Appendix 2-8. Chapter 3. *Exploiting genomic data to identify proteins involved in abalone reproduction* . Omar Mendoza-Porras; Natasha A. Botwright; Sean M. McWilliams; Mathew T. Cook; James O. Harris; Gene Wijffels and Michelle L. Colgrave

Abalone gonadic proteins involved in the structure of spermatozoa and egg, sexual maturation and fertilization

Proteins identified in searches against custom and public database						NCBI BLASTp results		
Male proteins	**Peptide sequence**	**Custom-DB**	**NCBI**	**UniprotKB**	**Search DB-species**	**Accession**	***Species***	**Reference**
	APTVNDYSR	x	gi|538394	Q25078	*H. laevigata*	-	-	
	DLIAKPVQDIPIR	x	gi|538394	Q25078	*H. laevigata*	-	-	
	HLSAIQK	x	gi|538394	Q25078	*H. laevigata*	-	-	
	IGAEIGR	x	gi|538394	Q25078	*H. laevigata*	-	-	
	IPLEVTYSFLVR	x	gi|538394	Q25078	*H. laevigata*	-	-	[17, 18]
Lysin	LVAWLQR	x	gi|538394	Q25078	*H. laevigata*	-	-	
	SLYFVNR	x	gi|538394	Q25078	*H. laevigata*	-	-	
	VQIIAGFDR	yes	gi|538394	Q25078	*H. laevigata*	-	-	
	WHFVPHTHVAR	x	gi|538394	Q25078	*H. laevigata*	-	-	
	QLASWLQR	x	gi|538394	Q25078	*H. laevigata*	-	-	
	EQFDDFVR	CX726714[a]	x	x	*H. discus*	gi|556094603	*L. gigantea*	
	LEMPIASQK	CX726714	x	x	*H. discus*	gi|524867119	*A. californica*	
	TDTGLTPGLLHVLNR	CX726714	x	x	*H. discus*	gi|405966318	*C. gigas*	
Ropporin-1-like protein	WLDLALPK	CX726714	x	x	*H. discus*	gi|524867119	*A. californica*	[56-58]
	DVLAWSAAYFR	yes	gi|405966318	K1Q4X4	*C. gigas*	-	-	
	IPFPLFK	yes	gi|405966318	K1Q4X4	*C. gigas*	-	-	
	TVCEILTADPEGGPAR	yes	gi|405966318	K1Q4X4	*C. gigas*	-	-	
	IGSFGGEVEWNK	yes	gi|405966318	K1Q4X4	*C. gigas*	-	-	
	ALAETENPLHISQECLYNR	CX726057[a]	x	x	*H. discus*	gi|524899863	*A. californica*	
	IGDITFWK	CX726057	x	x	*H. discus*	gi|556096296	*L. gigantea*	
Tektin-3-like	TELNHETDNMITETNSLQEAK	CX726057	x	x	*H. discus*	gi|556096297	*L. gigantea*	[8, 18, 37-39]
	VPPVFAAAR	CX726057	x	x	*H. discus*	gi|524899863	*A. californica*	
	VLQEIFDMEK	CU998137[b]	x	x	*C. gigas*	gi|524889528	*A. californica*	
	MAENSLQELLR	CU998137	x	x	*C. gigas*	gi|556096296	*L. gigantea*	
	YYSPGYSDALLER	FK719814[g]	x	x	*I. obsoleta*	gi|524882912	*A. californica*	
	AVAVANVLPK	yes	gi|405950834	K1PQU8	*C. gigas*	-	-	
Sperm-associated antigen 6	GDILPQLVYSLAEQNR	yes	gi|405950834	K1PQU8	*C. gigas*	-	-	[8, 13, 30, 31]
	IAASALSDICK	yes	gi|405950834	K1PQU8	*C. gigas*	-	-	
	LANYNDDLAEAVVK	yes	gi|405950834	K1PQU8	*C. gigas*	-	-	
	AAQQTDEILNSILPPR	yes	gi|405957088	B6RB89	*H. discus discus*	-	-	
Axonemal dynein light chain p33 (Fragment)	ELYSQCFDELIR	yes	gi|405957088	B6RB89	*H. discus discus*	-	-	[8, 24, 43]
	TQLEGIIAPK	yes	gi|405957088	B6RB89	*H. discus discus*	-	-	
	YDNPVLVSR	yes	gi|405957088	B6RB89	*H. discus discus*	-	-	
	NLPMDVAIECLTLR	x	gi|405950079	K1PNN3	*C. gigas*	-	-	
Tektin-2	TYRPNVELCR	x	gi|405950079	K1PNN3	*C. gigas*	-	-	[8, 13, 30, 31]
	GSTTPQQWEDFSR	yes	gi|405950079	K1PNN3	*C. gigas*	-	-	
	VGDSFEQLCLLQEAR	x	gi|405950079	K1PNN3	*C. gigas*	-	-	
	FQSSAVMALQEASEAYLVGLFEDTNLCAIHAK	yes	gi|405962350	K1QGW2	*C. gigas*	-	-	
Histone H3 (Fragment)	STELLIR	yes	gi|405962350	K1QGW2	*C. gigas*	-	-	[82-84]
	YRPGTVALR	yes	gi|405962350	K1QGW2	*C. gigas*	-	-	
	VTIMPK	yes	gi|405962350	K1QGW2	*C. gigas*	-	-	
	GELIDNLVNMLEK	yes	gi|405958266	x	*C. gigas*	-	-	
Atrial natriuretic peptide receptor A	VETIGDAYMVVSGLPIR	yes	gi|405958266	x	*C. gigas*	-	-	[8, 24, 52-54]
	YCLFGDTVNTASR	yes	gi|405958266	x	*C. gigas*	-	-	
	YSSNLEAIVSDR	yes	gi|405958266	x	*C. gigas*	-	-	
Dynein heavy chain 5, axonemal	LVITPLTDR	yes	gi|405957466	K1PX83	*C. gigas*	-	-	[8, 43]
	YPLLIDPQGQGK	yes	gi|405957466	K1PX83	*C. gigas*	-	-	
71kDa heat shock protein	DAGTISGLNVLR	yes	gi|109689150	Q17UC1	*H. tuberculata*	-	-	[8]
	TTPSYVAFTDTER	yes	gi|109689150	Q17UC1	*H. tuberculata*	-	-	
Tektin A1	TLIDNILEDTSR	x	gi|211998646	C6F0B3	*H. asinina*	-	-	[8, 18]
	VQDINFWK	yes	gi|211998646	C6F0B3	*H. asinina*	-	-	
Radial spoke head protein 9-like protein	LASVITEIEDEVR	GT866847[c]	x	x	*H. diversicolor*	gi|524897833	*A. californica*	[8, 44, 45]
	NLDLPFML	x	gi|405959092	K1Q8G0	*C. gigas*	-	-	
Enkurin	VPVPTTDNFLK	CX726831[a]	x	x	*H. discus*	gi|405973830	*C. gigas*	[8, 34-36]
Parkin coregulated gene protein homolog-like	GDFPIALEHDTK	jgi|Lotgi1|235731|[d]	x	x	*L. gigantea*	gi|556097114	*L. gigantea*	[8, 46, 47]
Nucleoside diphosphate kinase homolog 5-like	FFPDSVVEPVAIGQAAK	GT867630[c]	x	x	*H. diversicolor*	gi|524910230	*A. californica*	[8, 49]
Sperm surface protein Sp17	GFQNILEGLAR	yes	gi|405961240	K1PZL4	*C. gigas*	-	-	[8, 13, 28-30]
Radial spoke head protein 4 homolog A-like isoform X2	STEVALAQVQR	GD240537[e]	x	x	*A. californica*	gi|524898469	*A. californica*	[8, 44, 45]
Cytosolic heat shock cognate protein 70	FEELNADLFR	FF076379[e]	x	x	*A. californica*	gi|77023193	*M. galloprovincialis*	[62, 63, 65]
Female proteins								
	ILEIIAEGK	CX726553[a]	x	x	*H. discus*	gi|164604844	*H. discus hannai*	
	MMLAPYLGPYLPTTCK	CX726553	x	x	*H. discus*	gi|164604844	*H. discus hannai*	
	YEFEAPDGSFFNTPGPFVASYIIK	CF805565[f]	x	x	*H. asinina*	gi|164604844	*H. discus hannai*	
	MAATTTGATLSVNGDPITVTR	yes	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	IAVFTVIMK	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	LAATAAIK	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	LAFQLFAR	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	LFGPEGVLTNR	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	LGLIATIDASSPTTPR	x	gi|164604844	x	*H. discus hannai*	-	-	[12-14]
	MLLDPSYHTQIQHDLQFMTR	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	NFVILSAVGESYVSFYPYSAER	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
Vitellogenin	TAAQSMTGPAVLAEGALFPVIAVTMMR	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	TALPAYPMPWEAITQVMLQYPNR	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	TCGLCSNMDGQQR	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	TEYTVEDPEFNAELPEGTMFVTK	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	TFDGVEYTLPNLGSCPAVLAMDCSPSK	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	TPENEAYWSVNIK	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	TSETIDLDIVPVLTK	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	VTPALFESR	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	YMAQTVTGIPGASK	x	gi|164604844	A9ZQ48	*H. discus hannai*	-	-	
	TDASVTIVTDLNAK	CF805559[f]	x	x	*H. asinina*	gi|91992398	*H. fulgens*	
	NVGFTTIGK	CF805559	x	x	*H. asinina*	gi|91992398	*H. fulgens*	
	AGCGDGIVFPK	yes	gi|91992392	A0MCR0	*H. corrugata*	-	-	
Vitelline envelope zona pellucida domain 10	AVNTGSEAGIRPISCDAIDGSSAR	yes	gi|91992392	A0MCR0	*H. corrugata*	-	-	[18-21]
	IADTLIAAK	yes	gi|91992392	A0MCR0	*H. corrugata*	-	-	
	MVDVLGHNLNGR	yes	gi|91992392	A0MCR0	*H. corrugata*	-	-	
	TQSPYFEAFTVNVNSALK	yes	gi|91992392	A0MCR0	*H. corrugata*	-	-	
	VIVSYGEPGSLVHQYTEEYTVTCTFSPK	yes	gi|91992392	A0MCR0	*H. corrugata*	-	-	
	AWSNGQGNEVGVR	GT868093[c]	x	x	*H. diversicolor*	gi|260408284	*H. rufescens*	
	EYQLSCTFGAK	GT868093	x	x	*H. diversicolor*	gi|260408284	*H. rufescens*	
Vitelline envelope zona pellucida domain protein 26	IEGNILAPQIVESFAGSGAR	GT868093	x	x	*H. diversicolor*	gi|260408284	*H. rufescens*	[18-21]
	LHLIDVLGR	yes	gi|260408284	D0EL64	*H. rufescens*	-	-	
	QVYALPIGR	yes	gi|260408284	D0EL64	*H. rufescens*	-	-	
	LTAPLPAYPVTPISDVK	x	gi|260408284	D0EL64	*H. rufescens*	-	-	
	AFSINTDTSLK	GT871243[c]	x	x	*H. diversicolor*	gi|260408278	*H. rufescens*	
Vitelline envelope zona pellucida domain 23	SSSSPIVLR	GT867712[c]	x	x	*H. diversicolor*	gi|260408278	*H. rufescens*	[18-21]
	GIRPVSCDAIDSSNAR	x	gi|260408278	D0EL61	*H. rufescens*	-	-	
	VIVSYGEPGSLLHQYTEEYTISCSFGAK	yes	gi|260408278	D0EL61	*H. rufescens*	-	-	
	QTIEESLAAPR	GT868910[c]	x	x	*H. diversicolor*	gi|260408268	*H. rufescens*	
Vitelline envelope zona pellucida domain protein 18	AGCGDGIVFPR	GT868910	x	x	*H. diversicolor*	gi|260408268	*H. rufescens*	[18-21]
	LPVSFPGK	GT868910	x	x	*H. diversicolor*	gi|260408268	*H. rufescens*	
Vitelline envelope zona pellucida domain 2 type 1 protein	YTITCSYQPGAK	yes	gi|315441238	E7CSC1	*H. asinina*	-	-	
	VFTVNGDPQLK	x	gi|315441239	E7CSC2	*H. asinina*	-	-	[18-21]
Vitelline envelope zona pellucida domain 2 type 7 protein	DGVNYALPVSYPGK	yes	gi|315441240	E7CSC2	*H. asinina*	-	-	[18-21]
	LHQSDEEYTITCTFQPK	yes	gi|315441240	E7CSC2	*H. asinina*	-	-	
Vitelline envelope zona pellucida domain protein 28	AVLDADPFTNCVFR	yes	gi|260408288	D0EL66	*H. rufescens*	-	-	
	EYQVTCVYGSR	x	gi|260408288	D0EL66	*H. rufescens*	-	-	[18-21]
Zona pellucida domain protein D	MVGVSPPFR	yes	gi|260408298	D0EL71	*H. rufescens*	-	-	[18-21]
	QTSAILADGCGTGYPFDK	GT868086[c]	x	x	*H. diversicolor*	gi|260408298	*H. rufescens*	
Nuclear autoantigenic sperm protein-like	NIICGEIPVAVNQFQEACAMLAK	GT867749[c]	x	x	*H. diversicolor*	gi|556116977	*L. gigantea*	[23, 25]
Vitelline envelope zona pellucida domain 5	ILLGGCGDGLVFR	x	gi|91992350	A0MCN6	*H. corrugata*	-	-	[18-21]
Vitelline envelope zona pellucida domain 6	IGDSLIAPSVLK	yes	gi|91992364	A0MCP6	*H. rufescens*	-	-	[18-21]

Vitelline envelope zona pellucida domain 7	LLLVDVLGR	GT868974[c]	gi|91992374	A0MCQ1	*H. rufescens*	-	-	[18-21]
Vitelline envelope zona pellucida domain 8	SAGFIVDGLTAR	GT866946[c]	x	x	*H. diversicolor*	gi|91992378	*H. discus hannai*	[18-21]
Vitelline envelope zona pellucida domain 9	SSFSPYFVSFHLGDDK	GT867720[c]	x	x	*H. diversicolor*	gi|91992390	*H. fulgens*	[18-21]
Vitelline envelope zona pellucida domain protein 13	GGCGDGIVFPR	x	gi|260408258	D0EL51	*H. fulgens*	-	-	[18-21]
Vitelline envelope zona pellucida domain protein 16	LFLTNVAGYK	GT870031[c]	x	x	*H. diversicolor*	gi|260408264	*H. rufescens*	[18-21]
Vitelline envelope zona pellucida domain protein 20	GVKPVSCDAIGVNTK	GT868311[c]	x	x	*H. diversicolor*	gi|260408272	*H. rufescens*	[18-21]
Vitelline envelope zona pellucida domain protein 21	EIQTNLGPLAK	x	gi|260408274	D0EL59	*H. rufescens*	-	-	[18-21]
Vitelline envelope zona pellucida domain protein 25	AGCGDGIVFAK	x	x	D0EL63	*H. rufescens*	-	-	[18-21]
Vitelline envelope zona pellucida domain protein 29	GLRPDGCDAIGTLTGR	GT870387[c]	x	x	*H. diversicolor*	gi|260408290	*H. rufescens*	[18-21]
Heat shock protein 60	NVIIEQSWGSPK	yes	gi|218683627	B8Y4I8	*B. glabrata*	-	-	[64, 65]
Zona pellucida domain A	LHSPEDEFK	GT866406[c]	gi|91992358	A0MCP3	*H. diversicolor*	-	-	[18-21]
Common proteins								
84kDa heat shock protein	ADLVNNLGTIAK	yes♀♂	gi|109689148♂♀	Q17UC2♀♂	*H. tuberculata*	-	-	[8]
	GVVDSEDLPLNISR	yes♀♂	gi|109689148♂♀	Q17UC2♀♂	*H. tuberculata*	-	-	
	SIYYITGESK	x♀♂	x♀♂	Q17UC2♂	*H. tuberculata*	-	-	
	NADDITQEEYAEFYK	yes♀; x♂	gi|109689148♀	Q17UC2♀	*H. tuberculata*	-	-	
	VFIMDNCEDLIPEYLNFVR	yes♀; x♂	gi|109689148♀	Q17UC2♀	*H. tuberculata*	-	-	
	YYTSQSGDEVTSLK	x♀♂	gi|109689148♀	Q17UC2♀	*H. tuberculata*	-	-	
Voltage-dependent anion channel 2-like protein	VNNSSQIGLGYTQK	x♀; yes♂	gi|298108443	D7RP02	*H. diversicolor*	-	-	[8, 36, 58, 77, 78]
	LTVSSLIEGK	x♀; yes♂	gi|298108443	D7RP02	*H. diversicolor*	-	-	
	TEITIEDQIAQGLK♀♂	yes	gi|298108443	D7RP02	*H. diversicolor*	-	-	
	LAFDTSFAPQTGK♀♂	yes	gi|298108443	D7RP02	*H. diversicolor*	-	-	
	WNEYGLTFTEK♀♂	yes	gi|298108443	D7RP02	*H. diversicolor*	-	-	
Histone H2A	AGLQFPVGR♀♂	yes	gi|126697370	B6RB38	*H. discus*	-	-	[82-84]
	VGAGAPVYLAAVLEYLAAEVLELAGNAAR♀♂	yes	gi|126697370	B6RB38	*H. discus*	-	-	
	GDEELDSLIK	x♀; yes♂	gi|126697370	B6RB38	*H. discus*	-	-	
	HLQLAIR	x♀; yes♂	gi|126697370	B6RB38	*H. discus*	-	-	
	LLSGVTIAQGGVLPNIQAVLLPK	x♀; yes♂	gi|126697370	B6RB38	*H. discus*	-	-	
Histone H4	DNIQGITKPAIR♀♂	yes	gi|51315709	Q6WV74	*M. chilensis*	-	-	[82-84]
	ISGLIYEETR♀♂	yes	gi|51315709	Q6WV74	*M. chilensis*	-	-	
	TVTAMDVVYALK♀♂	yes	gi|51315709	Q6WV74	*M. chilensis*	-	-	
	VFLENVIR	x♀; yes♂	gi|51315709	Q6WV74	*M. chilensis*			
Histone H2B	AMSIMNSFVNDIFER♀♂	yes	gi|89520692	A2CI31	*C. farreri*♀	-	-	[82-84]
	-		gi|158997667	I1SKJ4	*A. californica*♂	-	-	
	LLLPGELAK♀♂	yes	gi|89520692	A2CI31	*C. farreri*♀	-	-	
	-		gi|158997667	I1SKJ4	*A. californica*♂	-	-	
	ESYSIYIYK	x	gi|158997667	I1SKJ4	*A. californica*♂	-	-	
	QVHPDTGISSK	x	gi|158997667	I1SKJ4	*A. californica*♂	-	-	
Ubiquitin	TITLEVEPSDTIENVK♀♂	yes	gi|388571222	J9Q7E3	*O. edulis*♀	-	-	[70-74]
	-	yes	gi|12240042	Q9BH32	*B. glabrata*♂	-	-	

Footnote. (x): peptide absent in database; (yes): peptide present in database; (♀): female; (♂): male. italics: protein name obtained from BLASTp. When peptides were present in custom database: a, GenBank (CX725921.1- CX726204.1). b, GenBank (CU984078.1-CU998429.1). c, GenBank (GT866281.1-GT873349.1). d, Joint Genome institute (http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html). e, GenBank (LIBEST_022923). f, GenBank (LIBEST_014519). g, Lambert et al, 2010. (Dashed rows): Excluding ESYSIYIYK and QVHPDTGISSK the rest of peptides were also identified in a L. gigantea EST found in the custom-built database. This EST is associated to the NCBI accession 556098074 that encodes a hypothetical protein termed histone cluster that contains regions of H4 and H2B.