pelodiscus sinensis*	102463215	1

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
cltrn	collectrin, amino acid transport regulator	*Anolis carolinensis*	100555401	11
clu	clusterin	*Gallus gallus*	395722	4
cry2	cryptochrome circadian regulator 2	*Falco cherrug*	102057234	2
cryab	crystallin alpha B	*Gallus gallus*	396089	1
csrp1	cysteine and glycine rich protein 1	*Gallus gallus*	396176	1
ctps2	CTP synthase 2	*Gallus gallus*	418620	3
ctu2	cytosolic thiouridylase subunit 2	*Python bivittatus*	103050806	2
cyp1a1	cytochrome P450, family 1, subfamily A, polypeptide 1	*Gallus gallus*	396052	1
cystm1	cysteine rich transmembrane module containing 1	*Protobothrops mucrosquamatus*	107284143	2
daglb	diacylglycerol lipase beta	*Anolis carolinensis*	103280799	1
dcxr	dicarbonyl and L-xylulose reductase	*Gallus gallus*	374066	3
ddi2	DNA damage inducible 1 homolog 2	*Chrysemys picta*	101937500	1
dhfr	dihydrofolate reductase	*Gallus gallus*	427317	1
dhtkd1	dehydrogenase E1 and transketolase domain containing 1	*Colius striatus*	104549268	3
dkk3	dickkopf WNT signaling pathway inhibitor 3	*Anolis carolinensis*	100564662	1
dnajc3	DnaJ heat shock protein family (Hsp40) member C3	*Gallus gallus*	418787	5
dpep1	dipeptidase 1	*Anolis carolinensis*	100552685	5
dym	dymeclin	*Acanthisitta chloris*	103808864	1
eif2s1	eukaryotic translation initiation factor 2 subunit alpha	*Gallus gallus*	423279	2
entpd5	ectonucleoside triphosphate diphosphohydrolase 5 (inactive)	*Gallus gallus*	423343	6
erp44	endoplasmic reticulum protein 44	*Pogona vitticeps*	110086071	5
etfa	electron transfer flavoprotein subunit alpha	*Antrostomus carolinensis*	104528613	2
etfb	electron transfer flavoprotein subunit beta	*Alligator sinensis*	112548425	2
far1	fatty acyl-CoA reductase 1	*Gallus gallus*	423028	4
fasn	fatty acid synthase	*Gallus gallus*	396061	1
fdx1	ferredoxin 1	*Gallus gallus*	373947	1
flad1	flavin adenine dinucleotide synthetase 1	*Protobothrops mucrosquamatus*	107296703	1
flt1	fms related receptor tyrosine kinase 1	*Gallus gallus*	374100	1
g3bp1	G3BP stress granule assembly factor 1	*Anolis carolinensis*	100560866	3
g3bp2	G3BP stress granule assembly factor 2	*Anolis carolinensis*	100561178	2
gaa	glucosidase alpha, acid	*Zonotrichia albicollis*	102073734	1

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
gdpd2	glycerophosphodiester phosphodiesterase domain containing 2	*Anas platyrhynchos*	101792816	1
gja1	gap junction protein alpha 1	*Gallus gallus*	395278	2
gk5	glycerol kinase 5 (putative)	*Gallus gallus*	424779	1
gmps	guanine monophosphate synthase	*Cariama cristata*	104160878	2
got1	glutamic-oxaloacetic transaminase 1	*Gallus gallus*	396261	2
got2	glutamic-oxaloacetic transaminase 2	*Gallus gallus*	396533	3
gskip	GSK3B interacting protein	*Gallus gallus*	423442	2
gsta3	glutathione S-transferase alpha 3	*Gallus gallus*	414896	1
gstt1	glutathione S-transferase theta 1	*Gallus gallus*	396322	2
habp4	hyaluronan binding protein 4	*Gallus gallus*	395444	1
herpud1	homocysteine inducible ER protein with ubiquitin like domain 1	*Anolis carolinensis*	100560750	3
herpud2	HERPUD family member 2	*Podarcis muralis*	114607247	1
hid1	HID1 domain containing	*Anolis carolinensis*	100566016	1
hif1a	hypoxia inducible factor 1 subunit alpha	*Gallus gallus*	374177	1
hmgcl	3-hydroxy-3-methylglutaryl-CoA lyase	*Gallus gallus*	396316	1
hmgcs1	3-hydroxy-3-methylglutaryl-CoA synthase 1	*Gallus gallus*	396379	5
hmox1	heme oxygenase 1	*Gallus gallus*	396287	1
hprt1	hypoxanthine phosphoribosyltransferase 1	*Gallus gallus*	395653	1
hras	HRas proto-oncogene, GTPase	*Gallus gallus*	396229	3
hsf2	heat shock transcription factor 2	*Gallus gallus*	421724	4
hsf3	Heat shock factor protein 3	*Gallus gallus*	422169	1
hsp90ab1	heat shock protein 90 alpha family class B member 1	*Gallus gallus*	396188	9
hspa4l	heat shock protein family A (Hsp70) member 4 like	*Anas platyrhynchos*	101798984	2
hspa5	heat shock 70kDa protein 5 (glucose-regulated protein, 78kDa)	*Gallus gallus*	396487	2
hspa9	heat shock protein family A (Hsp70) member 9	*Anolis carolinensis*	100556293	5
hspb1	heat shock protein family B (small) member 1	*Pogona vitticeps*	110085158	1
hspb8	heat shock protein family B (small) member 8	*Chelonia mydas*	102942714	1
hykk	hydroxylysine kinase	*Python bivittatus*	103058107	1
igf1	insulin like growth factor 1	*Gallus gallus*	418090	5
ilvbl	ilvB acetolactate synthase like	*Anser cygnoides domesticus*	106049442	1
insig1	insulin induced gene 1	*Gallus gallus*	420442	1

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
ireb2	iron responsive element binding protein 2	*Gallus gallus*	427490	1
itpk1	inositol-tetrakisphosphate 1-kinase	*Gallus gallus*	423421	2
jun	Jun proto-oncogene, AP-1 transcription factor subunit	*Gallus gallus*	424673	2
kcnk1	potassium two pore domain channel subfamily K member 1	*Anolis carolinensis*	100558649	3
klf2	Kruppel like factor 2	*Gallus gallus*	420148	1
lcat	lecithin-cholesterol acyltransferase	*Gallus gallus*	396136	1
ldha	lactate dehydrogenase A	*Gallus gallus*	396221	1
ldhb	lactate dehydrogenase B	*Gallus gallus*	373997	4
lgr4	leucine rich repeat containing G protein-coupled receptor 4	*Protobothrops mucrosquamatus*	107286935	1
lhx1	LIM homeobox 1	*Gallus gallus*	396381	1
liph	lipase H	*Chrysemys picta*	101946556	1
lpl	lipoprotein lipase	*Anolis carolinensis*	100556298	2
ltv1	LTV1 ribosome biogenesis factor	*Anolis carolinensis*	100551914	2
man1b1	mannosidase alpha class 1B member 1	*Coturnix japonica*	107321628	2
map7	microtubule-associated protein 7	*Alligator sinensis*	102375976	3
mapkap1	MAPK associated protein 1	*Pogona vitticeps*	110082952	1
mcm3	minichromosome maintenance complex component 3	*Anser cygnoides domesticus*	106038116	1
mcrip1	MAPK regulated corepressor interacting protein 1	*Gallus gallus*	101749776	1
mdh1	malate dehydrogenase 1	*Gallus gallus*	421281	6
mif	macrophage migration inhibitory factor (glycosylation-inhibiting factor)	*Gallus gallus*	100857237	2
mmachc	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria	*Gallus gallus*	424597	1
mmp2	matrix metallopeptidase 2	*Gallus gallus*	386583	1
mov10	Mov10 RISC complex RNA helicase	*Gallus gallus*	419872	1
msrb1	methionine sulfoxide reductase B1	*Gallus gallus*	416540	2
mthfd2	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2, methenyltetrahydrofolate cyclohydrolase	*Gallus gallus*	426126	1
mtpn	myotrophin	*Gallus gallus*	395702	2
mtres1	mitochondrial transcription rescue factor 1	*Gekko japonicus*	107113572	2
myh9	myosin, heavy chain 9, non-muscle	*Gallus gallus*	396469	3
myl9	myosin, light chain 9, regulatory	*Gallus gallus*	396215	4
naga	alpha-N-acetylgalactosaminidase	*Gallus gallus*	396547	2

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
nfe2l1	nuclear factor, erythroid 2 like 1	*Gallus gallus*	417987	1
nfkb2	nuclear factor kappa B subunit 2	*Anolis carolinensis*	100566624	1
npc1	NPC intracellular cholesterol transporter 1	*Protobothrops mucrosquamatus*	107292297	1
npl	N-acetylneuraminate pyruvate lyase	*Gallus gallus*	429074	2
nr3c1	nuclear receptor subfamily 3 group C member 1	*Coturnix japonica*	107320036	1
nt5c2	5'-nucleotidase, cytosolic II	*Gallus gallus*	423871	6
nt5c3a	5'-nucleotidase, cytosolic IIIA	*Gallus gallus*	395080	1
ntn1	netrin 1	*Gallus gallus*	396389	1
ormdl2	ORMDL sphingolipid biosynthesis regulator 2	*Gallus gallus*	425059	5
oscp1	organic solute carrier partner 1	*Anolis carolinensis*	100558376	1
oser1	oxidative stress responsive serine rich 1	*Gekko japonicus*	107121310	2
osgin1	oxidative stress induced growth inhibitor 1	*Anolis carolinensis*	103279343	1
otc	ornithine carbamoyltransferase	*Gallus gallus*	395735	1
oxsr1	oxidative stress responsive kinase 1	*Anolis carolinensis*	100561570	1
p2rx7	purinergic receptor P2X 7	*Gallus gallus*	771952	1
p3h2	prolyl 3-hydroxylase 2	*Gallus gallus*	414143	1
p4hb	prolyl 4-hydroxylase subunit beta	*Gallus gallus*	374091	17
park7	Parkinsonism associated deglycase	*Gallus gallus*	395277	2
pck2	phosphoenolpyruvate carboxykinase 2, mitochondrial	*Gallus gallus*	396457	1
pdcd10	programmed cell death 10	*Gallus gallus*	425003	2
pdha1	pyruvate dehydrogenase E1 subunit alpha 1	*Thamnophis sirtalis*	106554165	3
pdia3	protein disulfide isomerase family A member 3	*Gallus gallus*	373899	4
pdlim3	PDZ and LIM domain 3	*Gallus gallus*	414873	3
pdlim4	PDZ and LIM domain 4	*Gallus gallus*	395643	2
pdlim7	PDZ and LIM domain 7	*Gallus gallus*	416362	2
pgm1	phosphoglucomutase 1	*Anolis carolinensis*	100562877	2
pgp	phosphoglycolate phosphatase	*Gallus gallus*	416559	1
pltp	phospholipid transfer protein	*Anolis carolinensis*	100560016	1
pparg	peroxisome proliferator-activated receptor gamma	*Gallus gallus*	373928	2
ppat	phosphoribosyl pyrophosphate amidotransferase	*Gallus gallus*	422743	1
ppm1b	protein phosphatase, Mg2+/Mn2+ dependent 1B	*Apteryx rowi*	112969203	3

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
ppp1cb	protein phosphatase 1 catalytic subunit beta	*Gallus gallus*	396019	3
prcc	proline rich mitotic checkpoint control factor	*Anolis carolinensis*	100553675	1
prdx1	peroxiredoxin 1	*Gallus gallus*	424598	4
prkrip1	PRKR interacting protein 1 (IL11 inducible)	*Gallus gallus*	417508	1
prnp	prion protein	*Gallus gallus*	396452	1
prpf4	pre-mRNA processing factor 4	*Falco cherrug*	102045807	1
prps1	phosphoribosyl pyrophosphate synthetase 1	*Alligator mississippiensis*	106736931	1
prps2	phosphoribosyl pyrophosphate synthetase 2	*Gallus gallus*	418639	3
prpsap2	phosphoribosyl pyrophosphate synthetase associated protein 2	*Gallus gallus*	416521	2
psap	prosaposin	*Gallus gallus*	395602	3
psen1	presenilin 1	*Gallus gallus*	373977	1
psen2	presenilin 2	*Gallus gallus*	374188	1
psma4	proteasome 20S subunit alpha 4	*Opisthocomus hoazin*	104337423	3
ptgs2	prostaglandin-endoperoxide synthase 2	*Gallus gallus*	396451	2
ptpn11	protein tyrosine phosphatase non-receptor type 11	*Gallus gallus*	395815	2
pum1	pumilio RNA binding family member 1	*Gallus gallus*	419554	2
qars1	glutaminyl-tRNA synthetase 1	*Apteryx mantelli mantelli*	106489403	2
ran	RAN, member RAS oncogene family	*Gallus gallus*	396193	1
rhoa	ras homolog family member A	*Gallus gallus*	395442	6
rnf11	ring finger protein 11	*Gallus gallus*	424631	2
rnf13	ring finger protein 13	*Gallus gallus*	396303	2
s100a10	S100 calcium binding protein A10	*Gallus gallus*	396506	1
scnn1a	sodium channel epithelial 1 alpha subunit	*Gallus gallus*	396050	3
scp2	sterol carrier protein 2	*Gallus gallus*	396550	6
sde2	SDE2 telomere maintenance homolog	*Podarcis muralis*	114594616	3
sdha	succinate dehydrogenase complex flavoprotein subunit A	*Gallus gallus*	395758	3
selenok	selenoprotein K	*Gallus gallus*	415995	1
selenot	selenoprotein T	*Gallus gallus*	425041	1
sephs1	selenophosphate synthetase 1	*Anolis carolinensis*	100559593	1
sephs2	selenophosphate synthetase 2	*Ficedula albicollis*	101820847	3
sepsecs	Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase	*Gallus gallus*	422808	1

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
serp1	stress associated endoplasmic reticulum protein 1	*Anolis carolinensis*	100560634	3
shq1	SHQ1, H/ACA ribonucleoprotein assembly factor	*Pogona vitticeps*	110077707	1
sirt5	sirtuin 5	*Gallus gallus*	420834	3
slc10a2	solute carrier family 10 member 2	*Anolis carolinensis*	100553566	1
slc10a3	solute carrier family 10 member 3	*Anolis carolinensis*	100565155	1
slc12a1	solute carrier family 12 member 1	*Anolis carolinensis*	100554780	6
slc12a3	solute carrier family 12 member 3	*Anolis carolinensis*	100562387	2
slc12a4	solute carrier family 12 member 4	*Anolis carolinensis*	100563811	1
slc12a7	solute carrier family 12 member 7	*Anolis carolinensis*	100558933	2
slc12a8	solute carrier family 12 member 8	*Anolis carolinensis*	100556947	7
slc13a1	solute carrier family 13 member 1	*Anolis carolinensis*	100552464	1
slc13a2	solute carrier family 13 member 2	*Anolis carolinensis*	100564744	1
slc13a3	solute carrier family 13 member 3	*Anolis carolinensis*	100563148	5
slc13a5	solute carrier family 13 member 5	*Pogona vitticeps*	110083294	2
slc15a2	solute carrier family 15 member 2	*Anolis carolinensis*	100564894	2
slc16a1	solute carrier family 16 member 1	*Anolis carolinensis*	100557889	1
slc16a13	solute carrier family 16 member 13	*Anolis carolinensis*	100557007	4
slc16a14	solute carrier family 16 member 14	*Anolis carolinensis*	100562937	1
slc16a4	solute carrier family 16 member 4	*Anolis carolinensis*	100564319	1
slc16a6	solute carrier family 16 member 6	*Anolis carolinensis*	100555486	5
slc16a9	solute carrier family 16 member 9	*Anolis carolinensis*	103279757	7
slc17a5	solute carrier family 17 member 5	*Anolis carolinensis*	100566964	4
slc18b1	solute carrier family 18 member B1	*Anolis carolinensis*	100553000	1
slc19a1	solute carrier family 19 member 1	*Anolis carolinensis*	100562129	1
slc1a5	solute carrier family 1 member 5	*Anolis carolinensis*	100562974	1
slc22a13	solute carrier family 22 member 13	*Anolis carolinensis*	100561176	3
slc22a16	solute carrier family 22 member 16	*Anolis carolinensis*	100564304	3
slc22a18	solute carrier family 22 member 18	*Anolis carolinensis*	100560561	2
slc22a2	solute carrier family 22 member 2	*Anolis carolinensis*	100560224	4
slc22a7	solute carrier family 22 member 7	*Anolis carolinensis*	100560506	3
slc23a1	solute carrier family 23 member 1	*Anolis carolinensis*	100556489	2

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
slc23a2	solute carrier family 23 member 2	*Anolis carolinensis*	100566000	2
slc23a3	solute carrier family 23 member 3	*Anolis carolinensis*	100567833	3
slc25a10	solute carrier family 25 member 10	*Anolis carolinensis*	100553900	6
slc25a11	solute carrier family 25 member 11	*Anolis carolinensis*	100554396	1
slc25a12	solute carrier family 25 member 12	*Anolis carolinensis*	100556466	1
slc25a13	solute carrier family 25 member 13	*Anolis carolinensis*	100561177	4
slc25a15	solute carrier family 25 member 15	*Anolis carolinensis*	100551983	7
slc25a17	solute carrier family 25 member 17	*Anolis carolinensis*	100556588	3
slc25a20	solute carrier family 25 member 20	*Anolis carolinensis*	100560902	4
slc25a21	solute carrier family 25 member 21	*Anolis carolinensis*	100557800	1
slc25a24	solute carrier family 25 member 24	*Anolis carolinensis*	100565644	2
slc25a26	solute carrier family 25 member 26	*Anolis carolinensis*	100567252	2
slc25a29	solute carrier family 25 member 29	*Anolis carolinensis*	100567211	3
slc25a3	solute carrier family 25 member 3	*Anolis carolinensis*	100557497	1
slc25a33	solute carrier family 25 member 33	*Anolis carolinensis*	100551968	4
slc25a34	solute carrier family 25 member 34	*Anolis carolinensis*	100556998	1
slc25a35	solute carrier family 25 member 35	*Anolis carolinensis*	100558631	1
slc25a36	solute carrier family 25 member 36	*Anolis carolinensis*	100566324	4
slc25a39	solute carrier family 25 member 39	*Anolis carolinensis*	100563736	6
slc25a4	solute carrier family 25 member 4	*Anolis carolinensis*	100563345	1
slc25a42	solute carrier family 25 member 42	*Pogona vitticeps*	110072543	1
slc25a43	solute carrier family 25 member 43	*Anolis carolinensis*	100559368	2
slc25a44	solute carrier family 25 member 44	*Pogona vitticeps*	110091224	1
slc25a45	solute carrier family 25 member 45	*Anolis carolinensis*	100552953	10
slc25a46	solute carrier family 25 member 46	*Anolis carolinensis*	100557900	1
slc25a48	solute carrier family 25 member 48	*Anolis carolinensis*	100566178	3
slc25a5	solute carrier family 25 member 5	*Anolis carolinensis*	100558779	2
slc25a51	solute carrier family 25 member 51	*Anolis carolinensis*	100556131	1
slc25a6	solute carrier family 25 member 6	*Anolis carolinensis*	100554158	3
slc26a1	solute carrier family 26 member 1	*Anolis carolinensis*	103279447	1
slc26a11	solute carrier family 26 member 11	*Anolis carolinensis*	100561225	1

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
slc26a6	solute carrier family 26 member 6	*Anolis carolinensis*	107982461	2
slc27a2	solute carrier family 27 member 2	*Gekko japonicus*	107122718	7
slc27a6	solute carrier family 27 member 6	*Anolis carolinensis*	100555843	1
slc28a2	solute carrier family 28 member 2	*Protobothrops mucrosquamatus*	107294065	1
slc29a1	solute carrier family 29 member 1 (Augustine blood group)	*Anolis carolinensis*	100552110	1
slc2a1	solute carrier family 2 member 1	*Pogona vitticeps*	110072700	1
slc2a11	solute carrier family 2 member 11	*Gekko japonicus*	107108542	11
slc2a4	solute carrier family 2 member 4	*Anolis carolinensis*	100566978	1
slc2a8	solute carrier family 2 member 8	*Anolis carolinensis*	100566203	2
slc2a9	solute carrier family 2 member 9	*Anolis carolinensis*	100558223	1
slc30a2	solute carrier family 30 member 2	*Anolis carolinensis*	100553996	3
slc30a4	solute carrier family 30 member 4	*Anolis carolinensis*	100559557	1
slc30a5	solute carrier family 30 member 5	*Anolis carolinensis*	100557079	3
slc30a6	solute carrier family 30 member 6	*Anolis carolinensis*	100563452	2
slc30a7	solute carrier family 30 member 7	*Anolis carolinensis*	100563076	4
slc30a9	solute carrier family 30 member 9	*Anolis carolinensis*	100563539	1
slc31a1	solute carrier family 31 member 1	*Anolis carolinensis*	100559623	3
slc31a2	solute carrier family 31 member 2	*Anolis carolinensis*	100565139	1
slc33a1	solute carrier family 33 member 1	*Anolis carolinensis*	100567707	2
slc34a1	solute carrier family 34 member 1	*Anolis carolinensis*	100556106	11
slc35a1	solute carrier family 35 member A1	*Anolis carolinensis*	100556879	2
slc35a2	solute carrier family 35 member A2	*Anolis carolinensis*	100564410	1
slc35a3	solute carrier family 35 member A3	*Anolis carolinensis*	100564189	2
slc35a4	solute carrier family 35 member A4	*Anolis carolinensis*	100561168	4
slc35b1	solute carrier family 35 member B1	*Anolis carolinensis*	100553382	1
slc35b2	solute carrier family 35 member B2	*Anolis carolinensis*	100552306	1
slc35b3	solute carrier family 35 member B3	*Anolis carolinensis*	100557631	1
slc35b4	solute carrier family 35 member B4	*Anolis carolinensis*	100553056	1
slc35c1	solute carrier family 35 member C1	*Anolis carolinensis*	100557069	2
slc35d2	solute carrier family 35 member D2	*Anolis carolinensis*	100555518	1
slc35e2a	solute carrier family 35 member E2A	*Chelonia mydas*	102938286	1

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
slc35e2b	solute carrier family 35 member E2B	*Anolis carolinensis*	100559505	1
slc35f2	solute carrier family 35 member F2	*Anolis carolinensis*	100561082	2
slc35f5	solute carrier family 35 member F5	*Anolis carolinensis*	100558912	6
slc35f6	solute carrier family 35 member F6	*Anolis carolinensis*	100559818	1
slc36a1	solute carrier family 36 member 1	*Gallus gallus*	770250	1
slc37a1	solute carrier family 37 member 1	*Anolis carolinensis*	100567321	1
slc37a3	solute carrier family 37 member 3	*Anolis carolinensis*	100555927	1
slc37a4	solute carrier family 37 member 4	*Anolis carolinensis*	100557460	2
slc38a10	solute carrier family 38 member 10	*Anolis carolinensis*	100559840	1
slc38a2	solute carrier family 38 member 2	*Anolis carolinensis*	100555856	1
slc38a7	solute carrier family 38 member 7	*Anolis carolinensis*	100551838	2
slc39a11	solute carrier family 39 member 11	*Anolis carolinensis*	100556888	1
slc39a13	solute carrier family 39 member 13	*Anolis carolinensis*	100554988	2
slc39a2	solute carrier family 39 member 2	*Anolis carolinensis*	100552868	1
slc39a6	solute carrier family 39 member 6	*Anolis carolinensis*	100562021	3
slc39a7	solute carrier family 39 member 7	*Pogona vitticeps*	110091138	2
slc3a1	solute carrier family 3 member 1	*Anolis carolinensis*	100554178	2
slc3a2	solute carrier family 3 member 2	*Anolis carolinensis*	100552621	2
slc40a1	solute carrier family 40 member 1	*Anolis carolinensis*	100555173	1
slc41a1	solute carrier family 41 member 1	*Anolis carolinensis*	100561116	1
slc41a2	solute carrier family 41 member 2	*Anolis carolinensis*	100551735	3
slc43a2	solute carrier family 43 member 2	*Anolis carolinensis*	100557317	2
slc43a3	solute carrier family 43 member 3	*Anolis carolinensis*	100567636	5
slc44a2	solute carrier family 44 member 2	*Anolis carolinensis*	100556497	1
slc44a3	solute carrier family 44 member 3	*Anolis carolinensis*	100565696	1
slc44a4	solute carrier family 44 member 4	*Anolis carolinensis*	100556571	1
slc46a2	solute carrier family 46 member 2	*Anolis carolinensis*	100561292	1
slc46a3	solute carrier family 46 member 3	*Anolis carolinensis*	100561301	2
slc47a1	solute carrier family 47 member 1	*Pogona vitticeps*	110087421	4
slc47a2	solute carrier family 47 member 2	*Gallus gallus*	417616	2
slc48a1	solute carrier family 48 member 1	*Anolis carolinensis*	100556459	1

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
slc49a3	solute carrier family 49 member 3	*Anolis carolinensis*	100561505	1
slc4a2	solute carrier family 4 member 2	*Anolis carolinensis*	100559030	1
slc4a4	solute carrier family 4 member 4	*Anolis carolinensis*	100559398	4
slc50a1	solute carrier family 50 member 1	*Anolis carolinensis*	103277537	3
slc51a	solute carrier family 51 subunit alpha	*Anolis carolinensis*	100560747	1
slc5a1	solute carrier family 5 member 1	*Pogona vitticeps*	110082003	2
slc5a12	solute carrier family 5 member 12	*Anolis carolinensis*	100560865	9
slc5a2	solute carrier family 5 member 2	*Anolis carolinensis*	100557391	2
slc5a6	solute carrier family 5 member 6	*Anolis carolinensis*	100565212	3
slc5a9	solute carrier family 5 member 9	*Anolis carolinensis*	100563480	4
slc66a1	solute carrier family 66 member 1	*Pogona vitticeps*	110073205	3
slc66a3	solute carrier family 66 member 3	*Anolis carolinensis*	100566854	1
slc6a1	solute carrier family 6 member 1	*Anolis carolinensis*	100556107	6
slc6a18	solute carrier family 6 member 18	*Alligator mississippiensis*	106736619	4
slc6a19	solute carrier family 6 member 19	*Anolis carolinensis*	100559327	9
slc7a13	solute carrier family 7 member 13	*Anolis carolinensis*	100558734	2
slc7a2	solute carrier family 7 member 2	*Anolis carolinensis*	100562227	1
slc7a3	solute carrier family 7 member 3	*Anolis carolinensis*	100566333	14
slc7a6	solute carrier family 7 member 6	*Anolis carolinensis*	100565050	7
slc7a8	solute carrier family 7 member 8	*Anolis carolinensis*	100553607	2
slc7a9	solute carrier family 7 member 9	*Anolis carolinensis*	100560204	2
slc9a2	solute carrier family 9 member A2	*Anolis carolinensis*	100552210	1
slc9a3r1	SLC9A3 regulator 1	*Anolis carolinensis*	100567187	5
slc9a6	solute carrier family 9 member A6	*Anolis carolinensis*	100554712	1
slc9a8	solute carrier family 9 member A8	*Anolis carolinensis*	100566100	1
slco2a1	solute carrier organic anion transporter family member 2A1	*Anolis carolinensis*	100554939	1
sord	sorbitol dehydrogenase	*Gallus gallus*	415332	3
sox9	SRY-box 9	*Gallus gallus*	374148	2
sptan1	spectrin alpha, non-erythrocytic 1	*Gallus gallus*	374234	1
stat3	signal transducer and activator of transcription 3	*Gallus gallus*	420027	2
stip1	stress induced phosphoprotein 1	*Anolis carolinensis*	100563364	3

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
stk4	serine/threonine kinase 4	*Gallus gallus*	419187	1
sugt1	SGT1 homolog, MIS12 kinetochore complex assembly cochaperone	*Anolis carolinensis*	100552516	1
sult1b1	sulfotransferase family cytosolic 1B member 1	*Gallus gallus*	395227	3
sumo1	small ubiquitin-like modifier 1	*Gallus gallus*	373930	2
suox	sulfite oxidase	*Gallus gallus*	107055404	2
taf1b	TATA-box binding protein associated factor, RNA polymerase I subunit B	*Ficedula albicollis*	101814864	1
taok3	TAO kinase 3	*Gallus gallus*	395499	2
tars1	threonyl-tRNA synthetase 1	*Numida meleagris*	110389937	2
tgfb1i1	transforming growth factor beta 1 induced transcript 1	*Gallus gallus*	395832	1
tgfbr2	transforming growth factor beta receptor 2	*Gallus gallus*	396399	1
tipin	TIMELESS interacting protein	*Gallus gallus*	415548	1
tjp2	tight junction protein 2	*Apteryx mantelli mantelli*	106484412	3
tk1	thymidine kinase 1	*Gallus gallus*	395719	1
tmem104	transmembrane protein 104	*Anolis carolinensis*	100566413	2
tmem184a	transmembrane protein 184A	*Anolis carolinensis*	100553463	3
tmem184c	transmembrane protein 184C	*Anolis carolinensis*	100566749	2
tnfaip1	TNF alpha induced protein 1	*Gallus gallus*	417672	1
tpcn1	two pore segment channel 1	*Anolis carolinensis*	100560676	1
tpcn2	two pore segment channel 2	*Anolis carolinensis*	100565414	1
trpm6	transient receptor potential cation channel subfamily M member 6	*Pogona vitticeps*	110075995	1
ttc33	tetratricopeptide repeat domain 33	*Gallus gallus*	427184	2
txn	thioredoxin	*Gallus gallus*	396437	1
ufl1	UFM1 specific ligase 1	*Gallus gallus*	421804	3
ugdh	UDP-glucose 6-dehydrogenase	*Gallus gallus*	422792	8
usp19	ubiquitin specific peptidase 19	*Pygoscelis adeliae*	103918920	1
utp18	UTP18 small subunit processome component	*Podarcis muralis*	114589903	1
vcl		*Gallus gallus*	396422	1
wdr73	WD repeat domain 73	*Thamnophis sirtalis*	106551202	2
wfs1	wolframin ER transmembrane glycoprotein	*Alligator mississippiensis*	102558683	2
wnk1	WNK lysine deficient protein kinase 1	*Pogona vitticeps*	110076681	2
wscd1	WSC domain containing 1	*Pogona vitticeps*	110089414	1

Symbol	Description	Organism	NCBI Gene symbol	**Present in transcripts**
Xdh	xanthine dehydrogenase	*Gallus gallus*	396025	6
Zyx	zyxin	*Gallus gallus*	418300	6

6.4 Appendix 4: Chapter 1: Supplementary Information - Initial Literature Reviews

The two reviews summarised below were conducted as a preliminary step in this study, and later expanded. They were used in the formation of the thesis introduction. The body text has been removed to avoid repetition, but methods and literature analysis are included here to provide the full search terms and methods for the original reviews. This was deemed necessary to qualify a number of statements where certain information (re: genomic research into pygmy bluetongues, or the use of RNA-seq methods specifically on squamates) is stated to be absent or lacking in the literature. Review methodologies are based on Moher et al. (2009) and Pullin and Gavin (2006).

6.4.1 A History of Research into the Australian Pygmy Bluetongue Lizard

This review provides background on the study species *Tiliqua adelaidensis*. A summary of research into the Australian pygmy bluetongue lizard since its rediscovery in 1992 shows that a focus on functional genomics is a logical next step, and one that will benefit the long-term conservation goals for this species.

6.4.2 A Review of Transcript and Gene Expression Studies in Non-model Squamates and other Reptiles

Differential gene expression and transcript sequencing is becoming increasingly popular as the expense and availability of technology becomes more favourable. Annotation and assembly of these data, however, remains problematic for many species that are the focus of ecological or conservation studies, where closely related references are not available and sequencing funds are minimal. This review used a systematic approach to scope and summarise genomic methodologies used over time, and how they have been applied to non-model organisms; specifically, squamates and other reptiles.

Full reference lists for records included in final visualisations in the sections below are available in additional supplementary material.

6.4.1 A History of Research into the Australian Pygmy Bluetongue Lizard

6.4.1.a Introduction

This review served as an overview of studies into the ecology and conservation of the Australian pygmy bluetongue lizard overall, but also demonstrates how few studies have been done that have a direct focus on genetics in this species. It also highlighted the lack of functional genomic information available to inform conservation decisions for the species.

6.4.1.b Search Parameters

Search Query

A systematic approach was used to supplement references and to ensure that all published articles on *T. adelaidensis* were included here. A literature search was carried out across multiple databases using the term:

> ["*Tiliqua adelaidensis*" OR "pygmy bluetongue" OR "pygmy blue tongue" OR "pygmy blue-tongue"]

Databases included Web of Science (Topic search), SCOPUS (Title, Abstract, Keyword search) and Science Direct (search all terms in any field). These terms and alternate hyphenation of the common name were used to ensure that any usage were returned by the search to include any articles that included direct mention of this species. The search was also conducted in PubMed, but the initial search returned only seven results—all duplicate records. Due to the initial balance of duplicate or irrelevant records, only Web of Science and SCOPUS were used to update the search. The latest update of literature included here was conducted on 24 March 2022. Additional articles such as the original 1992 rediscovery by Armstrong and Reid (1992) have been added to the results from this search.

Literature Search Terms

A systematic approach (Figure 6-9) was used to ensure a complete list of publications was included to generate Figure 6-10, which shows common research areas and terms that occur throughout the history of research and publications on this species. Specific search terms tailored for each database's syntax are provided below.

Web of Science Core Collection:

> TOPIC: ("*Tiliqua adelaidensis*" OR "pygmy bluetongue" OR "pygmy blue tongue" OR "pygmy blue-tongue")

Science Direct, where the search term was entered directly into the main search 'find articles with these terms' to search all fields:

> "*Tiliqua adelaidensis*" OR "pygmy bluetongue" OR "pygmy blue tongue" OR "pygmy blue-tongue"

Scopus:

TITLE-ABS-KEY("*Tiliqua adelaidensis*" OR "pygmy bluetongue" OR "pygmy blue tongue" OR "pygmy blue-tongue")

Database Record Filtering

All citations were exported from the respective databases with 'full records' and include title, keyword and abstract fields filled as completely as possible. Filtering of search results was conducted in Endnote X9 according to a PRISMA-style filtering process (Moher et al. 2009; Figure 6-9). The sole most important criterion for retaining a record in the final list was a specific focus on the Australian pygmy bluetongue lizard (*T. adelaidensis*). Records with a focus on other related species with only a cursory mention of *T. adelaidensis* for comparison, or to suggest that findings or methods could also be applied in the conservation of this species, were excluded.

Data Visualisation

A bibliographic file of final filtered reference information was imported into VOSviewer version 1.6.14 (van Eck & Waltman 2010). VOSviewer was used to produce a co-occurrence network map of identified keywords, based on terms that appeared in bibliographic data. Full counts were used to measure the strength of network connections, and connections were limited to 63 keywords with a minimum of three occurrences (out of a possible 466 identified terms).

VOSviewer Network Analysis

Co-occurrence of 63 key terms (separated by a comma) was identified by limiting to keywords with a minimum of three occurrences from a list of 466 potential keywords, from among bibliographic data.

Cluster 1 (15 items): Araneae, body condition, burrow, communities, endangered species, evolution, grassland, grazing, impact, lizard, pygmy bluetongue lizard, spider, Squamata, *Tiliqua adelaidensis*, vegetation

Cluster 2 (14 items): Agriculture, amphibians, behaviour/behavior, behavioural sciences, blue tongue lizard, dispersal, New-Zealand, reintroduction, reptile, success, survival, translocation, veterinary sciences

Cluster 3 (13 items): blue-tongue lizard, burrows, diet, endangered lizard, gene flow, models, movement, persistence, predation, responses, science and technology – other topics, skink, *Tiliqua-adelaidensis*

Cluster 4 (11 items): abundance, Australia, biodiversity & conservation, burrow choice, climate-change, conservation, environmental sciences, lizards, reproduction, reptiles, South Australia

Cluster 5 (10 items): *adelaidensis,* Australian lizard, *Egernia-stokesii,* population, recognition, *rugosa,* scats Scincidae, *Tiliqua,* zoology

6.4.1.c Results

The above search methods returned 73 records from Web of Science, 75 from Scopus, 21 from Science Direct and 7 from PubMed (none of which were unique). After combining exports using EndNote and removing duplicate records, 112 unique records remained. One historical record not found in this search despite 'pygmy bluetongue' appearing in the title was added to make the total 113. After limiting the records to peer-reviewed articles, 85 remained. The 85 peer-reviewed records were assessed for relevance using title and abstract, resulting in the exclusion of a further 14 records that were articles about other species that may have used *T. adelaidensis* as a keyword or note for future applications or comparison. Full text for all records was searched for using both Flinders University and Adelaide University credentials, and 8 records did not return an accessible full text download. The remaining 63 records were again assessed using the full text (Figure 6-9)

and no further records were removed. The final 63 records may be given additional focus in text because of accessibility but since VOSviewer uses only bibliographic data and abstracts to build network maps, all 72 records were used in the creation of Figure 6-10; that is, the 73 retained after filtering title and abstract minus one record removed on assessment of full text (Figure 6-9).

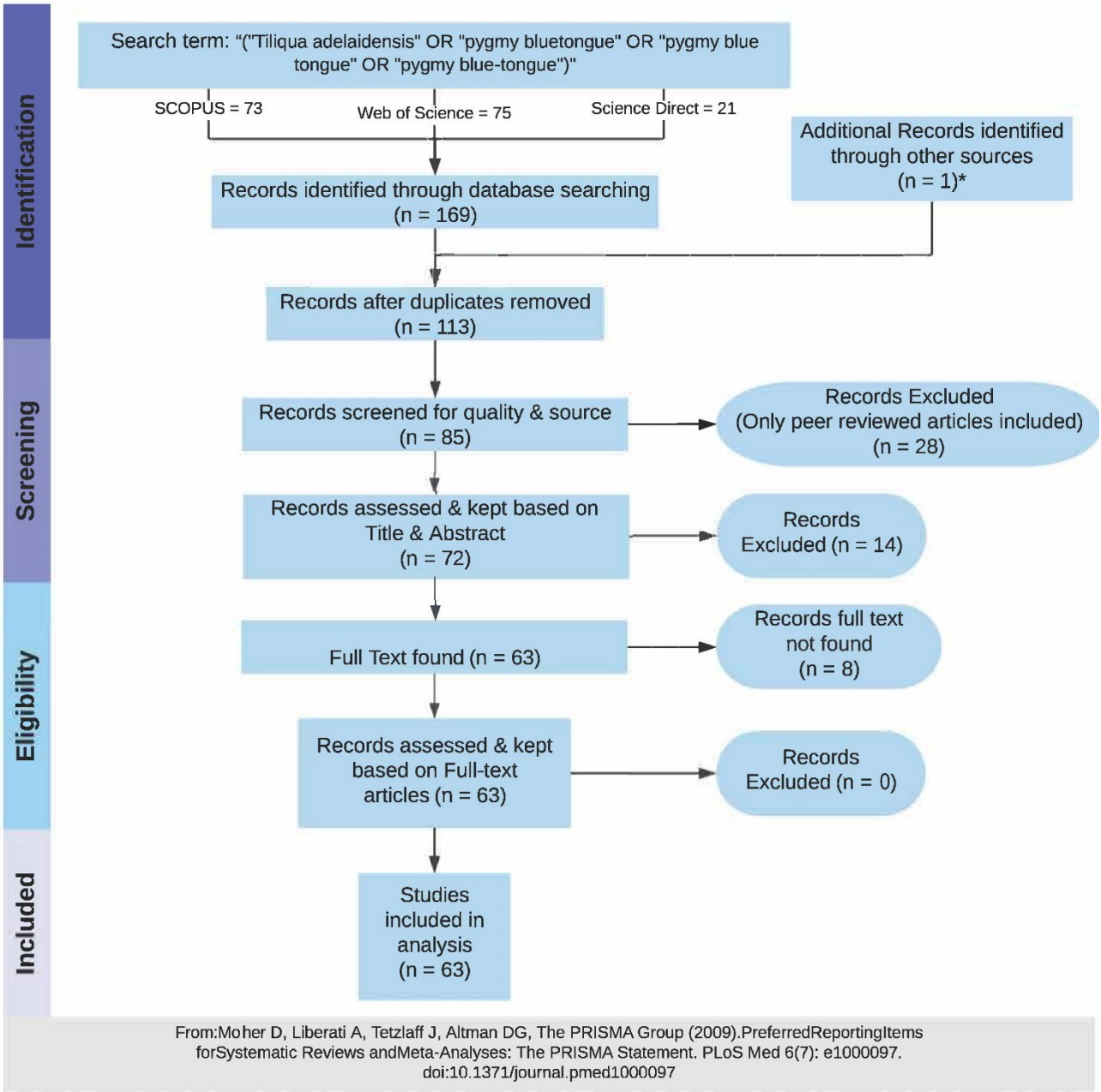


Figure 6-9: PRISMA style review filtering flow diagram for analysis of articles on the subject of *Tiliqua adelaidensis*. Literature databases searched using the term" ["*Tiliqua adelaidensis*" OR "pygmy bluetongue" OR "pygmy blue tongue" OR "pygmy blue-tongue"]". This filtering of articles determined their presence in the creation of Figure 6-10 (Moher et al. 2009).

There were 63 key terms identified in VOSviewer and included in the network map (Figure 6-10); 862 links for these keywords were found between literature records, and they were sorted into five clusters. These clusters are colour coordinated in Figure 6-10. A much smaller number of clusters have been directly linked to the cluster "gene flow" indicating a direct link to genetic studies in this species (Figure 6-14).

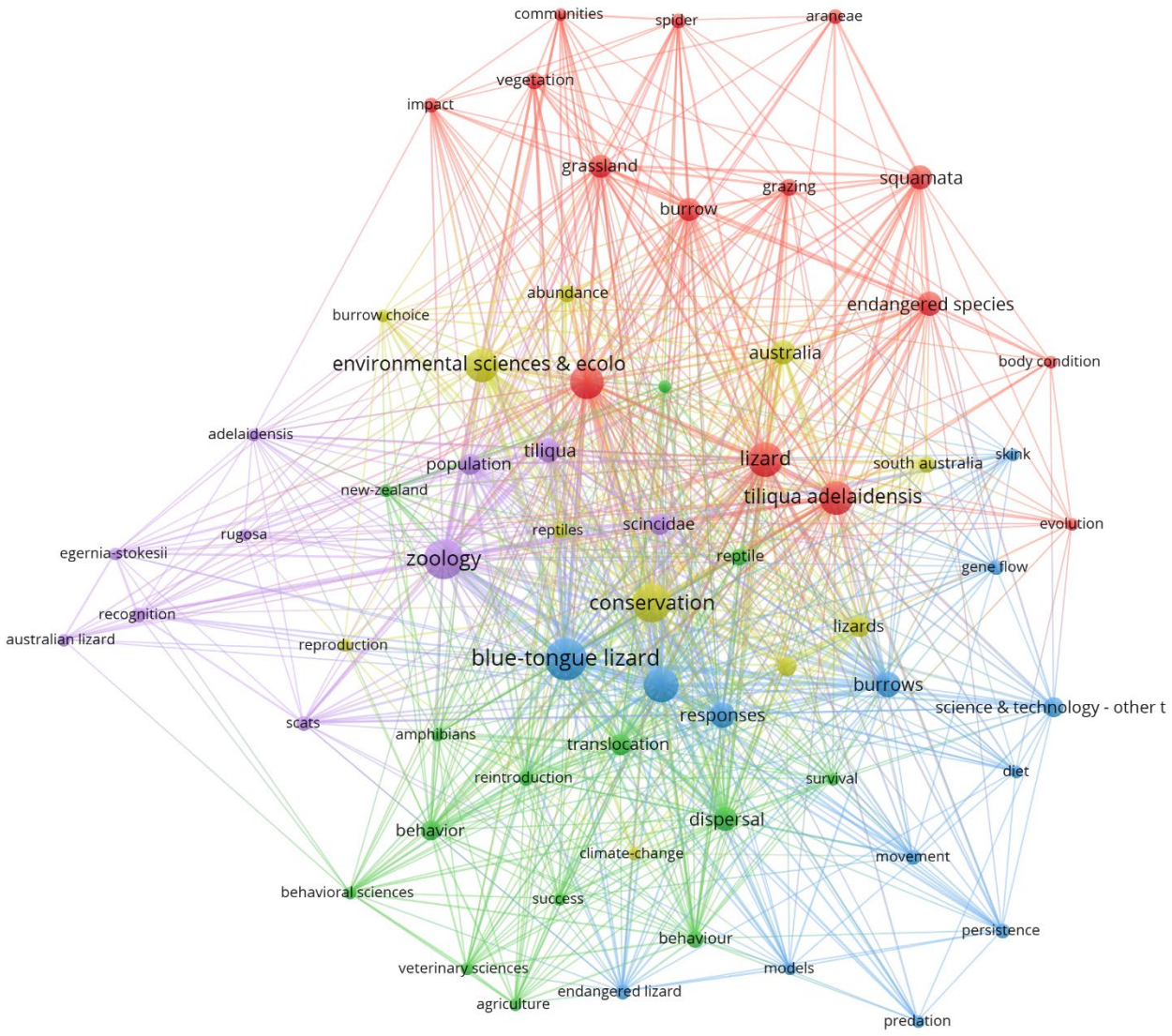


Figure 6-10: Network diagram showing common terms and research areas common among publications about *Tiliqua adelaidensis*, the Australian pygmy bluetongue lizard (1992-2022). Literature databases searched using the term "("*Tiliqua adelaidensis*" OR "pygmy bluetongue" OR "pygmy blue tongue" OR "pygmy blue-tongue")". As VOS uses bibliographic data with a focus on keywords and abstract, 72 records were used in this visualisation (Figure 6-9). Colours indicate 5 clusters of keywords in the network analysis. Note blue keyword "gene flow" on the right hand side with a small number of connections is the only keyword referencing genetic analyses (van Eck and Waltman 2010).

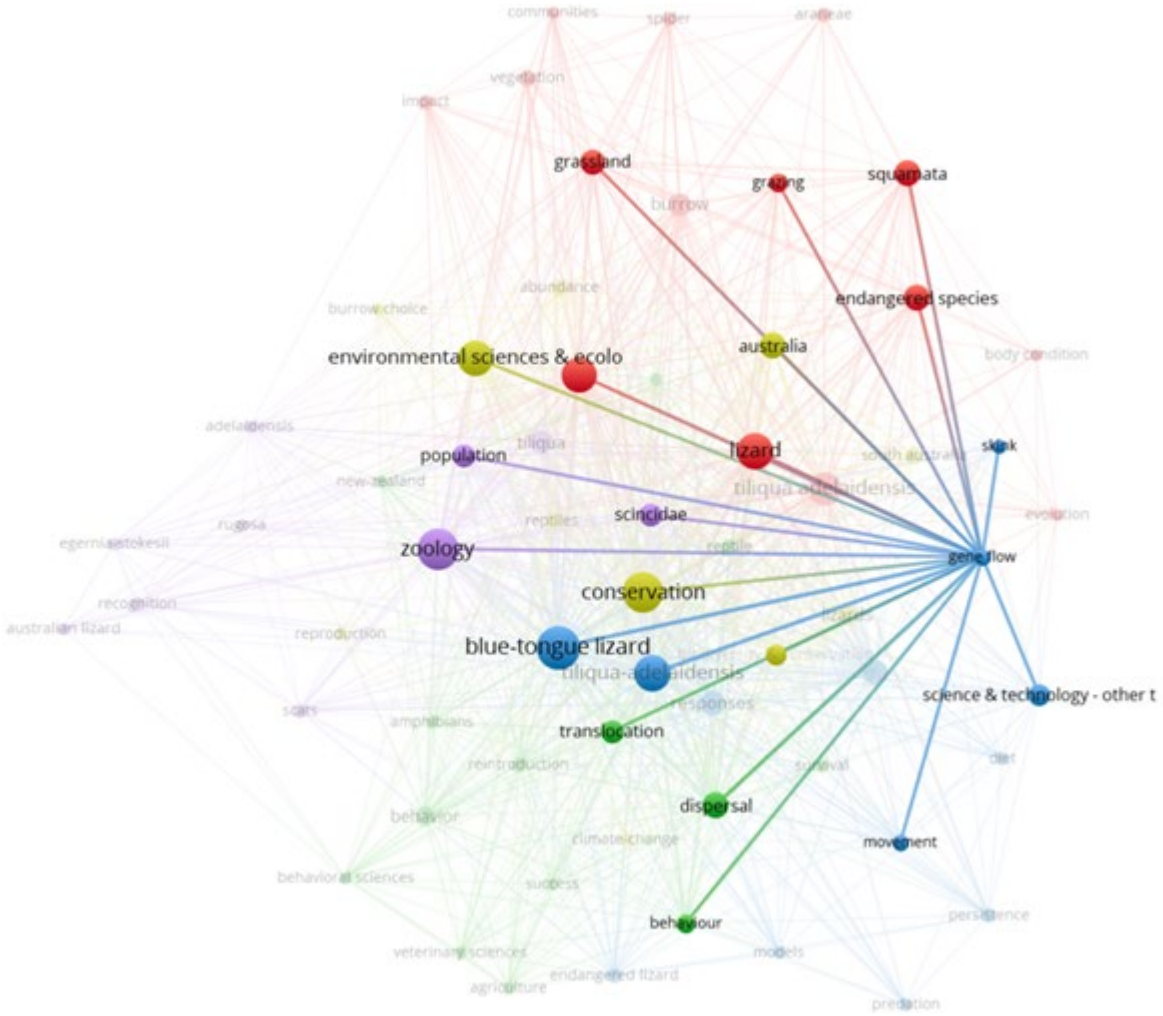


Figure 6-11: Network diagram showing common terms and research areas common among publications about *Tiliqua adelaidensis*, the Australian pygmy bluetongue lizard (1992-2022). Literature databases searched using the term "("*Tiliqua adelaidensis*" OR "pygmy bluetongue" OR "pygmy blue tongue" OR "pygmy blue-tongue")". As VOS uses bibliographic data with a focus on keywords and abstract, 72 records were used in this visualisation (Figure 6-9). Colours indicate 5 clusters of keywords in the network analysis. This is an alternate version of Figure 6-10, highlighting the connections of the keyword "gene flow" (van Eck and Waltman 2010).

6.4.2 A Review of Transcript and Gene Expression Studies in Non-model Squamates and Other Reptiles

6.4.2.a Introduction

This review originally served the purpose of a scoping literature review on RNA-seq, and transcript analysis and gene expression in conservation of reptiles and squamates. There is no single standard approach to transcriptomic analysis, and the lack of genomic reference sequences greatly inhibits the application of these tools in many species. This review provided a snapshot of the different methods used and how these technologies have been applied to non-model reptile organisms, particularly in the absence of a reference genome.

6.4.2.b Search Parameters

Initial Database Assessment

An initial search was conducted using the phrase ((“systematic review”) AND (“gene expression” OR RNA-seq OR transcript) AND (“non model” OR non-model)) to determine if any specific and systematic analysis of the development and changes in sequencing and bioinformatic tool use had been conducted on non-model organisms, as many organisms face the same challenges. Then A literature search was carried out across multiple databases using the term:

> ((transcript OR "long-read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR "non-model" OR squamat*))

Databases included Web of Science (Topic search), SCOPUS (Title, Abstract, keyword search), and Science Direct (Title, Abstract, keyword search). Google Scholar was also searched, however due to the vast amount of content and limited control over fields an open search returned over 14,600 results. This search was limited to terms appearing in the title and produced only 13 duplicate results.

To estimate the level of overlap between records labelled as different organisms, further searches were conducted altering the parameters in the Web of Science database:

> TOPIC: ((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND ("non-model" AND (reptil* OR squamat*)))
>
> TOPIC: ("non-model" AND (reptil* OR squamat*))
>
> TOPIC: ("Non-model" NEAR/10(reptil* or squamat*))
>
> TOPIC: (("novel species” OR “novel organism") AND (reptil* OR squamat*))

Because of the very low overlap between articles that refer to the phrase “non-model” organisms as well as reptiles or more specifically squamates, “non model” was excluded to limit the number of papers discussed here in more depth.

Final Search Query

A final search using the following phrase was conducted in Web of Science (Topic search), SCOPUS (Title, Abstract, Keyword search) and Science Direct (Title, Abstract, Keyword search):

((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR squamat*))

Search results were not constrained by any other fields such as publication year or type. Specific search terms as edited by the respective database syntax were as follows:

Web of Science Core Collection:

TOPIC: ((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR "non-model" OR squamat*))

TOPIC: ((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR squamat*))

Scopus:

TITLE-ABS-KEY((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR "non-model" OR squamat*))

TITLE-ABS-KEY((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR squamat*))

Science Direct, with the search term entered directly into the field for "Title, abstract or author-specified keywords"; therefore this does not appear as part of the search term:

TOPIC: ((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptile OR Reptilia OR "non-model" OR squamate OR Squamata))

TOPIC: ((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptile OR Reptilia OR squamate OR Squamata))

Google Scholar:

allintitle: ((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptile OR Reptilia OR "non-model" OR squamate OR Squamata))

Wildcard "*" terms are not supported in Science Direct or Google Scholar's title search, so reptile, Reptilia, squamate or Squamata were the terms given.

These search terms were then added to email alert lists and periodically re-run to check for newly published articles. The latest update of literature included here was conducted on 24 March 2022.

Database Record Filtering

Duplicates were removed in Endnote X9, favouring Web of Science records over others, and Scopus records over Science Direct records. This was done to keep as much consistency as possible in text formatting. In some cases, Web of Science favours a simplified text, and there are database differences in the inclusion of letter accents, as well as addenda to names like Jr. and II. In cases where an author may be attributed to a Web of Science record and to a different Science Direct record (without a Web of Science duplicate), these fields were not always identical. Efforts were made to manually change references where this occurs.

Flinders University and Adelaide University credentials were used to source the full text records of filtered references from databases, and records that could not be sourced were removed from further analysis. Peer-reviewed articles with the full text available for assessment were included in further consideration using a PRISMA-style filtering flow (Moher et al. 2009; Figure 6-12). Records were discarded for having either a main focus that did not include a reptile species or not focusing on transcript analysis or gene expression data.

Exclusion criteria:

1. studies that did not focus on transcript annotation, characterisation, or expression
2. studies that did not focus on a reptile organism
3. records that were reviews that only mention reptile transcript data as a secondary source.

Remaining records were split into four categories based on organism and methods of study:

- studies focused on squamates with a core focus on differential gene expression analysis
- studies focused on squamates with a core focus on transcript or isoform characterisation and annotation, but no specific focus on differential expression
- studies focused on other non-squamate reptiles (including birds) with a core focus on differential gene expression analysis
- studies focused on other non-squamate reptiles with a core focus on transcript or isoform characterisation and annotation, but no specific focus on differential expression.

Although many sources from each of these groups are discussed in the text and all four categories are shown in Figure 6-13, the studies focused on squamates with a core focus on differential gene expression analysis were subject to more in-depth assessment (Figure 6-14).

Data Visualisation

The sum of these four categories, as well as the singular category for gene expression studies on squamates only were summarised in network analyses run in VOSviewer version 1.6.14 (van Eck & Waltman 2010), similar to the review in Appendix 4.1. Co-occurrence of keywords was analysed using full counts among bibliographic records.

A network was created for the combined four groups filtered on assessment of title and abstract; 111 keywords were selected from an identified list of 1,863 by VOSviewer, by limiting to keywords with a minimum of five occurrences (Figure 6-13).

For the filtered group of studies on squamates and gene expression that included 33 fully assessed records, an overlay diagram was created where the network of keywords in papers is overlaid by a colour gradient indicating publication year over time. Eighty-one keywords were selected from an identified list of 491 by VOSviewer, by limiting to keywords with a minimum of two occurrences (Figure 6-14).

Both these visualisations are included to demonstrate the change in focus of methods and study design over time.

6.4.2.c Results

Initial Database Assessment

Initial search results for texts containing the terms ((transcript or "long read" or "differential gene expression" or "transcript assembly" or RNA-Seq) AND ("non-model" reptile* or squamat*)) resulted in 754 records in Web of Science (These results were consistent if the initial search was carried out with the refined phrase or whether the “search within results” function was used for “reptile OR squamate”), 828 records from Scopus and 109 records from Science Direct, in an initial search in 2020. Including both “non-model” and “non model” did not impact the returned results so only “non-model” was used in all initial searches.

Searching the topic field of the Web of Science database for records with all the above genomic terms but requiring that “non model” appears along with either “reptil*” or “squamat*”, yielded only 2 articles. A further search eliminating reference to genomic methods to determine how many articles overall reference both “non model” organisms as well as either “reptil*” or “squamat*” yielded only 21 records, and limiting the separation of these terms to ten words narrowed this again to 4. Similarly, altering these terms to search instead for "novel species” or “novel organism” and either “reptil*” or “squamat*” also yielded only 20 records in Web of Science.

Final Search Query

These numbers were greatly reduced after exclusion of the term “non-model” for a more specific focus on squamates and other reptiles. When the search was refined to include “reptil*” or “squamat*”, but not “non-model” (in 2020), Web of Science topic search results were reduced to 185, Scopus to 293 and Science Direct to 40. On removal of “non-model” as a term, the search phrase returned no results in Google Scholar. The final number of updated search results from the final search conducted on 24 March 2022 included 221 records from Web of Science, 343 records from SCOPUS and 40 records from Science Direct, for a total list of 604 records.

Database Record Filtering

After removal of duplicates from the final search, 369 unique records remained for further analysis and filtering (Figure 6-12). Records where full text could not be sourced were removed as they could not be properly assessed (N = 71). The remaining 290 peer-reviewed records with the full text available were assessed based on title and abstract according to the following criteria.

Records that matched the exclusion criteria were completely excluded from further analysis, which resulted in the exclusion of 138 records, although 70 of these did focus on transcript analysis of some kind in amphibian species and may be referenced elsewhere in this thesis.

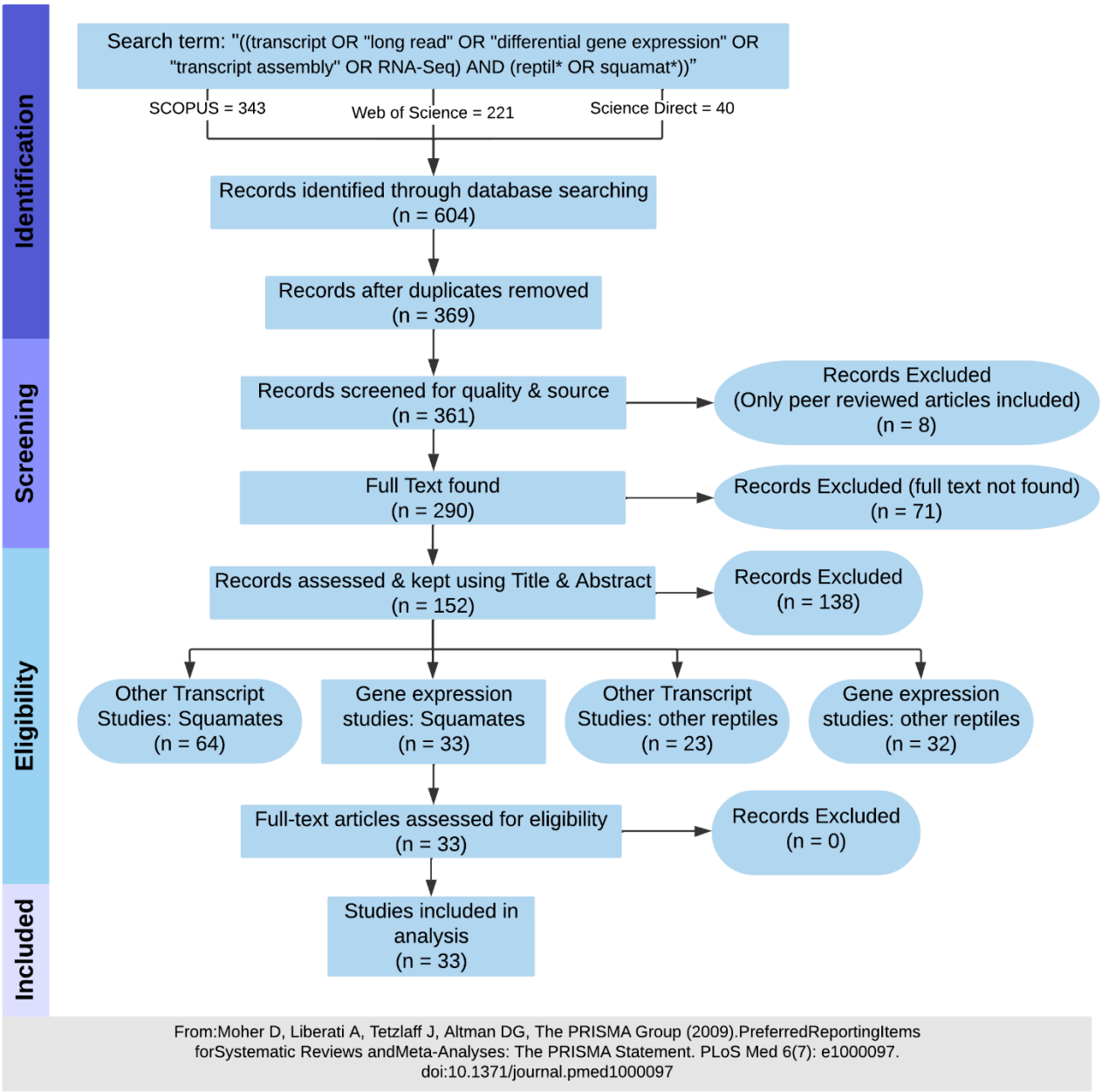


Figure 6-12: PRISMA style filtering of search term record results for scientific journal databases using the terms "(transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR squamat*)" (Moher et al. 2009).

The records retained up until this step were further assessed and split into four categories to narrow the focus of the analysis:

- studies focused on Squamates with a core focus on transcript or isoform characterisation and annotation, but no specific focus on differential expression (64 records)
- studies focused on Squamates with a core focus on differential gene expression analysis (33 records)
- studies focused on other non-squamate reptiles with a core focus on transcript or isoform characterisation and annotation, but no specific focus on differential expression (23 records)

- studies focused on other non-squamate reptiles (including birds) with a core focus on differential gene expression analysis (32 records)

All remaining 152 records were visualised in Figure 6-13 , and a further analysis with a specific focus on differential gene expression in a squamate organism using the final list of 33 records was conducted (Figure 6-14).

Many of the results from the initial, wider search including any reference to non-model organisms, rightfully, have been cited in this thesis. These articles generally have a focus on methodology, troubleshooting and demonstrating how methods can serve studies in these organisms through their own studies, something that becomes a focus here. As this is not strictly a systematic review of the sub-set of papers analysed further, many are cited in this passage for context; however, for the purpose of exploring how these tools and methods have been employed in other squamates and reptiles more closely related to my own focus species the results discussed in more depth were those from the search restricted to reptiles and squamates.

Many of the papers returned by the broader search term were those discussing methods to mitigate or solve alignment and assembly difficulties in organisms with no closely related reference genome and minimal other genomic references, or those using established protocols with their species as a case study for expanded application. Seventy of the excluded records do discuss transcript analysis or gene expression but are focused on amphibian species, and flag reptiles in keywords or as a comparative discussion. The majority of these records were on *Xenopus* spp. Many of the 55 excluded records on reptiles but not specifically squamates, and that do discuss transcripts or gene expression, focus on turtle species, and the majority of squamate records that were excluded for not specifically focusing on gene expression analysis are about venom composition and transcript characterisation of venom proteins.

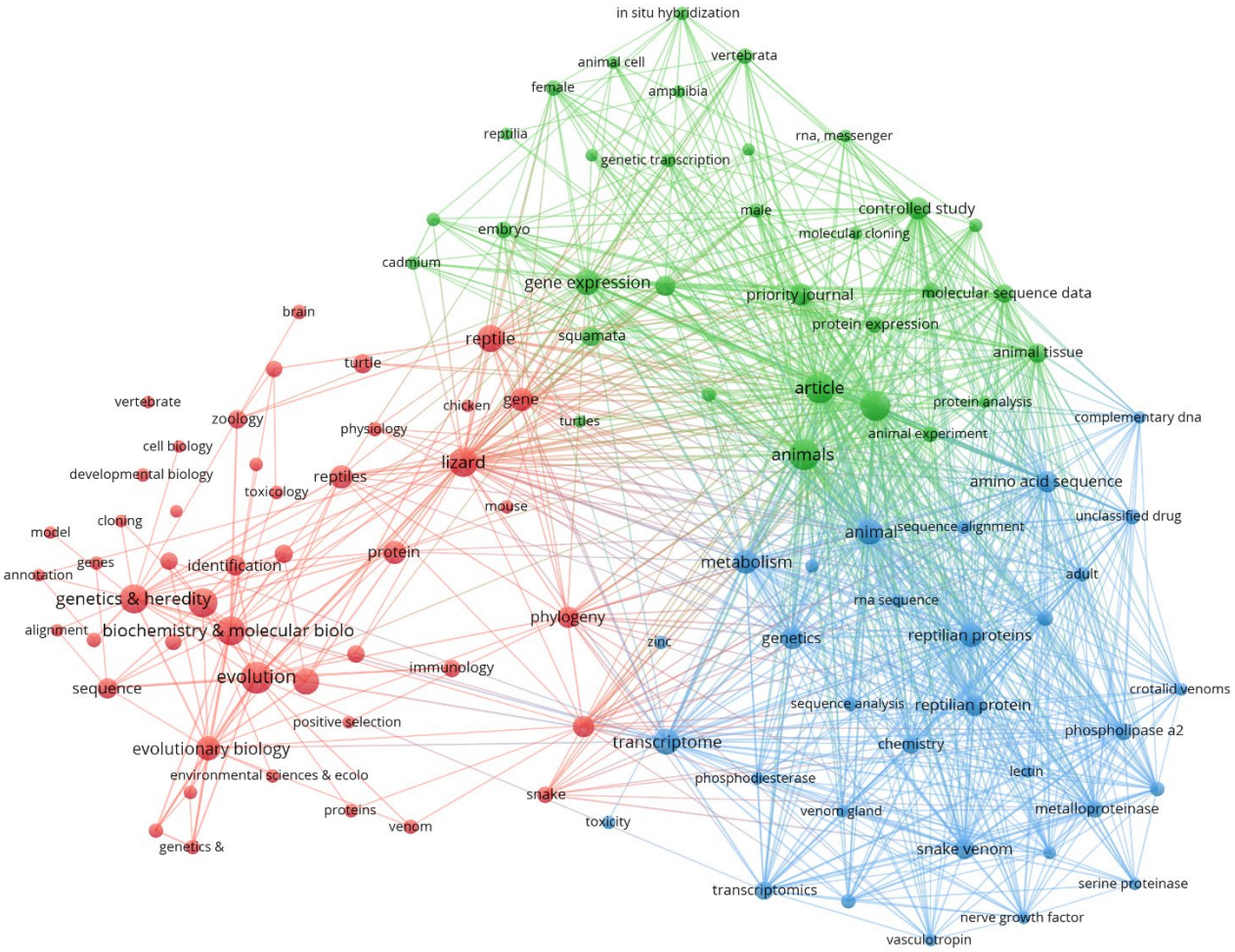


Figure 6-13: Network diagram showing key terms among publications about differential gene expression transcript analysis in Squamate reptiles. Literature databases searched using the term ”((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR squamat*))”. All counted categories as per Figure 6-12, including 152 records of transcript annotation or gene expression studies for both squamates, and non-squamate reptiles were included in this broader network map. Colours indicate 3 clusters of keywords in the network analysis (van Eck and Waltman 2010).

6.4.2.d Data Visualisation

VOSviewer Network Analysis: Larger Group of 152 Records as Visualised in Figure 6-13.

In VOSviewer version 1.6.14, similar to Appendix 4.1. co-occurrence of keywords (separated by a comma) was analysed using full counts among bibliographic records. For the combined four categories, 111 keywords were selected from an identified list of 1,863, by limiting to keywords with a minimum of five occurrences.

Cluster 1 (47 items): adaptation, alignment, annotation, biochemistry & molecular biology, biotechnology & applied microbiology, brain, cell biology, chicken, cloning, developmental biology, endocrinology & metabolism, environmental sciences, evolution, evolutionary biology, expression,

gene, gene-expression, genes, genetics &, genetics & heredity, genome, heredity, identification, immunology, lizard, mechanisms, messenger-RNA, model, molecular evolution, mouse, phylogenetic analysis, phylogeny, physiology, positive selection, protein, proteins, reptile, reptiles, rna-seq, science & technology, sequence, snake, toxicology, turtle, venom, vertebrate, zoology

Cluster 2 (32 items): amphibia, animal cell, animal experiment, animal tissue, animalia, animals, article, cadmium, controlled study, embryo, female, gene expression, gene expression regulation, genetic transcription, in situ hybridization, lizards, male, messenger RNA, metallothionein, molecular cloning, molecular sequence data, nonhuman, northern blotting, nucleotide sequence, priority journal, protein analysis, protein expression, Reptilia, rna- messenger, Squamata, turtles, vertebrata

Cluster 3 (32 items): adult, amino acid oxidase, amino acid sequence, animal, chemistry, complementary DNA, crotalid venoms, gene expression profiling, genetics, lectin, metabolism, metalloproteinase, nerve growth factor, phosphodiesterase, phospholipase a2, polymerase chain reaction, proteome, proteomics, reptilian protein, reptilian proteins, rna sequence, sequence alignment, sequence analysis, serine proteinase, snake venom, toxicity, transcriptome, transcriptomics, unclassified drug, vasculotropin, venom gland, zinc

VOSviewer Network Analysis: Filtered Group of 33 Records as Visualised in Figure 6-14.

For this category of 33 fully assessed records, 81 keywords were selected from an identified list of 491, by limiting to keywords with a minimum of two occurrences.

Cluster 1 (32 items): accelerated evolution, biochemistry & molecular biology, complex, differential expression, evolution, evolutionary biology, expression, gene-expression, genetics & heredity, genome, identification, inhibition, mass-spectrometry, model, molecular evolution, placenta, pregnancy, protein, *pseudemonia-entrecasteeauxii*, receptor, reptiles, rna-seq, sequence, squamates, transcriptome, transport, uterus, venom, vertebrate, viviparity, x-chromosome, zoology

Cluster 2 (18 items): animals, article, cloning- molecular, controlled study, lizards, male, molecular cloning, molecular sequence data, mouse, nonhuman, northern blotting, ovary, protein expression, regeneration, reptile, Reptilia, spinal cord, tail

Cluster 3 (11 items): amino acid sequence, animal, animal experiment, animal tissue, genetics, phylogeny, polymerase chain reaction, reptilian protein, reptilian proteins, rna extraction, transcriptomics

Cluster 4 (10 items): copper, copper transporter, gene, lizard, Mammalia, oogenesis, *Podarcis sicula*, Squamata, toxicity, zinc

Cluster 5 (10 items): cadmium, exposure, female, gene expression, gene induction, in situ hybridization, lizard embryo, metallothionein, priority journal, rna

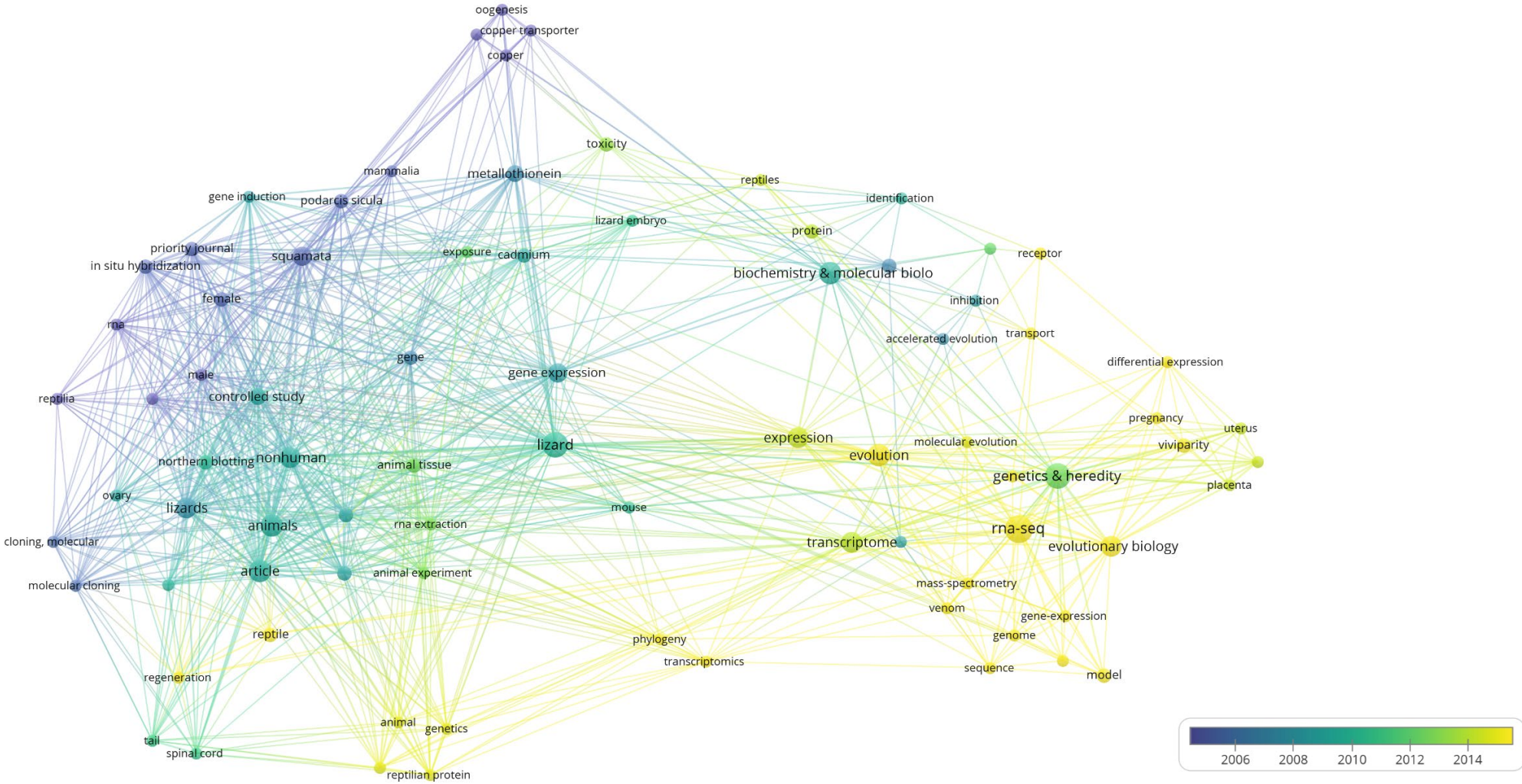


Figure 6-14: Overlay diagram showing key terms among 33 final filtered publications about differential gene expression transcript analysis in Squamate reptiles. Literature databases searched using the term ”((transcript OR "long read" OR "differential gene expression" OR "transcript assembly" OR RNA-Seq) AND (reptil* OR squamat*))”. Colours indicate year of publication of linked references in the network analysis to show the shift in methods and focus over time (van Eck and Waltman 2010). Note transcriptomics and related keywords coloured in yellow indicates inclusion at or after 2015.

6.4.3 References

Moher, D, Liberati, A, Tetzlaff, J & Altman, DG 2009, 'Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement', *PLoS Med*, vol. 6, no. 7, p. e1000097.

Pullin, AS & Gavin, BS 2006, 'Guidelines for Systematic Review in Conservation and Environmental Management', *Conservation Biology*, vol. 20, no. 6, pp. 1647-56.

van Eck, NJ & Waltman, L 2010, 'Software survey: VOSviewer, a computer program for bibliometric mapping', *Scientometrics*, vol. 84, no. 2, pp. 523-38.

Tiliqua adelaidensis juvenile - C. Maher