

This file is part of the following work:

**Whitelaw, Brooke (2021) *The iridescent enigma: genome evolution and species boundaries of the blue-ringed octopus species complex (Octopodidae: Hapalochlaena)*. PhD Thesis, James Cook University.**

Access to this file is available from:

<https://doi.org/10.25903/691j%2Daa97>

Copyright © 2021 Brooke Whitelaw.

The author has certified to JCU that they have made a reasonable effort to gain permission and acknowledge the owners of any third party copyright material included in this document. If you believe that this is not the case, please email

[researchonline@jcu.edu.au](mailto:researchonline@jcu.edu.au)

**The iridescent enigma: Genome evolution and species boundaries of the blue-ringed octopus species complex (Octopodidae: *Hapalochlaena*)**

Thesis submitted by

Brooke Whitelaw

Date: 27/7/21

For the degree of Doctor of Philosophy

In the Centre of Sustainable Tropical Fisheries and Aquaculture  
(CSTFA)

College of Science and Engineering

James Cook University

# ACKNOWLEDGEMENTS

Thank you.

To the funding bodies, collaborators and institutions (museums: WAM, AM, MV, NTM) for making these projects possible.

To Professor Jan Strugnell, for being such an amazing academic role model and always providing guidance, support and encouragement, and for believing in me when I was doubtful myself. It means more than I can say.

To Dr Ira Cooke, for being the best bioinformatic mentor I could have asked for and encouraging me to take on new challenges, like a casual non-model genome annotation project. I have learnt so much from you since the days you explained what a coding language was.

To Dr Julian Finn, for sharing your enthusiasm and expertise on the amazing blue-ringed octopus genus and giving me somewhere to retreat to in chilly Melbourne.

To Professor Kyall Zenger, for your support and guidance and linkage map expertise.

To Professor Oleg Simakov, for your generosity in sharing your expertise in genome annotation and analysis and Elena for your support and kindness when I came to Vienna.

To Dr Dave Jones and Jarrod Guppy, for all your support and assistance throughout the linkage map analysis. I was very lucky to have your guidance.

To Legana, Shannon, Maria, Alyssa and the marine omics group, for your support and encouragement.

To my friends Anna, Kai and Mana for your encouragement and support throughout the years. You are amazing!

To my partner Daniel for your unending support and patiently listening to my bioinformatic rambling.

To my family, for their unfailing encouragement and confidence throughout these long few years.

## STATEMENT OF THE CONTRIBUTION OF OTHERS

Thesis chapter	Publication and/or conference output on which the chapter is based	Nature and extent of contribution of co-contributors
1	<p>The evolution and origin of tetrodotoxin acquisition in the blue-ringed octopus (genus <i>Hapalochlaena</i>)</p> <p><b>Brooke L Whitelaw</b>, Ira R Cooke, Julian Finn, Kyall Zenger, Jan M Strugnell - <i>Aquatic Toxicology</i>, Volume 206, January 2019, Pages 114-122,  <a href="https://doi.org/10.1016/j.aquatox.2018.10.012">https://doi.org/10.1016/j.aquatox.2018.10.012</a></p>	<p><b>Brooke L. Whitelaw</b> Primary author who conducted extensive reading and consolidation of literature.</p> <p><b>Ira R. Cooke</b> Advisor on manuscript structure and aided with editing/proofing</p> <p><b>Julian Finn</b> Advisor on manuscript structure and aided with editing/proofing</p> <p><b>Kyall Zenger</b> Advisor on manuscript structure and aided with editing/proofing</p> <p><b>Jan M. Strugnell</b> Advisor on manuscript structure and aided with editing/proofing</p>
2	<p>Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion and loss</p> <p><b>Brooke L Whitelaw</b>, Ira R Cooke, Julian Finn, Rute R da Fonseca, Elena A Ritschard, M T P Gilbert, Oleg Simakov, Jan M Strugnell - <i>GigaScience</i>, Volume 9, Issue 11, November</p>	<p><b>Brooke L. Whitelaw</b> Project co-conception, genome annotation, investigative analyses, computational analyses, results interpretation, figure generation, primary manuscript author</p> <p><b>Ira R. Cooke</b> Project co-conception, genome sequencing and assembly, bioinformatics support, manuscript structure and aided with editing/proofing</p>

	<p>2020, giaa120,  <a href="https://doi.org/10.1093/gigascience/giaa120">https://doi.org/10.1093/gigascience/giaa120</a></p>	<p><b>Julian Finn</b> Project co-conception, manuscript structure and aided with editing/proofing</p> <p><b>Rute R. da Fonseca</b> Project co-conception, manuscript structure and aided with editing/proofing</p> <p><b>Elena A. Ritschard</b> Project co-conception, manuscript structure and aided with editing/proofing</p> <p><b>M. T. P. Gilbert</b> Project co-conception, manuscript structure and aided with editing/proofing</p> <p><b>Oleg Simakov</b> Project co-conception, bioinformatics support, interpretation of results, manuscript structure and aided with editing/proofing</p> <p><b>Jan M. Strugnell</b> Project co-conception, supervision, funding, genome sequencing and assembly, results interpretation, mentorship, manuscript structure and editing/proofing</p>
<p><b>3</b></p>	<p>High density linkage map of the southern blue-ringed octopus (<i>Hapalochlaena maculosa</i>)</p> <p>Brooke L. Whitelaw, David B. Jones, Jarrod Guppy, Peter Morse, Jan M. Strugnell, Ira. R Cooke, Kyall Zenger</p> <p>Target journal: Genome Biology and Evolution</p>	<p><b>Brooke L. Whitelaw</b> Project development, computational analyses, results interpretation, figure generation, primary manuscript author</p> <p><b>David B. Jones</b> Mentorship, bioinformatics support, results interpretation and aided with editing/proofing</p>

		<p><b>Jarrold Guppy</b> Mentorship, bioinformatics support, results interpretation and aided with editing/proofing</p> <p><b>Peter Morse</b> Conducted an experiment examining multiple paternity and mating dynamics in <i>Hapalochlaena maculosa</i>. The resulting genotype data generated was used for linkage map construction in this study.</p> <p><b>Jan M Strugnell</b> Manuscript structure and aided with editing/proofing</p> <p><b>Ira R Cooke</b> Manuscript structure and aided with editing/proofing</p> <p><b>Kyall Zenger</b> supervision, funding, results interpretation, mentorship and editing/proofing</p>
4	<p>SNP data reveals the complex and diverse evolutionary history of the blue-ringed octopus genus (Octopodidae: <i>Hapalochlaena</i>) in the Asia-Pacific</p> <p>Brooke L. Whitelaw, Julian Finn, Kyall Zenger, Ira R Cooke, Jan M Strugnell</p> <p>Target Journal: Molecular Phylogenetics and Evolution</p>	<p><b>Brooke L. Whitelaw</b> Project co-conception, DNA extractions, computational analyses, results interpretation, figure generation, primary manuscript author</p> <p><b>Julian Finn</b> Project co-conception, supervision, funding, sampling, results interpretation, mentorship and editing/proofing</p> <p><b>Kyall Zenger</b> Results interpretation, manuscript structure and aided with editing/proofing</p> <p><b>Ira R Cooke</b> Supervision, results interpretation, mentorship and editing/proofing</p>

		<b>Jan M Strugnell</b> Project co-conception, supervision, funding, results interpretation, mentorship and editing/proofing
--	--	---



## ABSTRACT

Understanding the evolution of unique evolutionary traits in octopods provides an interesting challenge due to their large and complex genomes. The genus of focus for this thesis is the blue-ringed octopus genus (*Hapalochlaena*), currently composed of three accepted species (*H. maculosa*, *H. lunulata*, and *H. fasciata*). Members of this genus are identifiable by their iridescent blue lines &/or rings, which are flashed in an aposematic display advertising their toxicity. *Hapalochlaena* are the only octopods known to sequester the potent neurotoxin tetrodotoxin (TTX) within their tissues and venom. Inclusion of this toxin within their venom is responsible for human fatalities and hospitalisations. Given this negative impact on human health, understanding the evolution of this genus at a genomic and taxonomic level is crucial. Genomic resources for the genus are currently limited with no published genomes available. Additionally, the impact of TTX acquisition on the evolution of *Hapalochlaena* at a genomic level has not been elucidated.

In order to compare the genome evolution of the TTX bearing octopod lineage (*Hapalochlaena*) to other non-TTX-bearing octopods (*Octopus bimaculoides*

& *Callistoctopus minor*) the southern blue-ringed octopus (*H. maculosa*) genome was annotated and comparative analyses performed. Annotation identified 29,328 genes and a high repeat content with repetitive elements comprising 37.09% of the *H. maculosa* genome. Divergence times between lineages were calculated using a concatenated alignment of 2,108 genes, placing the divergence of *H. maculosa* from the genus *Octopus* at ~59 mya. Additionally, gene families associated with neural development and function (zinc finger and cadherin) were identified as exhibiting the largest octopod specific gene family expansions. Examination of tissue specific expression in the posterior salivary glands (venom glands) revealed a dominance of serine protease expression in non-TTX octopods relative to the TTX bearing *H. maculosa* in accordance with an overall reduction in gene family size exhibited in *H. maculosa*. Mutations, which confer tetrodotoxin resistance in pufferfish (*Tetraodon nigroviridis*) and newt (*Taricha granulosa*) were identified in the sodium channel of *H. maculosa* along with additional mutations exclusive to the genus, which are candidates for resistance. The source of tetrodotoxin within *Hapalochlaena* is yet to be confirmed, however bacterial production is one hypothesis. Analysis of the microbiome of the posterior salivary gland revealed a diverse array of bacterial families with no single dominant family as a candidate for production. This study provides the first annotated blue-ringed octopus genome and large-scale genomic comparisons between three octopod genomes.

In the absence of a published genetic linkage map for any cephalopod, a high-density genetic linkage map for the southern blue-ringed octopus *H. maculosa* was generated. Linkage maps have high versatility in genetic studies and can be used to aid in the reassembly and refinement of non-model genomes. The linkage map was generated using 2,166 single nucleotide polymorphisms (SNP) and 2,455 presence absence variant (PAV) loci from 10 families comprised of 276 individuals. A total of 47 linkage groups (LG) spanning 2016cM were resolved for an estimated 30n chromosomes. The *H. maculosa* genome was composed of 47K scaffolds and reassembly using the generated linkage map allowed for the placement of 1,151 scaffolds using 1,278 markers covering 34.7% of the total genome. The updated assembly was used to examine placement of HOX genes, a highly conserved cluster in metazoa involved in development. Genes within the HOX gene cluster are usually found within close proximity within the Phylum Mollusca. However, in *Octopus bimaculoides* all HOX genes were demonstrated to occur on separate scaffolds, with no greater resolution of their placement in relation to each other possible. Due to the combination of both genomic and linkage data in *H. maculosa*, three (SCR, LOX4 and POST1) of the eight genes identified in cephalopod genomes were able to be placed within a single pseudo-chromosome (LG 9). Contrary to other molluscan clusters such as *Lottia gigantea* the order of genes differs in *H. maculosa* with LOX4

not located between SCR and POST1. Implications of these changes have yet to be investigated.

Delineation of species within the blue-ringed octopus (*Hapalochlaena*) species complex was achieved using a combination of SNPs (10,346 resolved) and mitochondrial sequence data (12S rRNA, 16S rRNA, cytochrome c oxidase subunit 1 [COI], cytochrome c oxidase subunit 3 [COIII] and cytochrome b [Cytb]). Inferred relationships indicated the current species diversity observed is incongruent with the number of accepted species. A minimum of 11 distinct organisational taxonomic units (OTU) were identified from 21 sampling locations throughout the Asia Pacific from depths between 3m-100m using 10,346 resolved SNPs. Currently, only three species are accepted within the complex (*H. maculosa*, *H. fasciata* and *H. lunulata*). Support was present for *H. maculosa* as currently defined, spanning across southern Australian waters (Victoria to Western Australia). Likewise, the blue-lined octopus (*H. fasciata*) located off NSW, Australia also formed a single defined unit. However, lined octopus collected from the North Pacific, also recognised as *H. fasciata*, were not found to be a sister taxon to NSW populations and determined to be a distinct OTU. Where samples were shared between data sets, delineation methods (ABGD, single rate-PTP and GYMC) corroborated OTU assignments. Bayesian analysis of the COI gene estimated divergence of the *Hapalochlaena* genus

from *Amphioctopus* to have occurred ~50mya, with major *Hapalochlaena* lineages diverging between ~50-25mya. The point of origin for the genus was estimated by RASP (Reconstruct Ancestral State in Phylogenies) analysis using COI genes to be located within the Central Indo Pacific. In order to fully resolve systematics within the genus and describe new species, additional morphological work is required to be used in conjunction with this study.

This thesis investigated the evolution of the *Hapalochlaena* species complex using a range of genetic tools including whole genomic comparisons, genetic linkage and reduced representation sequencing (GBS), while also providing genetic tools and resources for future research. The current state of *Hapalochlaena* systematics was revealed to be insufficient for the species diversity observed. Furthermore, evolution of the *Hapalochlaena* genome revealed distinct differences to the genomes of non-TTX-bearing octopods.

# TABLE OF CONTENTS

<b>Chapter 1</b>	<b>General Introduction.....</b>	<b>1</b>
1.1	Abstract.....	1
1.2	Introduction.....	1
2	Tetrodotoxin (TTX).....	2
2.1	Basis of TTX toxicity.....	2
2.2	Convergent occurrence of TTX in distant taxa.....	2
3	Evolution of tetrodotoxin (TTX) and saxitoxin (STX) acquisition.....	2
3.1	TTX and STX resistant channels.....	2
3.1.1	Domain I (DI).....	3
3.1.2	Domain II (DII).....	4
3.1.3	Domain III (DIII).....	4
3.1.4	Domain IV (DIV).....	4
3.2	Transport of TTX and STX.....	4
3.3	Sequestering of TTX and STX.....	5
3.4	Cost of TTX exploitation.....	5
4	TTX and STX acquisition in Octopoda.....	5
4.1	STX accumulation in <i>O. vulgaris</i> .....	5
4.2	Recruitment of TTX into <i>Hapalochlaena</i> venom.....	6
4.3	Microbiome of the <i>Hapalohclaena</i> posterior salivary gland.....	6
5	Role of TTX in <i>Hapalochlaena</i> .....	6
5.1	Role of TTX defence and offence.....	6

5.2	Impact of TTX on venom evolution.....	7
6	Conclusion.....	7
7	Conflict of Interest.....	7
8	References.....	7

**Chapter 2      Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion, and loss.....10**

2.1	Abstract.....	10
2.2	Background.....	11
2.3	Data Description.....	11
2.3.1	Genome assembly and annotation.....	11
2.3.2	PSMC and mutation rate.....	11
2.3.3	Phylogenomics.....	12
2.3.4	Organismal impact of novel genes and gene family expansions.....	12
2.3.5	Novel patterns of gene expression.....	12
2.3.6	Evolution of the octopus non-coding genome.....	14
2.3.7	Dynamics of gene expression in the PSG.....	14
2.3.8	TTX resistance of the Na <sub>v</sub> channels.....	15
2.3.9	Microbiome of the PSG.....	15
2.4	Conclusions.....	16
2.5	Methods.....	16
2.5.1	Genome sequencing and assembly.....	16
2.5.2	Transcriptome sequencing.....	16
2.5.3	De novo transcriptome assembly.....	17

2.5.4	Genome annotation.....	18
2.5.5	Heterozygosity.....	18
2.5.6	Repetitive and transposable elements.....	18
2.5.7	Calibration of sequence divergence with respect to time.....	19
2.5.8	Effective population size (PSMC).....	19
2.5.9	Mutation rate.....	19
2.5.10	Quantifying gene expression/specificity.....	19
2.5.11	Gene model expression dynamics.....	19
2.5.12	Dynamics of PSG gene expression.....	19
2.5.13	The role of the Nav in TTX resistance.....	20
2.6	Availability of Source Code and requirements.....	20
2.7	Availability of Supporting Data and Materials.....	20
2.8	Additional files.....	20
2.9	Abbreviations.....	20
2.10	Ethics Declaration.....	21
2.11	Competing Interests.....	21
2.12	Funding.....	21
2.13	Acknowledgements.....	21
2.14	References.....	21
<b>Chapter 3</b>	<b>High density Genetic Linkage map of the southern blue-ringed octopus (Octopodidae: <i>Hapalochlaena maculosa</i>).....</b>	<b>25</b>
3.1.	Abstract.....	25
3.2.	Introduction.....	26
3.3.	Methods.....	31



3.3.1. Sample collection and family structure.....	31
3.3.2. DNA extraction and SNP generation.....	31
3.3.3. SNP selection, quality control and genotyping.....	31
3.3.4. Paternity and family designation.....	32
3.3.5. Map construction.....	33
3.3.6. Re-orientation of genomic scaffolds and genes of interest.....	34
3.3.7 Genome Coverage.....	35
3.3.8. Segregation distortion and sex specific recombination.....	36
3.4 Results.....	37
3.4.1. Genotyping and paternity.....	37
3.4.2. Linkage map construction and genome coverage.....	41
3.4.3. Genome mapping and scaffold re-orientation.....	45
3.4.4. Genes of interest (HOX).....	49
3.5. Discussion.....	53
3.5.1. Paternity and linkage map generation.....	53
3.5.2. Cephalopod genome evolution and structure.....	57
3.5.3. Evolution of the HOX gene cluster in cephalopods.....	59
3.5.4. QTL mapping and future work.....	60
3.6. Conclusion.....	62

**Chapter 4 SNP data reveals the complex and diverse evolutionary history of the blue-ringed octopus genus (*Octopodidae: Hapalochlaena*) in the Asia-Pacific.....63**

4.1. Abstract.....	63
4.2. Introduction.....	64

4.3.	Methods.....	71
4.3.1.	Sample collection.....	71
4.3.2.	Mitochondrial DNA extraction and amplification.....	72
4.3.3.	DNA extraction and Genotyping By Sequencing.....	72
4.3.4.	SNP identification.....	73
4.3.5.	Genetic structure/divergence of SNP data.....	73
4.3.6.	Signatures of selection.....	74
4.3.7.	Putative species limit estimation using mitochondrial genes.....	75
4.3.8.	Divergence time estimation.....	76
4.4.	Results.....	78
4.4.1.	Species boundaries OTU identification.....	78
4.4.1.1.	Temperate Australasia (TAUS): OTU K.....	78
4.4.1.2.	Lined Blue-ringed octopus: OTU D, E & I.....	85
4.4.1.3.	North Western Australia and Timor Leste: OTU G & F.....	86
4.4.1.4.	Great Barrier Reef: OTU H (H).....	87
4.4.1.5.	North QLD and Western Australia: OTU J.....	88
4.4.1.6.	“Deep water”: OTU A.....	88
4.4.1.7.	Singletons WA: OTU B & C .....	89
4.4.2.	Divergence time estimation/Ancstral state reconstruction (RASP)....	93
4.5	Discussion.....	94
4.5.1	Species boundaries and OTUs.....	94

4.5.2	Dating and diversification of the genus.....	100
4.5.3	Mechanisms of divergence and speciation.....	102
4.6	Conclusion.....	104
<b>Chapter 5</b>	<b>General Discussion.....</b>	<b>105</b>
5.1	Summary.....	105
5.2	Major outcomes.....	105
5.3	Significance and contributions to a greater understanding.....	112
5.4	Future directions.....	117
REFERENCES.....		123
APPENDICES.....		149

## LIST OF TABLES

<b>Table 1.1</b> Substitutions conferring tetrodotoxin and saxitoxin resistance in the Na <sub>v</sub> P-loop regions of domains I, II, III and IV.....	3
<b>Table 3.1</b> Summary of <i>Haplochlauena maculosa</i> families generated in this study. Bolded rows represent families which passed filtering and were included in linkage map generation. n = size of family.....	38
<b>Table 3.2</b> Filtering and quality assessment of SNPs produced using DARTQC. Rep average is a metric specific to DArT (Diversity Arrays) specifying the proportion of alleles that give repeatable results over 30 replicates.....	39
<b>Table 3.3</b> SNPs removed through filtering in ten <i>Haplochlauena maculosa</i> families using PLINK.....	40
<b>Table 3.4</b> Statistical summary of <i>Haplochlauena maculosa</i> sex averaged, male and female linkage maps.....	43
<b>Table 3.5</b> Genome coverage estimation of sex average, female and male linkage maps.....	45
<b>Table 3.6</b> Characteristics and summary of <i>Haplochlauena maculosa</i> chromonomer assembly for the 47 linkage groups (LG)/pseudo-chromosomes.....	46
<b>Table 3.7</b> Assembly statistics for the original <i>Haplochlauena maculosa</i> assembly and the new chromonomer generated assembly.....	49

## LIST OF FIGURES

**Figure 1.1.** Schematic of voltage-gated sodium channel (Nav) alpha subunits (DI, DII, DIII and DIV). Each unit is composed of six subunits 1–4 (orange) and 5–6 (green). Alternating extra and intracellular loops are shown in black with the p-loops between subunits 5 and 6 highlighted in red (p-loops are the site of TTX binding, sequences shown in Fig.2). **a)** A lateral view of the units embedded within the cellular membrane (grey). **b)** Top down view of the alpha subunits (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).....3

**Figure 1.2.** Alignment of the four p-loop domains (DI, DII, DIII and DIV) in the NaV1.4 sequences from pufferfish (blue), salamanders (green), garter snakes (orange) and the soft-shelled clam (red). Mutations known to confer tetrodotoxin and saxitoxin resistance are shown in bold and coloured appropriately. Mutations suspected to provide resistance are shown underlined in bold (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).....4

**Figure 2.1.** Comparisons of molluscan genomes and gene families. (A) Time-calibrated maximum likelihood phylogeny of 7 molluscan genomes (*Aplysia californica*, *Lottia gigantea*, *Crassostrea gigas*, *Euprymna scolopes*, *Octopus bimaculoides*, *Callistoctopus minor*, and *Hapalochlaena maculosa*) and 4 transcriptomes (*Octopus kaurna*, *Octopus vulgaris*, *Sepia officinalis*, and *Idiosepius notoides*) using 2,108 single-copy orthologous sequence clusters. Node labels show divergence times in millions of years (mya); blue (divergence to octopods) and orange bars (decapods) represent standard error within a 95% confidence interval. Octopodiformes lineages are highlighted in blue and decapod orange. Scale bar

represents mya. (B) Expansions of octopod gene families relative to molluscan genomes *Aplysia californica* (*A. cali*), *Biomphalaria glabrata* (*B. glab*), *C. gigas* (*C. gig*), *L. gigantea* (*L. gig*), *E. scolopes* (*E. scol*), *C. minor* (*C. min*), *O. bimaculoides* (*O. bim*), and *H. maculosa* (*H. mac*). (C) Lineage-specific gene expansions in the octopod genomes *C. minor* (*C. min*), *O. bimaculoides* (*O. bim*), and *H. maculosa* (*H. mac*). CHGN: chondroitin N-acetylgalactosaminyltransferase; C2H2: Cys2-His2; SPRR: small proline-rich proteins.....13

**Figure 2.2.** Dynamics of gene expression in octopod genomes. Proportion of gene expression across levels of specificity from not specific to octopods or an octopus species (left) to octopod-specific (middle) and lineage-specific (right). Donut plots show gene expression as some expression in any tissue (purple), no expression (blue), or expression that has been lost (dark blue). Loss of expression requires an ortholog of the gene to be expressed in  $\geq 1$  species and not expressed in the other species. Heatmaps at each specificity level show average expression of genes within their respective tissues, low expression (cream) to high expression (dark red).....14

**Figure 2.3.** Dynamics of gene expression in neural and venom-producing tissues of octopods. Tissue-specific expression of genes within the brain (red) and posterior salivary gland (PSG) (blue) of *H. maculosa* (*H.mac*), *O. bimaculoides* (*O.bim*), and *C. minor* (*C.min*). A) Venn diagram shows numbers of shared and exclusive genes between species (left). B) Bar chart of the top 5 Pfams and their contribution to overall expression in the brain (right).....16

**Figure 2.4.** Examination of posterior salivary gland (PSG) gene expression between 3 octopod genomes. (A) Heat map of genes expressed specifically in the PSG of *H. maculosa* ( $\tau > 0.8$ ) and their orthologs in *O. bimaculoides* and *C. minor* lacking specific expression to the PSG ( $\tau < 0.8$ ). Genes with an ortholog lacking expression

are coloured in grey while the absence of an ortholog is white. (B) Heat map of genes expressed specifically in the PSG of both *O. bimaculoides* and *C. minor* ( $\tau > 0.8$ ) and their orthologs in *H. maculosa* lacking specific expression to the PSG.....17

**Figure 2.5.** Mechanism of tetrodotoxin resistance within the posterior salivary gland of *H. maculosa* (PSG). (A) Alignment of voltage-gated sodium channel  $\alpha$ -subunits (DI, DII, DIII, and DIV) p-loop regions. Mutations conferring resistance are coloured in green (pufferfish), orange (salamander), purple (clam), and blue (octopus). Susceptible mutations at the same site are black and boldface. Sites that may be involved with resistance are in boldface. (B) Schematic of voltage-gated sodium channel (Nav)  $\alpha$ -subunits (DI, DII, DIII, and DIV). Each unit is composed of 6 subunits, 1–4 (blue) and 5–6 (yellow). Alternating extra- and intracellular loops are shown in black with the p-loops between subunits 5 and 6 highlighted in red. Mutations conferring resistance are shown within black circles on p-loops.....18

**Figure 2.6.** Assessment of bacteria within the posterior salivary gland of *H. maculosa* (PSG). (A) Bacterial composition at the phylum level of an *H. maculosa* posterior salivary/venom gland. (B) Composition of the largest phylum, Protobacteria, of an *H. maculosa* posterior salivary/venom gland.....19

**Figure 3.1.** Genetic linkage map generated using LEPMAP3 for *Haplochlæna maculosa* a) sex averaged map b) maternal map c) paternal map. Marker density (cm/Locus) is visualised in a scale from red (high) to blue (low).....42

**Figure 3.2.** Comparison of HOX gene arrangement in the **a)** owl limpet (*Lottia gigantea*) **b)** California two spot octopus (*Octopus bimaculoides*) and the southern blue-ringed octopus (*Haplochlæna maculosa*). Genes are coloured consistently

between species. Scaffold lengths are not to scale.....51

**Figure 3.3.** Maximum likelihood phylogeny of the HOX genes (aa) in six cephalopods. Bootstrap values >70 present on branches. Taxa are coloured in accordance with HOX genes.....52

**Figure 4.1.** Published distributions of *Hapalochlaena* species. Distributions of *Hapalochlaena* sp. in the Indo Pacific. **a)** A guide to Squid, Cuttlefish and Octopuses of Australasia (Norman and Reid, 2000). *Hapalochlaena* are shown as follows, *H. maculosa* (orange), *H. fasciata* (red) and *H. lunulata* (mustard). **b)** Cephalopods a World Guide (Norman, 2000). Additional reported localities, *Hapalochlaena* sp. 1 (green), *Hapalochlaena* sp. 2 (purple), *Hapalochlaena* sp. 3 (blue), *Hapalochlaena* sp. 4 (pink) and *Hapalochlaena* sp. 5 (brown). **c)** FAO. Cephalopods of the world. Volume 3. Octopods and Vampire Squids (Norman et al., 2016). *Hapalochlaena* are shown as follows, *H. maculosa* (orange), *H. fasciata* (red) and *H. lunulata* (mustard). **d)** Cephalopods of Australia and Sub-Antarctic Territories (Reid, 2016). *Hapalochlaena* are shown as follows, *H. maculosa* (orange), *H. fasciata* (red) and *H. cf. lunulata* (mustard).....80

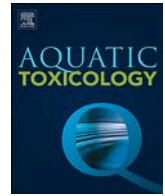
**Figure 4.2.** Figure 4.2. Delineation of *Hapalochlaena* species boundaries and genetic structure throughout the Indo Pacific using 10,346 SNPs: a) SVDQuartet phylogeny of *Hapalochlaena* throughout the Indo Pacific generated using 10,346 SNPs, coloured branches represent putative taxonomic units A-K: light blue (A/Deep water Yeppoon, QLD & North West, WA), dark blue (B/North West Shelf, WA), light green (C/Ningaloo, WA), apple green (D/Taiwan lined), pink (E/Taiwan ringed), red (F/Timor Leste), light orange (G/Great Barrier Reef, QLD), dark orange



(H/H\*/Darwin, NT & Kimberly & Exmouth, WA ), lilac (I/NSW), purple (J/Cape York, QLD & Shark Bay, QLD) and brown (K/Southern coast of Australia). Posterior support values > 0.90 present on nodes. Bars at terminal branches indicate admixture of OTUs inferred using STRUCTURE, colours approximately correspond to OTUs. Lined box adjacent to OTU indicated lined markings while, OTUs without a box exhibit ringed markings; b) Species delineation using the mitochondrial COI gene. Bayesian phylogeny (MrBayes) of *Hapalochlaena* throughout the Indo Pacific is coloured according to OTU with black used to represent taxa included from NCBI. Grey boxes represent putative species in accordance to sPTP, GYMC SC (strict clock) and GYMC RC (relaxed clock) methods; c) Map of sample locations coloured by organisational taxonomic units A-K; d) Arrangement of samples according to the first two principal components of a PCoA based on SNP data generated using the R dartR.....81

**Figure 4.3.** Divergence and radiation of *Hapalochlaena* throughout the Asia Pacific:

a) Divergence time estimates of *Hapalochlaena* throughout the Indo Pacific generated using Bayesian methods. Bars (blue) represent standard error within a 95% confidence interval; b) Reconstructed Ancestral State Phylogeny constructed using RASP BBM. Most likely points of origin inferred for each node in the tree are shown as pie charts coloured according to major geographic regions using the colour scheme shown in c; c) Map of oceanic zones: temperate Australasia (TAUS) = Green, North East Australia (N\_QLD) = Pink, North West Australia (NW\_AUS) = Red, Central Indo Pacific (CIP) = Blue and temperate north Pacific (TNP) = beige.....90



# The evolution and origin of tetrodotoxin acquisition in the blue-ringed octopus (genus *Hapalochlaena*)

Brooke L. Whitelaw<sup>a,\*</sup>, Ira R. Cooke<sup>b,d</sup>, Julian Finn<sup>e</sup>, Kyall Zenger<sup>a</sup>, J.M. Strugnell<sup>a,c</sup>

<sup>a</sup> Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Townsville, Queensland, 4811, Australia

<sup>b</sup> College of Public Health, Medical and Vet Sciences, James Cook University, Townsville, Queensland, 4811, Australia

<sup>c</sup> Department of Ecology, Environment and Evolution, La Trobe University, Melbourne, 3086, Vic. Australia

<sup>d</sup> La Trobe Institute of Molecular Science, La Trobe University, Melbourne, 3086, Vic. Australia

<sup>e</sup> Sciences, Museum Victoria, Carlton, Victoria 3053, Australia

## ARTICLE INFO

### Keywords:

Tetrodotoxin  
Octopod  
Hapalochlaena  
Toxin  
Venom  
Sodium channel

## ABSTRACT

Tetrodotoxin is a potent non-proteinaceous neurotoxin, which is commonly found in the marine environment. Synthesised by bacteria, tetrodotoxin has been isolated from the tissues of several genera including pufferfish, salamanders and octopus. Believed to provide a defensive function, the independent evolution of tetrodotoxin sequestration is poorly understood in most species. Two mechanisms of tetrodotoxin resistance have been identified to date, tetrodotoxin binding proteins in the circulatory system and mutations to voltage gated sodium channels, the binding target of tetrodotoxin with the former potentially succeeding the latter in evolutionary time. This review focuses on the evolution of tetrodotoxin acquisition, in particular how it may have occurred within the blue-ringed octopus genus (*Hapalochlaena*) and the subsequent impact on venom evolution.

## 1. Introduction

Toxicity is a trait found among a diverse array of phyla. The ability to produce, contain and/or secrete toxins may provide a defensive advantage against predators (Nelsen et al., 2014), or a competitive advantage by enabling a predator to tackle larger or more challenging prey (Mebs, 2001; Casewell et al., 2013). Endogenous production of proteinaceous toxins is a common feature among both poisonous and venomous species. Alternatively, non-proteinaceous toxins may be acquired from an exogenous source (Magarlamov et al., 2017), but the ability to utilise these freely available resources requires: i) sufficient resistance to the toxin to prevent harm to the host (Jost et al., 2008; Soong and Venkatesh, 2006); ii) availability of the toxin in sufficient quantities from an exogenous source; iii) transport and sequestration mechanisms to store adequate quantities of toxin within specific tissues (Matsumoto et al., 2007, 2010).

The marine environment contains free-living bacteria and dinoflagellates, which produce potent neurotoxins such as tetrodotoxin (TTX) and saxitoxin (STX) (Kem, 2014). For many marine species, exposure to these toxins can be fatal, with mass strandings of the Humboldt squid *Dosidicus gigas* a potential result of STX produced from dinoflagellate blooms (Reyero et al., 1999; Geraci et al., 1989). However, a select group of species have evolved the ability to use these potent

toxins to their advantage. This ability is likely to be the result of convergent evolution as it is spread throughout many taxonomic groups across both semi-terrestrial and marine phyla including Mollusca (Cheng et al., 1995; Crone et al., 1976; Kudo et al., 2014), Amphibia (Kotaki and Shimizu, 1993), Arthropoda (Noguchi et al., 1984), Chaetognatha (Thuesen and Kogure, 1989), Chordata (Yasumoto et al., 1988), Platyhelminthes (Miyazawa et al., 1987) and Echinodermata (Narita et al., 1987).

The presence of TTX and related toxins within species is often associated with a defensive function. A notable exception is the blue-ringed octopus genus (*Hapalochlaena*) (Williams and Caldwell, 2009; Yotsu-Yamashita et al., 2007), which has a distinct (most likely aposomatic) appearance with iridescent blue lines and/or rings advertising the sequestration of the potent neurotoxin TTX within tissues and venom (Freeman and Turner, 1970; Gage et al., 1975). Evolution of toxicity in this unique lineage has received little attention; as a result the methods of TTX attainment, transport and storage are largely unresolved (Williams et al., 2012a). Although *Hapalochlaena* are the most widely known octopod lineage to use an exogenous toxin other rare cases of toxicity within related octopus species have been observed and are associated with ingestion of STX contaminated bivalves. *Octopus vulgaris* was found to retain STX within the digestive gland post-ingestion of bivalves exposed to a toxic algal bloom and exhibit no symptoms

\* Corresponding author.

E-mail address: [brooke.whitelaw@my.jcu.edu.au](mailto:brooke.whitelaw@my.jcu.edu.au) (B.L. Whitelaw).

<https://doi.org/10.1016/j.aquatox.2018.10.012>

Received 8 June 2018; Received in revised form 22 October 2018; Accepted 22 October 2018

Available online 24 October 2018

0166-445X/© 2018 Elsevier B.V. All rights reserved.

of intoxication (Lopes et al., 2014, 2013). This review details the current understanding of the evolutionary processes involved in TTX acquisition and the current state of research on toxicity in *Haplochromis*.

## 2. Tetrodotoxin (TTX)

### 2.1. Basis of TTX toxicity

TTX intoxication results in the inhibition of signal transduction by nerve cells inducing paralysis and death in susceptible taxa (Cestèle and Catterall, 2000; How et al., 2003; Noguchi et al., 2011a). TTX is a potent non-proteinaceous neurotoxin composed of a guanidinium moiety, pyrimidine ring and six hydroxyl groups on a highly-oxygenated carbon backbone (C-4, C-6, C-8, C-9, C-10 and C-11) (Lee and Ruben, 2008). TTX derives its name from the Tetraodontidae; the family of fishes from which it was first isolated (Yasumoto et al., 1986). TTX analogues have been isolated from all TTX bearing taxa, among the most common are 4-epiTTX and anhydroTTX, which are putative conversion or equilibrium products (Hanifin, 2010; Yasumoto and Yotsu-Yamashita, 1996).

TTX inhibits action potential propagation by binding to voltage gated sodium channels (Na<sub>v</sub>) in muscle and nerve tissue resulting in paralysis (Catterall et al., 2007). Na<sub>v</sub> are membrane bound proteins responsible for the regulation of sodium channel flow between the intra and extracellular environment of cells in muscle and nerve tissues. Propagation of an action potential in a nerve cell is reliant on Na<sub>v</sub> channel activity; inhibition of sodium ion flow can result in paralysis of affected tissues (Frank and Catterall, 2003). Na<sub>v</sub> channels are composed of a large alpha subunit flanked by one or more accessory beta subunits. Functionally, the beta subunits influence kinetics of Na<sub>v</sub> gating, voltage dependency signal transduction, cell adhesion and channel expression (Lee and Ruben, 2008). The main alpha unit is a transmembrane bound protein composed of four homologous domains (DI–DIV), each domain is composed of six subdomains (S1–S6) and the subdomains are connected by alternating extra and intercellular loops. The P-loop region is located on the extracellular loop between the S5 and S6 subdomains and is crucial in the binding affinity of TTX to the Na<sub>v</sub> channel.

The biochemical processes underlying TTX binding affinity to Na<sub>v</sub> have been investigated through inhibition of TTX binding with carboxyl-modifying reagents (Schrager and Profera, 1973), pH titration (Hille, 1968), specific monovalent cations, divalent protons and metal ions (Henderson et al., 1974). These experiments provide evidence for interactions of the guanidinium and hydroxyl groups with the P-loop domains of Na<sub>v</sub>. Specifically, the hydroxyl groups of TTX C9, C10 and C11 are proposed to form hydrogen bonds with the Glu at position 945 in DII and Asp at position 1532 in DIV respectively, while negatively charged functional groups located in the DI domain form ion pairs with the guanidinium group (Lipkind and Fozzard, 1994; Hille, 1975; Choudhary et al., 2003).

### 2.2. Convergent occurrence of TTX in distant taxa

TTX is widespread throughout the animal kingdom and has been isolated from both marine and terrestrial species (Bane et al., 2014). The occurrence of TTX among phylogenetically unrelated species has been attributed to a bacterial origin (Chau et al., 2011; Simidu et al., 1987; Yan et al., 2005). TTX is abundant in the marine environment. Sediments examined from the western Pacific coast of Japan contained on average one lethal mouse unit of TTX within 10 g of sediment. Depth did not have a significant impact on TTX concentration with 21 m, 82 m and 4033 m samples containing similar quantities of the toxin (Kogure et al., 1988). Examination of bacterial communities in deep-sea marine sediments (4033 m) has revealed a diverse array of 22 TTX-producing strains from ten genera: *Bacillus*, *Micrococcus*, *Acinetobacter*, *Aeromonas*, *Alcaligenes*, *Altermonas*, *Flavobacterium*, *Moraxella*, *Pseudomonas* and *Vibrio* (Do et al., 1991, 1990). TTX-producing bacteria are not exclusive

to marine environments with five genera isolated from freshwater sediments (Do et al., 1993).

Bioaccumulation of TTX through direct ingestion of TTX-producing bacteria or TTX contaminated food sources at higher trophic levels has been supported for several members of the family Tetraodontidae. Non-toxic *Takifugu niphobles* and *T. rubripes* have been observed to acquire toxicity when fed a TTX containing diet (Kono et al., 2008a; Honda et al., 2005a), absence of toxicity was also observed in *T. rubripes* when fed a TTX-free diet (Noguchi et al., 2006a). Similar cases of bioaccumulation of the structurally and functionally similar toxin STX have been observed in bivalve molluscs through ingestion of toxic dinoflagellates (Bricelj and Shumway, 1998; Kvittek, 1991).

Not all TTX-producing bacteria have been isolated as free-living with *Pseudoaltermonas tetradonis* exclusively isolated from the pufferfish, *T. poecilonotus* and red algae (Ivanova et al., 2001; Simisu et al., 1990). Commensal associations with bacteria have been proposed as a source of TTX for TTX-bearing genera with high abundances of TTX-producing strains isolated from their tissues. The first culture of a TTX-producing strain from a toxic species was a *Vibrio* from the floral egg crab, *Atergatis floridus* (Noguchi et al., 1984). Subsequent studies have cultured TTX-producing bacterial strains from over 20 genera from a diverse range of TTX-bearing taxa including pufferfish (Noguchi et al., 1987), gastropods (Cheng et al., 1995; Wang et al., 2008), octopus (Hwang et al., 1989) and crustaceans (Noguchi et al., 1984). While impressive, this may be an underrepresentation of the true diversity and prevalence of TTX-producing endosymbiotic bacteria for two main reasons: i) Not all bacteria can be cultured in laboratory conditions; ii) The production of secondary metabolites by a host can interact with endosymbiotic bacteria and result in the up regulation of specific compounds (Chau et al., 2011).

A unique case of endogenous TTX production has been described from the newt, *Taricha granulosa* (Lehman and Brodie, 2004). Within *T. granulosa*, TTX is distributed throughout the skin, ovaries and muscle, yet none of these tissues have been found to contain TTX-producing bacteria (Lehman and Brodie, 2004). Self-production of TTX has been suggested for this species as subsequent studies found TTX was able to be generated in this species when held in a captive environment and fed a TTX free diet (Cardall et al., 2004; Hanifin and Brodie, 2002). Despite the wide occurrence of TTX among phyla, the biosynthetic pathway has yet to be elucidated.

## 3. Evolution of tetrodotoxin (TTX) and saxitoxin (STX) acquisition

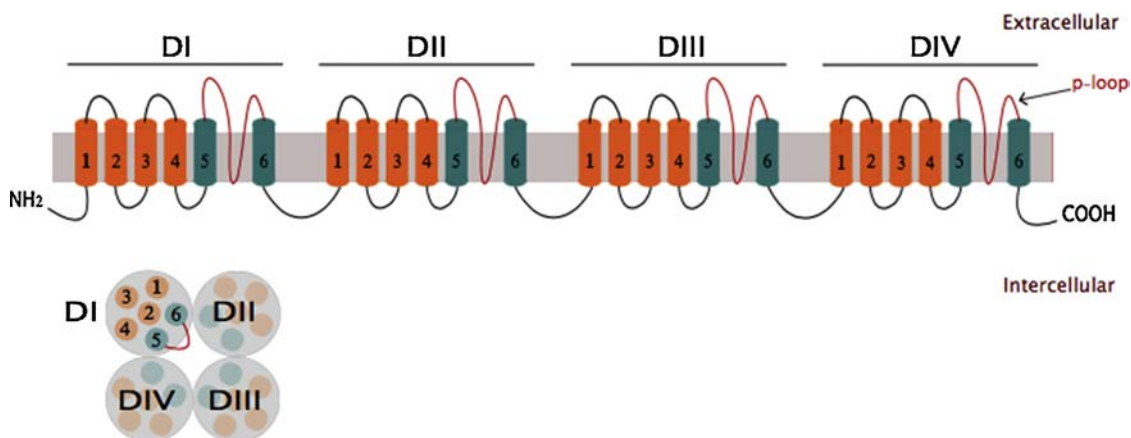
### 3.1. TTX and STX resistant channels

Resistance to exogenous toxins is an initial requirement for the evolution of toxicity. Exposure to paralytic shellfish toxins and TTX from algae blooms and bacterial sources respectively is common in marine ecosystems and is likely to have been a powerful driver in the evolution of toxin resistance (Kem, 2014). TTX intoxication for a susceptible organism is severely compromising resulting in death directly through action potential (AP) inhibition and paralysis or indirectly through-inhibited responses to other organisms (predators or prey) (Brodie, 1990). Since resistance of Na<sub>v</sub> channels to TTX and STX is present through a diverse array of taxa it is likely to have evolved many times independently, the parallel evolution of isoforms is a common phenomenon among closely related species (McGlothlin et al., 2014). This has been well-documented in the TTX resistant garter snakes which prey on tetraodon newts of the genus *Taricha*, where the Na<sub>v</sub>1.4 member of the sodium channel family has acquired independent substitutions conferring TTX resistance (Table 1) (Brodie, 1990; McGlothlin et al., 2014; Williams et al., 2003).

A single P-loop integral to TTX binding affinity is located within each of the four domains in the alpha subunit (DI, DII, DIII and DIV) (Goldin et al., 2000). Resistance can be conferred through substitutions in one or several of the P-loop regions with the most common being an

**Table 1**  
Substitutions conferring tetrodotoxin and saxitoxin resistance in the Na<sub>v</sub> P-loop regions of domains I, II, III, IV.

Organism	Channel	Substitution	Resistance	Ref
<b>DI</b> Salamander ( <i>Cynops pyrrhogaster</i> ) Mammalian channels Pufferfish ( <i>Takifugu</i> and <i>Tetraodon</i> )	Na <sub>v</sub> 1.5Lb and Na <sub>v</sub> 1.6b  Na <sub>v</sub> 1.5, Na <sub>v</sub> 1.8 and Na <sub>v</sub> 1.9  Na <sub>v</sub> 1.4a,	Phe-Tyr or a nonaromatic residue	190-2500 fold resistance to TTX	(Kaneko et al., 1997)  (Satin et al., 1992; Cummins et al., 1999) (Soong and Venkatesh, 2006; Venkatesh et al., 2005; Yotsu-Yamashita et al., 2000) (Jost et al., 2008)
Pufferfish ( <i>Arothron</i> , <i>Canthigaster</i> , <i>Takifugu</i> and <i>Tetraodon</i> )	Na <sub>v</sub> 1.1La, Na <sub>v</sub> 1.4a, Na <sub>v</sub> 1.5La, Na <sub>v</sub> 1.5Lb and Na <sub>v</sub> 1.6b	Phe-Asn/Cys or Ala	NA	(Jost et al., 2008)
<b>DII</b> Pufferfish ( <i>Tetraodon</i> ) Soft shelled clam ( <i>Mya arenaria</i> ) Pufferfish ( <i>Arothron</i> , <i>Canthigaster</i> , <i>Takifugu</i> and <i>Tetraodon</i> )	Na <sub>v</sub> 1.4b  Na <sub>v</sub> 1.4a, Na <sub>v</sub> 1.5Lb and Na <sub>v</sub> 1.6b	Glu-Asp  Thr-Ser or Asn	1500 and 3000 fold resistance to STX and TTX respectively  NA	(Venkatesh et al., 2005) (Bricej et al., 2010, 2005) (Jost et al., 2008)
<b>DIII</b> Garter snakes ( <i>Thamnophis couchii</i> ) Pufferfish ( <i>Arothron</i> , <i>Canthigaster</i> , <i>Takifugu</i> and <i>Tetraodon</i> ) Salamanders ( <i>Taricha granulosa</i> , <i>Taricha torosa</i> , <i>Cynops pyrrhogaster</i> , <i>Pachyrhynchus labiatus</i> , <i>Triturus dobrogicus</i> and <i>Notophthalmus viridescens</i> )	Na <sub>v</sub> 1.4  Na <sub>v</sub> 1.1La, Na <sub>v</sub> 1.1Lb, Na <sub>v</sub> 1.4a, Na <sub>v</sub> 1.4b, and Na <sub>v</sub> 1.5Lb and Na <sub>v</sub> 1.4 saISCN4a (Na <sub>v</sub> 1.4)	Met to Thr	15 fold resistance to STX and TTX	(Feldman et al., 2012)  (Jost et al., 2008)  (Hanifin and Gilly, 2015)
<b>DIV</b> Garter snake ( <i>Thamnophis sirtalis</i> ) Garter snake ( <i>Thamnophis sirtalis</i> W, WC and B) Pufferfish ( <i>Arothron</i> , <i>Canthigaster</i> and <i>Takifugu</i> ) Garter snake ( <i>Thamnophis sirtalis</i> WC) Pufferfish ( <i>Arothron</i> , <i>Canthigaster</i> , <i>Takifugu</i> and <i>Tetraodon</i> ) Garter snake ( <i>Thamnophis atratus</i> and <i>Thamnophis sirtalis</i> B) Pufferfish ( <i>Canthigaster</i> )	Na <sub>v</sub> 1.4 Na <sub>v</sub> 1.4 Na <sub>v</sub> 1.4a and Na <sub>v</sub> 1.6b Na <sub>v</sub> 1.4 Na <sub>v</sub> 1.1Lb, Na <sub>v</sub> 1.4a and Na <sub>v</sub> 1.6a Na <sub>v</sub> 1.4 Na <sub>v</sub> 1.4b	Ile to Leu Ile to Val Ile to Met (same site as above) Gly to Ala  Asp to Asn	NA Halves TTX binding affinity May increase resistance by 5 fold (unconfirmed) 1.5 fold  30 to 40 fold resistance to TTX	(Feldman et al., 2012) (Geffeney et al., 2005) (Jost et al., 2008) (Geffeney et al., 2005) (Jost et al., 2008)  (Choudhary et al., 2003; Penzotti et al., 1998) (Jost et al., 2008; Choudhary et al., 2003)



**Fig. 1.** Schematic of voltage-gated sodium channel (Na<sub>v</sub>) alpha subunits (DI, DII, DIII and DIV). Each unit is composed of six subunits 1–4 (orange) and 5–6 (green). Alternating extra and intercellular loops are shown in black with the p-loops between subunits 5 and 6 highlighted in red (p-loops are site of TTX binding, sequences shown in Fig.2). A) A lateral view of the units imbedded within the cellular membrane (grey). B) Top down view of the alpha subunits (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

aromatic to non-aromatic substitution believed to inhibit binding of the guanidinium moiety (Lee and Ruben, 2008). Alternate substitutions have also been discovered which contribute to TTX resistance, however the direct mechanisms for most have not been elucidated. A list of TTX resistance associated mutations is shown in Table 1 (Figs. 1 and 2).

### 3.1.1. Domain I (DI)

Substitutions providing TTX resistance within the DI region were the first to be discovered and are the most well studied. One

substitution in particular of Phe or Tyr to a non-aromatic residue can inhibit binding of the guanidinium moiety resulting in between 190 to 2500 fold TTX and STX resistance (Soong and Venkatesh, 2006; Kaneko et al., 1997; Satin et al., 1992; Venkatesh et al., 2005). This substitution is responsible for TTX resistance in the mammalian cardiac channel Na<sub>v</sub>1.5 and nervous system channels Na<sub>v</sub>1.8 and Na<sub>v</sub>1.9 as well as one salamander species, *Cynops pyrrhogaster* (Na<sub>v</sub> 1.5Lb and Na<sub>v</sub> 1.6b) (Satin et al., 1992; Cummins et al., 1999). Jost et al. (2008) identified a further three substitutions conferring resistance (Asn, Cys or Ala) in five

	DI	DII	DIII	DIV
Consensus	FXNAFLALFRMLTQDYWENLFQTLT	FFHSFLIVFRILCGEWIETMW	GLGLYSLLQVATFKGWMIMYA	FGSSIIICLFQXTTSAGWDGLLNP
<i>Mya arenaria</i>	.G..L.CS...TQD....YM.V.	.L...M...V.....DS..	LNS..A.F...Y...I..IRD	.FK...T...MC.....V.KG
<i>Tetraodon nigroviridis</i> Nav 1.4a	.G.....C.....QT.	..N.....	.K.....I.....TA..	..I...G...QILL.
<i>Tetraodon nigroviridis</i> Nav 1.4b	.G.....F.....	..A...I..V.....D...	.....E.....	..N.M...MI.....S.
<i>Takifugu rubripes</i> Nav 1.4a	.G.....N..S..LT.	.....	AK.....I.....P	.....I.....T..L.
<i>Thamnophis sirtalis</i> W	.S...S.....	.....	.....	..N.....EV.....
<i>Thamnophis sirtalis</i> WC	.S...S.....	.....	.....	..N.....EV...A.....
<i>Thamnophis sirtalis</i> B	.S...S.....	.....	.....	..N...L...EV.....NV.....
<i>Thamnophis atratus</i>	.S...S.....	.....	.....E...P	..N.....EI.....N.....
<i>Taricha granulosa</i>	.N...S.....	.....	.....T.....	.....S.....SD..IP
<i>Pachytriron labiatus</i>	.N...A.....	.....	.....T.....	.....L...S.....SD..
<i>Notophthalmus viridescens</i>	.N...A.....	.....Y.....	.....T.....	.....L...S.....SD.....
<i>Cynops pyrrhogaster</i>	.N...A.....	.....	.....T.....	.....L...S.....SD.....

Fig. 2. Alignment of the four p-loop domains (DI, DII, DIII and DIV) in the Na<sub>v</sub>1.4 sequences from pufferfish (blue), salamanders (green), garter snakes (orange) and the soft shelled clam (red). Mutations known to confer tetrodotoxin and saxitoxin resistance are shown in bold and coloured appropriately. Mutations suspected to provide resistance are shown underlined in bold (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

independently evolving Na<sub>v</sub> genes from *Takifugu*, *Tetraodon* and *Arothron*. Tissues included cardiac, neural and skeletal (Na<sub>v</sub> 1.1La, Na<sub>v</sub>1.4a, Na<sub>v</sub>1.5La, Na<sub>v</sub>1.5Lb and Na<sub>v</sub>1.6b) (Jost et al., 2008).

### 3.1.2. Domain II (DII)

The only characterized STX resistance substitutions for molluscs are found in Domain II. *Mya arenaria* (soft shelled clam) contains a single substitution of Glu-Asp at position 758, which confers 1500 and 3000 fold resistance to STX and TTX respectively (Bricelj et al., 2010). An identical substitution occurs in the pufferfish, *T. nigroviridis*, Na<sub>v</sub>1.4b channel (Venkatesh et al., 2005). Substitution of the adjacent site Thr-Ser or Asn are observed in three independent events (*Tetraodon* and *Arothron* [1.4a], *Takifugu* [1.5Lb], *Cathigaster* [1.6b]). Additionally, the mammalian channels Na<sub>v</sub>1.8 and Na<sub>v</sub>1.9 also exhibit this substitution. The functional significance of this remains untested. However, substitutions at the same site to Ile, Asp or Lys are known to increase TTX resistance by 2000 fold. An alternate hypothesis relates to substitutions at DI, which is also present in all of the aforementioned sequences. It has been suggested that fitness lost by the initial DI mutation may be compensated for by the subsequent DII substitution (Jost et al., 2008).

### 3.1.3. Domain III (DIII)

Convergent evolution of the DIII domain has resulted in a Met to Thr mutation conferring 15 fold resistance to both TTX and STX occurring in the garter snake, *T. couchii* (Feldman et al., 2012), six genera of the family Salamandridae (Hanifin and Gilly, 2015) and four genera within the family Tetraodontidae (Jost et al., 2008). Higher resistance of 6,000–10,000 fold was observed through a synthesised Glu or Lys replacement at this site. However, impairment of channel conductance was strong and likely results in negative selection for this substitution (Terlau et al., 1991).

### 3.1.4. Domain (DIV)

DIV is the only channel to contain a mutation in the inner pore region of Na<sub>v</sub>, which does not directly interact with sodium ions. In addition, mutational changes at this site have been identified, which are unique to specific populations of *T. sirtalis* W (Warrenton), WC (Willow Creek) and B (Benton) (Geffeney et al., 2005). Three independent occurrences of an Ala to Gly substitution at site 1529 were discovered; Na<sub>v</sub>1.1Lb, Na<sub>v</sub>1.4a and Na<sub>v</sub>1.6a in pufferfish (Jost et al., 2008) along with the WC population of *T. sirtalis* (Geffeney et al., 2005). This mutation has is believed to have an indirect impact of the binding of TTX conferring 1.5 fold resistance (Jost et al., 2008). An Asp to Asn conferring 30–40 fold TTX resistance has also been found in the garter snake, *T. sirtalis* B, *T. aratus* and in the pufferfish genus *Canthigaster* (Magarlamov et al., 2017; Geffeney et al., 2005). All three *T. sirtalis*

populations also possess a Ile to Val substitution that reduces the binding affinity of TTX by half (Geffeney et al., 2005).

### 3.2. Transport of TTX and STX

Mechanisms of toxin transport are required for the safe excretion or sequestration of toxins from an exogenous source (Monteiro and Costa, 2011). Several mechanisms are employed for toxin neutralization, transport and excretion, the most prominent being toxin binding proteins. TTX binding proteins (TBP) have been isolated from non-toxic organisms including shore crabs (Nagashima et al., 2002) and toxic species including gastropods (Hwang et al., 2007), pufferfish (Matsumoto et al., 2007), newts and horseshoe crabs (Ho et al., 1994). Given the widespread presence of TTX it seems likely that these proteins may originally have evolved as a defence against environmental TTX for organisms with TTX sensitive Na<sub>v</sub> channels. Proteins able to bind and neutralize toxins provide protection for intermittent or low exposure. However, high doses of TTX can still be lethal to organisms with TTX resistant Na<sub>v</sub> and TPB may perform a homeostatic role by regulating the quantity and distribution of TTX throughout the animal. It has also been suggested that TBP and STX Binding Proteins (SBP) could inhibit bioaccumulation of toxins (Llewellyn, 1997). Divergent lineages and independent evolutionary histories could result in alternate applications of TBP. TBP as a defence against TTX intoxication is known to occur within the haemolymph of the shore crab, *Hemigrapsus sanguineus*. The neutralizing ability of TBP in *H. sanguineus* provides protection from ingestion of or exposure to TTX (Shiomi et al., 1992). Likewise, SBP isolated from the non-toxic shore crabs, *Thalamita crenata*, provide resistance against STX produced during algal blooms (Lin et al., 2015). A shortfall of reliance on TBP or SBP lies in their binding specificity, with the TPB in *H. sanguineus* providing no resistance to other Na<sub>v</sub> targeting toxins such as STX (Shiomi et al., 1992). Only a few proteins have been identified as TTX binding and implicated as transport or defensive proteins. Of the TBP characterised to date, the Pufferfish STX and TTX Binding Protein (PSTBP) remains the best-documented (Hashiguchi et al., 2015).

Originally isolated from the plasma of *Takifugu niphobles* by Matsui et al. (2000), PSTBP was suggested to be involved in TTX transport and sequestration (Matsui et al., 2000). A protein of similar weight and homologous to the PSTBP characterised by Matsui et al. (2000) was isolated from *Takifugu pardalis* and sequenced as a partial N-terminus peptide and complete cDNA by Yotsu-Yamashita et al. (2002). Proteins homologous to PSTBP have since been isolated from an additional five members of the family Tetraodontidae (*T. poecilonotus*, *T. synderi*, *T. vermicularis*, *T. porphyreus* and *T. rubripes*) (Yotsu-Yamashita et al., 2010). Investigation into the evolutionary history of the protein found

that PSTBP are glycoproteins with high sequence similarity (47%) to the tributyltin-binding protein type 2 (TBT-bp2) found in the plasma of the Japanese flounder, *Paralichthys olivaceus* (Oba et al., 2007). PSTBP is believed to have originated through fusions of pre-existing TBT-bp2 proteins due to the occurrence of two tandemly repeating domains homologous to TBT-bp2 within the protein (Hashiguchi et al., 2015). While both non-toxic and toxic pufferfish species contain copies of TBT-bp2 genes, PSTBP was found to be exclusive to the *Takifugu* genus. Phylogenetic analysis of PSTBP suggests a common origin of the protein followed by subsequent duplication events.

Studies have confirmed the liver as the site of PSTBP synthesis in *Takifugu* through examination of PSTBP gene expression (Yotsu-Yamashita et al., 2013). Originating in the liver, PSTBP is distributed via the circulatory system throughout bodily tissues. The circulatory system plays a key role in the transport of toxins for elimination or sequestration purposes (Lopes et al., 2014). Therefore, it is not surprising that all known TBP proteins have been isolated from either blood or haemolymph in addition to tissues such as the liver in members of the family Tetraodontidae (Matsumoto et al., 2010; Nagashima et al., 2002; Hwang et al., 2007; Matsui et al., 2000; Yotsu-Yamashita et al., 2013, 2002). Cultured non-toxic juvenile *T. niphobles* first accumulate the toxin within the liver and later transport TTX to the skin when they are fed a TTX containing diet (Kono et al., 2008a). Similarly intramuscular administration of TTX in *T. rubripes* was transported to the skin over time (Ikeda et al., 2009). The distribution of PSTBP supports the role of the protein in TTX transport, however PSTBP does not play a role in the sequestration of TTX. Evidence against a sequestering function for PSTBP has been observed in *T. rubripes* where PSTBP was restricted to the ovarian wall and not found within the yolk where TTX is stored (Yotsu-Yamashita et al., 2013). Similar patterns were observed within the skin where PSTBP was abundant in the surrounding dermis of secretory glands and absent from tissues where TTX was accumulated. The opposite was true for *T. vermicularis* with PSTBP only found within the secretory glands (Mahmud et al., 2003).

The complete biological pathways of TBP and SBP have not been elucidated in any species to date and no studies to date have examined the possibility of a defensive origin for toxin transport proteins

### 3.3. Sequestering of TTX and STX

The ability to sequester and concentrate toxins is a key feature among TTX bearing organisms (Noguchi et al., 2006b). The potent neurotoxin can act as an effective defence against predators and is often actively concentrated in specific tissues (Tanu et al., 2002; Tsuruda et al., 2002). This targeted strategy allows for effective use of an evolutionarily costly resource. Bioaccumulation requires active uptake of a compound into the intracellular environment where the concentration can be increased. As opposed to passive biosorption in which the compound is bound to the cellular membrane and conforms to the concentration of the environment (Mebs, 2001). Additional processes are required for the uptake of compounds, however mechanisms underlying TTX bioaccumulation have yet to be fully elucidated in any TTX bearing organisms.

### 3.4. Cost of TTX exploitation

The ability to exploit TTX confers several advantages but the ability to sequester TTX is not without cost. Resistance to the toxin in the form of mutations to Na<sub>v</sub>, the production of TBP and the biological implication of a venom use, all incur varying degrees of cost on the animal. Mutations to the Na<sub>v</sub> channels in TTX bearing taxa affect the ability of these pores to function resulting in a reduction of action potential generation. Na<sub>v</sub> resistance is constrained in order to maintain adequate function of these crucial channels (Lee and Ruben, 2008; Lee et al., 2011). Convergent evolution of the DIII domain has resulted in a Met to Thr mutation conferring 15 fold resistance to both TTX and STX

occurring in the two garter snakes *T. atratus* and *T. couchii* (Geffeney et al., 2005) and four genera within the family Tetraodontidae (Jost et al., 2008). Higher resistance of 6,000–10,000 fold was observed through a synthesised Glu or Lys replacement at this site. However, impairment of channel conductance was strong and resulted in negative selection for this substitution (Terlau et al., 1991). It should be noted that amino acid substitution to Na<sub>v</sub> channels provides resistance not immunity. TTX resistant garter snakes still exhibit symptoms of TTX intoxication post-ingestion of a TTX-bearing newt. Impairment of physical activity of up to 197 min post-ingestion was observed in resistant *T. sirtalis*, leaving them vulnerable to predators and reducing their ability to thermoregulate (Brodie, 1990; Williams et al., 2003).

Alternate methods of TTX resistance, which may or may not function in tandem with resistant Na<sub>v</sub> involve the production of TBP. These proteins have been suggested to play a role in the resistance to and transport of TTX, however their production incurs a metabolic cost. The degree by which TBP production negatively impacts the animal metabolically or through other unknown interactions has yet to be studied.

In the case of the blue-ringed octopus genus the inclusion of TTX within venom may incur or subsidise costs. Venom production is believed to be a metabolically costly activity with many organisms displaying venom-conserving behaviour (Morgenstern and King, 2013). Some contention remains as to the extent this impacts venom evolution in different lineages with some evidence suggesting a lower cost than previously hypothesised (Pintor et al., 2010; Smith et al., 2014). Regardless, the direct cost of toxin production is partially circumvented by acquiring toxins through environmental or bacterial sources. However, the speed of venom restoration is crucial to the fitness of venomous organisms. TTX accumulation rates in *Hapalochlaena* have not yet been investigated. Restoration of TTX levels through proposed self-production in the newt *T. granulosa* required up to nine months for most individuals (Cardall et al., 2004). Rates of TTX accumulation through dietary sources can also vary.

## 4. TTX and STX acquisition in Octopoda

### 4.1. STX accumulation in *O. vulgaris*

STX is a potent neurotoxin, which has the same mode of action as TTX and enters the food chain through dinoflagellates ingested by bivalve molluscs and crustaceans (Kem, 2014; Noguchi et al., 2011b). Bioaccumulation of STX has been well-documented in key fishery species such as bivalve molluscs due to their potential to cause paralytic shellfish poisoning (Kvitek, 1991; Noguchi et al., 2011b). Higher trophic organisms are also at risk of poisoning or accumulation of STX including (Atlantic mackerel) (Castonguay et al., 1997) and octopods. Octopods are opportunistic predators, which prey on diverse organisms including bivalve molluscs, crustaceans, sea birds and small fish (Cornet et al., 2014; Fiorito and Gherardi, 1999; Grisley, 1993; Sazima and de Almeida, 2008). To date STX has been isolated from two octopod species (*Octopus vulgaris* and *Octopus australis*) (Monteiro and Costa, 2011; Robertson et al., 2004; Lopes et al., 2014) and one squid (*Dosidicus gigas*) (Braid et al., 2012).

*Octopus vulgaris* show no symptoms of intoxication post-ingestion of STX contaminated bivalves (Lopes et al., 2014). The ability to consume contaminated prey allows for exploitation of an additional resource. Pharmacokinetics of STX uptake, transfer and elimination in *O. vulgaris* revealed high retention in the digestive gland (DG) followed by a three fold decrease in the kidneys and minimal concentrations within other tissues including the posterior salivary gland (PSG). Low elimination and depuration rates results in accumulation of STX for long periods of time within the DG of *O. vulgaris* (Lopes et al., 2014). The inability to transfer high quantities of STX into tissues apart from the DG suggests STX does not provide a deterrent to predators. In contrast, STX sequestered by *O. australis* is contained within the arms and may confer a defensive advantage (Robertson et al., 2004). No sequencing of sodium

channels has been conducted for either species or discovery of a STX binding protein and the mechanism of resistance is unknown.

Unlike octopods, no evidence for resistance to TTX or STX has been found for squids. Early investigations of TTX function found squid mantle tissue was susceptible, however the species of squid was not noted in the study (Gage et al., 1975, 1976). Additional studies have supported this finding for the veined squid, *Loligo forbesi* and longfin inshore squid, *Loligo pealei* with Na<sub>v</sub> channels of the giant axons binding to TTX and STX respectively (Strichartz et al., 1979; Keynes and Ritchie, 1984). STX has been implicated in several mass strandings of Humboldt squid, *Dosidicus gigas* on Vancouver Island in 2009. Stomach contents of stranded specimens were found to contain paralytic shellfish toxin contaminated Pacific herring and Pacific sardines (Braid et al., 2012). Further dietary/feeding studies are required to confirm the hypothesis of STX intoxication as the cause of *D. gigas* strandings.

#### 4.2. Recruitment of TTX into *Hapalochlaena venom*

Octopods are efficient predators and are known to prey on a variety of organisms including crustaceans and bivalve molluscs (Norman and Reid, 2000). The strategic use of venom injected via vulnerable chinks (eye stalks) has allowed soft bodied octopods to target comparatively well-armored crustaceans (Fiorito and Gherardi, 1999; Hanlon and Messenger, 1998). Venom use in octopods was observed as early as 1888 where the paralytic effect against crabs was noted by Lo Bainco (Ghiretti, 1960). Subsequent studies which injected pure posterior salivary gland extract into crabs confirmed the toxic effect and gland of production within *Octopus vulgaris* (Ghiretti, 1959). The primary lethal component was later identified as a highly glycosylated protein (Ghiretti, 1959; Cariello and Zanetti, 1977). Proteinaceous toxins are the primary lethal component of all studied octopod venom; the exception is the genus *Hapalochlaena*, the only known octopods to requisition the non-proteinaceous compound TTX for a venomous purpose (Freeman and Turner, 1970; Gage et al., 1975, 1976; Gage and Dulhunty, 1973). Non-proteinaceous toxins often serve defensive functions through incorporation into tissue (Fuhrman, 1986) or excretion from skin glands (Tsuruda et al., 2002). In comparison to other octopods, *Hapalochlaena* venom is capable of targeting a diverse range of large invertebrate and vertebrate predators or prey. Most notably, species within the genus *Hapalochlaena* are the only octopods whose bite has resulted in human fatalities (Jacups and Currie, 2008; Flecker and Cotton, 1955). Venomous applications for TTX have been also suggested for four species of arrow worm (*Chaetognatha*), due to the presence of TTX within the head of captured animals (Thuesen and Kogure, 1989; Thuesen et al., 1988).

The duality of TTX use for defensive and offensive purposes in *Hapalochlaena* has not been exploited in related octopod lineages despite its incorporation into other venomous taxa. Sequestration of TTX from ingestion of toxic newts has been observed in garter snake livers, however no TTX has been isolated from the venom gland (Williams et al., 2012b). Similarly hognose snakes that prey on TTX-producing newts do not sequester TTX within their venom glands (Feldman et al., 2016).

#### 4.3. Microbiome of the *Hapalochlaena* posterior salivary gland

The origin of TTX in *Hapalochlaena*, as in many other marine species, remains contentious (Chau et al., 2013). Due to the wide occurrence of TTX among unrelated taxa, two main mechanisms of recruitment are often cited, bioaccumulation via dietary sources and production by bacteria living within the host tissues (Chau et al., 2011). It has been suggested that due to the absence of TTX-producing bacterial strains isolated from *Taricha* newts and the increase of TTX level while in captivity that self-production is responsible. However, more research needs to be conducted as this is unlikely to be a common strategy (Lehman and Brodie, 2004; Cardall et al., 2004; Hanifin and

Brodie, 2002) (Gall et al., 2012), and no evidence has been found for endogenous production in *Hapalochlaena*. Bacteria are known producers of TTX and occur throughout most marine environments entering the food chain via detritivores and planktonic feeders (Kem, 2014; Do et al., 1990; Jal and Khora, 2015). Evidence for bioaccumulation has been suggested for several species of pufferfish including *T. rubripes* and *Takifugu niphobles* (Honda et al., 2005b; Kono et al., 2008b). No evidence for a dietary origin of TTX in *Hapalochlaena* has been investigated.

The bacterial origin has been suggested and is often accepted for *Hapalochlaena* however, few studies have been conducted on the subject and no conclusive evidence exists. An early study by Hwang et al. (1989) isolated and successfully cultured TTX-producing strains from three specimens collected from the Philippines including *Vibrio*, *Alteromonas*, *Bacillus* and *Pseudomonas*. The production of TTX was assessed using five separate methods (mouse bioassay, thin-layer chromatography, gas chromatography-mass spectrometry [GC-MS], electrophoresis and high performance liquid chromatography [HPLC]) (Hwang et al., 1989) negating the issues of relying on each method individually (Magarlamov et al., 2017) and providing some assurance of TTX detection. Methods tended to agree with some variance in the strength of TTX detection between HPLC and GC-MS the later providing a stronger signal. The strains isolated are known TTX producers and have been found in deep-sea sediments (Do et al., 1990) as well as a range of TTX-bearing taxa.

A more recent study on a blue-ringed octopus specimen from eastern Australia (*H. sp.*) was unable to repeat these results (Chau et al., 2013). Similar strains were isolated from the PSG, however several key differences should be noted. Strains were selected based on the presence of polyketide synthase (PKS), non-ribosomal peptide synthase (NRPS) and aminotransferase (AMT) genes introducing a bias and potentially excluding some TTX-producing strains (Chau et al., 2013). Additionally, while the three genes are hypothesised to play a role in TTX synthesis the biosynthetic pathway of TTX is unknown and no evidence exists to support that these genes are true indicators of a TTX-producing organism (Chau et al., 2011).

The dynamics of the bacterial associations in *Hapalochlaena* have yet to be addressed. The aforementioned studies provide a snapshot of bacterial communities within the particular animal examined, during a given time. Limitations due to sampling, one and three animals homogenised together in the case of Chau et al. (2013) and Hwang et al. (1989) respectively prevent some key questions from being addressed. While some bacterial families were found to be shared between the two studies the relationship of bacterial communities as transient or co-evolving cannot be established. Furthermore impacts of age, gender, location and species have yet to be examined. If the bacteria are responsible for TTX production in *Hapalochlaena* it may not be possible to culture these bacteria outside of the host. Additionally, secondary inducers provided by the host may be required to trigger or up regulate production of TTX. Absence of these triggers could result in only basal levels of TTX production outside of the host (Chau et al., 2011).

### 5. Role of TTX in *Hapalochlaena*

#### 5.1. Role of TTX in defence and offense

The incorporation of TTX into tissues unrelated to venom production coupled with the aposematic colouration suggests a defensive application in the genus *Hapalochlaena* (Mäthger et al., 2012). Accumulation of TTX into body tissues of all studied species has been observed, however distribution and quantity appear to differ between species and individuals (Williams and Caldwell, 2009; Yotsu-Yamashita et al., 2007). While three accepted species have been examined, discrepancies between studies have led to incomparable results between species.

Yotsu-Yamashita et al. (2007) sub-sectioned *H. maculosa* juveniles into arms, cephalothorax and abdomen and quantified TTX with a post

column fluorescent-HPLC system. All sections contained TTX, however due to the groupings of tissues the fine scale distribution could not be determined (Yotsu-Yamashita et al., 2007). A similar study by Williams et al. (2009) examined adult *Hapalochlaena* and were not able to account for age specific impacts (Williams and Caldwell, 2009). Ontogenetic effects of TTX accumulation and distribution have been observed in pufferfish when injected with TTX. Juvenile *T. rubripes* transferred and sequestered TTX primarily within skin while adults retained TTX within liver tissues (Kiriake et al., 2016).

Williams and Caldwell (2009) thoroughly examined five and four adult *H. lunulata* and *H. fasciata* respectively for TTX accumulation within 12 tissues through TTX quantification via HPLC (Williams and Caldwell, 2009). *Hapalochlaena fasciata* contained TTX within all 11 tissues (posterior salivary gland [PSG], anterior salivary gland, nephridia, gill, ventral mantle, dorsal mantle, arm, digestive gland, oviduct gland, brachial heart and testes) while *H. lunulata* had a limited distribution of TTX isolated only from the PSG, mantle and ink. However, unlike the wild caught *H. fasciata*, the *H. lunulata* used in this study were obtained through the pet trade and were in captivity for an undetermined period of time (Williams and Caldwell, 2009). Loss of toxicity has been previously observed in captive pufferfish and has been attributed to the absence of a TTX food source or loss of TTX-producing bacteria (Kono et al., 2008a).

Sex specific TTX accumulation occurs within *Hapalochlaena*, with females storing large quantities of TTX within reproductive tissues prior to laying eggs (Sheumack et al., 1984). *Hapalochlaena lunulata* eggs are imbued with TTX initially and evidence has been found for potential self-production of TTX post-laying (Williams et al., 2011a). TTX has also been isolated from *H. maculosa* eggs (Sheumack et al., 1984). The presence of TTX within eggs and juveniles was initially believed to provide protection from predators, however some stomatopod predators were found to be unaffected by a similar amount of TTX within a food source suggesting protection is not due to TTX alone (Williams et al., 2011b).

Ink is a primary defensive strategy for coleoid cephalopods. Reduction of the ink sac has been observed in *Hapalochlaena* (Huffard and Caldwell, 2002). While TTX has been incorporated into the ink sac of *H. lunulata* (Williams and Caldwell, 2009) this is not believed to be an adaptive strategy as inking behavior is the result of incidental muscle contractions of the mantle when stressed (Huffard and Caldwell, 2002).

## 5.2. Impact of TTX on venom evolution

Proteinaceous toxins dominate the study of venom evolution. One prominent hypothesis proposes that proteinaceous venom evolution occurs through a process of duplication and neofunctionalization where by a copy of an innocuous key-processing gene is requisitioned for a new purpose (Casewell et al., 2013). Proteinaceous toxins form the lethal component of several octopod and cuttlefish venoms examined to date (Fry et al., 2009a; Ruder et al., 2013a, b; Undheim et al., 2010). *Hapalochlaena* represents an exception with the inclusion of the potent non-proteinaceous toxin TTX. The inclusion of TTX in *Hapalochlaena* introduces redundancy to ancestrally crucial protein neurotoxins. The impact of TTX on venom evolution is poorly understood. This is due to two main factors: (i) The inclusion of a compound neurotoxin potent enough to impact proteinaceous venom evolution is uncommon and understudied with only one other phyla (*Chaetognatha*) believed to contain TTX within its venom; (ii) Venom evolution in octopods is not well-known with only two species containing expression information at the protein level and three venom proteins functionally annotated.

One of the best-studied examples of an octopus neurotoxin is the tachykinin related peptide, Eledoisin. Tachykinins are present within the neural tissue of *O. vulgaris* and act as functional neurotransmitters (Kanda et al., 2003, 2007). A related peptide serves as a potent short neurotoxin able to affect invertebrate and vertebrate prey in the PSG of

the octopods *O. vulgaris* and *O. kaurina* (Ruder et al., 2013a). Tachykinins serve a key neurotoxic role in several octopod species, however this role may be usurped in *Hapalochlaena* by TTX as expression is notably absent in the PSG of *H. maculosa* despite retention of tachykinin sequences in the genome (Whitelaw et al., 2016). Loss of redundant neurotoxins is not the only impact of TTX. Venom is not solely composed of neurotoxins and contains proteins known as dispersal factors, which aid in the transport of toxic components (Escoubas et al., 2008; Fry et al., 2009b). Hyaluronidase is a well-known dispersal factor isolated from rattlesnakes (Bordon et al., 2012; Kemparaju and Girish, 2006) shows relatively high expression in the PSG of *Hapalochlaena maculosa* (Whitelaw et al., 2016).

## 6. Conclusion

The ability to exploit toxic compounds is a strategy which extends through a diverse array of unrelated phyla and provides a powerful offense or defence. Evolution of this strategy is complex and varies between taxa, however the key features of resistance, sequestration, retention and acquisition are required in each system. TTX and STX are commonly sourced throughout the marine environment and have become integral to the survival of taxa. Octopods such as *O. vulgaris* have been exposed to STX via contaminated bivalves during algal blooms and have acquired a measure of resistance. Similarly, *Hapalochlaena* possesses resistance to TTX and is known to incorporate the toxin within its tissues and venom. The origin of TTX acquisition is unknown, however the evolution of resistance through ingestion of contaminated food sources may be the first step in the evolutionary process. We propose that the evolution of TTX use in *Hapalochlaena* originated with resistance to TTX, which could either be an ancestral state resulting from exposure to contaminated food sources or an independent event followed by the evolution of further processes allowing for sequestration of TTX in these species.

## Conflict of interest

No authors involved in this manuscript have any competing interests to declare.

## References

- Bane, V., Lehane, M., Dikshit, M., O'Riordan, A., Furey, A., 2014. Tetrodotoxin: chemistry, toxicity, source, distribution and detection. *Toxins (Basel)* 6, 693–755.
- Bordon, K.C.F., Perino, M.G., Giglio, J.R., Arantes, E.C., 2012. Isolation, enzymatic characterization and anti-edematogenic activity of the first reported rattlesnake hyaluronidase from *Crotalus durissus terrificus* venom. *Biochimie* 94, 2740–2748.
- Braid, H.E., et al., 2012. Preying on commercial fisheries and accumulating paralytic shellfish toxins: a dietary analysis of invasive *Dosidicus gigas* (Cephalopoda Ommastrephidae) stranded in Pacific Canada. *Mar. Biol.* 159, 25–31.
- Bricelj, V.M., Shumway, S.E., 1998. Paralytic shellfish toxins in bivalve molluscs: occurrence, transfer kinetics, and biotransformation. *Rev. Fish. Sci.* 6, 315–383.
- Bricelj, V.M., et al., 2005. Sodium channel mutation leading to saxitoxin resistance in clams increases risk of PSP. *Nature* 434, 763–767.
- Bricelj, V.M., MacQuarrie, S.P., Doane, J.A.E., Connell, L.B., 2010. Evidence of selection for resistance to paralytic shellfish toxins during the early life history of soft-shell clam, *Mya arenaria*, populations. *Limnol. Oceanogr.* 55, 2463–2475.
- Brodie, E.D., 1990. Tetrodotoxin resistance in garter snakes: an evolutionary response of predators to dangerous prey. *Evolution* 44, 651–659.
- Cardall, B.L., Brodie, E.D., Hanifin, C.T., 2004. Secretion and regeneration of tetrodotoxin in the rough-skin newt (*Taricha granulosa*). *Toxicol.* 44, 933–938.
- Cariello, L., Zanetti, L., 1977. A- and  $\beta$ -cephalotoxin: two paralyzing proteins from posterior salivary glands of *Octopus vulgaris*. *Comp. Biochem. Physiol. C, Comp. Pharmacol.* 57, 169–173.
- Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A., Fry, B.G., 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol. Evol.* 28, 219–229.
- Castonguay, M., et al., 1997. Accumulation of PSP toxins in Atlantic mackerel: seasonal and ontogenetic variations. *J. Fish Biol.* 50, 1203–1213.
- Catterall, W.A., et al., 2007. Voltage-gated ion channels and gating modifier toxins. *Toxicol.* 49, 124–141.
- Cestèle, S., Catterall, W.A., 2000. Molecular mechanisms of neurotoxin action on voltage-gated sodium channels. *Biochimie* 82, 883–892.
- Chau, R., Kalaitzis, J.A., Neilan, B.A., 2011. On the origins and biosynthesis of tetrodotoxin. *Aquat. Toxicol.* 104, 61–72.











- Chau, R., Kalaitzis, J.A., Wood, S.A., Neilan, B.A., 2013. Diversity and biosynthetic potential of culturable microbes associated with toxic marine animals. *Mar. Drugs* 11, 2695–2712.
- Cheng, C.A., et al., 1995. Microflora and tetrodotoxin-producing bacteria in a gastropod, *Niotoha clathrata*. *Food Chem. Toxicol.* 33, 929–934.
- Choudhary, G., Yotsu-Yamashita, M., Shang, L., Yasumoto, T., Dudley, S.C., 2003. Interactions of the C-11 hydroxyl of tetrodotoxin with the sodium channel outer vestibule. *Biophys. J.* 84, 287–294.
- Cornet, V., et al., 2014. Dual role of the cuttlefish salivary proteome in defense and predation. *J. Proteomics* 108, 209–222.
- Crone, H.D., Leake, B., Jarvis, M.W., Freeman, S.E., 1976. On the nature of “Maculotoxin”, a toxin from the blue-ringed octopus (*Haplochlaua maculosa*). *Toxicol.* 14, 423–426.
- Cummins, T.R., et al., 1999. A novel persistent tetrodotoxin-resistant sodium current in SNS-null and wild-type small primary sensory neurons. *J. Neurosci.* 19 RC43-RC43.
- Do, H.K., Kogure, K., Simidu, U., 1990. Identification of deep-sea-sediment bacteria which produce tetrodotoxin. *Appl. Environ. Microbiol.* 56, 1162–1163.
- Do, H.K., et al., 1991. Tetrodotoxin production of actinomycetes isolated from marine sediment. *J. Appl. Bacteriol.* 70, 464–468.
- Do, H.K., et al., 1993. Presence of tetrodotoxin and tetrodotoxin-producing bacteria in freshwater sediments. *Appl. Environ. Microbiol.* 59, 3934–3937.
- Escoubas, P., Quinton, L., Nicholson, G.M., 2008. Venomics: unravelling the complexity of animal venoms with mass spectrometry. *J. Mass Spectrom.* 43, 279–295.
- Feldman, C.R., Brodie, E.D., Pfrender, M.E., 2012. Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. *Proc. Natl. Acad. Sci.* 109, 4556–4561.
- Feldman, C., et al., 2016. Is there more than one way to skin a newt? Convergent toxin resistance in snakes is not due to a common genetic mechanism. *Heredity* 116, 84–91.
- Fiorito, G., Gherardi, F., 1999. Prey-handling behaviour of *Octopus vulgaris* (Mollusca, Cephalopoda) on Bivalve preys. *Behav. Process.* 46, 75–88.
- Flecker, H., Cotton, B., 1955. Fatal bite from octopus. *Med. J. Aust.* 42, 329–331.
- Frank, H.Y., Catterall, W.A., 2003. Overview of the voltage-gated sodium channel family. *Genome Biol.* 4, 207.
- Freeman, S.E., Turner, R., 1970. Maculotoxin, a potent toxin secreted by *Octopus maculosus* Hoyle. *Toxicol. Appl. Pharmacol.* 16, 681–690.
- Fry, B., Roelants, K., Norman, J., 2009a. Tentacles of venom: toxic protein convergence in the Kingdom Animalia. *J. Mol. Evol.* 68, 311–321.
- Fry, B.G., et al., 2009b. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu. Rev. Genom. Hum. G* 10, 483–511.
- Fuhrman, F.A., 1986. Tetrodotoxin, tarichatoxin, and chiriquitoxin: historical perspectives. *Ann. N. Y. Acad. Sci.* 479, 1–14.
- Gage, P.W., Dulhunty, A.F., 1973. Effects of toxin from the Blue-ringed octopus (*Haplochlaua maculosa*). *Mar. Pharmacogn.* 85–106.
- Gage, P.W., Moore, J.W. & Westerfield, M. Octopus toxin blocks sodium conductance in squid axon. *Biophys. J.* 15 A260-A260 (Biophysical Society 9650 Rockville Pike, Bethesda, MD 20814-3998, 1975).
- Gage, P.W., Moore, J.W., Westerfield, M., 1976. An octopus toxin, maculotoxin, selectively blocks sodium current in squid axons. *J. Physiol.* 259, 427.
- Gall, B.G., Stokes, A.N., French, S.S., Brodie, E.D., 2012. Female newts (*Taricha granulosa*) produce tetrodotoxin laden eggs after long term captivity. *Toxicol.* 60, 1057–1062.
- Geffeney, S.L., Fujimoto, E., Brodie, E.D., Brodie, E.D., Ruben, P.C., 2005. Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. *Nature* 434, 759–763.
- Geraci, J.R., et al., 1989. Humpback whales (*Megaptera novaeangliae*) fatally poisoned by dinoflagellate toxin. *Can. J. Fish. Aquat. Sci.* 46, 1895–1898.
- Ghiretti, F., 1959. Cephalotoxin: the crab-paralysing agent of the posterior salivary glands of cephalopods. *Nature* 183, 1192–1193.
- Ghiretti, F., 1960. Toxicity of octopus saliva against crustacea. *Ann. N.Y. Acad. Sci.* 90, 726–741.
- Goldin, A.L., et al., 2000. Nomenclature of voltage-gated sodium channels. *Neuron* 28, 365–368.
- Grisley, M.S., 1993. Separation and partial characterization of salivary enzymes expressed during prey handling in the octopus *Eledone cirrhosa*. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* 105, 183–192.
- Hanifin, C.T., 2010. The chemical and evolutionary ecology of tetrodotoxin (TTX) toxicity in terrestrial vertebrates. *Mar. Drugs* 8, 577–593.
- Hanifin, C.T., Brodie, E.D., 2002. Tetrodotoxin levels of the rough-skin newt, *Taricha granulosa*, increase in long-term captivity. *Toxicol.* 40, 1149–1153.
- Hanifin, C.T., Gilly, W.F., 2015. Evolutionary history of a complex adaptation: tetrodotoxin resistance in salamanders. *Evolution* 69, 232–244.
- Hanlon, R.T., Messenger, J.B., 1998. *Cephalopod Behaviour*. Cambridge University Press.
- Hashiguchi, Y., et al., 2015. Characterization and evolutionary analysis of tributyltin-binding protein and pufferfish saxitoxin and tetrodotoxin-binding protein genes in toxic and nontoxic pufferfishes. *J. Evol. Biol.* 28, 1103–1118.
- Henderson, R., Ritchie, J., Strichartz, G., 1974. Evidence that tetrodotoxin and saxitoxin act at a metal cation binding site in the sodium channels of nerve membrane. *Proc. Natl. Acad. Sci.* 71, 3936–3940.
- Hille, B., 1968. Charges and potentials at the nerve surface. *J. Gen. Physiol.* 51, 221–236.
- Hille, B., 1975. The receptor for tetrodotoxin and saxitoxin. A structural hypothesis. *Biophys. J.* 15, 615–619.
- Ho, B., Yeo, D.S.A., Ding, J.L., 1994. A tetrodotoxin neutralizing system in the haemolymph of the horseshoe crab, *Carcinoscorpius rotundicauda*. *Toxicol.* 32, 755–762.
- Honda, S., et al., 2005a. Toxication of cultured puffer fish *Takifugu rubripes* by feeding on tetrodotoxin-containing diet. *Nippon. Suisan Gakkaishi* 71, 815–820.
- Honda, S., et al., 2005b. Toxication of cultured puffer fish *Takifugu rubripes* by feeding on tetrodotoxin-containing diet. *Bull. Jpn. Soc. Sci. Fish. (Japan)*.
- How, C.-K., Chern, C.-H., Huang, Y.-C., Wang, L.-M., Lee, C.-H., 2003. Tetrodotoxin poisoning. *Am. J. Emerg. Med.* 21, 51–54.
- Huffard, C.L., Caldwell, R.L., 2002. Inking in a blue-ringed octopus, *Haplochlaua lunulata*, with a vestigial ink sac. *Pac. Sci.* 56, 255–257.
- Hwang, D.F., et al., 1989. Tetrodotoxin-producing bacteria from the blue-ringed octopus *Octopus maculosus*. *Mar. Biol.* 100, 327–332.
- Hwang, P.-A., Tsai, Y.-H., Lin, H.-P., Hwang, D.-F., 2007. Tetrodotoxin-binding proteins isolated from five species of toxic gastropods. *Food Chem.* 103, 1153–1158.
- Ikeda, K., et al., 2009. Transfer profile of intramuscularly administered tetrodotoxin to non-toxic cultured specimens of the pufferfish *Takifugu rubripes*. *Toxicol.* 53, 99–103.
- Ivanova, E.P., et al., 2001. Retrieval of the species *Alteromonas tetraodonis* Simidu et al. 1990 as *Pseudoalteromonas tetraodonis* comb. nov. and emendation of description. *Int. J. Syst. Evol. Microbiol.* 51, 1071–1078.
- Jacups, S., Currie, B.J., 2008. Blue-ringed octopuses: a brief review of their toxicology. *North. Territ. Nat.* 20, 50–57.
- Jal, S., Khora, S., 2015. An overview on the origin and production of tetrodotoxin, a potent neurotoxin. *J. Appl. Microbiol.* 119, 907–916.
- Jost, M.C., et al., 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Mol. Biol. Evol.* 25, 1016–1024.
- Kanda, A., Iwakoshi-Ukena, E., Takuwa-Kuroda, K., Minakata, H., 2003. Isolation and characterization of novel tachykinins from the posterior salivary gland of the common octopus *Octopus vulgaris*. *Peptides* 24, 35–43.
- Kanda, A., Takuwa-Kuroda, K., Aoyama, M., Satake, H., 2007. A novel tachykinin-related peptide receptor of *Octopus vulgaris*—evolutionary aspects of invertebrate tachykinin and tachykinin-related peptide. *FEBS J.* 274, 2229–2239.
- Kaneko, Y., Matsumoto, G., Hanyu, Y., 1997. TTX resistivity of Na<sup>+</sup> channel in newt retinal neuron. *Biochem. Biophys. Res. Commun.* 240, 651–656.
- Kem, W.R., 2014. Marine venoms and toxins. In: Wexler, P. (Ed.), *Encyclopedia of Toxicology*, third edition. Academic Press, Oxford, pp. 160–163.
- Kemparaju, K., Girish, K., 2006. Snake venom hyaluronidase: a therapeutic target. *Cell Biochem. Funct.* 24, 7–12.
- Keynes, R., Ritchie, J., 1984. On the binding of labelled saxitoxin to the squid giant axon. *Proc. R. Soc. London B: Biol. Sci.* 222, 147–153.
- Kiriake, A., et al., 2016. Comparison of tetrodotoxin uptake and gene expression in the liver between juvenile and adult tiger pufferfish, *Takifugu rubripes*. *Toxicol.* 111, 6–12.
- Kogure, K., et al., 1988. Accumulation of tetrodotoxin in marine sediment. *Marine ecology progress series*. Oldendorf 45, 303–305.
- Kono, M., Matsui, T., Furukawa, K., Yotsu-Yamashita, M., Yamamori, K., 2008a. Accumulation of tetrodotoxin and 4,9-anhydrotetrodotoxin in cultured juvenile kusafugu *Fugu niphobles* by dietary administration of natural toxic komonfugu *Fugu poecilonotus* liver. *Toxicol.* 51, 1269–1273.
- Kono, M., Matsui, T., Furukawa, K., Yotsu-Yamashita, M., Yamamori, K., 2008b. Accumulation of tetrodotoxin and 4, 9-anhydrotetrodotoxin in cultured juvenile kusafugu *Fugu niphobles* by dietary administration of natural toxic komonfugu *Fugu poecilonotus* liver. *Toxicol.* 51, 1269–1273.
- Kotaki, Y., Shimizu, Y., 1993. 1-Hydroxy-5, 11-dideoxytetrodotoxin, the first N-hydroxy and ring-deoxy derivative of tetrodotoxin found in the newt *Taricha granulosa*. *J. Am. Chem. Soc.* 115, 827–830.
- Kudo, Y., et al., 2014. Isolation of 6-deoxytetrodotoxin from the pufferfish, *Takifugu pardalis*, and a comparison of the effects of the C-6 and C-11 hydroxy groups of tetrodotoxin on its activity. *J. Nat. Prod.* 77, 1000–1004.
- Kvitek, R.G., 1991. Paralytic shellfish toxins sequestered by bivalves as a defense against siphon-nipping fish. *Mar. Biol.* 111, 369–374.
- Lee, C.H., Ruben, P.C., 2008. Interaction between voltage-gated sodium channels and the neurotoxin, tetrodotoxin. *Channels* 2, 407–412.
- Lee, C.H., Jones, D.K., Ahern, C., Sarhan, M.F., Ruben, P.C., 2011. Biophysical costs associated with tetrodotoxin resistance in the sodium channel pore of the garter snake, *Thamnophis sirtalis*. *J. Comp. Physiol. A* 197, 33–43.
- Lehman, E.M., Brodie, E.D., 2004. No evidence for an endosymbiotic bacterial origin of tetrodotoxin in the newt *Taricha granulosa*. *Toxicol.* 44, 243–249.
- Lin, H., et al., 2015. Neutralizing effect of hemolymph from the shore crab, *Thalamita crenata*, on paralytic shellfish toxins. *Toxicol.* 99, 51–57.
- Lipkind, G.M., Fozzard, H.A., 1994. A structural model of the tetrodotoxin and saxitoxin binding site of the Na<sup>+</sup> channel. *Biophys. J.* 66, 1–13.
- Llewellyn, L., 1997. Haemolymph protein in xanthid crabs: its selective binding of saxitoxin and possible role in toxin bioaccumulation. *Mar. Biol.* 128, 599–606.
- Lopes, V.M., Lopes, A.R., Costa, P., Rosa, R., 2013. Cephalopods as vectors of harmful algal bloom toxins in marine food webs. *Mar. Drugs* 11, 3381–3409.
- Lopes, V.M., Baptista, M., Repolho, T., Rosa, R., Costa, P.R., 2014. Uptake, transfer and elimination kinetics of paralytic shellfish toxins in common octopus (*Octopus vulgaris*). *Aquat. Toxicol.* 146, 205–211.
- Magarlamov, T.Y., Melnikova, D.I., Chernyshev, A.V., 2017. Tetrodotoxin-Producing Bacteria: Detection, Distribution and Migration of the Toxin in Aquatic Systems. *Toxins (Basel)* 9.
- Mahmud, Y., et al., 2003. Intra-tissue distribution of tetrodotoxin in two marine puffers *Takifugu vermicularis* and *Chelonodon patoca*. *Toxicol.* 41, 13–18.
- Mäthger, L.M., Bell, G.R., Kuzirian, A.M., Allen, J.J., Hanlon, R.T., 2012. How does the blue-ringed octopus (*Haplochlaua lunulata*) flash its blue rings? *J. Exp. Biol.* 215, 3752–3757.
- Matsui, T., Yamamori, K., Furukawa, K., Kono, M., 2000. Purification and some properties of a tetrodotoxin binding protein from the blood plasma of kusafugu, *Takifugu niphobles*. *Toxicol.* 38, 463–468.
- Matsumoto, T., et al., 2007. Involvement of carrier-mediated transport system in uptake

- of tetrodotoxin into liver tissue slices of puffer fish *Takifugu rubripes*. *Toxicon* 50, 173–179.
- Matsumoto, T., et al., 2010. Plasma protein binding of tetrodotoxin in the marine puffer fish *Takifugu rubripes*. *Toxicon* 55, 415–420.
- McGlothlin, J.W., et al., 2014. Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake *Thamnophis sirtalis*. *Mol. Biol. Evol.*
- Mebs, D., 2001. Toxicity in animals. Trends in evolution? *Toxicon* 39, 87–96.
- Miyazawa, K., Jeon, J.K., Noguchi, T., Ito, K., Hashimoto, K., 1987. Distribution of tetrodotoxin in the tissues of the flatworm *Planocera multitentaculata* (Platyhelminthes). *Toxicon* 25, 975–980.
- Monteiro, A., Costa, P.R., 2011. Distribution and selective elimination of paralytic shellfish toxins in different tissues of *Octopus vulgaris*. *Harmful Algae* 10, 732–737.
- Morgenstern, D., King, G.F., 2013. The venom optimization hypothesis revisited. *Toxicon* 63, 120–128.
- Nagashima, Y., Yamamoto, K., Shimakura, K., Shiomi, K., 2002. A tetrodotoxin-binding protein in the hemolymph of shore crab *Hemigrapsus sanguineus*: purification and properties. *Toxicon* 40, 753–760.
- Narita, H., et al., 1987. *Vibrio alginolyticus*, a TTX-producing bacterium isolated from the starfish *Astropecten polyacanthus*. *Nippon. Suisan Gakkaishi* 53, 617–621.
- Nelsen, D.R., et al., 2014. Poisons, toxins, and venoms: redefining and classifying toxic biological secretions and the organisms that employ them. *Biol. Rev. Camb. Philos. Soc.* 89, 450–465.
- Noguchi, T., Uzu, A., Daigo, K., Shida, Y., Hashimoto, K., 1984. A tetrodotoxin-like substance as a minor toxin in the xanthid crab *Atergatis floridus*. *Toxicon* 22, 425–432.
- Noguchi, T., et al., 1987. *Vibrio alginolyticus*, a tetrodotoxin-producing bacterium, in the intestines of the fish *Fugu vermicularis vermicularis*. *Mar. Biol.* 94, 625–630.
- Noguchi, T., Arakawa, O., Takatani, T., 2006a. Toxicity of pufferfish *Takifugu rubripes* cultured in netcages at sea or aquaria on land. *Comp. Biochem. Physiol. Part D Genomics Proteomics* 1, 153–157.
- Noguchi, T., Arakawa, O., Takatani, T., 2006b. TTX accumulation in pufferfish. Comparative biochemistry and physiology part d. *Genom. Proteom.* 1, 145–152.
- Noguchi, T., Onuki, K., Arakawa, O., 2011a. Tetrodotoxin poisoning due to pufferfish and gastropods, and their intoxication mechanism. *ISRN Toxicol.* 2011, 276939.
- Noguchi, T., Miyazawa, K., Daigo, K., Arakawa, O., 2011b. Paralytic shellfish poisoning (PSP) toxin-and/or tetrodotoxin-contaminated crabs and food poisoning by them. *Toxin Rev.* 30, 91–102.
- Norman, M., Reid, A., 2000. Guide to Squid, Cuttlefish and Octopuses of Australasia. CSIRO publishing.
- Oba, Y., et al., 2007. Purification and characterization of tributyltin-binding protein type 2 from plasma of Japanese flounder, *Paralichthys olivaceus*. *J. Biochem.* 142, 229–238.
- Penzotti, J.L., Fozzard, H.A., Lipkind, G.M., Dudley, S.C., 1998. Differences in saxitoxin and tetrodotoxin binding revealed by mutagenesis of the Na<sup>+</sup> channel outer vestibule. *Biophys. J.* 75, 2647–2657.
- Pintor, A.F.V., Krockenberger, A.K., Seymour, J.E., 2010. Costs of venom production in the common death adder (*Acanthophis antarcticus*). *Toxicon* 56, 1035–1042.
- Reyero, M., et al., 1999. Evidence of saxitoxin derivatives as causative agents in the 1997 mass mortality of monk seals in the Cape Blanc Peninsula. *Nat. Toxins* 7, 311–315.
- Robertson, A., Stirling, D., Robillot, C., Llewellyn, L., Negri, A., 2004. First report of saxitoxin in octopi. *Toxicon* 44, 765–771.
- Ruder, T., et al., 2013a. Functional characterization on invertebrate and vertebrate tissues of tachykinin peptides from octopus venoms. *Peptides* 47, 71–76.
- Ruder, T., et al., 2013b. Molecular phylogeny and evolution of the proteins encoded by coleoid (Cuttlefish, Octopus, and squid) posterior venom glands. *J. Mol. Evol.* 76, 192–204.
- Satin, J., et al., 1992. A mutant of TTX-resistant cardiac sodium channels with TTX-sensitive properties. *Science* 256, 1202–1205.
- Sazima, I., de Almeida, L.B., 2008. The bird kraken: octopus preys on a sea bird at an oceanic island in the tropical West Atlantic. *Mar. Biodivers. Rec.* 1, e47.
- Schrager, P., Profera, C., 1973. Inhibition of the receptor for tetrodotoxin in nerve membranes by reagents modifying carboxyl groups. *Biochim. Biophys. Acta (BBA)-Biomembranes* 318, 141–146.
- Sheumack, D.D., Howden, M.E.H., Spence, I., 1984. Occurrence of a tetrodotoxin-like compound in the eggs of the venomous blue-ringed octopus (*Hapalochlaena maculosa*). *Toxicon* 22, 811–812.
- Shiomi, K., Yamaguchi, S., Kikuchi, T., Yamamori, K., Matsui, T., 1992. Occurrence of tetrodotoxin-binding high molecular weight substances in the body fluid of shore crab (*Hemigrapsus sanguineus*). *Toxicon* 30, 1529–1537.
- Simidu, U., Noguchi, T., Hwang, D.F., Shida, Y., Hashimoto, K., 1987. Marine bacteria which produce tetrodotoxin. *Appl. Environ. Microbiol.* 53, 1714–1715.
- Simisu, U., Kita-Tsukamoto, K., Yasumoto, T., Yotsu-Yamashita, M., 1990. Taxonomy of four marine bacterial strains that produce tetrodotoxin. *Int. J. Syst. Evol. Microbiol.* 40, 331–336.
- Smith, M.T., Ortega, J., Beaupre, S.J., 2014. Metabolic cost of venom replenishment by Prairie Rattlesnakes (*Crotalus viridis viridis*). *Toxicon* 86, 1–7.
- Soong, T.W., Venkatesh, B., 2006. Adaptive evolution of tetrodotoxin resistance in animals. *Trends Genet.* 22, 621–626.
- Strichartz, G., Rogart, R., Ritchie, J., 1979. Binding of radioactively labeled saxitoxin to the squid giant axon. *J. Membr. Biol.* 48, 357–364.
- Tanu, M.B., et al., 2002. Localization of tetrodotoxin in the skin of a brackishwater puffer *Tetraodon steindachneri* on the basis of immunohistological study. *Toxicon* 40, 103–106.
- Terlau, H., et al., 1991. Mapping the site of block by tetrodotoxin and saxitoxin of sodium channel II. *FEBS Lett.* 293, 93–96.
- Thuesen, E.V., Kogure, K., 1989. Bacterial production of tetrodotoxin in four species of Chaetognatha. *Biol. Bull.* 176, 191–194.
- Thuesen, E.V., Kogure, K., Hashimoto, K., Nemoto, T., 1988. Poison arrowworms: a tetrodotoxin venom in the marine phylum Chaetognatha. *J. Exp. Mar. Biol. Ecol.* 116, 249–256.
- Tsuruda, K., et al., 2002. Secretory glands of tetrodotoxin in the skin of the Japanese newt *Cynops pyrrhogaster*. *Toxicon* 40, 131–136.
- Undheim, E.A.B., et al., 2010. Venom on ice: first insights into Antarctic octopus venoms. *Toxicon* 56, 897–913.
- Venkatesh, B., et al., 2005. Genetic basis of tetrodotoxin resistance in Pufferfishes. *Curr. Biol.* 15, 2069–2072.
- Wang, X.-J., Yu, R.-C., Luo, X., Zhou, M.-J., Lin, X.-T., 2008. Toxin-screening and identification of bacteria isolated from highly toxic marine gastropod *Nassarius semiplicatus*. *Toxicon* 52, 55–61.
- Whitelaw, B.L., et al., 2016. Combined transcriptomic and proteomic analysis of the posterior salivary gland from the southern blue-ringed octopus and the southern sand octopus. *J. Proteome Res.* 15, 3284–3297.
- Williams, B.L., Caldwell, R.L., 2009. Intra-organismal distribution of tetrodotoxin in two species of blue-ringed octopuses (*Hapalochlaena fasciata* and *H. lunulata*). *Toxicon* 54, 345–353.
- Williams, B.L., Brodie Jr, E.D., Brodie III, E.D., 2003. Coevolution of deadly toxins and predator resistance: self-assessment of resistance by garter snakes leads to behavioral rejection of toxic newt prey. *Herpetologica* 59, 155–163.
- Williams, B.L., Hanifin, C.T., Brodie, E.D., Caldwell, R.L., 2011a. Ontogeny of tetrodotoxin levels in blue-ringed octopuses: maternal investment and apparent independent production in offspring of *Hapalochlaena lunulata*. *J. Chem. Ecol.* 37, 10–17.
- Williams, B., Lovenburg, V., Huffard, C., Caldwell, R., 2011b. Chemical defense in pelagic octopus paralarvae: tetrodotoxin alone does not protect individual paralarvae of the greater blue-ringed octopus (*Hapalochlaena lunulata*) from common reef predators. *Chemoecology* 21, 131–141.
- Williams, B.L., Stark, M.R., Caldwell, R.L., 2012a. Microdistribution of tetrodotoxin in two species of blue-ringed octopuses (*Hapalochlaena lunulata* and *Hapalochlaena fasciata*) detected by fluorescent immunolabeling. *Toxicon* 60, 1307–1313.
- Williams, B.L., Hanifin, C.T., Brodie, E.D., 2012b. Predators usurp prey defenses? Toxicokinetics of tetrodotoxin in common garter snakes after consumption of rough-skinned newts. *Chemoecology* 22, 179–185.
- Yan, Q., Yu, P.H.-F., Li, H.-Z., 2005. Detection of tetrodotoxin and bacterial production by *Serratia marcescens*. *World J. Microbiol. Biotechnol.* 21, 1255–1258.
- Yasumoto, T., Yotsu-Yamashita, M., 1996. Chemical and etiological studies on tetrodotoxin and its analogs. *J. Toxicol. Toxin Rev.* 15, 81–90.
- Yasumoto, T., et al., 1986. Interspecies distribution and possible origin of tetrodotoxin. *Ann. N. Y. Acad. Sci.* 479, 44–51.
- Yasumoto, T., Yotsu, M., Murata, M., Naoki, H., 1988. New tetrodotoxin analogs from the newt *Cynops ensicauda*. *J. Am. Chem. Soc.* 110, 2344–2345.
- Yotsu-Yamashita, M., et al., 2000. Binding properties of 3H-PbTx-3 and 3H-Saxitoxin to brain membranes and to skeletal muscle membranes of puffer fish *Fugu pardalis* and the primary structure of a voltage-gated Na<sup>+</sup> channel  $\alpha$ -Subunit ( $\alpha$ MNa1) from skeletal muscle of *F. pardalis*. *Biochem. Biophys. Res. Commun.* 267, 403–412.
- Yotsu-Yamashita, M., et al., 2002. Mutual binding inhibition of tetrodotoxin and saxitoxin to their binding protein from the plasma of the puffer fish, *Fugu pardalis*. *Biosci. Biotechnol. Biochem.* 66, 2520–2524.
- Yotsu-Yamashita, M., Mebs, D., Flachsenberger, W., 2007. Distribution of tetrodotoxin in the body of the blue-ringed octopus (*Hapalochlaena maculosa*). *Toxicon* 49, 410–412.
- Yotsu-Yamashita, M., Yamaki, H., Okoshi, N., Araki, N., 2010. Distribution of homologous proteins to puffer fish saxitoxin and tetrodotoxin binding protein in the plasma of puffer fish and among the tissues of *Fugu pardalis* examined by Western blot analysis. *Toxicon* 55, 1119–1124.
- Yotsu-Yamashita, M., et al., 2013. Localization of pufferfish saxitoxin and tetrodotoxin binding protein (PSTBP) in the tissues of the pufferfish, *Takifugu pardalis*, analyzed by immunohistochemical staining. *Toxicon* 72, 23–28.

## RESEARCH

# Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion, and loss

Brooke L. Whitelaw <sup>1,2,\*</sup>, Ira R. Cooke <sup>3,4</sup>, Julian Finn <sup>2</sup>, Rute R. da Fonseca <sup>5</sup>, Elena A. Ritschard <sup>6,7</sup>, M. T. P. Gilbert <sup>8</sup>, Oleg Simakov <sup>6</sup> and Jan M. Strugnell <sup>1,9</sup>

<sup>1</sup>Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, 1 James Cook Dr, Douglas QLD 4811, Australia; <sup>2</sup>Sciences, Museum Victoria, 11 Nicholson St, Carlton, Victoria 3053, Australia; <sup>3</sup>College of Public Health, Medical and Vet Sciences, James Cook University, 1 James Cook Dr, Douglas QLD 4811, Australia; <sup>4</sup>La Trobe Institute of Molecular Science, La Trobe University, Plenty Rd & Kingsbury Dr, Bundoora, Melbourne, Victoria 3086, Australia; <sup>5</sup>Center for Macroecology, Evolution and Climate (CMEC), GLOBE Institute, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark; <sup>6</sup>Department of Neurosciences and Developmental Biology, University of Vienna, Universitätsring 1, 1010 Wien, Vienna, Austria; <sup>7</sup>Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Naples, Italy; <sup>8</sup>Center for Evolutionary Hologenomics, GLOBE Institute, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen, Denmark and <sup>9</sup>Department of Ecology, Environment and Evolution, La Trobe University, Plenty Rd & Kingsbury Dr, Bundoora, Melbourne, Victoria 3086, Australia

\*Correspondence address. Brooke Whitelaw, Tel: +0424642621; Email: [1blwhitelaw8@gmail.com](mailto:1blwhitelaw8@gmail.com)  <http://orcid.org/0000-0002-5555-4612>, Address: 1 James Cook Dr, Douglas QLD 4811

## Abstract

**Background:** Cephalopods represent a rich system for investigating the genetic basis underlying organismal novelties. This diverse group of specialized predators has evolved many adaptations including proteinaceous venom. Of particular interest is the blue-ringed octopus genus (*Hapalochlaena*), which are the only octopods known to store large quantities of the potent neurotoxin, tetrodotoxin, within their tissues and venom gland. **Findings:** To reveal genomic correlates of organismal novelties, we conducted a comparative study of 3 octopod genomes, including the Southern blue-ringed octopus (*Hapalochlaena maculosa*). We present the genome of this species and reveal highly dynamic evolutionary patterns at both non-coding and coding organizational levels. Gene family expansions previously reported in *Octopus bimaculoides* (e.g., zinc finger and cadherins, both associated with neural functions), as well as formation of novel gene families, dominate the genomic landscape in all octopods. Examination of tissue-specific genes in the posterior salivary gland revealed that expression was dominated by serine proteases in non-tetrodotoxin-bearing octopods, while this family was a minor component in *H. maculosa*. Moreover, voltage-gated sodium channels in *H. maculosa* contain a resistance mutation found in pufferfish and garter snakes, which is exclusive to the genus. Analysis of the posterior salivary gland microbiome revealed a

Received: 11 May 2020; Revised: 10 August 2020; Accepted: 6 October 2020

© The Author(s) 2020. Published by Oxford University Press GigaScience. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

diverse array of bacterial species, including genera that can produce tetrodotoxin, suggestive of a possible production source. **Conclusions:** We present the first tetrodotoxin-bearing octopod genome *H. maculosa*, which displays lineage-specific adaptations to tetrodotoxin acquisition. This genome, along with other recently published cephalopod genomes, represents a valuable resource from which future work could advance our understanding of the evolution of genomic novelty in this family.

**Keywords:** cephalopod genome; comparative genomics; gene family expansions; transposable elements; venom evolution

## Background

Reconstructing the evolution of novelties at the genomic level is becoming an increasingly viable approach to elucidate their origin. The recent publication of octopod genomes provides an opportunity to investigate the link between genomic and organismal evolution in this unique lineage for which genomic resources have been lacking [1]. From their emergence 275 million years ago (mya) [2], octopods have diversified into >300 species, inhabiting tropical to polar regions, from the deep sea to shallow intertidal zones [3]. As a highly diverse group, octopods show remarkable variation in body form and function. They are specialized soft-bodied predators that are well adapted to their environment with prehensile limbs lined with chemosensory suckers [4], the ability to manipulate skin texture and colour using specialized chromatophores [5], the largest invertebrate nervous systems (excluding those of other cephalopods) [6], and a relatively large circumesophageal brain allowing for complex problem solving and retention of information [7]. Furthermore, the cephalopods have independently evolved proteinaceous venom, which is produced and stored within a specialized gland known as the posterior salivary gland (PSG). All octopods are believed to possess a form of proteinaceous venom used to subdue prey [8–10]. Serine proteases are a common component of cephalopod venoms and have been observed in the PSG of squids, cuttlefish, and octopods [10–13]. Convergent recruitment of serine proteases has been observed between many vertebrate (Squamata [14–16] and Monotremata [17]) and invertebrate (Hymenoptera [18], Arachnida [19], Gastropoda [20], Remipedia [21], and Cnidaria [22]) venomous lineages.

In addition to these proteinaceous venoms, the blue-ringed octopus (genus *Hapalochlaena*) is the only group that also contains the potent non-proteinaceous neurotoxin tetrodotoxin (TTX) [12, 23]. The mechanism of TTX resistance, which allows for safe sequestration of TTX, has been attributed to several substitutions in the p-loop regions of voltage-gated sodium channels ( $\text{Na}_v$ ) in *Hapalochlaena lunulata* [24]. However, these channels have yet to be examined in *Hapalochlaena maculosa* and *Hapalochlaena fasciata*. TTX resistance has also been studied in a range of other genera including pufferfish [25], newts [26, 27], arachnids [28], snakes [29], and gastropods [30].

The blue-ringed octopus is easily identified by iridescent blue rings, which advertise its toxicity in an aposematic display [31–33]. Sequestration of the TTX within bodily tissues is unique to this genus among cephalopods [32, 34]. While other unrelated TTX-bearing species primarily use TTX for defense, *Hapalochlaena* is the only known taxon to utilize TTX in venom [23, 35]. The effect of TTX inclusion on venom composition and function has been previously investigated in the southern blue-ringed octopus (*H. maculosa*) [9]. Relative to the non-TTX-bearing species *Octopus kaurna*, *H. maculosa* exhibited greater expression of putative dispersal factors such as hyaluronidase, which serve to aid in the dispersal of toxic venom components [9]. Conversely, tachykinins—neurotoxins known from other octopods [36, 37]—were absent from the *H. maculosa* PSG [9]. Further in-

vestigation into the broader impact of TTX on the evolutionary trajectory of the species has yet to be addressed owing to the absence of a genome.

This study presents the genome of the southern blue-ringed octopus (*H. maculosa*, NCBI:txid61716; marinespecies.org: tax-name:342334), the first from the genus *Hapalochlaena*. By using a comparative genomic approach we are able to examine the emergence of octopod novelties, at a molecular level between *H. maculosa* and the 2 non-TTX-bearing octopods: the California 2-spot octopus (*Octopus bimaculoides*) and the long-armed octopus (*Callistoctopus minor*). We also address unique features of venom evolution in octopods while also addressing the species-specific evolution of tetrodotoxin acquisition and resistance in *H. maculosa*.

## Data Description

### Genome assembly and annotation

The southern blue-ringed octopus genome was sequenced using Illumina paired-end and Dovetail sequencing from a single female collected at Beaumaris Sea Scout Boat Shed, Beaumaris, Port Phillip Bay, Victoria, Australia. The assembly was composed of 48,285 scaffolds with an N50 of 0.93 Mb and total size of 4.08 Gb. A total of 29,328 inferred protein-coding genes were predicted using a PASA [38] and an Augustus [39] pipeline and supplemented with zinc finger and cadherin genes obtained from aligning *H. maculosa* transcripts to *O. bimaculoides* gene models (Supplementary Note S1.1–S1.4). Completeness of the genome was estimated using BUSCO [40], which identified 87.7% complete and 7.5% fragmented genes against the metazoan database of 978 groups (Supplementary Note S3.2).

*H. maculosa* has a highly heterozygous genome (0.95%), similar to *Octopus vulgaris* (1.1%) [41] but far higher than *O. bimaculoides* (0.08%) [42]. While the low heterozygosity of *O. bimaculoides* is surprising, other molluscs also have highly heterozygous genomes in accordance with *H. maculosa*, including the gastropods (1–3.66%) [43, 44] and bivalves (0.51–3%) [45–51] (Supplementary Table S5).

### PSMC and mutation rate

The mutation rate for *H. maculosa* was estimated to be  $2.4 \times 10^{-9}$  per site per generation on the basis of analysis of synonymous differences with *O. bimaculoides* (Supplementary Note S1.5). The mutation rate is comparable to the average mammalian mutation rate of  $2.2 \times 10^{-9}$  per site per generation, and *Drosophila*,  $2.8 \times 10^{-9}$  [52, 53]. Owing to the unavailability of a suitable closely related and comprehensive genome until the publication of *O. bimaculoides* in 2015 [42], this is the first genome-wide mutation rate estimated for any cephalopod genome.

The historic effective population size ( $N_e$ ) of *H. maculosa* was estimated using the pairwise sequentially Markovian coalescent (PSMC) model (Supplementary Fig. S2). Population size was found to initially increase during the early Pleistocene, followed

by a steady decline that slows slightly at ~100 kya. It should be noted that PSMC estimates are not reliable at very recent times owing to a scarcity of genomic blocks that share a recent common ancestor in this highly heterozygous genome. A decline in population size started during the mid-Pleistocene ~1 mya, a time of unstable environmental conditions with fluctuations in both temperature and glaciation events [54–56]. Corals in the genus *Acropora* show a similar pattern of expansion and contraction attributed to niche availability after the mass extinction of shallow-water marine organisms 2–3 mya, followed by the unstable mid-Pleistocene climate [57, 58]. A similar pattern of expansion and decline in effective population size has also been observed in the Antarctic icefish among other marine organisms distributed in the Southern Hemisphere [59].

### Phylogenomics

A total of 2,108 (single copy/1-to-1) orthologous clusters were identified between the molluscan genomes and transcriptomes of 11 species and used to construct a time-calibrated maximum likelihood tree (Fig. 1a). The phylogenetic reconstruction estimated the divergence time between *H. maculosa* and its nearest relative, *O. bimaculoides*, to be ~59 mya. *C. minor* diverged from this clade much earlier at ~183 mya. Previous phylogenies using a combination of a small number of mitochondrial and nuclear genes [60–62] and orthologs derived from transcriptomes [63] support this topology. Likewise, estimates by Tanner et al. 2017, using a concatenated alignment of 197 genes with a Bayesian approach, placed divergence of *H. maculosa* from *Abdopus aculeatus* at ~59 mya [2].

Inference of “shared” phenotypic traits can be difficult to resolve with the current literature. For example, false eye spots/ocelli observed in both *O. bimaculoides* and *H. maculosa* are structurally very different. Each ocellus in *H. maculosa* is composed of a continuous single blue ring [33], while *O. bimaculoides* has a blue ring composed of multiple small rings. Morphological variations of ocelli structure and colour, in conjunction with the taxonomically sporadic occurrence of this trait across species within *Octopus* and *Amphioctopus*, limit our interpretation as to the evolutionary history of this trait in octopods [3]. Large gaps remain in the literature between phenotypic traits in cephalopods and their genomic source [1]. This study aims to provide a genomic framework to enable resolution of these features by profiling changes in several genomic characters: (i) gene duplications, (ii) novel gene formation, and (iii) non-coding element evolution.

### Organismal impact of novel genes and gene family expansions

Gene family expansions between octopods (*O. bimaculoides*, *C. minor*, and *H. maculosa*) and 3 other molluscan genomes (*Aplysia californica*, *Lottia gigantea*, and *Crassostrea gigas*) were examined using Pfam annotations. A total of 5,565 Pfam domains were identified among 6 molluscan genomes. *H. maculosa* and *C. minor* exhibit expansions in the cadherin gene family, characteristic of other octopod genomes, including *O. bimaculoides* (Fig. 1B) [42, 64]. *C. minor*, in particular, shows the greatest expansion of this family within octopods. Expansions of protocadherins, a subset of the cadherin family, have also occurred independently in squid [42], with the octopod expansions occurring after divergence ~135 mya [42]. The shared ancestry of octopod cadherins was also documented by Styfals et al. [64] using phylogenetic inference between *O. bimaculoides* and *O. vulgaris*. Cadherins,

specifically protocadherins, play crucial roles in synapse formation, elimination, and axon targeting within mammals and are essential mediators of short-range neuronal connections [65–68]. It should be noted that octopods lack a myelin sheath; as a result short-range connections are integral to maintaining signal fidelity over distance [6]. The independent expansions of protocadherins within chordate and cephalopod lineages are believed to be associated with increased neuronal complexity [42, 64]. Elevated expression of protocadherins within neural tissues has been observed in *O. vulgaris* and *O. bimaculoides* by Styfals et al. [64] and Albertin et al. [42], respectively. In particular Styfals et al. [64] noted differential expression across neural tissues including supra-esophageal mass, sub-esophageal mass, optic lobe, and the stellate ganglion [64]. However, functional implications of observed expression patterns remain speculative without further study.

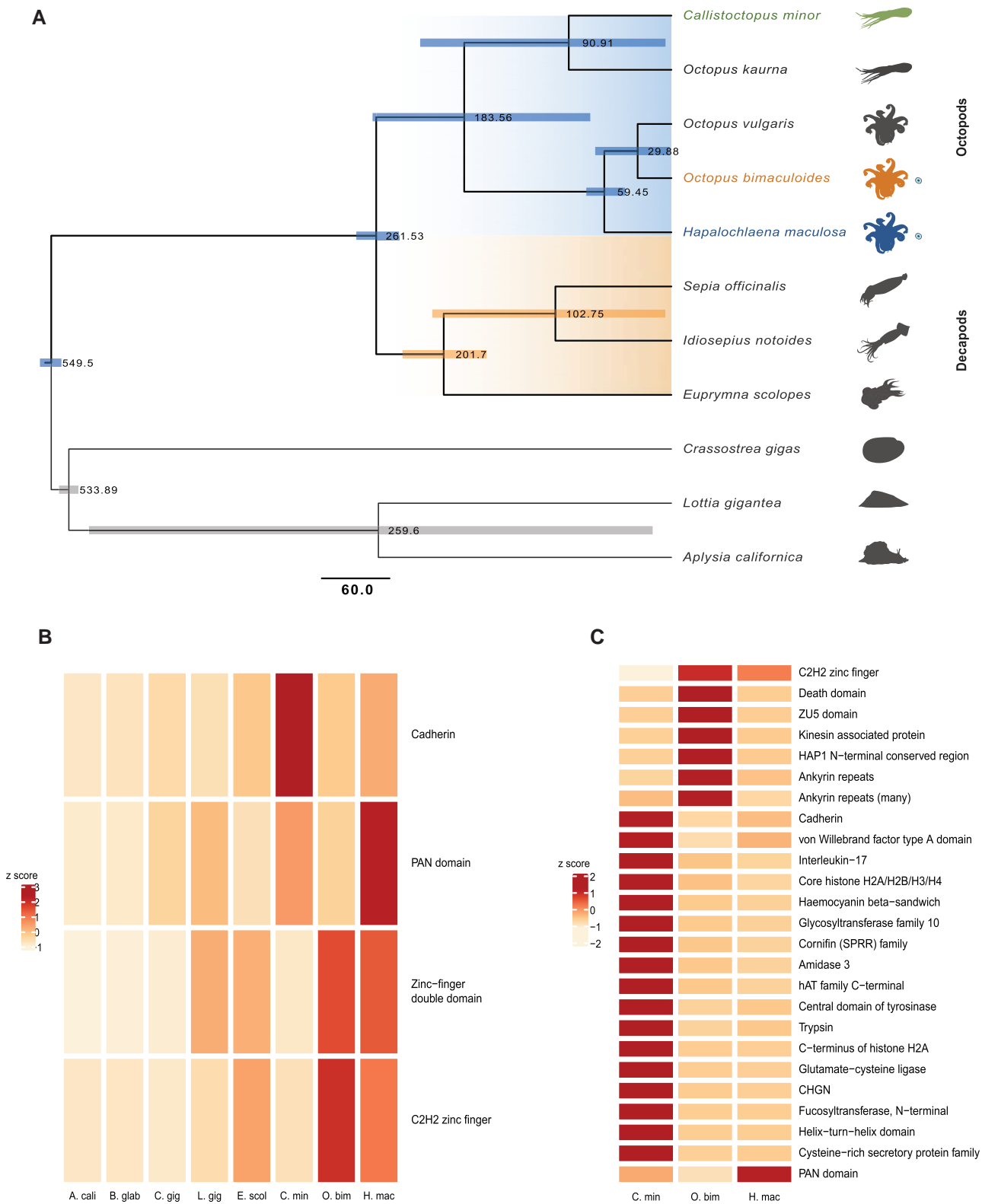
*H. maculosa* also shows expansions in the C2H2-type zinc finger family. Zinc fingers form an ancient family of transcription factors, which among other roles serve to regulate transposon splicing, as well as embryonic and neural development [69, 70]. Expansion of this type of zinc finger in *O. bimaculoides* has been associated with neural tissues. It should be noted that due to the inherent difficulty in fully annotating the zinc finger family, alternative methods were used to examine the number of exons in *C. minor* with high similarity to annotated zinc finger genes in *O. bimaculoides* (Supplementary Note S5.1). A total of 609 exons (not captured by published gene models) from *C. minor* were found with high similarity to accepted zinc finger genes in *O. bimaculoides*, suggesting that this family is larger than that which the genome annotation infers.

Examination of genes specifically expressed within neural tissues found that cadherins were among the most highly expressed gene families of all octopod species. Particularly in *C. minor*, relative to the other octopods, such a trend reflects the gene family expansions found in this species (Fig. 2). Zinc fingers were less pronounced, representing 1.1% of overall expression in *C. minor* compared to cadherins at 11.3%. Overall, neural tissues express a large diversity of Pfams with each species, exhibiting a similar profile and proportion of orthologous to lineage-specific genes.

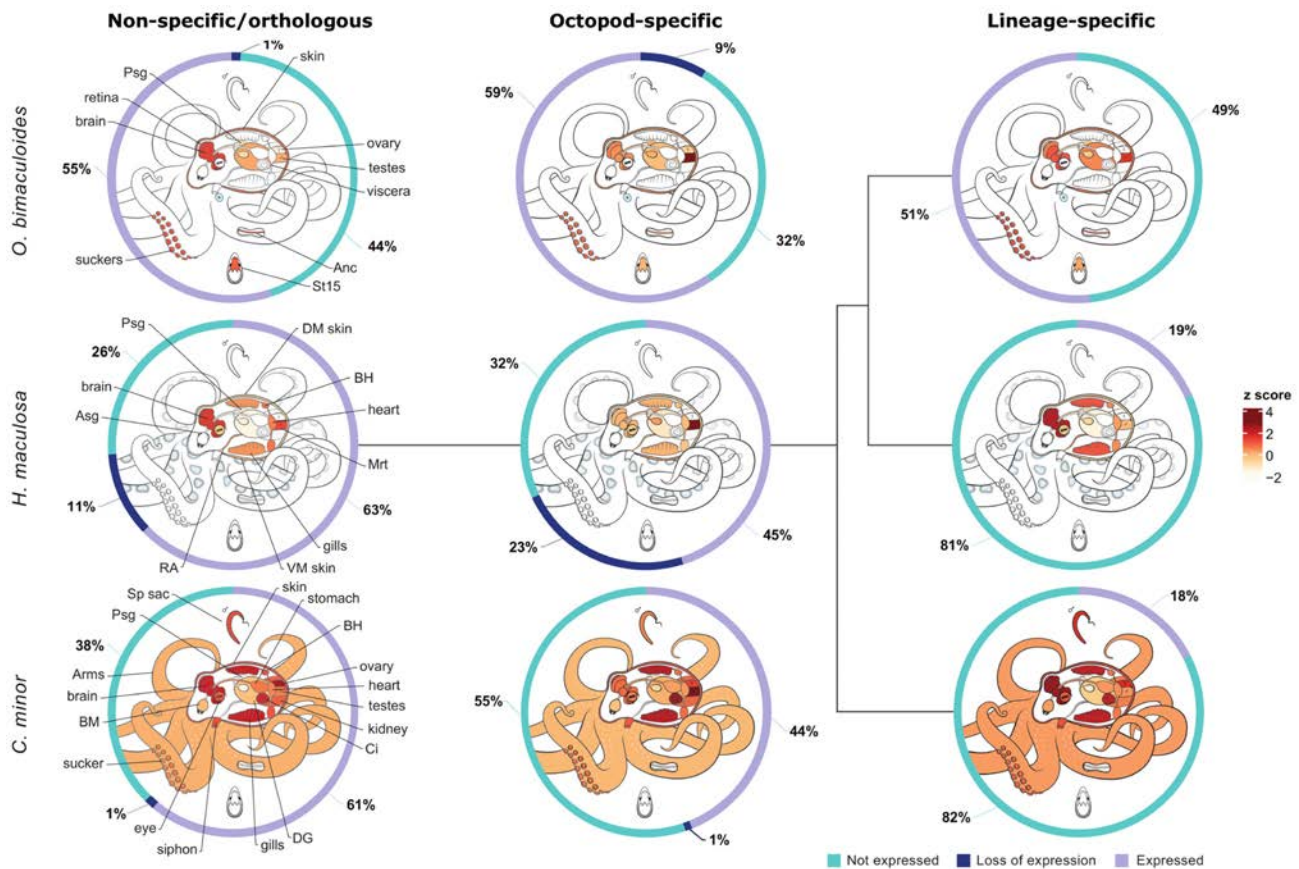
### Novel patterns of gene expression

High-level examination of gene dynamics (expression, loss of expression, and absence of expression) between octopods across different levels of orthology provides insight into large-scale expression patterns and highlights lineage-specific loss of expression.

The greatest proportion of genes in each species examined were not specific to octopods or an octopus lineage (ancient genes) (Fig. 2). Expression of these genes was enriched in neural tissues across all species, indicating the core conservation of neural development and function. However, we also find that genes specific to each octopod species also show this expression pattern. The overall elevated expression of genes within neural tissues could be reflective of the extensive neural network present in cephalopods, which comprises ~520 million nerve cells [71], rivalling vertebrates/mammals in size [6]. Expression of many novel genes in the nervous system may also indicate contribution of those genes to lineage-specific neural network evolution. In contrast, genes that date back to the shared octopod ancestor show highest expression in male reproductive tissues in all species.



**Figure 1:** Comparisons of molluscan genomes and gene families. (A) Time-calibrated maximum likelihood phylogeny of 7 molluscan genomes (*Aplysia californica*, *Lottia gigantea*, *Crassostrea gigas*, *Euprymna scolopes*, *Octopus bimaculoides*, *Callistoctopus minor*, and *Hapalochlaena maculosa*) and 4 transcriptomes (*Octopus karna*, *Octopus vulgaris*, *Sepia officinalis*, and *Idiosepius notoides*) using 2,108 single-copy orthologous sequence clusters. Node labels show divergence times in millions of years (mya); blue (divergence to octopods) and orange bars (decapods) represent standard error within a 95% confidence interval. Octopodiformes lineages are highlighted in blue and decapod orange. Scale bar represents mya. (B) Expansions of octopod gene families relative to molluscan genomes *Aplysia californica* (A. cali), *Biomphalaria glabrata* (B. glab), *C. gigas* (C. gig), *L. gigantea* (L. gig), *E. scolopes* (E. scol), *C. minor* (C. min), *O. bimaculoides* (O. bim), and *H. maculosa* (H. mac). (C) Lineage-specific gene expansions in the octopod genomes *C. minor* (C. min), *O. bimaculoides* (O. bim), and *H. maculosa* (H. mac). CHGN: chondroitin N-acetylgalactosaminyltransferase; C2H2: Cys2-His2; SPRR: small proline-rich proteins.



**Figure 2:** Dynamics of gene expression in octopod genomes. Proportion of gene expression across levels of specificity from not specific to octopods or an octopus species (left) to octopod-specific (middle) and lineage-specific (right). Donut plots show gene expression as some expression in any tissue (purple), no expression (blue), or expression that has been lost (dark blue). Loss of expression requires an ortholog of the gene to be expressed in  $\geq 1$  species and not expressed in the other species. Heatmaps at each specificity level show average expression of genes within their respective tissues, low expression (cream) to high expression (dark red).

Loss of expression between octopod genomes is exhibited most clearly in *H. maculosa*, with 11% (1,993 genes) of all ancient genes having no expression, compared to 1% in both *O. bimaculoides* and *C. minor*. Absence of gene expression for genes whose orthologs have retained expression in  $\geq 1$  other species suggests a unique evolutionary trajectory from other octopods. It should be noted that differences in tissue sampling may in part influence these values and owing to the limited sampling of species, loss of expression cannot be inferred at a species level and may have occurred at any point in the lineage. To fully understand the implications of the gene family contractions and loss of expression in *H. maculosa*, relative to other octopods, further investigation is required.

### Evolution of the octopod non-coding genome

Similar to other cephalopod genomes, the *H. maculosa* genome has a high repeat content of 37.09% (bases masked). *O. bimaculoides* and *C. minor* are also highly repetitive, with 46% and 44% of their genomes composed of transposable elements (TE), respectively. Of the repetitive elements, LINES dominate the decapodiform *Euprymna scolopes* genome, accounting for its larger genome size [72], while SINES are expanded in all 4 octopod genomes. SINES have been previously documented in *O. bimaculoides* (7.86%) [42], comparable with *H. maculosa* (7.53%), while fewer SINES were previously reported for *C. minor* (4.7%) [73]. SINE elements also dominate the *O. vulgaris* genome, with

an expansion occurring after divergence from *O. bimaculoides* [41]. Rolling circle elements are a prominent minor component in octopods, particularly in *H. maculosa*. Rolling circle transposons have been isolated from plant (*Zea mays*) and mammalian genomes. They depend greatly on proteins used in host DNA replication and are the only known class of eukaryotic mobile element (transposon) to have this dependence [74]. TE elements in cephalopod lineages show differing expansions between most of the genomes currently available, suggesting that they are highly active and play a strong role in cephalopod evolution.

Enrichment of transposable elements associated with genes (flanking regions 10 kb up- and downstream) was not observed compared to the whole genome for any species examined. More notable were differences between species; in particular *C. minor* shows a greater proportion of LINE to SINE elements relative to both *O. bimaculoides* and *H. maculosa*.

Together, this highlights a very dynamic evolutionary composition of repeats in cephalopods that requires further study to test for any potential association with changes in gene expression or genome evolution.

### Dynamics of gene expression in the PSG

The PSG is the primary venom-producing gland in octopods. Venom composition in the majority of octopods is primarily composed of proteinaceous toxins. *Hapalochlaena* is an excep-

tion, containing an additional non-proteinaceous neurotoxin, TTX, within their venom. We hypothesize that the *Hapalochlaena* PSG will exhibit a loss of redundant proteinaceous toxins due to the presence of TTX.

Examination of all PSG-specific genes from the 3 octopods revealed a disproportionate number of genes exclusive to *H. maculosa* (Fig. 3A). A total of 623 genes were exclusive to *H. maculosa* PSG compared with only 230 and 164 exclusive to *O. bimaculoides* and *C. minor* PSGs, respectively. Additionally, we predict that the *H. maculosa* PSG is functionally more diverse on the basis of the number of Pfam families detected, 532 in total. Comparatively, the PSG genes in *O. bimaculoides* and *C. minor* are fewer and more specialized. Gene family expansions of serine proteases dominate expression, comprising >30% of total PSG-specific expression in *C. minor* and 17–20% in *O. bimaculoides* (Fig. 3B). Serine proteases were also among genes whose expression seems to have shifted between octopod species. Several serine proteases show specific expression to the PSG of *O. bimaculoides* and *C. minor* while being expressed in a non-specific pattern among brain, skin, muscle, and anterior salivary gland tissues in *H. maculosa* (Fig. 4B). Most notable is the absence of many paralogs in both *H. maculosa* and *O. bimaculoides*, suggesting a lineage-specific expansion of this cluster in *C. minor*. Fewer serine protease genes can also be observed in *H. maculosa* (Fig. 4). Similarly, repolysin (M12B) exhibits shifting expression in *H. maculosa*, presumably from the PSG to the branchial heart, and a complete loss of paralogs from the genome. While the function of this protein has not been assessed in octopus, members of this protein family exhibit anticoagulant properties in snake venom [75–78].

Serine proteases have been previously documented in cephalopod venom and are prime candidates for conserved toxins in octopods. Cephalopod-specific expansions have been identified with strong association to the PSG in 11 cephalopods (7 octopus, 2 squid, and 2 cuttlefish) [8, 13]. All serine proteases identified from the PSG of these species were found to belong to the cephalopod-specific clade. Functionally, cephalopod venom serine proteases have yet to be assessed. However, octopod venom has been observed to have strong digestive and hemolytic properties, which may be in part due to this crucial protein family [79–81]. The reduced number and expression of serine proteases in *H. maculosa* suggests a change in function of the PSG for this species. These results support the hypothesis of toxin redundancy in the *H. maculosa* PSG due to the incorporation of tetrodotoxin. Previous proteomic analysis of the *H. maculosa* PSG revealed high expression of hyaluronidase, which often serves as a dispersal factor within snake venom, facilitating the spread of toxin while not being directly toxic to their prey [9, 82]. While further investigation is required, the incorporation of TTX within *H. maculosa* venom may have contributed to a shift in function, with proteins present acting to support the spread of venom and digestion of tissues.

### TTX resistance of the Na<sub>v</sub> channels

To identify the mechanism of TTX resistance in *H. maculosa*, the voltage-gated sodium channel (Na<sub>v</sub>) sequences were compared between susceptible (human) and resistant (pufferfish, salamanders, and garter snakes) species. TTX binds to the p-loop regions of sodium channels, inhibiting the flow of sodium ions in neurons, resulting in paralysis [83, 84]. Inhibition of TTX binding has been observed in species that either ingest TTX via prey, such as garter snakes [85], and in those that retain TTX within their tissues like pufferfish [86].

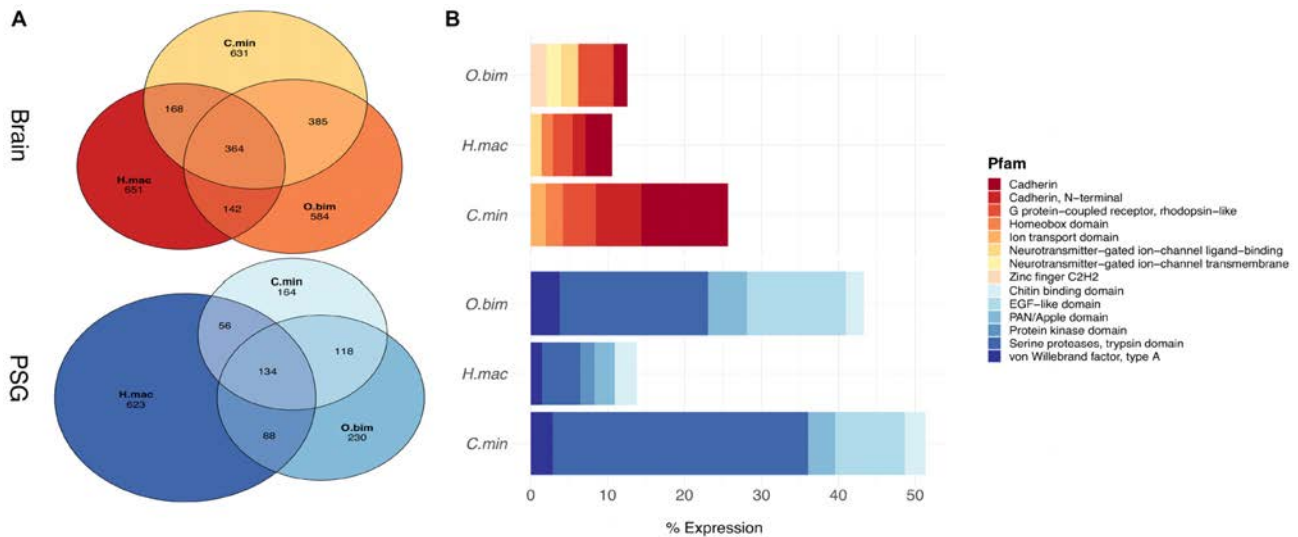
Two Na<sub>v</sub> genes were identified in the *H. maculosa* genome (Na<sub>v</sub>1 and Na<sub>v</sub>2); this is congruent with the recent identification of 2 Na<sub>v</sub> isoforms in *H. lunulata* [24] (Supplementary Figs S8 and S9). Among cephalopods with sequenced Na<sub>v</sub>1 channels, p-loop regions are highly conserved, with both DI and DII shared between all species. The regions DIII and DIV closer to the C-terminal end of the protein in *Hapalochlaena* sp. contain mutations, which may affect TTX binding and differ between families and species as follows. Similar to the pufferfish (*Arothron*, *Canthigaster*, *Takifugu*, and *Tetraodon*) [87] and garter snake *Thamnophis couchii* [88], *H. maculosa* Nav1 has a mutation within the third p-loop at site (DIII) from M1406T, while all other cephalopods have an Ile(I) at this position (Fig. 5A). The dumbo octopus (*Grimpoteuthis*) is the only exception, retaining the susceptible M at this site similar to humans and other non-resistant mammals [83]. Additionally, the fourth p-loop (DIV) in *H. maculosa* exhibits 2 substitutions at known TTX binding sites: D1669H and H1670S. In a previous study a Met to Thr substitution into a TTX-sensitive Nav1.4 channel decreased binding affinity to TTX by 15-fold [87]. Likewise, a 10-fold increase in sensitivity was observed from a T1674M substitution in a mite (*Varroa destructor*) channel VdNav1 [28]. However, resistance is often a result of multiple substitutions and when I1674T/D1967S occur together in VdNav1, resistance is multiplicative, resulting in “super-resistant” channels with binding inhibition of 1,000-fold. The combination of M1406T/D1669H in *H. maculosa* also occurs in the turbellarian flatworm *Bdelloura candida* (BcNav1) [87, 89]. While it has yet to be assessed for TTX resistance, the replacement of aspartic acid in *B. candida* with a neutral amino acid has been predicted to disrupt TTX binding by preventing formation of a salt bridge or hydrogen bond [89, 90]. These 3 substitutions (M1406T, D1669H, and H1670S) in *H. maculosa*, with the potential to inhibit TTX binding, have also been identified by Geffeney et al. [24] in *H. lunulata*. It has yet to be established whether these mutations are derived from a shared ancestor or have occurred independently.

While *Hapalochlaena* remains the best-documented example of TTX resistance among cephalopods, other species may contain some level of TTX resistance (e.g., *O. vulgaris*) [91, 92]. Saxitoxin (STX) is a similar toxin in structure and function, and mutations resistant to TTX are often also STX inhibiting [93]. *O. vulgaris* has been observed consuming STX-contaminated bivalves with no negative effects and as such is believed to be resistant [92]. However, no mutations known to reduce TTX/STX binding affinity occur in its Nav1 [92, 94]. The selective pressure facilitating the evolution of STX/TTX resistance in these shallow-water benthic octopods may be toxic prey, similar to garter snakes. STX is also known as a paralytic shellfish poison. Produced by photosynthetic dinoflagellates and bioaccumulated in bivalves [95], this toxin contaminates a common octopus food source. Pelagic squids such as the Humboldt (*Doryteuthis gigas*) and longfin inshore squid (*Doryteuthis pealeii*) do not appear to be TTX/STX resistant; mass strandings of Humboldt squid have been associated with ingestion of STX-contaminated fish [96]. Likewise, no evidence of resistance was found in the sodium channel of the dumbo octopus (*Grimpoteuthis*). This species typically inhabits depths of 2,000–5,000 m and is unlikely to encounter STX-contaminated food sources [97].

### Microbiome of the PSG

TTX is produced through a wide variety of bacteria, which are common in marine sediments and have been isolated from organisms such as pufferfish [25, 98, 99]. Sequestration of TTX





**Figure 3:** Dynamics of gene expression in neural and venom-producing tissues of octopods. Tissue-specific expression of genes within the brain (red) and posterior salivary gland (PSG) (blue) of *H. maculosa* (*H.mac*), *O. bimaculoides* (*O.bim*), and *C. minor* (*C.min*). A) Venn diagram shows numbers of shared and exclusive genes between species (left). B) Bar chart of the top 5 Pfams and their contribution to overall expression in the brain (right).

is not exclusive to the blue-ringed octopus among molluscs. Gastropods such as *Pleurobranchaea maculata* [100] and *Niotha clathrata* [30], as well as some bivalves, are also capable of sequestering TTX [95]. The commonly held hypothesis for TTX acquisition within *Hapalochlaena* is that it is bacterial in origin and is either ingested or endosymbiotic [100, 101]. Analysis of a ribo-depleted RNA sample from the PSG of *H. maculosa* revealed a highly diverse composition of bacterial genera with Simpson and Shannon diversity indices of 4.77 and 0.94, respectively. The dominant phyla were Proteobacteria and Firmicutes, composing, respectively, 41% and 22% of overall bacterial species detected (Fig. 6). To date, 151 strains of TTX-producing bacteria have been identified from 31 genera. Of these, 104 are members of Proteobacteria [102]. The genera *Pseudomonas* and *Bacillus* belonging to the phyla Proteobacteria and Firmicutes, respectively, have been previously identified in the PSG of *Hapalochlaena* sp. (*Octopus maculosus*) [101]. Examination of these bacterial strains revealed TTX production, and extracts injected into mice proved to be lethal [101]. A more recent study on the bacterial composition of *H. maculosa* PSG did not identify TTX-producing strains [100]. However, only a small subset of the many strains identified were tested. Congruent with our findings the diversity of bacterial genera was high and this may complicate identification of species responsible for TTX production. The biosynthetic pathway of TTX has yet to be elucidated, and as a result, only culturable bacterial species can be tested for TTX production.

## Conclusions

This work describes the genome of a unique TTX-bearing mollusc, the southern blue-ringed octopus (*H. maculosa*). Much of cephalopod evolution is barely understood owing to sparseness of genomic data. Our analysis provides the first glimpse into genomic changes underlying genome evolution of closely related octopod species. While the size, heterozygosity, and repetitiveness of the blue ring genome is congruent with previously published octopod genomes, we find similar yet independent expansions of key neuronal gene families across all 3 species and show

evidence for the involvement of gene novelty in the evolution of key neuronal, reproductive, and sensory tissues. The evolution of venom in octopods also differs between species, with *H. maculosa* showing a reduction in the number and expression of serine proteases in their venom gland relative to the other octopods in this study. Inclusion of TTX in *H. maculosa* distinguishes this species from related octopods and is believed to affect toxin recruitment and retention because the highly potent TTX is sufficient to subdue common octopod prey without additional toxins.

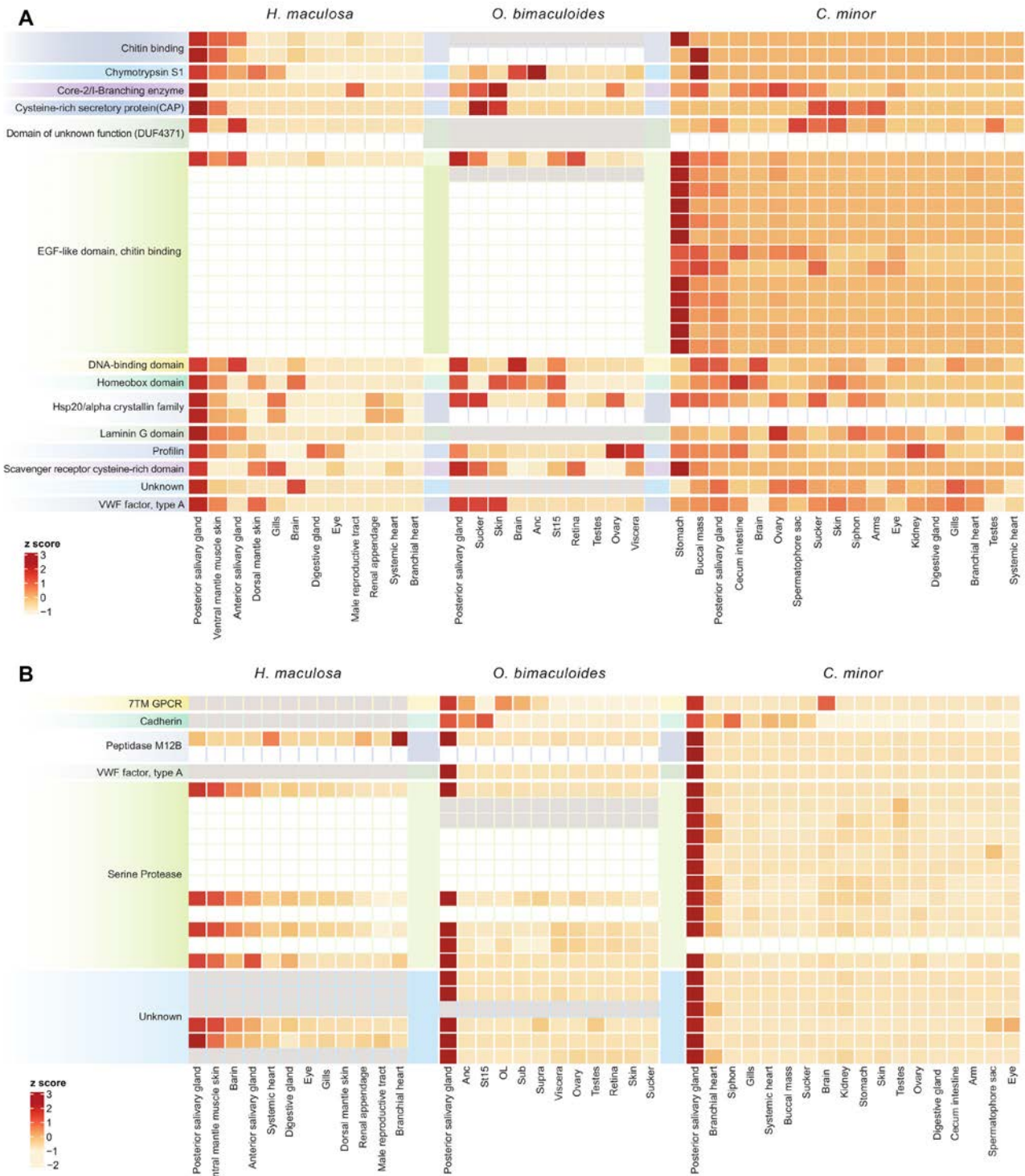
## Methods

### Genome sequencing and assembly

DNA was extracted from a single *H. maculosa* female collected at Port Phillip Bay, Victoria, Australia. Two types of Illumina libraries were constructed, standard paired end and Illumina mate pairs (Supplementary Data S2). Dovetail sequencing, Chicago libraries improved upon original sequencing, resulting in an overall coverage of 71 $\times$ . Assembly-stats [103] was used to ascertain the quality of the assembly and relevant metrics (Supplementary Note S1).

### Transcriptome sequencing

The *H. maculosa* transcriptome was generated using 12 tissues (brain, anterior salivary gland, digestive gland, renal, brachial heart, male reproductive tract, systemic heart, eyeballs, gills, posterior salivary gland, dorsal mantle, and ventral mantle tissue). RNA was extracted using the Qiagen RNeasy kit. Construction of complementary DNA libraries was outsourced to AGRF (Australian Genome Research Facility), Melbourne, and conducted using their TruSeq mRNA Library Prep with polyA selection and unique dual indexing method. Libraries were constructed using 3  $\mu$ g of RNA at a concentration of >100 ng/ $\mu$ L. Each tissue was sequenced on 1/12th of an Illumina HiSeq2000 lane with 1 lane used in total.

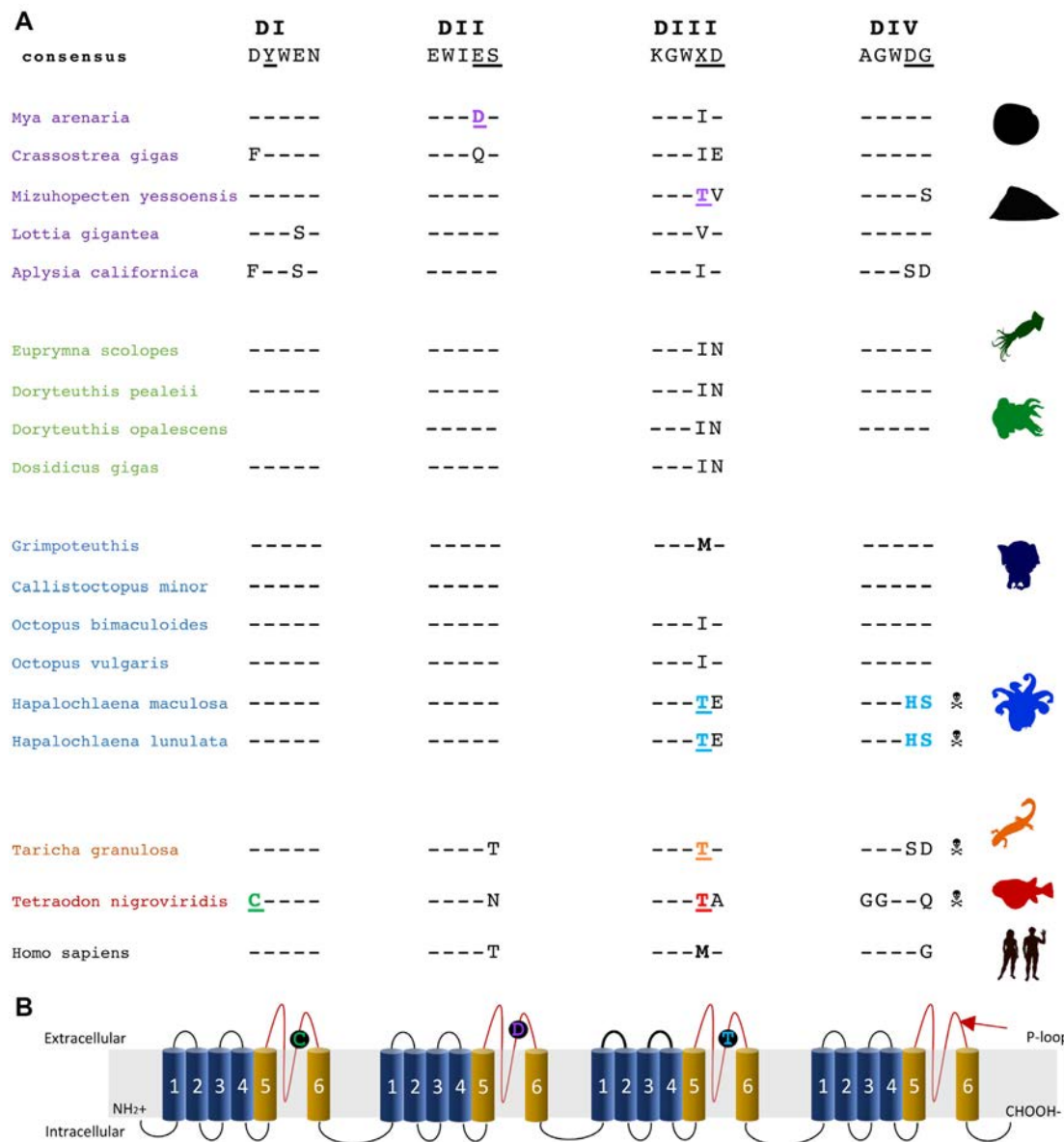


**Figure 4:** Examination of posterior salivary gland (PSG) gene expression between 3 octopod genomes. **(A)** Heat map of genes expressed specifically in the PSG of *H. maculosa* ( $\tau > 0.8$ ) and their orthologs in *O. bimaculoides* and *C. minor* lacking specific expression to the PSG ( $\tau < 0.8$ ). Genes with an ortholog lacking expression are coloured in grey while the absence of an ortholog is white. **(B)** Heat map of genes expressed specifically in the PSG of both *O. bimaculoides* and *C. minor* ( $\tau > 0.8$ ) and their orthologs in *H. maculosa* lacking specific expression to the PSG.

### De novo transcriptome assembly

*De novo* assembly of the *H. maculosa* transcriptome was conducted using sequencing data from 11 tissues (as listed above) and Trinity v10.11.201 (Trinity, [RRID:SCR.013048](https://doi.org/10.1093/bioinformatics/btq119)) [104]. Default

parameters were used aside from k-mer coverage, which was set to 3 to account for the large data volume. Protein-coding sequences were identified using Trinotate (Trinotate, [RRID:SCR.018930](https://doi.org/10.1093/bioinformatics/btq119)) [105] and domains assigned by Interpro v72.0 (InterPro, [RRID:SCR.006695](https://doi.org/10.1093/bioinformatics/btq119)) [106].



**Figure 5:** Mechanism of tetrodotoxin resistance within the posterior salivary gland of *H. maculosa* (PSG). (A) Alignment of voltage-gated sodium channel  $\alpha$ -subunits (DI, DII, DIII, and DIV) p-loop regions. Mutations conferring resistance are coloured in green (pufferfish), orange (salamander), purple (clam), and blue (octopus). Susceptible mutations at the same site are black and boldface. Sites that may be involved with resistance are in boldface. (B) Schematic of voltage-gated sodium channel ( $\text{Na}_v$ )  $\alpha$ -subunits (DI, DII, DIII, and DIV). Each unit is composed of 6 subunits, 1-4 (blue) and 5-6 (yellow). Alternating extra- and intracellular loops are shown in black with the p-loops between subunits 5 and 6 highlighted in red. Mutations conferring resistance are shown within black circles on p-loops.

### Genome annotation

Genes were annotated using a *de novo* predictor supplemented with transcriptomic evidence. Training models were produced by PASA (PASA, [RRID:SCR\\_014656](#)) [38] using a transcriptome composed of 12 tissues (as listed above) and supplied to the *de novo* predictor Augustus (Augustus, [RRID:SCR\\_008417](#)) [39] along with intron, exon, and repeat hints (generated by repeatmasker). Alternative splicing of gene models was also predicted using PASA (PASA, [RRID:SCR\\_014656](#)). Methods used for annotation have been documented in an online git repository [107]. Additional genes were predicted by mapping raw expressed reads against the genome. Functional annotation of gene models was achieved using InterPro v72.0 (InterPro, [RRID:SCR\\_006695](#)) [106]. Completeness of genes was assessed us-

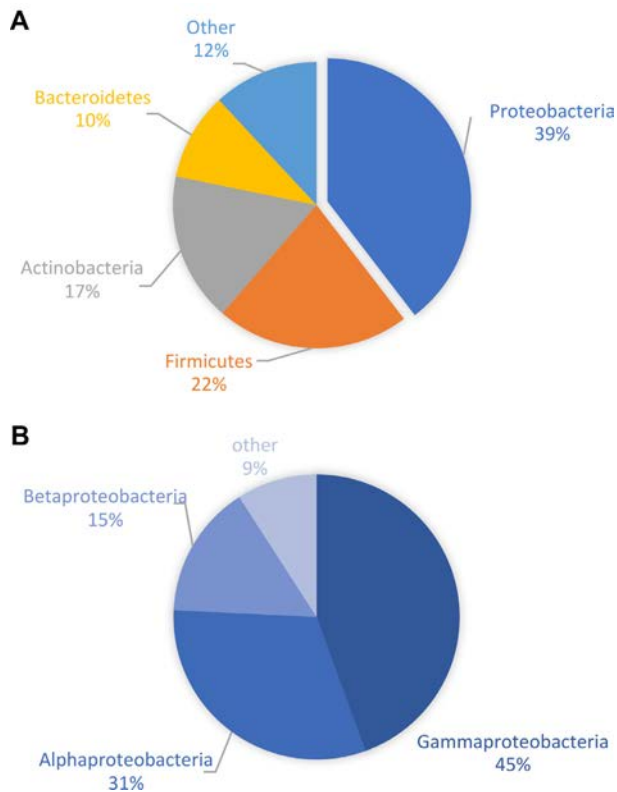
ing BUSCO v3 Metazoan database (BUSCO, [RRID:SCR\\_015008](#)) [40].

### Heterozygosity

JELLYFISH v2.2.1 (Jellyfish, [RRID:SCR\\_005491](#)) was used in conjunction with GenomeScope (GenomeScope, [RRID:SCR\\_017014](#)) [108] to calculate heterozygosity in *H. maculosa* using a *k*-mer frequency of 21 (Supplementary Table S5).

### Repetitive and transposable elements

Repetitive and transposable elements were annotated using RepeatModeler v1.0.9 (RepeatScout) (RepeatModeler, [RRID:SCR\\_015027](#)) and masking performed with RepeatMasker v4.0.8 (Re-



**Figure 6:** Assessment of bacteria within the posterior salivary gland of *H. maculosa* (PSG). (A) Bacterial composition at the phylum level of an *H. maculosa* posterior salivary/venom gland. (B) Composition of the largest phylum, Proteobacteria, of an *H. maculosa* posterior salivary/venom gland.

peatMasker, [RRID:SCR.012954](#) [109] (Supplementary Note S3.3). Analysis of gene-associated TEs was conducted by extracting TEs within flanking regions 10 kb upstream and downstream of genes using Bedtools v2.27.1 (BEDTools, [RRID:SCR.006646](#)) [110].

### Calibration of sequence divergence with respect to time

Divergence times between the molluscan genomes (*C. gigas*, *L. gigantea*, *A. californica*, *E. scolopes*, *O. bimaculoides*, *C. minor*, and *H. maculosa*) and transcriptomes (*Sepia officinalis*, *Idiosepius notoides*, *O. kaurna*, and *O. vulgaris*) was obtained using a mutual best hit approach. Bioprojects for each genome used are as follows: *C. gigas* (PRJNA629593 and PRJEB3535), *L. gigantea* (PRJNA259762 and PRJNA175706), *A. californica* (PRJNA629593 and PRJNA13635), and *E. scolopes* (PRJNA47095). *O. bimaculoides* was obtained from [111]. The *I. notoides* (BioProject: PRJNA302677) transcriptome was sequenced and assembled using the same method previously described for the *H. maculosa* transcriptome. Whole genomes and transcriptomes were BLASTed against *O. bimaculoides*. The resulting hits were filtered, and alignments shared between all species extracted. A maximum likelihood phylogeny was generated using RAxML v8.0 (RAxML, [RRID:SCR.006086](#)) [112]. Phylobayes v3.3 (PhyloBayes, [RRID:SCR.006402](#)) [113] was used to calculate divergence times (Supplementary NoteS4.1).

### Effective population size (PSMC)

Historical changes in effective population size were estimated using PSMC implemented in the software MSMC [114, 115]. To generate inputs for MSMC we selected a subset of the reads

used for genome assembly corresponding to 38× coverage of reads from libraries with short (500 bp) insert sizes. These were pre-processed according to GATK best practices; briefly, adapters were marked with Picard 2.2.1, reads were mapped to the *H. maculosa* genome using bwa mem v 0.7.17 (BWA, [RRID:SCR.010910](#)) [116], and PCR duplicates identified using Picard v2.2.1. To avoid inaccuracies due to poor coverage or ambiguous read mapping we masked regions where short reads would be unable to find unique matches using SNPable [117] and where coverage was more than double or less than half the genome-wide average of 38×. Variant sites were called within unmasked regions and results converted to MSMC input format using msmc-tools [118]. All data for *H. maculosa* scaffolds of length >1 Mb were then used to generate 100 bootstrap replicates by dividing data into 500-kb chunks and assembling them into 20 chromosomes with 100 chunks each. We then ran msmc2 on each bootstrap replicate and assembled and imported the resulting data into R for plotting. A mutation rate of 2.4e−9 per base per year and a generation time of 1 year were assumed in order to set a timescale in years and convert coalescence rates to effective population size.

### Mutation rate

Mutation rate was calculated by extracting orthologous genes from *O. bimaculoides* and *H. maculosa*. Neutrality was assumed for genes with very low expression (<10 TPM across all tissues). Neutral genes were aligned using MAFFT v7.407 [119] and codeml (PAML, [RRID:SCR.014932](#)) [120] was used to calculate substitution metrics (dS). Per base neutral substitution between lineages was determined using the mean dS value divided by divergence time (refer to "Calibration of sequence divergence with respect to time") over the number of generations. Because octopus are diploid the rate was divided by 2. Divergence between species was calculated using Phylobayes v3.3 (PhyloBayes, [RRID:SCR.006402](#)) [113].

### Quantifying gene expression/specificity

Gene expression (as TPM) within individual tissues was calculated using Kallisto (kallisto, [RRID:SCR.016582](#)) [121] for the transcriptomic data sets of *H. maculosa*, *O. bimaculoides*, and *C. minor*. Defaults were used and counts of specific genes were calculated as TPM defined as any gene with  $\tau > 0.80$ .

### Gene model expression dynamics

Patterns of gene expression and loss were assessed across octopod genomes at differing taxonomic/organismal levels. Gene models were classified as lineage-specific, octopod specific, or non-specific (orthologous to a gene outside of octopods). Expression at each level was determined using whole transcriptomes from all tissues of each species. Genes with expression within  $\geq 1$  tissue were determined to be expressed; loss of expression was classified as a gene with a single ortholog in each species, which is expressed in  $\geq 1$  species and not expressed in the remaining species.

### Dynamics of PSG gene expression

To identify patterns of PSG-specific gene expression (losses and shifts) between the 3 available octopod genomes, genes with expression specific to the PSG of each species were examined separately. Specific gene expression was defined as  $\tau > 0.8$ . Orthologous groups were identified between species using Orthovenn2

[122] and sequences that were identified as lineage specific were confirmed using BLAST. Types of expressions were categorized as follows: a loss of expression requires a gene to be present in all 3 octopods and expressed in  $\geq 1$  species while having no detectable expression in  $\geq 1$  species. A shift in expression occurs when an ortholog present in all species is expressed in different tissues.

### The role of the $\text{Na}_v$ in TTX resistance

Sodium channels for the 3 octopus genomes along with all available in-house cephalopod transcriptomes were extracted manually using a series of BLAST searches against the nr database. Annotation was achieved using Interpro v72.0 (InterPro, RRID: SCR.006695) [106] and identification and extraction of p-loop regions of the sodium channel  $\alpha$ -subunit were manually performed. Where sodium channels were incomplete, alignment against related complete channels was used to extract the p-loop regions. Individual mutations with potential to confer resistance were identified manually in Geneious v10.1 [123].

### Microbiome of PSG

A single ribo-depleted RNA sample of *H. maculosa* PSG was examined using the SAMSA2 pipeline [124] to identify the bacterial composition and corresponding molecular functions. Two databases were used, Subsys and NCBI RefBac. The Krona package [125] was used to produce visualizations of each dataset.

### Availability of Source Code and requirements

Project name: BRO\_annotation  
 Project home page: [https://github.com/blwhitelaw/BRO\\_annotation](https://github.com/blwhitelaw/BRO_annotation)  
 Operating system(s): Linux  
 Programming language: Unix/Bash  
 Other requirements: high-performance computing  
 License: GPL-2.0 License  
 Any restrictions to use by non-academics: none  
 RRID:SCR.019072

### Availability of Supporting Data and Materials

Genomic and transcriptomic data produced and used in this article have been made available in the NCBI BioProject: PRJNA602771 under the following accession numbers: raw transcriptome (SAMN13930963–SAMN13930975), genome assembly (SAMN13906985), raw genome reads (SAMN13906958), and gene models (SAMN13942395). Voucher specimen for the transcriptome is stored at Melbourne Museum. All supporting data and materials are available in the GigaScience GigaDB database [126]. This includes expression data for the transcriptome, raw transcriptome reads, gene models, gene annotation gff and assembled genome, as well as files used in figure generation (i.e., trees, heat maps).

### Additional Files

Supplementary Note S1. GENOME SEQUENCING, ASSEMBLY AND ANALYSES  
 Supplementary Note S2. TRANSCRIPTOME SEQUENCING AND ANALYSIS  
 Supplementary Note S3. ANNOTATION OF TRANSPOSABLE ELEMENTS AND PROTEIN CODING GENES

Supplementary Note S4. MULTI-GENE PHYLOGENY AND GENE FAMILY EXPANSION ANALYSES

Supplementary Note S5. ANALYSIS OF NEURAL ASSOCIATED GENE FAMILIES

Supplementary Note S6. EVOLUTION OF THE VENOM/POSTERIOR SALIVARY GLAND IN OCTOPODS

Supplementary Note S7. EVOLUTION OF TETRODOTOXIN RESISTANCE IN *H. MACULOSA*

Supplementary Note S8. MICROBIOME OF THE *H. MACULOSA* POSTERIOR SALIVARY GLAND

Supplementary Table S1. Summary for Illumina libraries

Supplementary Table S2. Comparison of original Illumina and Dovetail augmented assemblies.

Supplementary Table S3. Statistical comparisons between original Illumina and Dovetail augmented assemblies

Supplementary Table S4. Assembly statistics for the three octopod genomes used in this study

Supplementary Table S5. GenomeScope version 1.0 *H. maculosa* results

Supplementary Table S6. Heterozygosity for published molluscan genomes

Supplementary Table S7. *H. maculosa* assembly assessed for completeness against the BUSCO Metazoan database.

Supplementary Table S8. Summary for *H. maculosa* repeat annotation

Supplementary Figure S1. Comparison of assembly continuity between original Illumina (input scaffolds) and Dovetail (Final scaffolds) augmented assemblies

Supplementary Figure S2. PSMC estimation of effective population size in *H. maculosa*

Supplementary Figure S3. QI-TREE Maximum-likelihood tree

Supplementary Figure S4. Phylogenetic tree of cadherins in *H. maculosa* (blue), *O. bimaculoides* (orange) and *C. minor* (green)

Supplementary Figure S5. Phylogenetic tree of protocadherins in *H. maculosa* (blue), *O. bimaculoides* (orange) and *C. minor* (green)

Supplementary Figure S6. Distribution of tau values for genes in *H. maculosa*, *C. minor* and *O. bimaculoides*

Supplementary Figure S7. Orthologous genes specifically expressed in the PSG of *O. bimaculoides* and *C. minor* which have no ortholog in *H. maculosa*

Supplementary Figure S8. Alignment of  $\text{Na}_v1$  p-loop regions

Supplementary Figure S9. Alignment of  $\text{Nav}2$  p-loop regions

Supplementary Figure S10. Expression of  $\text{Nav}1$  and  $\text{Nav}2$  channels across shared tissues of *H. maculosa* (*Hmac*), *O. bimaculoides* (*Obim*) and *C. minor* (*Cmin*)

Supplementary Data 1 : Table of genomic Illumina library insert sizes

### Abbreviations

BLAST: Basic Local Alignment Search Tool; bp: base pairs; BUSCO: Benchmarking Universal Single-Copy Orthologs; BWA: Burrows-Wheeler Aligner; CHGN: chondroitin N-acetylgalactosaminyltransferase, C2H2: Cys2-His2; GATK: Genome Analysis Toolkit; Gb: gigabase pairs; kb: kilobase pairs; LINE: long interspersed nuclear element; MAFFT: Multiple Alignment using Fast Fourier Transform; Mb: megabase pairs; MSMC: multiple sequentially Markovian coalescent; mya: million years ago; NCBI: National Center for Biotechnology Information; PASA: Program to Assemble Spliced Alignments; PSG: posterior salivary gland; PSMC: pairwise sequentially Markovian coalescent; RAXML: Randomized Axelerated Maximum Likelihood;

SINE: short interspersed nuclear element; SPRR: small proline-rich proteins; STX: saxitoxin; TE: transposable element; TPM: transcripts per million; TTX: tetrodotoxin.

## Ethics Declaration

Animal Ethics Approval: field collection of fishes, cephalopods (nautilus, squids, cuttlefishes, and octopuses), and decapod crustaceans (crabs, lobsters, crayfishes, and their allies) was conducted for Museums Victoria (Animal Ethics Committee: Museums Victoria; AEC Approval No. 10006).

## Competing Interests

The authors declare that they have no competing interests.

## Funding

This work was supported by an Australian Biological Resources Study (ABRS) grant (ref: RF211–41). O.S. was supported by the Austrian Science Fund (FWF) grant P30686-B29.

## Acknowledgements

We thank Dr. Mark Norman and Colin Silvey for aiding J.F. in collection of the 2 blue-ringed octopus specimens. We thank Jacqui Stuart for her brilliant illustrations and work on Figure beautification.

## References

- Ritschard EA, Whitelaw B, Albertin CB, et al. Coupled genomic evolutionary histories as signatures of organismal innovations in cephalopods: Co-evolutionary signatures across levels of genome organization may shed light on functional linkage and origin of cephalopod novelties. *Bioessays* 2019;**41**(12):e1900073.
- Tanner AR, Fuchs D, Winkelmann IE, et al. Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. *Proc R Soc B Biol Sci* 2017;**284**(1850):2016–818.
- Jereb P, Roper CFE, Norman MD, et al. Cephalopods of the world. An annotated and illustrated catalogue of cephalopod species known to date. In: *Octopods and Vampire Squids*, Vol. 3. Rome: FAO; 2014.
- Graziadei PPC, Gagne HT. Sensory innervation in the rim of the octopus sucker. *J Morphol* 1976;**150**(3):639–79.
- Froesch D. Projection of chromatophore nerves on the body surface of *Octopus vulgaris*. *Mar Biol* 1973;**19**(2):153–5.
- Budelmann BU. The cephalopod nervous system: what evolution has made of the molluscan design. In: *The Nervous Systems of Invertebrates: An Evolutionary and Comparative Approach*. Springer; 1995:115–38.
- Gray EG, Young JZ. Electron microscopy of synaptic structure of octopus brain. *J Cell Biol* 1964;**21**(1):87–103.
- Fingerhut L, Strugnell JM, Faou P, et al. Shotgun proteomics analysis of saliva and salivary gland tissue from the common octopus *Octopus vulgaris*. *J Proteome Res* 2018;**17**(11):3866–76.
- Whitelaw BL, Cooke IR, Finn J, et al. The evolution and origin of tetrodotoxin acquisition in the blue-ringed octopus (genus *Hapalochlaena*). *Aquatic Toxicol* 2019;**206**:114–22.
- Cooke IR, Whitelaw B, Norman M, et al. Toxicity in cephalopods. In: Gopalakrishnakone P, Malhotra A. *Evolution of Venomous Animals and Their Toxins*. Dordrecht: Springer; 2015:125–43.
- Whitelaw BL, Strugnell JM, Faou P, et al. Combined transcriptomic and proteomic analysis of the posterior salivary gland from the southern blue-ringed octopus and the southern sand octopus. *J Proteome Res* 2016;**15**(9):3284–97.
- Fry BG, Roelants K, Norman JA. Tentacles of venom: toxic protein convergence in the Kingdom Animalia. *J Mol Evol* 2009;**68**(4):311–21.
- Ruder T, Sunagar K, Undheim EAB, et al. Molecular phylogeny and evolution of the proteins encoded by coleoid (cuttlefish, octopus, and squid) posterior venom glands. *J Mol Evol* 2013;**76**(4):192–204.
- Fry BG, Winter K, Norman JA, et al. Functional and structural diversification of the anguimorpha lizard venom system. *Mol Cell Proteomics* 2010;**9**(11):2369–90.
- Hendon RA, Tu AT. Biochemical characterization of the lizard toxin gilatoxin. *Biochemistry* 1981;**20**(12):3517–22.
- Matsui T, Fujimura Y, Titani K. Snake venom proteases affecting hemostasis and thrombosis. *Biochim Biophys Acta* 2000;**1477**:146–56.
- Mitreva M, Papenfuss AT, Whittington CM, et al. Novel venom gene discovery in the platypus. *Genome Biol* 2010;**11**(9):R95.
- Choo YM, Lee KS, Yoon HJ, et al. Antifibrinolytic role of a bee venom serine protease inhibitor that acts as a plasmin inhibitor. *PLoS One* 2012;**7**(2):e32269.
- Veiga SS, Da Silveira RB, Dreyfuss JL, et al. Identification of high molecular weight serine-proteases in *Loxosceles intermedia* (brown spider) venom. *Toxicon* 2000;**38**(6):825–39.
- Modica MV, Lombardo F, Franchini P, et al. The venomous cocktail of the vampire snail *Colubraria reticulata* (Mollusca, Gastropoda). *BMC Genomics* 2015;**16**(1):441.
- von Reumont BM, Undheim EAB, Jauss RT, et al. Venomics of remipede crustaceans reveals novel peptide diversity and illuminates the venom's biological role. *Toxins* 2017;**9**(8):234.
- Jaimes-Becerra A, Chung R, Morandini AC, et al. Comparative proteomics reveals recruitment patterns of some protein families in the venoms of Cnidaria. *Toxicon* 2017;**137**:19–26.
- Crone HD, Leake B, Jarvis MW, et al. On the nature of “Maculotoxin,” a toxin from the blue-ringed octopus (*Hapalochlaena maculosa*). *Toxicon* 1976;**14**(6):423–6.
- Geffeney SL, Williams BL, Rosenthal JJC, et al. Convergent and parallel evolution in a voltage-gated sodium channel underlies TTX-resistance in the greater blue-ringed octopus: *Hapalochlaena lunulata*. *Toxicon* 2019;**170**:77–84.
- Wu Z, Yang Y, Xie L, et al. Toxicity and distribution of tetrodotoxin-producing bacteria in puffer fish *Fugu rubripes* collected from the Bohai Sea of China. *Toxicon* 2005;**46**(4):471–6.
- Tsuruda K, Arakawa O, Kawatsu K, et al. Secretory glands of tetrodotoxin in the skin of the Japanese newt *Cynops pyrrhogaster*. *Toxicon* 2002;**40**(2):131–6.
- Wakely JF, Fuhrman GJ, Fuhrman FA, et al. The occurrence of tetrodotoxin (tarichatoxin) in Amphibia and the distribution of the toxin in the organs of newts (*Taricha*). *Toxicon* 1966;**3**(3):195–203.
- Du Y, Nomura Y, Liu Z, et al. Functional expression of an arachnid sodium channel reveals residues responsible for

- tetrodotoxin resistance in invertebrate sodium channels. *J Biol Chem* 2009;**284**(49):33869–75.
29. McGlothlin JW, Chuckalovcak JP, Janes DE, et al. Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake *Thamnophis sirtalis*. *Mol Biol Evol* 2014;**31**(11):2836–46.
  30. Cheng CA, Hwang DF, Tsai YH, et al. Microflora and tetrodotoxin-producing bacteria in a gastropod, *Niotha clathrata*. *Food Chem Toxicol* 1995;**33**(11):929–34.
  31. Williams BL, Caldwell RL. Intra-organismal distribution of tetrodotoxin in two species of blue-ringed octopuses (*Hapalochlaena fasciata* and *H. lunulata*). *Toxicon* 2009;**54**(3):345–53.
  32. Yotsu-Yamashita M, Mebs D, Flachsenberger W. Distribution of tetrodotoxin in the body of the blue-ringed octopus (*Hapalochlaena maculosa*). *Toxicon* 2007;**49**(3):410–2.
  33. Mäthger LM, Bell GRR, Kuzirian AM, et al. How does the blue-ringed octopus (*Hapalochlaena lunulata*) flash its blue rings? *J Exp Biol* 2012;**215**(21):3752–7.
  34. Williams BL, Stark MR, Caldwell RL. Microdistribution of tetrodotoxin in two species of blue-ringed octopuses (*Hapalochlaena lunulata* and *Hapalochlaena fasciata*) detected by fluorescent immunolabeling. *Toxicon* 2012;**60**(7):1307–13.
  35. Savage IVE, Howden MEH. Hapalotoxin, a second lethal toxin from the octopus *Hapalochlaena maculosa*. *Toxicon* 1977;**15**(5):463–6.
  36. Kanda A, Iwakoshi-Ukena E, Takuwa-Kuroda K, et al. Isolation and characterization of novel tachykinins from the posterior salivary gland of the common octopus *Octopus vulgaris*. *Peptides* 2003;**24**(1):35–43.
  37. Ruder T, Ali SA, Ormerod K, et al. Functional characterization on invertebrate and vertebrate tissues of tachykinin peptides from octopus venoms. *Peptides* 2013;**47**:71–6.
  38. Haas BJ, Delcher AL, Mount SM, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* 2003;**31**(19):5654–66.
  39. Stanke M. Gene Prediction with a Hidden Markov Model. Ph.D. Thesis. University of Göttingen; 2004.
  40. Waterhouse RM, Sepey M, Simão FA, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 2018;**35**(3):543–8.
  41. Zarrella I, Herten K, Maes GE, et al. The survey and reference assisted assembly of the *Octopus vulgaris* genome. *Sci Data* 2019;**6**(1):13.
  42. Albertin CB, Simakov O, Mitros T, et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 2015;**524**(7564):220–4.
  43. Cai H, Li Q, Fang X, et al. A draft genome assembly of the solar-powered sea slug *Elysia chlorotica*. *Sci Data* 2019;**6**:190022.
  44. Liu C, Zhang Y, Ren Y, et al. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. *Gigascience* 2018;**7**(9):giy101.
  45. Du X, Fan G, Jiao Y, et al. The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *Gigascience* 2017;**6**(8):gix059.
  46. Powell D, Subramanian S, Suwansa-ard S, et al. The genome of the oyster *Saccostrea* offers insight into the environmental resilience of bivalves. *DNA Res* 2018;**25**(6):655–65.
  47. Sun J, Zhang Y, Xu T, et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nat Ecol Evol* 2017;**1**(5):121.
  48. Uliano-Silva M, Dondero F, Dan Otto T, et al. A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*. *Gigascience* 2017;**7**(2):gix128.
  49. Zhang G, Fang X, Guo X, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012;**490**(7418):49.
  50. Jiao W, Fu X, Dou J, et al. High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA Res* 2014;**21**(1):85–101.
  51. Garton DW, Haag WR. Heterozygosity, shell length and metabolism in the European mussel, *Dreissena polymorpha*, from a recently established population in Lake Erie. *Comp Biochem Physiol A Physiol* 1991;**99**(1–2):45–8.
  52. Simmons MJ, Crow JF. Mutations affecting fitness in *Drosophila* populations. *Annu Rev Genet* 1977;**11**(1):49–78.
  53. Kumar S, Subramanian S. Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A* 2002;**99**(2):803–8.
  54. Hayward BW, Kawagata S, Grenfell HR, et al. Last global extinction in the deep sea during the mid-Pleistocene climate transition. *Paleoceanogr Paleoclimatol* 2007;**22**(3), doi.org/10.1029/2007PA001424.
  55. Hofreiter M, Stewart J. Ecological change, range fluctuations and population dynamics during the Pleistocene. *Curr Biol* 2009;**19**(14):R584–94.
  56. Reynolds TV, Matthee CA, Von Der Heyden S. The influence of Pleistocene climatic changes and ocean currents on the phylogeography of the southern African barnacle, *Tetraclita serrata* (Thoracica; Cirripedia). *PLoS One* 2014;**9**(7):e102115.
  57. Huber CD, DeGiorgio M, Hellmann I, et al. Detecting recent selective sweeps while controlling for mutation rate and background selection. *Mol Ecol* 2016;**25**(1):142–56.
  58. Mao Y, Economo EP, Satoh N. The roles of introgression and climate change in the rise to dominance of *Acropora* corals. *Curr Biol* 2018;**28**(21):3373–82.e5.
  59. Kim B-M, Amores A, Kang S, et al. Antarctic blackfin icefish genome reveals adaptations to extreme environments. *Nat Ecol Evol* 2019;**3**(3):469.
  60. Guzik MT, Norman MD, Crozier RH. Molecular phylogeny of the benthic shallow-water octopuses (Cephalopoda: Octopodinae). *Mol Phylogenet Evol* 2005;**37**(1):235–48.
  61. Strugnell J, Jackson J, Drummond AJ, et al. Divergence time estimates for major cephalopod groups: evidence from multiple genes. *Cladistics* 2006;**22**(1):89–96.
  62. Strugnell JM, Rogers AD, Prodöhl PA, et al. The thermohaline expressway: the Southern Ocean as a centre of origin for deep-sea octopuses. *Cladistics* 2008;**24**(6):853–60.
  63. Lindgren AR, Anderson FE. Assessing the utility of transcriptome data for inferring phylogenetic relationships among coleoid cephalopods. *Mol Phylogenet Evol* 2018;**118**:330–42.
  64. Styfhals R, Seuntjens E, Simakov O, et al. In silico identification and expression of protocadherin gene family in *Octopus vulgaris*. *Front Physiol* 2019;**9**:1905.
  65. Zipursky SL, Sanes JR. Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. *Cell* 2010;**143**(3):343–53.
  66. Tepass U, Truong K, Godt D, et al. Cadherins in embryonic and neural morphogenesis. *Nat Rev Mol Cell Biol* 2000;**1**(2):91.
  67. Peek SL, Mah KM, Weiner JA. Regulation of neural circuit formation by protocadherins. *Cell Mol Life Sci* 2017;**74**:4133–57.

68. De Wit J, Ghosh A. Specification of synaptic connectivity by cell surface interactions. *Nat Rev Neurosci* 2016;**17**:4.
69. Layden MJ, Meyer NP, Pang K, et al. Expression and phylogenetic analysis of the zic gene family in the evolution and development of metazoans. *Evodevo* 2010;**1**(1):12.
70. Liu H, Chang L-H, Sun Y, et al. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol Evol* 2014;**6**(3):510–25.
71. Young JZ. The number and sizes of nerve cells in Octopus. *J Zool* 1963;**140**(2):229–54.
72. Belcaid M, Casaburi G, McAnulty SJ, et al. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proc Natl Acad Sci U S A* 2019;**116**(8):3030–5.
73. Kim B-M, Kang S, Ahn D-H, et al. The genome of common long-arm octopus *Octopus minor*. *Gigascience* 2018;**7**(11):giy119.
74. Kapitonov VV, Jurka J. Rolling-circle transposons in eukaryotes. *Proc Natl Acad Sci U S A* 2001;**98**(15):8714–9.
75. Oulion B, Dobson JS, Zdenek CN, et al. Factor X activating *Atractaspis* snake venoms and the relative coagulotoxicity neutralising efficacy of African antivenoms. *Toxicol Lett* 2018;**288**:119–28.
76. Rogalski A, Soerensen C, op den Brouw B, et al. Differential procoagulant effects of saw-scaled viper (Serpentes: Viperidae: Echis) snake venoms on human plasma and the narrow taxonomic ranges of antivenom efficacies. *Toxicol Lett* 2017;**280**:159–70.
77. Sousa LF, Zdenek CN, Dobson JS, et al. Coagulotoxicity of *Bothrops* (lancehead pit-vipers) venoms from Brazil: Differential biochemistry and antivenom efficacy resulting from prey-driven venom variation. *Toxins* 2018;**10**(10):411.
78. Fox JW, Serrano SMT. Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. *Toxicon* 2005;**45**(8):969–85.
79. Grisley MS, Boyle PR. Bioassay and proteolytic activity of digestive enzymes from octopus saliva. *Comp Biochem Physiol B Comp Biochem* 1987;**88**(4):1117–23.
80. Key LN, Boyle PR, Jaspers M. Novel activities of saliva from the octopus *Eledone cirrhosa* (Mollusca; Cephalopoda). *Toxicon* 2002;**40**(6):677–83.
81. McDonald NM, Cottrell GA. Purification and mode of action of toxin from *Eledone cirrosa*. *Comp Gen Pharmacol* 1972;**3**(10):243–8.
82. Bordon KCF, Perino MG, Giglio JR, et al. Isolation, enzymatic characterization and antiedematogenic activity of the first reported rattlesnake hyaluronidase from *Crotalus durissus terrificus* venom. *Biochimie* 2012;**94**(12):2740–8.
83. Frank HY, Catterall WA. Overview of the voltage-gated sodium channel family. *Genome Biol* 2003;**4**(3):207.
84. How C-K, Chern C-H, Huang Y-C, et al. Tetrodotoxin poisoning. *Am J Emerg Med* 2003;**21**(1):51–4.
85. Brodie ED, III, Brodie ED, Jr. Tetrodotoxin resistance in garter snakes: an evolutionary response of predators to dangerous prey. *Evolution* 1990;**44**(3):651–9.
86. Nakamura M, Yasumoto T. Tetrodotoxin derivatives in puffer fish. *Toxicon* 1985;**23**(2):271–6.
87. Jost MC, Hillis DM, Lu Y, et al. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Mol Biol Evol* 2008;**25**(6):1016–24.
88. Feldman CR, Brodie ED, Pfreder ME. Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. *Proc Natl Acad Sci U S A* 2012;**109**(12):4556–61.
89. Jeziorski MC, Greenberg RM, Anderson PAV. Cloning of a putative voltage-gated sodium channel from the turbellarian flatworm *Bdelloura candida*. *Parasitology* 1997;**115**(3):289–96.
90. Shen H, Li Z, Jiang Y, et al. Structural basis for the modulation of voltage-gated sodium channels by animal toxins. *Science* 2018;**362**:eaau2596.
91. Flachsenberger W, Kerr DIB. Lack of effect of tetrodotoxin and of an extract from the posterior salivary gland of the blue-ringed octopus following injection into the octopus and following application to its brachial nerve. *Toxicon* 1985;**23**(6):997–9.
92. Lopes VM, Baptista M, Repolho T, et al. Uptake, transfer and elimination kinetics of paralytic shellfish toxins in common octopus (*Octopus vulgaris*). *Aquatic Toxicol* 2014;**146**:205–11.
93. Ulbricht W, Wagner H, Schmidtmayer J. Kinetics of TTX-STX block of sodium channels. *Ann NY Acad Sci* 1986;**479**(1):68–83.
94. Monteiro A, Costa PR. Distribution and selective elimination of paralytic shellfish toxins in different tissues of *Octopus vulgaris*. *Harmful Algae* 2011;**10**(6):732–7.
95. Li S-C, Wang W-X, Hsieh D. Feeding and absorption of the toxic dinoflagellate *Alexandrium tamarense* by two marine bivalves from the South China Sea. *Mar Biol* 2001;**139**(4):617–24.
96. Braid HE, Deeds J, DeGrasse SL, et al. Preying on commercial fisheries and accumulating paralytic shellfish toxins: a dietary analysis of invasive *Dosidicus gigas* (Cephalopoda Ommastrephidae) stranded in Pacific Canada. *Mar Biol* 2012;**159**(1):25–31.
97. Shea EK, Ziegler A, Faber C, et al. Dumbo octopus hatching provides insight into early cirrate life cycle. *Curr Biol* 2018;**28**(4):R144–5.
98. Lee M-J, Jeong D-Y, Kim W-S, et al. A tetrodotoxin-producing *Vibrio* strain, LM-1, from the puffer fish *Fugu vermicularis radiatus*. *Appl Environ Microbiol* 2000;**66**(4):1698–701.
99. Yotsu M, Yamazaki T, Meguro Y, et al. Production of tetrodotoxin and its derivatives by *Pseudomonas* sp. isolated from the skin of a pufferfish. *Toxicon* 1987;**25**(2):225–8.
100. Chau R, Kalaitzis J, Wood S, et al. Diversity and biosynthetic potential of culturable microbes associated with toxic marine animals. *Marine Drugs* 2013;**11**(8):2695–712.
101. Hwang DF, Arakawa O, Saito T, et al. Tetrodotoxin-producing bacteria from the blue-ringed octopus *Octopus maculosus*. *Mar Biol* 1989;**100**(3):327–32.
102. Magarlamov T, Melnikova D, Chernyshev A. Tetrodotoxin-producing bacteria: Detection, distribution and migration of the toxin in aquatic systems. *Toxins* 2017;**9**(5):166.
103. assembly-stats. <https://github.com/sanger-pathogens/assembly-stats>. Accessed 29 September 2020.
104. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;**29**(7):644–52.
105. Bryant DM, Johnson K, DiTommaso T, et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep* 2017;**18**(3):762–76.
106. Jones P, Binns D, Chang H-Y, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;**30**(9):1236–40.
107. BRO\_annotation. [https://github.com/blwhitelaw/BRO\\_annotation](https://github.com/blwhitelaw/BRO_annotation). Accessed 29 September 2020.



108. Vurtture GW, Sedlazeck FJ, Nattestad M, et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 2017;**33**(14):2202–4.
109. A.F.A. Smit Hubley R, Green P. RepeatMasker Open-4.0. <http://repeatmasker.org>. Accessed 29 September 2020.
110. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**(6):841–2.
111. Octopus bimaculoides genome files. <http://octopus.unit.ois.t.jp/OCTDATA/BASIC/Metazome/Obimaculoides.280.fa.gz>. Accessed 29 September 2020.
112. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;**22**(21):2688–90.
113. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;**52**(5):696–704.
114. Li H, Durbin R. Inference of human population history from whole genome sequence of a single individual. *Nature* 2011;**475**(7357):493.
115. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 2014;**46**(8):919.
116. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;**25**(14):1754–60.
117. SNPable. <http://lh3lh3.users.sourceforge.net/snppable.shtml>. Accessed 29 September 2020.
118. msmc-tools. <https://github.com/stschiff/msmc-tools>. Accessed 29 September 2020.
119. Rozewicki J, Li S, Amada KM, et al. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res* 2019;**47**(W1):W5–10.
120. Adachi J, Hasegawa M. MOLPHY Version 2.3: Programs for Molecular Phylogenetics Based on Maximum Likelihood. Tokyo: Institute of Statistical Mathematics; 1996.
121. Bray NL, Pimentel H, Melsted P, et al. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016;**34**(5):525–7.
122. Xu L, Dong Z, Fang L, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 2019;**47**(W1):W52–8.
123. Geneious v10.1. [https://assets.geneious.com/documentation/geneious/release\\_notes.html#v10.1](https://assets.geneious.com/documentation/geneious/release_notes.html#v10.1). Accessed 29 September 2020.
124. Westreich ST, Korf I, Mills DA, et al. SAMSA: a comprehensive metatranscriptome analysis pipeline. *BMC Bioinformatics* 2016;**17**(1):399.
125. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 2011;**12**(1):385.
126. Whitelaw BL, Cooke IR, Finn J, et al. Supporting data for “Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion and loss.” GigaScience Database 2020. <http://dx.doi.org/10.5524/100793>.

## CHAPTER 3

### **High density genetic linkage map of the southern blue-ringed octopus (Octopodidae: *Hapalochlaena maculosa*)**

Brooke L. Whitelaw, David B. Jones, Jarrod Guppy, Peter Morse, Jan M. Strugnell, Ira R. Cooke, Kyall Zenger

#### **3.1 ABSTRACT**

The southern blue-ringed octopus (*Hapalochlaena maculosa*) is distributed across the Southern Australian coast and, among other species within the genus, it is known to sequester the potent neurotoxin tetrodotoxin (TTX) within its venom and tissues. *Hapalochlaena* are the only octopods known to sequester TTX, however its biosynthetic origin remains unresolved. Members of the genus represent an interesting case study within octopods. A draft genome was first published in 2020 (Whitelaw et al., 2020), however no chromosome level assembly is available. Genetic linkage maps provide a useful resource for non-model genomes and can aid in genome reassembly to form more contiguous pseudo-chromosomes. We present the first linkage map of any cephalopod, the southern blue-ringed octopus, composed of 47 linkage groups (LG). A total of 2,166 single nucleotide polymorphisms (SNPs) and 2,455 presence absence variant (PAV) loci were utilised by Lep-Map3 in linkage map construction. The map length spans 2016.62cM with an average marker distance of 0.85cM. Integration of the recent *H. maculosa* genome allowed 1,151 scaffolds comprising 34% of the total genomic sequence to

be orientated and/or placed using 1,278 markers across all 47 LG. Long range genomic information provided by the linkage map provides a new perspective on previous findings that conserved HOX genes in Octopus genomes are distributed on separate scaffolds. In the *H. maculosa* linkage map three (SCR, LOX4 and POST1) of six identified HOX genes (HOX1/LAB, SCR, LOX2, LOX4, LOX5, POST1) were located within the same LG (LG 9). Generation of a linkage map for *H. maculosa* has provided a valuable resource for understanding the evolution of cephalopod genomes and will provide a base for future work including Qualitative trait loci (QTL) mapping.

### **3.2 INTRODUCTION**

Members of the blue-ringed octopus genus (*Hapalochlaena*) are the only octopods known to sequester the potent neurotoxin tetrodotoxin (TTX) within their tissues and venom (Sheumack et al., 1984). They are easily identified by their iridescent blue rings and/or lines which are flashed when threatened in an aposematic advertisement of their toxicity (Mäthger et al., 2012). They also remain the only octopod known to be capable of inflicting a lethal bite to humans (Jacups and Currie, 2008; White, 2018). Investigations into venom evolution at a transcriptomic and proteomic level have indicated lowered abundance of venom and venom-associated proteins in venom glands of *Hapalochlaena maculosa* relative to other octopods. This is hypothesised to reflect the inclusion of TTX resulting in the redundancy of other octopod venom components (Whitelaw et al., 2016). Additionally, analysis of the *Hapalochlaena*

*maculosa* genome detected a potential contraction of a key family of venom proteins (serine proteases) relative to related octopods (Whitelaw et al., 2020).

The current genome assembly for *H. maculosa* is highly fragmented (47K+ scaffolds) (Whitelaw et al., 2020). Similar to other cephalopods the *H. maculosa* genome exhibits a high proportion of repetitive elements (37%), in conjunction with high heterozygosity (0.95%), and a relatively large size overall (4Gb) (Whitelaw et al., 2020). These features are not conducive to generating a chromosome level assembly (Sohn and Nam, 2018). The first octopod genome (the California two-spot octopus, *Octopus bimaculoides*) was published in 2015 (Albertin et al., 2015) and since then an additional four octopod genomes have been published (*Callistoctopus minor*, *Octopus vulgaris*, *H. maculosa* and *Octopus sinensis*) (Kim et al., 2018; Li et al., 2020; Whitelaw et al., 2020; Zarrella et al., 2019). To date, the only chromosomal level octopus genome published is that of *O. sinensis*, (Li et al., 2020). Earlier karyotyping studies of cephalopods revealed a chromosomal number of 30n for four members of Octopodiformes (*C. minor*, *Cistopus chinensis*, *O. vulgaris* and *Amphioctopus fangsiao*) (Gao and Natsukari, 1990; Wang and Zheng, 2017) and 46n for five decapodiformes (*Sepia esculenta*, *Sepia lycidas*, *Sepioteuthis lessoniana*, *Heterololigo bleekeri* and *Photololigo edulis*) (Gao and Natsukari, 1990). While genetic resources such as genomes are gaining traction in the elucidation of cephalopod evolution they are difficult to generate and do not provide a complete picture.

Linkage maps are a valuable tool that facilitate the examination of key processes governing genome evolution by acting as a platform for analysing large (chromosomal rearrangements) and fine scale (gene specific) events (Leitwein et al., 2017). They form an

important component of comparative whole genomic analyses, providing support for genome assembly and annotation in non-model species (Velmurugan et al., 2016). Historically, genetic linkage maps have been used to elucidate the inheritance of complex disease traits and for marker-assisted selection in agricultural species via Qualitative Trait Loci (QTL) analyses (Choi et al., 2020). More recently linkage maps have been utilised in non-model species in an ecological context, aiding the untangling of genotypic and phenotypic traits and their modes of selection (Hagen et al., 2020; Manousaki et al., 2016).

Genetic linkage maps have great utility when applied to cephalopod genomics and could improve our understanding in three key ways, i) improvement of current fragmented genome assemblies, ii) supporting the assembly of new genomes, improving ease and reliability and iii) bridging the gap between genetics and phenotype. Chromosome level genome assemblies are also valuable because they allow for examination of the role of large scale structural variation in evolutionary processes such as speciation, while also providing the genomic context to understand the evolution of key gene families. The HOX gene cluster is a highly conserved set of developmental genes occurring throughout metazoans (Duboule, 2007). Cephalopods exhibit drastic changes to their body plan relative to other metazoans courtesy of unique expression patterns of recruited HOX genes (Lee et al., 2003). A total of nine genes were first identified in the Hawaiian bobtail squid *Euprymna scolopes* (Callaerts et al., 2002) and later found to be shared with *O. bimaculoides*. However, unlike related molluscan genomes (Biscotti et al., 2014) the HOX genes did not form a cluster and were instead scattered across different scaffolds ranging in length between 53kb and 751kb (Albertin et al., 2015). Although this suggests that there is no HOX cluster in cephalopods, potential co-location of cephalopod HOX genes at the

chromosome level has not yet been investigated. Furthermore, genetic linkage maps for one species can aid in the generation of maps for related taxa as has been demonstrated for the reed warbler (*Acrocephalus arundinaceus*) whose linkage map was derived from the common chicken (Dawson et al., 2006). Lastly, a genetic linkage map also provides a method to link octopod genomics and phenomics, facilitating the investigation of the genetic basis of biological traits. Such studies have applications in evolutionary biology (QTL associated with retention of pedomorphic traits in salamanders) (Voss et al., 2012) and aquaculture (body weight and length of turbot) (Sánchez-Molano et al., 2011). QTL studies could be valuable for commercial species such as *Octopus vulgaris* which are within the early stages of development as an aquaculture species (Estefanell et al., 2011).

Generation of a linkage map for cephalopods has not previously been attempted, in part due to the absence of genome assemblies and genetic marker resources prior to 2015 (Albertin et al., 2015), but also due to the difficulty in keeping and breeding cephalopods in captivity. This latter requirement is essential to track pedigree information and attain the large family sizes, which are required for robust linkage analysis (Kawakami et al., 2014). This chapter presents the first linkage map of any cephalopod, the southern blue-ringed octopus (*Hapalochlaena maculosa*) constructed using a combination of single nucleotide polymorphism (SNP) and presence absence variant (PAV) markers. The assembly generated in chapter 2 is incorporated here with the linkage map to produce an improved assembly leading to insights in HOX gene placement within *H. maculosa*.

## 3.3 METHODS

### *3.3.1. Sample collection and family structure*

Detailed descriptions of sample collection, mating structure and housing are available in (Morse et al., 2018a) and were performed by Dr Peter Morse. Briefly, specimens of *H. maculosa* were collected from Cockburn Sound, Western Australia using false shelter traps. Animals were individually housed at the Fremantle octopus facility in 1 litre aquaria. Collected animals (36 in total) were divided into 12 groups with each female being paired with one of the two males sequentially. Animals were paired in a larger separate 30L aquarium overnight and interactions recorded. Resulting families' clutch sizes are located in table 3.1.

### *3.3.2. DNA extraction and SNP generation*

Muscle tissue samples were taken from the arm tip of adults and the body of hatchlings by Dr Shannon Kjeldsen and submitted to Diversity Arrays Technology PL, Canberra ACT, Australia, where they underwent DNA extraction, restriction enzyme digestion, library preparation, genotype-by-sequencing data generation and QA/QC of sequences via DArTseq™ 1.0 technology (Sansaloni et al., 2011).

### *3.3.3. SNP selection, quality control and genotyping*

Filters were applied to the SNP data provided by Diversity Arrays using the program DARTQC (<https://github.com/esteinig/dartQC>). Retained SNPs met the following criteria, individual genotypes were called on a minimum of 5 read counts, average SNP repeatability > 0.9, SNP call rate > 0.6, and MAF (Minor Allele Frequency) > 0.01. In addition, sequence similarity searches of 95% identity were conducted within CD-HIT (Fu et al., 2012; Li and Godzik, 2006) and for any SNPs that were identified within the same flanking regions, only the SNP with the highest MAF was retained. Individual families were identified (see section 3.3.4), and PLINK (Purcell et al., 2007) was used to identify and remove mendelian errors at a threshold of 0.1 with any remaining errors set to missing. Presence-absence variant (PAV) loci were filtered for reproducibility > 0.9 and a chi-square test performed to identify mendelian errors ( $p > 0.05$ ). SNP markers were merged across ten families using bedtools merge (Quinlan and Hall, 2010), SNPs were included if they occurred in a single family. PAV and SNP markers for ten families were combined to form a single file, and only markers which occurred in at least one family were retained.

### *3.3.4. Paternity and family designation*

Due to the potential of copulations prior to capture and the ability for females to produce offspring fathered by multiple males (Morse et al., 2018a), paternity was evaluated using 7,100 of the filtered SNPs. Cervus was used to calculate allele frequencies across all samples, simulate parentage and assign parent to offspring as follows (parent pairs with known sexes, 1,600 offspring, 15 candidate mothers, probability mother sampled 0.95, 20 candidate



fathers, probability father sampled 0.95, probability of loci typed 0.95, probability of loci mistyped 0.01, and minimum loci typed 200). Colony was also used to validate families with the following parameter settings (female monogamy, male polygamy, no inbreeding, no clones, diecious, diploid, full-likelihood, medium precision, updating allele frequencies – yes, sibship scaling – yes, weak prior, paternal sibship size 20, maternal sibship size 20, 968 markers, 511 progeny, and respective parents) (Jones and Wang, 2010).

### *3.3.5. Map construction*

The Lep-MAP3 pipeline (Rastas, 2017) was used to construct the final map composed of 10 families with a total of 277 individuals. Markers were filtered for high levels of segregation distortion (parameter dataTolerance 0.001, X2 test, df = 1-2,  $P < 0.001$ ) using the Filtering2 module. Missing parental genotypes were then predicted using the ParentCall2 module with half siblings taken into account with the --halfsib flag. SeparateChromosomes2 was used to assign markers to linkage groups (LG) based on a LOD threshold of 10 and minimum group size of 10.

The resulting 47 LG were ordered using OrderMarkers to create a sex averaged map. Sex informative maps were generated using the preserved marker order obtained from the sex averaged map with intervals recalculated based on female and male informative meiotic events. Maps were visualised using the LinkageMapView package in R.

### 3.3.6. Re-orientation of genomic scaffolds and genes of interest

In order to improve contiguity of the *H. maculosa* genome, markers were mapped to their genomic locations and used in conjunction with the generated linkage map to re-orientate and/or place scaffolds. This produced 47 pseudo-chromosomes, each corresponding to a linkage group. SNP and PAV markers were mapped to the genome of the southern-blue-ringed octopus (NCBI: PRJNA602771, *Hapalochlaena maculosa*) (Whitelaw et al., 2020) using the `bwa mem` (Li, 2013) with a seed length of  $k = 20$ . Input files for chromonomer were generated using the output from Lep-Map3 in R in conjunction with the mapped loci locations. Chromonomer was run with default parameters (Catchen et al., 2020) including the following inputs: genomic locations for mapped markers and contigs, genome.fasta, and genetic linkage map structure. To determine the proportion of the genome included in the 47 pseudo-chromosomes the sum of base pairs from each pseudo-chromosome generated by chromonomer was divided by the total genome length (bp). Assembly-stats (<https://github.com/sanger-pathogens/assembly-stats>) was run to assess the new pseudo-chromosome assembly produced by chromonomer.

Conserved HOX genes were identified from the *H. maculosa* genome by conducting BLAST searches against a data base of known cephalopod HOX genes manually curated from NCBI. The mutual best hits were retained and annotated using InterPro (Finn et al., 2016). Genomic positions were ascertained by aligning genes to the genome using `est2genome` from `exonerate v2.2` (Slater and Birney, 2005). Cephalopod HOX genes were aligned using MAFFT

v7.407 (Rozewicki et al., 2019) and a consensus maximum likelihood phylogeny generated using RAxML v 8.0 with the WAG substitution model for 100 bootstraps (Stamatakis, 2006).

### *3.3.7 Genome Coverage*

Genome coverage was ascertained by estimated genome size/summed length of all linkage groups. Genome size was estimated in centimorgans (cM) using two methods (Yu et al., 2015) and the average taken for coverage estimation. It should be noted that genome size here does not refer to a physical distance but to a distance calculated using recombination frequencies similar to the linkage map. The average interval for each linkage group (LG) was calculated by dividing total LG length by the number of intervals. Genome size estimation 1 (Ge 1) multiplied the total LG length by the factor  $(m+1)/(m-1)$ , where by  $m$  = number of markers within each LG (Chakravarti et al., 1991). Genome size estimation 2 (Ge 2) was calculated by adding  $2 \times (\text{average marker interval})$  to the total LG length (Postlethwait et al., 1994) (Table 3.5). The sum across all LGs was calculated with each method and the average between the two used.

### *3.3.8. Segregation distortion and sex specific recombination*

Gametic and post zygotic selection can result in deviations from Mendelian inheritance of codominant alleles known as segregation distortion. In order to detect segregation distortion between families and across markers log-likelihood ratio tests were performed for goodness of fit to Mendelian expectations. Tests were performed using the software LINKMFEX v3.4 (Danzmann, 2006) which calculates G-values across all parents for each family. Insufficient

number of loci significantly deviated from expected segregation ( $p > 0.05$ ) within and between families to perform further testing with confidence (Supplementary table 2). Sex specific recombination rates were ascertained using LINKMFEX v3.4. Intervals shared between genders independent of family were identified in R and are located in Supplementary table 1. Due to the low number of intervals shared between sexes, sex specific recombination rates could not be calculated nor compared with confidence.

## **3.4 RESULTS**

### *3.4.1. Genotyping and paternity*

A total of 25,253 presence-absence variant (PAV) loci and 19,729 SNPs were discovered using DArTSeq™ genotyping (Diversity Arrays). Stringent filtering, as previously described (methods section 3.3.2), resulted in the retention of 5,656 SNPs and 11,179 PAV for linkage mapping analysis (Tables 3.2 - 3.3). Parentage analysis revealed multiple paternity in 12 clutches resulting in 28 families and 543 offspring in total. Additionally, of the 28 families, pre-capture copulation of five females resulted in five families fathered by a non-genotyped male, and as such, these were excluded from downstream analyses. Families with fewer than 10 individuals or with a parent which did not pass PLINK filtering thresholds (Mendelian errors  $> 10\%$  per marker and  $10\%$  per individual) were also excluded. The remaining 10 families were used in linkage map generation (Table 3.1.).

Table 3.1. Summary of *Haplochromis maculosa* families generated in this study. Bolded rows represent families which passed filtering and were included in linkage map generation. n = size of family.

<b>Female</b>	<b>Male</b>	<b>Family_ID</b>	<b>n (pre filtering)</b>	<b>Status</b>
<b>131</b>	<b>127</b>	<b>131_127</b>	<b>37</b>	<b>Included</b>
	128	131_128	10	excluded due to small size
<b>132</b>	<b>142</b>	<b>132_142</b>	<b>43</b>	<b>Included</b>
	143	132_143	5	excluded due to small size
133	<b>132</b>	<b>133_132</b>	<b>33</b>	<b>Included</b>
	Unknown male 2	133_02	16	excluded due to unknown male
	131	133_131	6	excluded due to small size
<b>136</b>	<b>134</b>	<b>136_134</b>	<b>26</b>	<b>Included</b>
	<b>145</b>	<b>136_145</b>	<b>26</b>	<b>Included</b>
137	Unknown male 1	137_01	35	excluded due to unknown male
	136	137_136	13	male removed by filtering
<b>138</b>	<b>133</b>	<b>138_133</b>	<b>37</b>	<b>Included</b>
	135	138_135	8	excluded due to small size
<b>140</b>	<b>155</b>	<b>140_155</b>	<b>27</b>	<b>Included</b>
	<b>149</b>	<b>140_149</b>	<b>14</b>	<b>Included</b>
<b>144</b>	<b>144</b>	<b>144_144</b>	<b>36</b>	<b>Included</b>
	151	144_151	5	excluded due to small size
<b>145</b>	<b>148</b>	<b>145_148</b>	<b>34</b>	<b>Included</b>
	138	145_138	3	excluded due to small size
146	139	146_139	24	Female removed by filtering
	Unknown male 3	146_03	15	excluded due to unknown male
	137	146_137	11	Female removed by filtering
	Unknown male 4	146_04	5	excluded due to unknown male
	148	146_148	11	Both parents removed by filtering
148	153	148_153	11	Both parents removed by filtering
	156	148_156	6	excluded due to small size
149	Unknown male 5	149_05	35	excluded due to unknown male
	154	149_154	9	excluded due to small size
	158	149_158	4	excluded due to small size

Table 3.2. Filtering and quality assessment of SNPs produced using DARTQC. Rep average is a metric specific to DArT (Diversity Arrays) specifying the proportion of alleles that give repeatable results over 30 replicates.

<b>FILTER</b>	<b>PARAMETERS</b>	<b>SILENCED SAMPLES</b>	<b>SILENCED SNPS</b>	<b>SILENCED CALLS</b>	<b>CHANGED CALLS</b>
Read Counts	5	0	0	4510437	0
Min SNP Call Rate	0.7	0	17542	0	0
Count Comparison	['0.05', '0.1']	0	0	11462	4353
DART SNP Metric	('RepAvg', '<', 0.9)	0	0	0	0
Cluster	0.95	0	1166	0	0
MAF	['0.02']	0	1460	0	0
<b>TOTAL SILENCED</b>		<b>0</b>	<b>20168</b>	<b>4521899</b>	<b>4353</b>
<b>TOTAL IN DATASET</b>		<b>548</b>	<b>24199</b>	<b>13261052</b>	<b>13261052</b>

Table 3.3. SNPs removed through filtering in ten *Haplochlaena maculosa* families using PLINK.

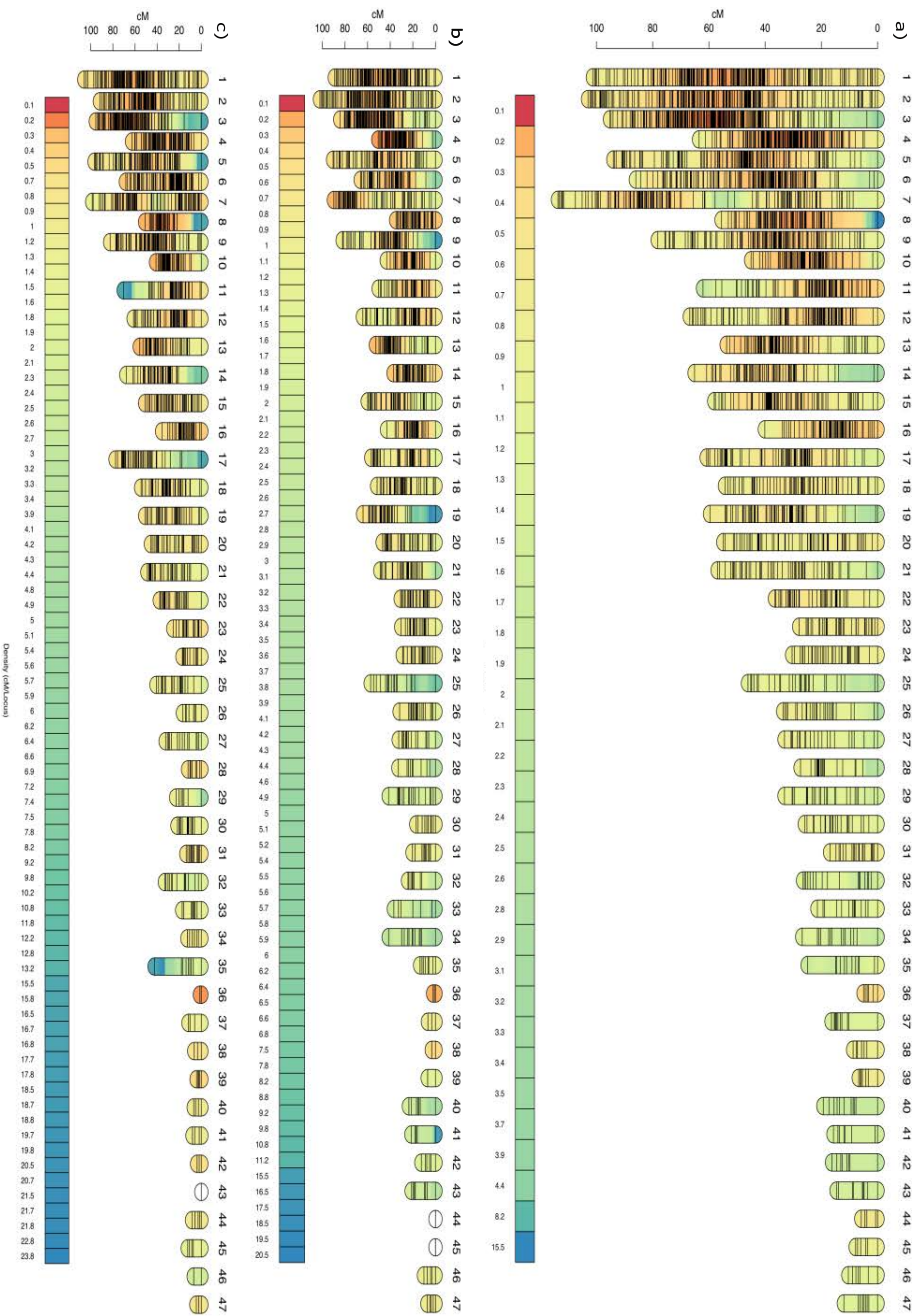
<b>Family</b>	<b>SNPs Post DARTQC filtering</b>	<b>family size</b>	<b>Missing genotype threshold (0.1)</b>	<b>SNPs removed (MAF 0.05)</b>	<b>Mendel errors excluded (SNPs 0.1, individuals 0.1)</b>	<b>individuals retained</b>	<b>SNPs retained</b>
F131_M127	7240	37	1656	3370	289	33	1925
F132_M142	7240	42	1275	3529	275	42	2161
F133_M132	7240	24	2331	2897	305	24	1707
F136_M134	7240	28	2476	3256	184	28	1324
F136_M145	7240	23	2828	2606	315	23	1491
F138_M133	7240	36	1126	3680	426	33	2008
F140_M149	7240	14	2198	2758	0	11	2284
F140_M155	7240	27	1987	3352	299	27	1602
F144_M144	7240	36	1125	3722	363	36	2030
F145_M148	7240	26	1615	3481	395	26	1749

### 3.4.2. Linkage map construction and genome coverage

A high-density sex averaged map was generated using 4,621 informative loci across 10 families and resulted in 47 linkage groups (LGs). Total map length was 2016.62cM with LGs ranging between 4.97cM (LG 36) and 113.68cM (LG 7) and averaging to 42.9cM. Intervals between markers averaged to 0.85cM (0.44cM, 0cM intervals inclusive) with the estimated chromosome number for *H. maculosa*, based on other octopods being 30 (Wang and Zheng, 2017). Markers per linkage group ranged between 370 (LG 1) to 10 (LG 47), with 20 LGs containing > 80 markers. Sex specific maps were relatively shorter with maternal and paternal maps totalling 1860.89cM and 1832.08cM in length respectively. Intervals between markers were slightly greater in the maternal map compared to the paternal map. Paternal maps exhibited an average interval of 1.38cM (0.53, 0cM intervals inclusive) while maternal maps were on average 1.46cM (0.51, 0cM intervals inclusive). More markers were found to be informative in the maternal (3,681) map relative to the paternal (3,478). Overall, linkage group length ranged from <1cM (LG 44 & 45) to 102.5cM (LG 2), and from <1cM (LG 43) - 105.93 (LG 1) for maternal and paternal maps respectively. Linkage group lengths between maternal and paternal maps were highly comparable. Genome coverage for sex average, maternal and paternal maps was estimated to be 96.4%, 94% and 97.2% respectively.



1



2

3

4 Figure 3.1. Genetic linkage map generated using LEPMAP3 for *H. maculosa* a) sex averaged map b) maternal map c) paternal map.

5 Marker density (cM/Locus) is visualised in a scale from red (high) to blue (low).

Table 3.4. Statistical summary of *Haplochromis maculosa* sex averaged, male and female linkage maps.

LG	Total Length (cM)			Number of Markers			Average Interval (cM)		
	Sex average	Female	Male	Sex average	Female	Male	Sex average	Female	Male
1	101.31	88.98	105.93	370	305	279	0.27	0.29	0.38
2	102.98	102.05	91.81	356	289	269	0.29	0.35	0.34
3	95.24	84.20	95.51	354	288	272	0.27	0.29	0.35
4	63.49	50.57	62.69	287	244	215	0.22	0.21	0.29
5	93.96	90.51	96.63	257	210	187	0.37	0.43	0.52
6	86.02	66.12	68.29	254	189	198	0.34	0.35	0.35
7	113.68	89.83	98.66	243	196	189	0.47	0.46	0.52
8	55.50	34.80	50.94	234	150	194	0.24	0.23	0.26
9	78.21	82.11	82.55	208	149	165	0.38	0.55	0.50
10	44.96	43.07	40.95	184	154	129	0.25	0.28	0.32
11	62.05	50.22	70.16	163	136	105	0.38	0.37	0.67
12	66.77	64.19	61.34	161	120	126	0.42	0.54	0.49
13	53.70	52.67	55.69	138	111	117	0.39	0.48	0.48
14	65.09	37.14	67.88	126	115	81	0.52	0.33	0.85
15	58.09	59.86	50.87	125	98	92	0.47	0.62	0.56
16	40.05	42.89	35.54	125	99	92	0.32	0.44	0.39
17	60.77	56.83	77.62	113	93	79	0.54	0.62	1.00
18	54.33	51.58	54.52	94	74	74	0.58	0.71	0.75
19	59.61	63.96	50.76	91	71	73	0.66	0.91	0.70
20	54.87	46.72	45.57	82	65	63	0.68	0.73	0.74
21	56.85	48.76	48.77	73	57	51	0.79	0.87	0.98

22	36.55	30.84	37.58	66	52	51	0.56	0.60	0.75
23	27.91	30.26	25.38	49	37	40	0.58	0.84	0.65
24	30.36	29.02	17.11	43	38	25	0.72	0.78	0.71
25	46.10	57.29	40.64	38	26	36	1.25	2.29	1.16
26	33.59	31.70	16.67	35	34	16	0.99	0.96	1.11
27	33.13	32.50	32.35	34	28	26	1.00	1.20	1.29
28	27.44	32.76	11.96	28	22	22	1.02	1.56	0.57
29	33.12	41.40	22.82	28	24	17	1.23	1.80	1.43
30	25.93	16.98	21.63	27	20	18	1.00	0.89	1.27
31	16.90	20.27	13.62	26	23	22	0.68	0.92	0.65
32	26.47	24.32	32.99	23	19	21	1.20	1.35	1.65
33	21.44	36.84	17.30	20	17	15	1.13	2.30	1.24
34	26.78	41.37	12.59	18	16	15	1.58	2.76	0.90
35	24.86	13.57	42.30	16	12	12	1.66	1.23	3.85
36	4.97	2.13	1.32	14	13	8	0.38	0.18	0.19
37	16.39	6.67	11.58	13	8	10	1.37	0.95	1.29
38	8.71	3.34	6.38	12	9	10	0.79	0.42	0.71
39	6.74	6.91	4.31	11	5	10	0.67	1.73	0.48
40	19.26	23.48	6.94	11	10	8	1.93	2.61	0.99
41	15.74	21.25	8.01	11	9	7	1.57	2.66	1.33
42	16.22	12.45	3.85	10	8	7	1.80	1.78	0.64
43	14.57	21.06	0	10	10	3	1.62	2.34	0.00
44	5.79	0	8.41	10	2	10	0.64	0.00	0.93
45	7.81	0	12.28	10	7	9	0.87	0.00	1.53
46	10.37	10.17	6.91	10	10	3	1.15	1.13	3.45
47	11.96	7.24	4.55	10	9	7	1.33	0.91	0.76

Table 3.5. Genome coverage estimation of sex average, female and male linkage maps.

Method	Map	Ge result	Average between methods Ge1+Ge2/2	% coverage
Ge1 = LG length * (marker number + 1/ marker number - 1)	Sex average	2091.75	2091.75	96.41
Ge2 = LG length +(2*average interval)	Sex average	2091.75		
Ge1 = LG length * (marker number + 1/ marker number - 1)	Male	1914.07	1914.06	97.22
Ge2 = LG length +(2* average interval)	Male	1914.06		
Ge1 = LG length * (marker number + 1/ marker number - 1)	Female	1949.44	1949.44	93.98
Ge2 = LG length +(2* average interval)	Female	1949.44		

### 3.4.3. Genome mapping and scaffold reorientation

Consolidation of genetic linkage and sequence maps was achieved using 1,278 markers across 47 LGs which successfully arranged 1,151 scaffolds covering 34% (1.39GB) of the sequenced genome for *H. maculosa*. Orientation was possible in 105 scaffolds with >1 SNPs mapped, the remaining 1,239 scaffolds with a single mapped marker were able to be placed within a linkage group. Splits occurred in 10 genomic scaffolds to maintain congruence of linkage map marker order, which was given precedence over genomic contig order, additionally 286 inconsistent markers were excluded.

Pseudo-chromosome lengths varied greatly with the largest scaffold LG 3 (150Mb) dwarfing the smallest LG39 (6,049bp), with an average length of 29.58Mb. Promising scaffolds containing a minimum of one mapped marker, were absent from the linkage map and therefore excluded (n=311). Conflicting markers were identified in 203 scaffolds indicating possible assembly errors within the genome. A total of 5,876 genes were annotated across all pseudo-chromosomes.

The most common Pfam domain within the gene set was cadherins, which were present in 53 genes (Supplementary table 4). The new assembly resulted in an overall reduction in the number of scaffolds from 48K to 47K and the largest scaffold increased from 11Mb in the original assembly to 150Mb. A summary of chromonomer results is available at this link (<http://203.101.230.130/chromonomer/index.php?v=hmac2020>).

Table 3.6. Characteristics and summary of *Haplochromis maculosa* Chromonomer assembly

LG	Length (Mb)	Markers per Scaffold			Total	Number of genes
		1	2	> 2		
1	112.76	95	7	4	106	444
2	126.98	91	11	2	104	609
3	150.54	101	12	2	115	606
4	79.88	75	12		87	496
5	70.24	71	4		75	255
6	82.68	66	6	2	74	296
7	80.66	74	2		76	366
8	78.33	64	4		68	312
9	64.34	49	4		53	198
10	61.35	57	3		60	271
11	67.60	39	2	3	44	181
12	48.97	49	4	1	54	252
13	40.98	31	3		34	79

14	39.80	30	2		32	80
15	24.50	33	2		35	192
16	76.57	42	2	2	46	246
17	35.90	27			27	207
18	10.72	24	1		25	106
19	21.35	30	1		31	78
20	20.09	32			32	70
21	8.54	17			17	31
22	13.18	23			23	114
23	5.68	11	1		12	29
24	2.39	9			9	17
25	0.63	7			7	6
26	9.91	6	2		8	18
27	9.52	6	1	1	8	66
28	6.48	7			7	8
29	3.45	7			7	30
30	2.67	7	1		8	50
31	5.07	10			10	8
32	0.72	7			7	7
33	1.91	6			6	27
34	1.71	5			5	10
35	1.23	3			3	12
36	0.91	2			2	12
37	0.59	5			5	10
38	4.76	1		1	2	27
39	0.01	1			1	NA

<b>40</b>	0.76	2			2	4
<b>41</b>	3.33	2			2	7
<b>42</b>	4.96	4			4	0
<b>43</b>	0.01	1			1	0
<b>44</b>	3.64	2			2	1
<b>45</b>	1.87	4			4	21
<b>46</b>	0.33	1			1	5
<b>47</b>	1.68	3			3	12
<b>Total</b>	<b>1,390</b>	<b>1239</b>	<b>87</b>	<b>18</b>	<b>1344</b>	<b>5876</b>

Table 3.7. Assembly statistics for the original *Haplochromis maculosa* assembly and the new chromonomer generated assembly.

<b>Metrics</b>	<b>Original genome assembly</b>	<b>Chromonomer assembly</b>
<b>total length (Mb)</b>	4009.60	4009.63
<b>number of Scaffolds</b>	48284.00	47190.00
<b>largest scaffold (Mb)</b>	11.01	150.55
<b>Average scaffold length (Mb)</b>	0.08	0.08
<b>N50 (Mb)</b>	0.93	1.25
<b>N60 (Mb)</b>	0.65	0.79
<b>N70 (Mb)</b>	0.44	0.49
<b>N80 (Mb)</b>	0.27	0.28
<b>N90 (Mb)</b>	0.12	0.13
<b>N100 (Mb)</b>	0.01	0.01

<b>N count</b>	1,574,315,944	1,574,344,251
<b>Gaps</b>	4,318,181	4,319,275

#### 3.4.4. Genes of interest (HOX)

HOX genes, which are integral to development, and are conserved throughout metazoans, were identified from the annotated *H. maculosa* genome and located within the updated assembly. Of the eight HOX genes identified within octopods six were present in *H. maculosa* assembly (HOX1/LAB, SCR, LOX2, LOX4, LOX5 and POST1). All HOX genes occurred on separate scaffolds within the original assembly (Supplementary table 6), however the updated assembly was able to place all three genes within LG 9 (SCR, LOX4 and POST1). The remaining three genes (HOX1/LAB, LOX2 and LOX5) were not present within a linkage group and were located on separate scaffolds (Figure 3.2). A total of four HOX genes (HOX1/LAB, LOX2, LOX5 and POST1) were also identified from the chromosomal level genome assembly of *O. sinensis* (Li et al., 2020). HOX genes identified in *O. sinensis* could not corroborate the arrangement observed in *H. maculosa* LG9 as only one of the three genes (POST1) was identified. Phylogenetic analysis of HOX genes identified from five cephalopod species, three octopods (*O. bimaculoides*, *O. sinensis* and *H. maculosa*), two bobtail squids (*Euprymna scolopes* and *Euprymna tasmanica*) and a cuttlefish (*Sepia officinalis*) displayed high bootstrap support (> 70) (Hillis and Bull, 1993) for all the gene clades observed (Figure 3.3) in accordance with findings by Hills and Bull, 1993.



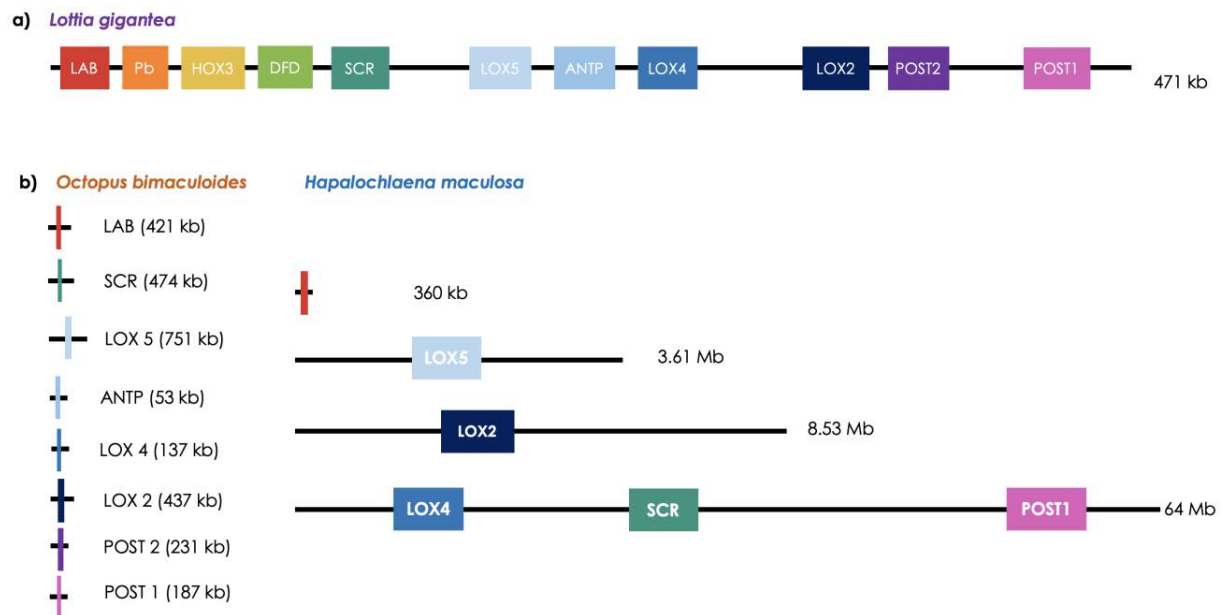


Figure 3.2. Comparison of HOX gene arrangement in the a) owl limpet (*Lottia gigantea*), b) California two spot octopus (*Octopus bimaculoides*) and the southern blue-ringed octopus (*Hapalochlaena maculosa*). Genes are coloured consistently between species. Scaffold lengths are not to scale.

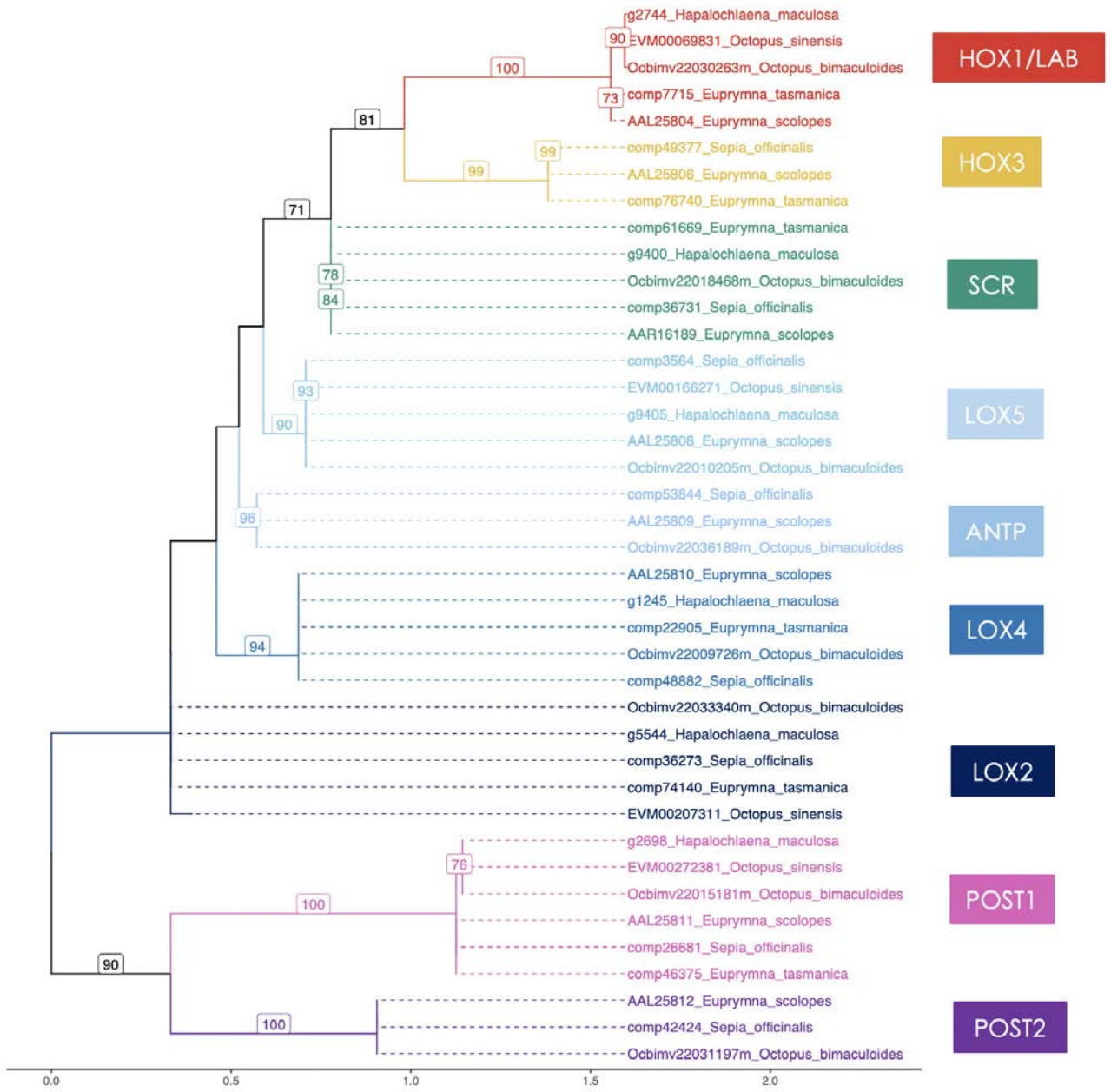


Figure 3.3. Maximum likelihood phylogeny of the HOX genes (aa) in six cephalopods. Bootstrap values >70 present on branches. Taxa are coloured in accordance with HOX genes.

## 3.5 DISCUSSION

This study presents a high-density genetic linkage map for the southern blue-ringed octopus (*Hapalochlaena maculosa*), the first generated for any cephalopod. Integration of the genetic linkage maps with the current genome assembly (Whitelaw et al., 2020) facilitated consolidation and reorientation of scaffolds into a more contiguous assembly. The improved assembly provided new insights into genome evolution of cephalopods, revealing the genomic placement of several key developmental HOX genes, which has previously not been possible in any other cephalopod genome assembly to date. Genetic resources generated in this study pave the way for future studies examining genomic and functional evolution across cephalopods.

### 3.5.1 Paternity and linkage map generation

Robust linkage map generation is highly dependent on the number of families, adequate family sizes and reliable pedigree reconstruction (Semagn et al., 2006). Larger families present greater opportunities to detect informative meiotic events, which are crucial in linkage analysis. Linkage maps are generated by calculating recombination events between loci and are thus dependant on the number of recombination events observed for the statistical strength of these calculations. The reduced family sizes present in this result in subsequently fewer recombination events observed between loci. This demonstrates some of the challenges associated with using wild caught octopods for linkage map generation. Even though mapping can be conducted across small families, it is estimated that a minimum of 200 individuals was ideal for map generation in several simulated data sets (recombinant inbred lines, F2 with

dominant and co-dominant markers, double haploid and backcrossing) (Ferreira et al., 2006).

However, standardization of linkage map generation is difficult due to the differences in genome structure (i.e. size and number of chromosomes), recombination rates, and variation of life history traits between organisms which impacts the size and structure of families available to be sampled (Kawakami et al., 2014). The data presented here provided the best genetic linkage map available to date in cephalopods, despite the small relative clutch sizes.

This study utilised data generated from a complimentary multiple paternity study on *H. maculosa* with parental samples sourced from the wild (Cockburn Sound, Western Australia) (Morse et al., 2018a), while this was the most successful captive breeding attempt to date in this genus and rendered our current work possible, the data was not without limitations. Before capture, and during the multiple paternity study, females were exposed to multiple males resulting in observed cases of multiple paternity. All females exhibited multiple paternity, including five unknown males (presumably from matings in the wild prior to the females capture). As a result, family sizes were reduced as female clutches were divided among the multiple males, which resulted in the loss of nine families, each containing < 10 individuals, in addition to the five families containing unknown paternity.

*Hapalochlaena maculosa* are semelparous and produce a single relatively small clutch (30-170) of large eggs (6-7mm) (Morse et al., 2018a; Tranter and Augustine, 1973), which they carry within their web. On hatching offspring are well-developed and assume a benthic lifestyle immediately. Other members of the genus such as *H. lunulata* exhibit a merobenthic lifestyle

and produce a larger quantity (270) of smaller eggs (3.5mm) (Overath and von Boletzky, 1974; Williams et al., 2011), which may make them more suitable for linkage mapping purposes. Due to the smaller family sizes used in this study resulting from limitations in clutch size and multiple paternity, a linkage mapping algorithm was selected, that was able to use a combined dataset integrating all families to maximise power in linkage group determination and marker order. Lep-Map3 was employed in this study because it allows not only for the integration of all families as one dataset, but also takes into account sibship between offspring (Rastas, 2017). This was of particular importance for this study since all *H. maculosa* females produced clutches with multiple paternity. To ensure quality of families analysed, paternity was verified using multiple methods and loci screened for Mendelian errors prior to map construction.

A total of 47 linkage groups were resolved using 4,621 markers with a sex average map length of 2,016cM (average interval length of 0.85). Map length in *H. maculosa* is larger than many other published molluscan linkage maps including of the bivalve *Pinctada maxima* (831.7cM) (Jones et al., 2013) and the gastropod *Biomphalaria glabrata* (746.7cM) (Adema et al., 2017) in line with their relatively smaller genomes 290Mb (*Neomenia per magna*) - 2.21Gb (*Bathymodiolus plantifrons*) (Takeuchi, 2017) compared to *H. maculosa* (4Gb). This suggests a similar recombination rate between molluscan lineages. Modelling conducted by Stapley et al. (2017) demonstrated a negative relationship between genome size and recombination rate within plant genomes. However, animals and fungi did not exhibit a reduction in recombination rate with an increase in genome size (Stapley et al., 2017). Genomic features may also impact recombination rates with the density of repetitive elements correlated negatively, while gene

density exhibits a positive relationship with recombination rate (Bartolomé et al., 2002; Boissinot et al., 2001; Fu et al., 2002; Tiley and Burleigh, 2015). Fine scale analysis of recombination rates along linkage groups was not possible with the current linkage map of *H. maculosa* generated in this study. However, future linkage maps could be utilized to examine patterns of recombination within the large and repeat-rich genomes characteristic of cephalopods.

### 3.5.2 Cephalopod genome evolution and structure

Prior to this study, no genetic linkage maps have been produced for any cephalopod, however, karyological studies conducted on a subset of 11 cephalopods estimate chromosome numbers for octopodiformes (n30), decapodiformes (n46) and nautiloids (n26) (Gao and Natsukari, 1990; Nakamura, 1985; Wang and Zheng, 2017). Structural variation was present between octopod karyotypes with the most divergent being *Amphioctopus fangsiao* (d'Obigny, 1839-1841), which exhibited no sub-telocentric chromosomes (Wang and Zheng, 2017). In composition, metacentric, submetacentric, telocentric and sub-telocentric chromosomes displayed slight variation between *Callistoctopus minor* (Sasaki, 1920) and *Cistopus chinensis* (Zheng et al., 2012). Karyograms revealed evolutionary distances between species to be congruent with molecular phylogenies conducted on the three octopods, *C. minor*, *A. fangsiao* and *C. chinensis* (Wang and Zheng, 2017). Linkage groups generated in this study do not correspond 1:1 with the expected chromosome number for *H. maculosa* suggesting the need for either a larger number of families, a larger number of progeny per family, or a larger set of informative markers to allow for chromosome level resolution and improved contiguity.

Cephalopods are characterised by their large and repetitive and heterozygous genomes ranging between 3Gb (*O. bimaculoides*) and 5Gb (*C. minor*) with 46% and 44% repetitive element composition (Albertin et al., 2015; Kim et al., 2018). These characteristics are compounded by the use of short read sequencing techniques to generate many older genomes (Sanger, Illumina) often resulting in difficult to assemble and highly fragmented genomes (Takeuchi, 2017). Long-read sequencing technologies such as Pac-Bio and Oxford Nanopore have improved the quality and contiguity of genomes, however factors such as cost and DNA quality can be prohibitive in some cases (Amarasinghe et al., 2020). The *H. maculosa* genome published recently was highly fragmented and was composed of 48K scaffolds (Whitelaw et al., 2020). After integration with the linkage map generated in this study the genome was reassembled into 47 pseudo-chromosomes covering 1.3 Gb and integrating 34.6% of the total genome. It should be noted that pseudo-chromosomes as used here denotes the closest to a chromosome assembly currently produced for a member of this genus. This improved genome assembly could be applied to examine evolutionary history of genomes, in addition to assisting in the generation of future octopod genomes (Simakov et al., 2020). Fragmented non-model genomes greatly benefit from complementary linkage maps, which allow for reorientation and placement of scaffolds into a chromosome level assembly (Fierst, 2015). Because linkage maps are independent of the primary assembly they can be used for continuous improvement of genomes as new primary assemblies become available (Hedgecock et al., 2015). This issue is highlighted by the results of this study where our highly fragmented primary assembly for *H. maculosa* meant that comparatively few scaffolds contained SNPs that could be incorporated

into the linkage map. Future improvements in the primary assembly (e.g. by long-read sequencing) could make use of the linkage map presented here to achieve a much more contiguous final result.

### 3.5.3 Evolution of the HOX gene cluster in cephalopods

Due to the improved contiguity of the *H. maculosa* genome, six of the eight expected genes from the HOX cluster were placed into a greater genomic context than has previously been possible in cephalopods. The HOX gene cluster forms a highly conserved set of developmental genes in metazoans (Duboule, 2007). A single set of nine genes was first observed in *E. scolopes*, and corroborated in the *O. bimaculoides* genome as a coleoid specific trait (Albertin et al., 2015). HOX genes in *E. scolopes* were assumed to form a cluster based on the pattern of expression along the central nervous system (CNS), which was found to be congruent with the ancestral role of axial patterning (Lee et al., 2003). Examination of homologous genes in *O. bimaculoides* revealed rearrangements of HOX genes inconsistent with other classes present within Mollusca as each gene was located on a separate scaffold (Albertin et al., 2015). However, greater resolution of each genes placement in relation to each other was not possible due to the fragmentation of the genome (Albertin et al., 2015). The improved assembly for *H. maculosa* generated in this study was able to place three HOX genes within the same scaffold (LOX4, SCR and POST1). Genomic placement of HOX genes within the chromosome level assembly of *O. sinensis* could not support or contradict the co-location of genes within *H. maculosa* as only a single gene (POST1) could be identified. Implications behind HOX gene placement in *H. maculosa* are difficult to infer without expression studies.



Unfortunately, patterns of HOX gene expression have not yet been examined in octopods and the implications of these gene placement rearrangements in *H. maculosa* remain unknown.

#### 3.5.4 QTL mapping and future work

Linkage maps provide a framework for understanding the evolution and inheritance of particular phenotypic traits through qualitative trait loci (QTL) studies (Shaw and Lesnick, 2009). Historically, such studies have been used to improve desirable traits in agricultural and aquacultural species including disease resistance (Han et al., 2018), fibre quality (cotton) (Zhang et al., 2009) and growth rate (Asian seabass) (Wang et al., 2019), to name a few. Aquacultural significance may not be applicable for *Hapalochlaena* directly, however an increasing number of octopod species are being raised for human consumption with 745,054 tonnes produced in 2003 within Japan alone (Berger, 2010). *Octopus vulgaris* is a prime target for aquacultural production due their relatively large clutch sizes and growth rates (Vaz-Pires et al., 2004). The *H. maculosa* linkage map produced in this study may aid in the prediction and construction of similar maps for related octopods, in addition to supporting the construction of additional octopod genomes. A similar application has been conducted when the draft chicken genome was used to predict the linkage map of passerine birds, which was then verified using the more closely related reed warbler (*Acrocephalus arundinaceus*) linkage map (Dawson et al., 2006). Despite the divergence times between the Galliformes (chicken) and Passeriformes lineages (~80-100mya) (Shetty et al., 1999; Van Tuinen and Hedges, 2001) sufficient synteny and microsatellite markers were conserved to facilitate map linkage map construction (Dawson et al., 2006). Furthermore, linkage maps and resultant QTL studies have applications in unravelling

the evolution of complex multigenic traits associated with adaptation and speciation (Price et al., 2018; Strasburg et al., 2012). A link was observed between QTL associated with male song and female song preference for two cricket species *Laupala kohalensis* and *Laupala paranigra* (Shaw and Lesnick, 2009). While the genes underlying the traits have yet to be identified, analysis of QTL allows for investigation of evolutionarily relevant traits and provides a guide for future genomic work (Shaw and Lesnick, 2009).

### **3.6 CONCLUSION**

This study successfully fulfilled the primary aim by producing the first linkage map of a cephalopod, the southern blue-ringed octopus (*Hapalochlaena maculosa*). Integration of the linkage map with the current genome assembly reduced fragmentation and enabled the placement of three HOX genes within the same linkage group providing a greater genomic context for this gene family compared to currently available cephalopod genomes. The linkage map produced in this study will provide a valuable resource for the generation of *Hapalochlaena* genomes while also providing a framework for understanding the inheritance of phenotypic traits through future work.

## CHAPTER 4

# SNP data reveals the complex and diverse evolutionary history of the blue-ringed octopus genus (*Octopodidae: Hapalochlaena*) in the Asia-Pacific

Brooke L. Whitelaw, Julian Finn, Kyall Zenger, Ira R. Cooke, Jan M. Strugnell

### 4.1 ABSTRACT

The blue-ringed octopus complex (*Hapalochlaena*), known to occur from Southern Australia to Japan, currently contains four species (*H. maculosa*, *H. fasciata*, *H. lunulata* and *H. nierstraszi*). These species are often distinguished using morphological characters (iridescent blue rings and/or lines) along with reproductive strategies. To examine species boundaries within the genus we used mitochondrial (12S rRNA, 16S rRNA, cytochrome c oxidase subunit 1 [COI], cytochrome c oxidase subunit 3 [COIII] and cytochrome b [Cytb]) and SNP data from specimens collected across its geographic range including variations in depth from 3m to >100m. Our investigation indicates greater species diversity present within the genus *Hapalochlaena* than is currently recognised. We identified 10,346 SNPs across all locations, which when analysed support a minimum of 11 distinct clades. Bayesian phylogenetic analysis

of the mitochondrial COI gene on a more limited sample set dates the diversification of the genus to ~30mya and corroborates eight of the lineages indicated by the SNP analyses.

Furthermore, the diagnostic lined patterning of *H. fasciata* found in North Pacific waters and NSW, Australia is likely the result of convergent evolution based on genetic distance between the clades. Several “deep water” (> 100m) lineages were also identified in this study with genetic convergence likely to be driven by external selective pressures. Additional examination of morphological traits is required to describe additional species within the complex.

## 4.2 INTRODUCTION

The blue-ringed octopus genus (*Hapalochlaena*) currently comprises of three species that are universally accepted and have been adequately described (*H. maculosa*, *H. lunulata* and *H. fasciata*), with a fourth species for which very little information is known (*Hapalochlaena nierstrazi*, Adam 1938) (Norman et al., 2016) (Jereb et al., 2014). The genus *Hapalochlaena* is dispersed throughout the Asia Pacific from Temperate Australasia (TAUS) to the Temperate North Pacific (TAUS) (Figure 4.3c) and is easily identified by their distinct iridescent blue rings and/or lines, which advertise their toxicity to would-be predators in an aposematic manner (Norman 2000) (Figure 4.1). All members of the genus thus far studied contain the potent neurotoxin tetrodotoxin (TTX) within their venom and tissues and are the only octopods known to inflict a lethal bite to humans (Flachsenberger and Kerr, 1985; Jacups and Currie, 2008). Human fatalities from both accidental ingestion of tissues and envenomation have been recorded (White, 2018; Wu et al., 2014). It should be noted that out of 12 documented bites

between 1954-1995 only half required medical care with the remaining six resulting in no symptoms (White, 2018). Dry bites (i.e. bites without the injection of venom) are a common mechanism to conserve valuable resources in many venomous taxa (Pucca et al., 2020), however it is unknown if *Hapalochlaena* are utilising dry bites or if venom potency fluctuates between or within individuals. Studies quantifying tetrodotoxin within tissues found variation both within and between species (Williams and Caldwell, 2009). *Hapalochlaena lunulata* collected from the aquarium trade exhibited lower overall quantities within fewer tissues compared to wild caught *H. fasciata* with minimum total amounts of TTX of 0-174ug to 60-405ug respectively (Williams and Caldwell, 2009). *Hapalochlaena maculosa* collected from South Australia exhibited the lowest total quantities at TTX 2.40 - 6.63ug per individual (Yotsu-Yamashita et al., 2007). The most recent study examined *H. fasciata* from Japanese waters and found large variation in TTX concentrations between individuals which may be associated with seasonality, gender and age, indicating the need for further research (Yamate et al., 2021).

The genus *Hapalochlaena* was introduced by Robson in 1929. *Hapalochlaena lunulata*, the Greater Blue-ringed Octopus, previously named *Octopus lunulatus* by Quoy and Gaimard (1832), based on a specimen from Papua New Guinea, was designated the type species. Robson (1929) also included in this genus, *Octopus maculosus* Hoyle 1883 (now known as *H. maculosa*) from "Australia", and *Octopus pictus* var. *fasciata* Hoyle, 1886 (now known as *H. fasciata*) from Port Jackson NSW Australia. Key diagnostic features used to distinguish these species include morphological characters (size, shape, and placement of rings and/or lines and papillae) along with behavioural and life history traits. *Hapalochlaena lunulata* is the only described member of

the genus to exhibit a merobenthic lifestyle producing relatively small eggs (3.5 mm in diameter) with paralarva able to disperse using water currents (Overath and von Boletzky, 1974). Of the four species it has the largest described range (Central Indo Pacific) (Figure 4.1c). In contrast, both *H. maculosa* and *H. fasciata* are holobenthic, producing relatively larger eggs (7-9 mm and 6-9 mm in diameter respectively) (Norman et al., 2016; Reid, 2016), which hatch into crawl-away (benthic) young (Tranter and Augustine, 1973).

Norman (2000) highlighted the potential existence of five additional putative *Hapalochlaena* species based on morphological and life history traits. The proposed species *Hapalochlaena* sp. 1, had previously been confused with *H. lunulata* based on general appearance, however in contrast to *H. lunulata*, this species exhibits a holobenthic lifestyle, attains a larger size, has a more muscular form and occupies a different habitat. Additionally, the anti-tropical distribution of *H. fasciata* led to the suggestion that the disjunct 'populations' distributed in New South Wales (NSW), Australia and Japan may represent sister taxa and was therefore noted as requiring further evaluation (Norman, 2000; Norman and Hochberg, 2005). As part of an ongoing morphological study, Finn (2015) reported a total of 12 blue-ringed octopus species occurring in Australia (Finn, 2015). While Finn and Lu (2015) reported that the two species from the Indo-West Pacific, historically treated under the Australian names *H. maculosa* and *H. fasciata* were morphologically distinct from the Australian fauna (Finn and Lu, 2015).

Clarifying the systematics of this genus has been subject to many of the same challenges as other cephalopods, particularly octopods. Upon death cephalopods exhibit a loss of many physical structures and pigmentation (Bookstein, 1985); this, in combination with few hard body parts and distortion upon preservation (Voight, 1994), can lead to difficulty distinguishing between closely related species (Roper, 1983). Museum collections contain invaluable specimens from a variety of locations across time, however the method of preservation can vary greatly and thus impact species identification and re-evaluation (Roper, 1983).

The phylogenetic status of proposed species within *Hapalochlaena* remains unresolved due to limited sampling across the geographic range and/or resolution of markers assessed in previous molecular studies (Acosta-Jofré et al., 2012; Guzik et al., 2005; Lindgren and Anderson, 2018; Takumiya et al., 2005; Tanner et al., 2017). Phylogenetic analysis of partial sequences of two mitochondrial (cytochrome c oxidase subunit 3 [COIII] and cytochrome b [Cytb]) and one nuclear gene (eukaryotic translation elongation factor 1 alpha 1 [EF-1 $\alpha$ ]) of a single representative of each of *H. fasciata* (NSW), *H. maculosa* (Victoria) and *Hapalochlaena* sp. 1 (Darwin, Northern Territory NT), in conjunction with 23 other octopod species and four octopod genera supported the monophyletic status of *Hapalochlaena* and found it to be distinct from the genus *Octopus* Cuvier 1797 (Guzik et al., 2005). This corroborated previous work conducted based on morphological features (Norman, 2000; Roper, 1983). Whitelaw et al. (2020) presented the genome of *H. maculosa* and estimated the divergence of this species from *Octopus bimaculoides* occurred 59mya. Additional studies analysing partial COI and COIII gene sequences from 36 coleoid cephalopods, including representatives from each of the three

accepted *Hapalochlaena* species in Japan and adjacent waters supported the monophyletic status of the genus (Kaneko et al., 2011). The identification of three distinct specimens each designated to a different species within the *Hapalochlaena* genus in this geographical range clearly indicates that the species diversity in this genus greatly exceeds that presented in popular guides (e.g. Norman, 2000; Reid, 2016). A large-scale phylogenetic study using genome wide SNP markers has been conducted on *H. maculosa* which discovered a clinal species pattern across its geographical range from southern Western Australia to Victoria/ Tasmania (Morse et al., 2018b). While genetic distances between populations were insufficient to warrant species status, phylogenetic reconstructions including the sister taxon, *H. fasciata* demonstrated genetic distances between distal populations of *H. maculosa* were equivalent to distances between *H. maculosa* and *H. fasciata*. Nevertheless, no systematic revisions were suggested or implemented. The advent of next generation sequencing technology provides a means to examine the underlying structure and evolutionary history of this species complex (Gueidan et al., 2007). While still dependent on particular preservation methods, genomic techniques can be utilised to reconcile species complexes and to provide a greater understanding of their evolutionary history, population dynamics, and genetic divergence (Richards et al., 2009; Schwentner et al., 2011).

Two known fatalities from *Hapalochlaena* bites have occurred in Australia, however identification of species involved in these cases lacked clarity due to the lack of taxonomic resolution at the time and the fact that neither victim retained the octopus. The first recorded fatality from an octopus bite occurred in Darwin, Australia (1954). The original published



medical report incorrectly identified the specimen as *Octopus rugosus* based on an “identical” specimen collected from the same location (Willan and Willan, 2008). Later examination of the preserved specimen resulted in a reclassification to *Hapalochlaena* sp. 1. (Jacups and Currie, 2008). Likewise the southern blue-ringed octopus (*H. maculosa*) was identified as being responsible for the second fatality in NSW 1967 (Lane et al., 1967; White, 2018). While the specimen was not retained, and therefore cannot be re-examined, the identification of *H. maculosa* is likely to be erroneous, as the range of this species is not currently considered to extend to NSW. Due to the potential risk to human health it is crucial to have clear systematics, which allow for confident identification of individuals involved in both medically significant cases and in academic studies. Misidentification has become a common problem in many sequence databases due to the current state of *Hapalochlaena* systematics. Currently, the observed morphological diversity across their range from southern Australia to Japan is not reflected in their taxonomy (Finn, 2015; Finn and Lu, 2015; Norman, 2000; Norman and Hochberg, 2005) and often leads to misidentification in both academia and popular media. Further confusion is introduced through collection of specimens from markets where the collection location is unknown.

The relationship between the geographically distant *H. maculosa* and *H. fasciata* populations in Temperate Australasia (TAUS) and their counterparts in the Temperate North Pacific (TNP) have yet to be thoroughly investigated. To date there has been no comprehensive phylogenetic study of the genus *Hapalochlaena*. Here we utilise genome-wide single nucleotide polymorphisms (SNP) in conjunction with mitochondrial genes to examine the genetic

structure, diversification, and putative species boundaries of the *Hapalochlaena* genus. Dating the point of origin and the subsequent radiation of the genus forms an additional avenue of investigation. This chapter builds on the previous by providing evolutionary context for *H. maculosa*, which was the focus of whole genomic comparisons with other octopus genera in chapter 2 and linkage map generation in chapter 3. Here the radiation and evolution of the genus is elucidated with the aim to provide a genetic context for ongoing and future species delineation and descriptions.

## 4.3 METHODS

### 4.3.1. Sample collection

In order to assess species boundaries within *Hapalochlaena* samples were sourced, throughout their range. As part of an ongoing morphological examination of the group, Julian Finn collected, identified and tissue sampled suitable material of representative morphological OTUs across the range. Tissue samples were lodged in and obtained from the collections of Museums Victoria (NMV, n = 35), Western Australian Museum (WAM, n = 7), Museum and Art Gallery of the Northern Territory (MAGNT, n = 2), Queensland Museum (QM, n = 11) and Australian Museum (AM, n = 7). Additional samples collected by Peter Morse from Woodman's Jetty and Albany WA were also included (Morse et al., 2018b). Storage methods of samples differed across museums and included preservation in 90% ethanol and frozen and maintained at -80 C (Supplementary table 1).

#### *4.3.2. Mitochondrial DNA extraction and amplification*

DNA was extracted from muscle tissue (n = 62) using the “High salt method” (Donnan Laboratories 2001). PCR amplification of partial mitochondrial genes 12S ribosomal RNA (12S) (Simon et al., 1991), 16S ribosomal RNA (16S) (Simon et al., 1991), cytochrome c oxidase subunit 1 (COI) (Folmer et al., 1994), cytochrome c oxidase subunit 3 (COIII) (Guzik et al., 2005) and cytochrome b (Cytb) (Guzik et al., 2005). Reactions were composed of 0.5 µL forward primer (10 µM), 0.5 µL reverse primer (10 µM), 0.1 µL Taq (*Onetaq, New England Biolabs*), 2.5 µL 10 x buffer (Paq5000™), 2 µL dNTP mix (10 µM, *Bioline*), 17.4 µL ddH<sub>2</sub>O and 2 µL DNA (diluted to between 1–5 ng/µL), totaling to 25µL. Detailed reaction conditions can be found in Allcock et al. 2008 (Allcock et al., 2008). PCR products were sequenced by Macrogen Inc, Seoul, Korea.

#### *4.3.3. DNA extraction and Genotyping by Sequencing*

Complete genomic DNA was extracted from all 62 samples using the “QIAamp DNA Micro Kit (Cat No./ID: 56304)” (Qiagen) according to the manufacturer's protocol. Concentration and quality of DNA extracted was assessed by visualisation on a 0.8% agarose gel DNA quantification and using NanoDrop 3300™. DNA extracts were transferred to Diversity Arrays Technology PL, Canberra ACT, Australia who conducted their DArTseq™ pipeline (i.e. restriction enzyme digestion, library preparation, genotype by sequencing (GBS), data generation and QC).

#### *4.3.4. SNP identification*

SNP data returned by Diversity Arrays were imported into R and filtered by loci call rate (> 80%) (removes loci), individual call rate (> 70%) (removes individuals), repeatability (the proportion of identical genotypes between technical replicates) (0.9) and minor allele frequency (1%) in this order using the dartR package (Gruber et al., 2018). Monomorphic loci resulting from the removal of individuals and/or populations were also removed using the dartR package. Filtering parameters were selected for maximum data retention while minimising errors and missingness due to the age and storage history of museum samples.

#### *4.3.5. Genetic structure/divergence of SNP data*

Genetic clusters were investigated using Principal Coordinates Analysis (PCoA) with the R package dartR (Gruber et al., 2018). SNP clusters were assessed for AIC and BIC from K =1 to 25 using the snap.clust.choose.k function in Adegenet (Jombart, 2008). K was selected based on the lowest value from both criteria and used to examine admixture of populations in ADMIXTURE v1.3.0. STRUCTURE v2.3.4 (Rosenberg et al., 2001) was run for K values between 8-14 and evaluated using the Evanno method implemented in STRUCTURE HARVESTER v0.6.94 (Earl and Vonholdt, 2012). In addition, a Discriminate Analysis of Principal Components (DAPC) (Jombart and Collins, 2015) using the best K was conducted to identify clusters of samples based on genetic similarity. Genetic divergence between populations and putative taxonomic groups was assessed using Weir and Cockerham's unbiased F-statistics ( $F_{st}$ ). Pairwise  $F_{st}$  values and their significance were calculated between populations and putative species in R using the Stampp v1.6.1 package (Pembleton et al., 2013).

#### *4.3.6. Signatures of selection*

Loci under putative selection were identified using three methods (PCAdapt v4.3.3 (Luu et al., 2017), OutFLANK v0.2 (Whitlock and Lotterhos, 2015), and Bayescan v2.1 (Foll and Gaggiotti, 2008) to reduce methodological bias. The R package, PCAdapt, is able to detect population structure using an initial principal component analysis (PCA) followed by identification of specific markers highly correlated with populations. The analysis was conducted on the full set of 10,346 SNPs. In order to reduce the impact of linkage disequilibrium (LD) on the detection of population structure a further thinning step was conducted with a window size of 200 and an  $r^2$  of 0.1. A scree plot was used to choose the optimal number of principal components ( $K=6$ ) and SNPs were considered to be outliers (under selection) if they had a  $q$ value  $< 0.01$  (estimated false discovery rate of 1%). Bayescan and Outflank required predefined populations and were conducted in R on all populations grouped by location (Supplementary table 1). A further analysis using each method was conducted on populations grouped by location, where appropriate, to minimize the impact of low population sizes (Supplementary table 1). The results were then contrasted.

#### *4.3.7. Putative species boundary estimation using mitochondrial genes*

Genetic structuring of locations was examined using five mitochondrial genes (12S,  $n=17$ ; 16S,  $n=16$ ; COI,  $n=36$ ; COIII,  $n=18$  and Cytb,  $n=13$ ) (Supplementary figure 4). Samples were obtained throughout the Asia Pacific (this study; see Supplementary table 2). Additional sequences were obtained from NCBI (Supplementary table 2). Alignments for each gene were

generated using MAFFT v7.45 (Rozewicki et al., 2019) and a median joining network calculated in POPART v1.0 (Population Analysis with Reticulate Trees) (Leigh and Bryant, 2015).

Species boundaries were estimated using the PTP (Poisson Tree Processes) (Kapli et al., 2017) method on all genes as single-rate, multi-rate and Bayesian. Additionally, sGYMC (single-threshold general mixed yule coalescent) and ABGD (Automated Barcode Gap Discovery) were also run on all genes. All PTP analyses accepted a RAxML tree in Newick format generated in Geneious v10.2.6 from a MAFFT alignment. Both single-rate and multi-rate PTP were calculated using the online server (<https://mptp.h-its.org>) with a p value of 0.001 specified for single-rate PTP. Specifications for the bPTP run were as follows, a total of 500,000 MCMC generations were run with the first 10% discarded as burn-in and thinning occurring every 100 iterations. sGYMC (Pons et al., 2006) was run using a Bayesian inference tree generated by BEAST2 from the same alignment used in Geneious v10.2.6 to generate the input RAXML trees used in previous analyses.

#### *4.3.8. Divergence time estimation*

Divergence times were estimated using a concatenated alignment of four mitochondrial genes (12S, 16S, COI and COIII) (n = 7) in BEAST2 (Bouckaert et al., 2014) (Cytb was not included as too few sequences were available). The alignment was generated using MAFFT v 1.4 in Geneious v10.2.6 on a subset of *Hapalochlaena* taxa. BEAST2 was run using a calibrated Yule model, relaxed clock log normal for 50 million generations, burnin of 10% and sampled every 1000 iterations. Additional taxa were included as outgroups (*Argonauta nodosus* [NCBI:

MK034303], *Amphioctopus aegina* [NCBI: KX108697], *Amphioctopus fangsiao* [NCBI: AB240156], *Amphioctopus marginatus* [NCBI: KY646153], *Callistoctopus luteus* [NCBI: NC\_039848], *Callistoctopus minor* [NCBI: HQ638215.1], *Octopus bimaculoides* [NCBI: KU295559] and *Octopus vulgaris* [NC\_006353]). Previously estimated divergence times using whole mitochondrial genomes for the node *O. bimaculoides/O. vulgaris* and calibrations for the node *A. nodosus*/(all octopods) were included to estimate divergence times within the phylogenetic tree (Uribe and Zardoya, 2017). In order to include a larger number of samples collected from a greater range of locations the analysis was also conducted using a dataset composed of only the COI gene, for which more data was available. Sequences were also included from the NCBI database resulting in 25 additional samples (Supplementary table 3).

The coalescent based approach, SVDQuartets (Chifman and Kubatko, 2014) was used to generate species trees based on 10,346 unlinked SNP loci and 62 individuals. Populations were renamed using their putative species groups for analysis and input alignments for each method generated using the functions `gl2snapp` or `gl2svdquartets` from the `dartR` package (Gruber et al., 2018). SVDQuartets was run using Paup v 4.0a (Swofford, 2002) using the following parameters: 100,000, tips assigned to species, randomly sampled quartets, 10,000 bootstraps, multispecies coalescent tree model.

## 4.4 RESULTS

### 4.4.1. Species boundaries OTU identification

After removal of 1,644 SNPs identified as outliers by PCAdapt the remaining 8,702 neutral loci were used to examine population structure. Support from mitochondrial and SNP analyses indicated between 8-11 operational taxonomic units (OTU) were present in this data set. For the purpose of this study when discussing results samples were classified into 11 OTUs from A-K (Figure 4.2).

#### 4.4.1.1. Temperate Australasia (TAUS): OTU K

Individuals found across southern Australia from VIC to southern WA formed a single OTU (K).  $F_{st}$  values increase relative to distance, with populations at the extremities of the range showing the greatest genetic divergence (Supplementary table 3). PCoA and DAPC analyses showed no clear structure between populations across southern Australia, likewise STRUCTURE analyses could not distinguish between populations (Figure 4.2). High bootstrap support for the clade (100) was observed in the SVDQuartets phylogeny. Delimitation analysis (ABGD, sPTP and GYMC) for the partial mitochondrial gene COI classified populations across southern Australia (currently classified as *H. maculosa*) as a single unit (OTU K) across all methods with the exception of a relaxed clock GYMC, which identified three distinct genetic groups (VIC, SA and WA). It should be noted that short branch lengths observed between the divergence of Western and Eastern localities suggest a more recent divergence compared to other OTUs identified in this study.



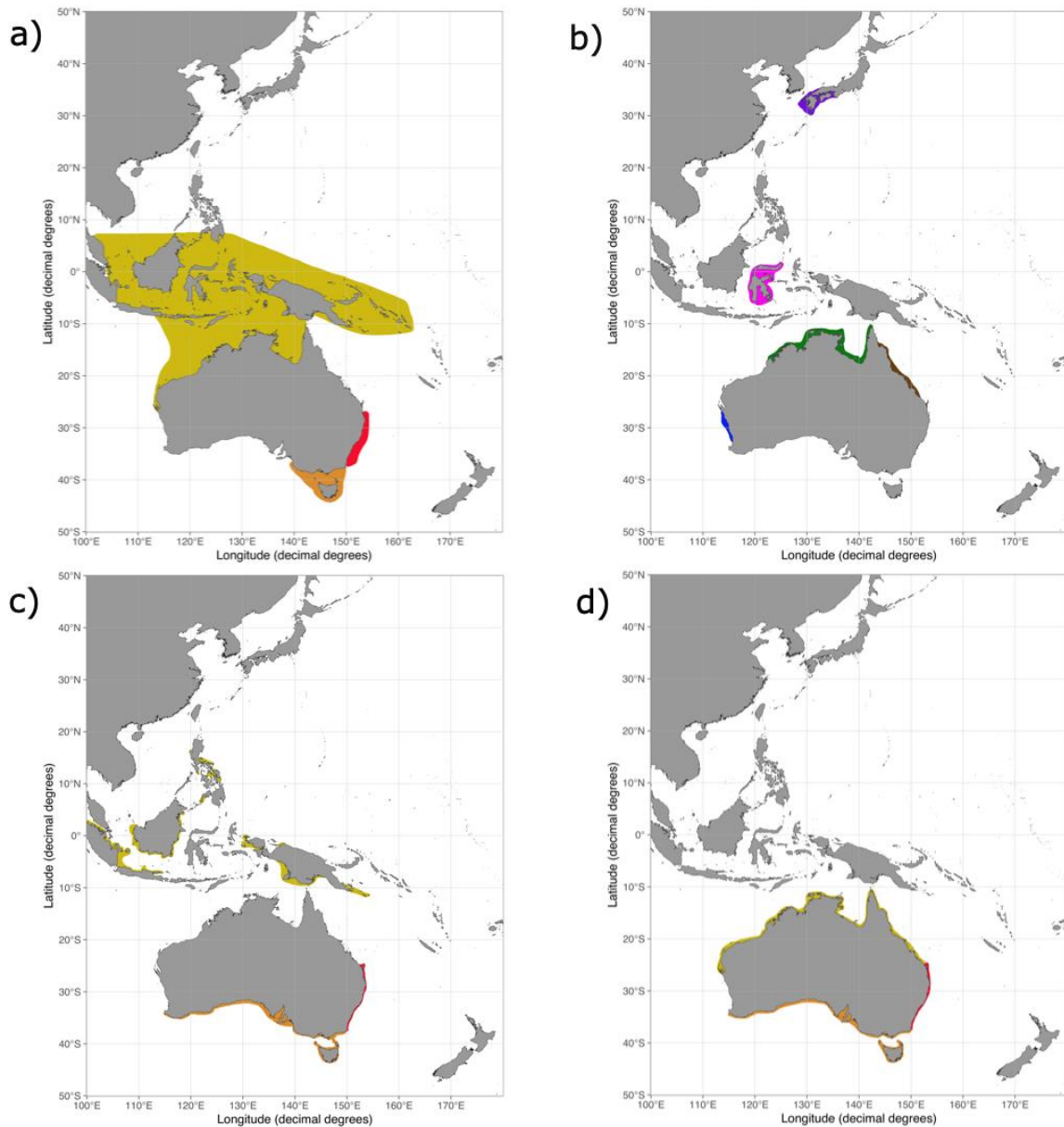
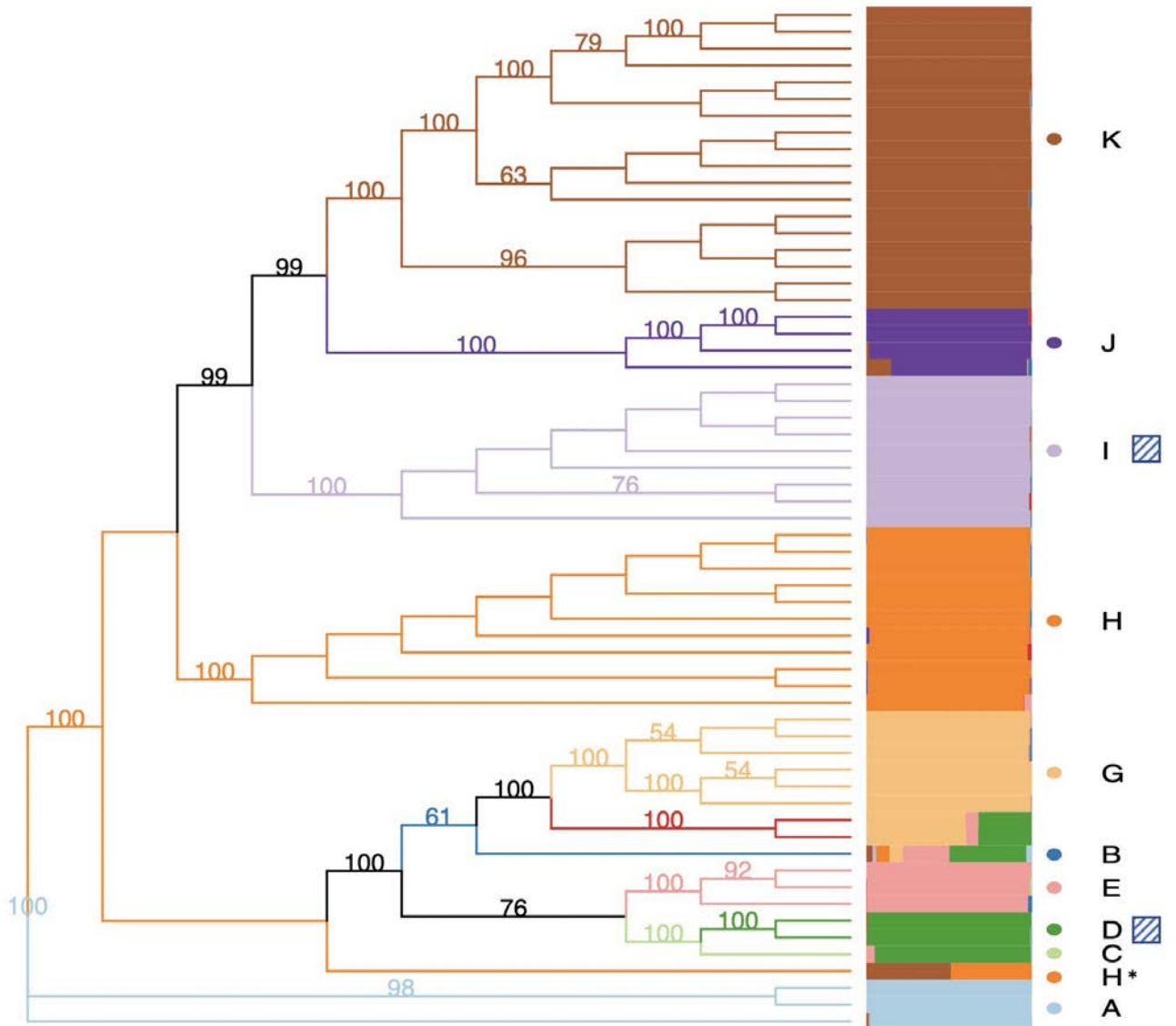


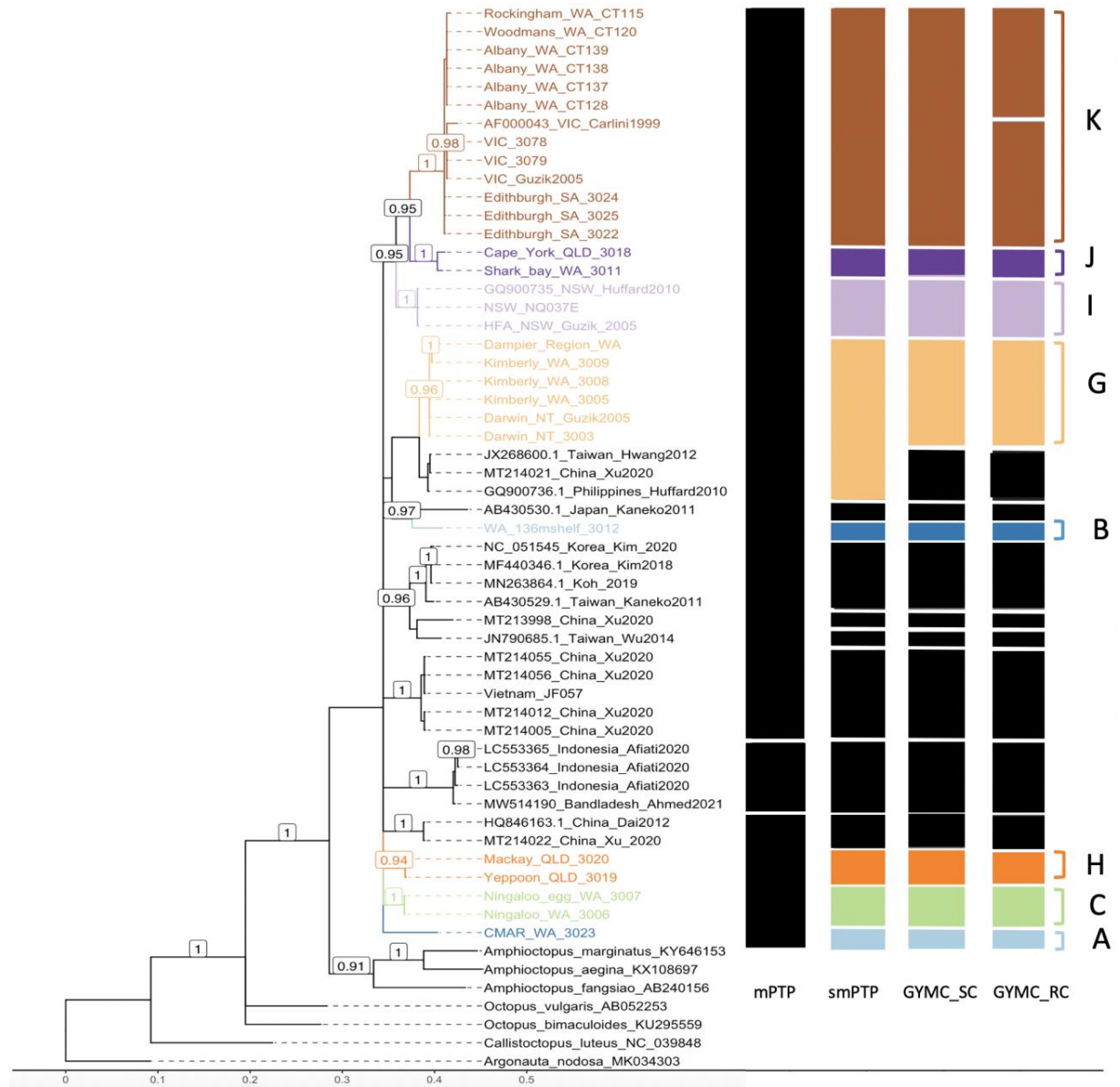
Figure 4.1. Published distributions of *Hapalochlaena* species. Distributions of *Hapalochlaena* sp. in the Indo Pacific. **a)** A guide to Squid, Cuttlefish and Octopuses of Australasia (Norman and Reid, 2000). *Hapalochlaena* are shown as follows, *H. maculosa* (orange), *H. fasciata* (red) and *H. lunulata* (mustard). **b)** Cephalopods a World Guide (Norman, 2000). Additional reported localities, *Hapalochlaena* sp. 1 (green), *Hapalochlaena* sp. 2 (purple), *Hapalochlaena* sp. 3 (blue), *Hapalochlaena* sp. 4 (pink) and *Hapalochlaena* sp. 5 (brown). **c)** FAO. Cephalopods of the world. Volume 3. Octopods and Vampire Squids (Norman et al., 2016). *Hapalochlaena* are shown as follows, *H. maculosa* (orange), *H. fasciata* (red) and *H. lunulata* (mustard). **d)**

Cephalopods of Australia and Sub-Antarctic Territories (Reid, 2016). *Hapalochlaena* are shown as follows, *H. maculosa* (orange), *H. fasciata* (red) and *H. cf. lunulata* (mustard).

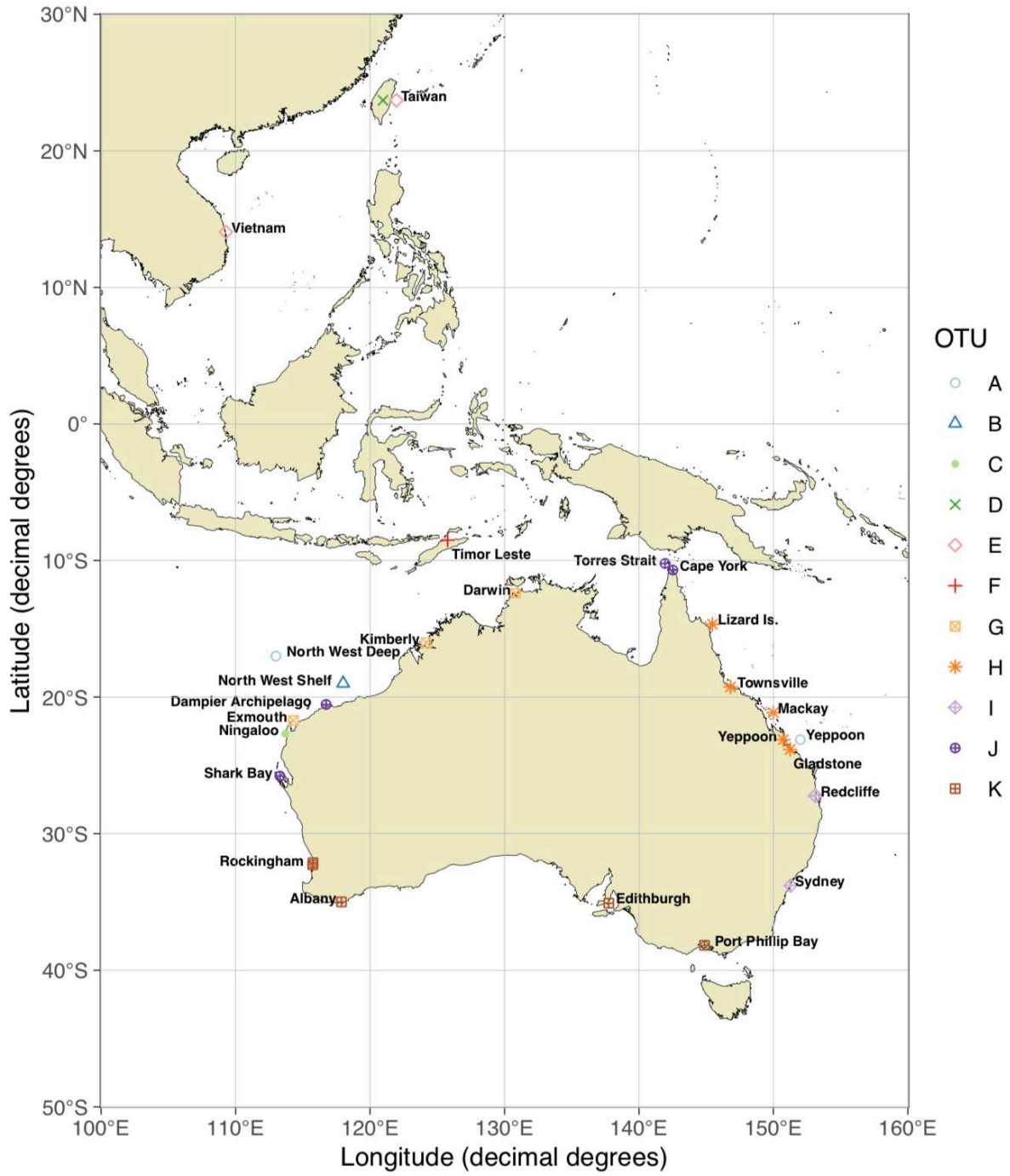
a)



b)



c)



d)

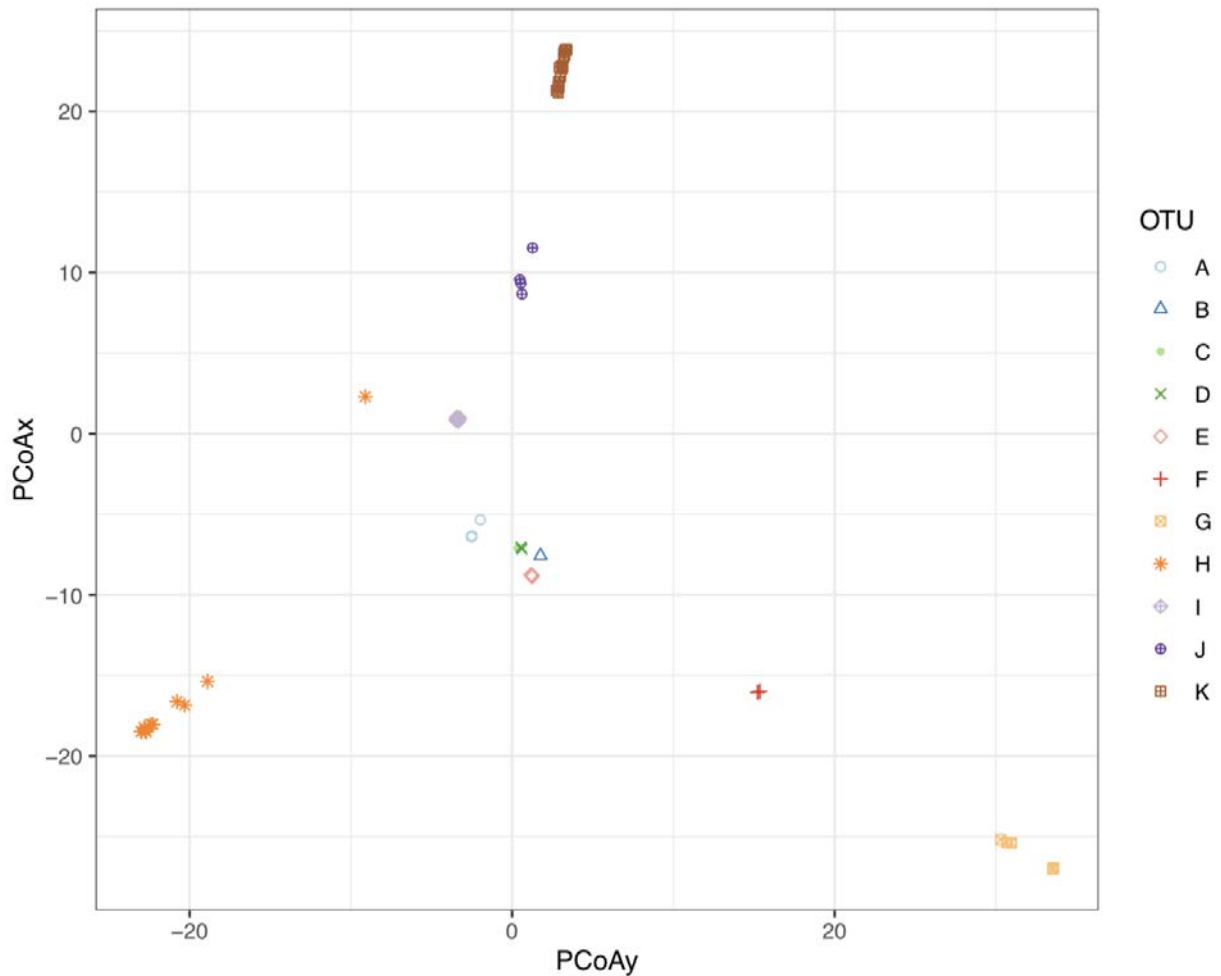


Figure 4.2. Delineation of *Haplochlana* species boundaries and genetic structure throughout the Indo Pacific using 10,346 SNPs: a) SVDQuartet phylogeny of *Haplochlana* throughout the Indo Pacific generated using 10,346 SNPs, coloured branches represent putative taxonomic units A-K: light blue (A/Deep water Yeppoon, QLD & North West, WA), dark blue (B/North West Shelf, WA), light green (C/Ningaloo, WA), apple green (D/Taiwan lined), pink (E/Taiwan ringed), red (F/Timor Leste), light orange (G/Great Barrier Reef, QLD), dark orange (H/H\*/Darwin, NT & Kimberly & Exmouth, WA ), lilac (I/NSW), purple (J/Cape York, QLD & Shark Bay, QLD) and

brown (K/Southern coast of Australia). Posterior support values > 0.90 present on nodes. Bars at terminal branches indicate admixture of OTUs inferred using STRUCTURE, colours approximately correspond to OTUs. Lined box adjacent to OTU indicated lined markings while, OTUs without a box exhibit ringed markings; b) Species delineation using the mitochondrial COI gene. Bayesian phylogeny (MrBayes) of *Hapalochlaena* throughout the Indo Pacific is coloured according to OTU with black used to represent taxa included from NCBI. Grey boxes represent putative species in accordance to sPTP, GYMC SC (strict clock) and GYMC RC (relaxed clock) methods; c) Map of sample locations coloured by organisational taxonomic units A-K; d) Arrangement of samples according to the first two principal components of a PCoA based on SNP data generated using the R dartR.

#### 4.4.1.2. Lined Blue-ringed octopus: OTU D, E & I

*Hapalochlaena* with lined markings from NSW (I) and Taiwan (D) were found to form separate and distinct OTU units. The NSW OTU (I) show the least similarity to other populations with  $F_{st}$  values > 0.78, while individuals with lined markers collected from Taiwan (D) displayed the greatest similarity to Mackay, QLD (0.68). Ringed individuals also sourced from Taiwan (E) did not display high similarity to OTU D (0.88) (Supplementary table 3). Taiwan OTUs D and E are loosely clustered with OTUs A, B, and C in PCoA and DAPC analyses while NSW OTU I formed a tight distinct cluster. STRUCTURE analysis showed little to no admixture in OTU I, similarly OTUs D and E also showed little admixture with each being prescribed its own unit. The SVDQuartet phylogeny displayed high bootstrap support for all OTUs (> 90) (Figure 4.2a). Delineation using mitochondrial genes (12S, 16S and COI) supported the OTU status of NSW (I) across methods: ABGD, sPTP and GYMC. Comparison of COI between the NSW OTU (I) and NCBI sequences

obtained from Japan and Korea does not support a sister taxa status (Figure 4.2a) suggesting the current description of *H. fasciata* for both populations may require re-evaluation.

#### 4.4.1.3. North Western Australia and Timor Leste: OTU G & F

Individuals collected from Darwin, NT, Kimberly, WA and Exmouth, WA formed a single OTU (G) spanning across northern Australia. Samples from these locations formed a single defined cluster in both PCoA and DAPC analyses congruent with lower  $F_{st}$  values between populations ( $< 0.12$ ) (Figure 4.2d, Supplementary table 3). STRUCTURE analysis indicates OTU G forms a distinct group with traces of admixture to nearby locations (Figure 4.2a), in particular admixture observed in Timor Leste samples (OTU F) suggests that this location may receive gene flow from both northern Australia and the Indo Pacific. This is also reflected in the placement of OTU F within the PCoA plot between OTU G and the B, C D and E cluster. Furthermore, high bootstrap support for the clade/OTU was observed in the SVDQuartet phylogeny (100). Mitochondrial sequences across all four genes (12S, 16S, COI and COIII) were examined using delimitation methods ABGD, single rate-PTP and GYMC which all supported the Darwin and Kimberley OTU (G), however the inclusion of a sample from the Dampier archipelago to this group contradicts the previous SNP analyses. It should be noted the SNP and mitochondrial data collected from Dampier Archipelago WA were obtained from different individuals collected at the same location. This could be indicative of sympatry of multiple OTUs.



#### 4.4.1.4. Great Barrier Reef: OTU H

Individual samples from along the Great Barrier Reef (GBR), QLD; Gladstone, Yeppoon, Townsville, Lizard Island and Mackay comprised a single OTU (H) with the exception of “deep water” Yeppoon samples (OTU A) (Figure 4.2). Relatively low  $F_{st}$  values were observed between populations along the GBR ( $< 0.12$ ) with clustering observed in both PCoA and DAPC analyses (Figure 4.2d, Supplementary table 3). Overall admixture across the GBR OTU (H) is low with no clear distinction between populations observed, however one individual collected from Mackay does show admixture with OTU K (Southern Australia/*H. maculosa*) (H\*) (Figure 4.2a). Additionally, placement of this sample within the SVDQuartets phylogeny suggests a distinction from other GBR samples and will warrant further investigation. COI could be successfully sequenced from two individuals collected from the GBR OTU had their COI genes successfully sequenced. Examination of COI genes using the delimitation methods ABGD, sPTP and GYMC unanimously supported the assignment of Yeppoon and Mackay individuals within the same OTU (H).

#### 4.4.1.5. North QLD and Western Australia: OTU J

Geographically distant populations ( $>2000\text{km}$ ) from QLD (Torres Strait and Cape York) and southern WA (Shark Bay and Dampier Archipelago) were found to form an OTU (J). Due to low sample numbers the Torres Strait + Cape York individuals and the Shark Bay + Dampier Archipelago were examined as two populations.  $F_{st}$  values were relatively low (0.15) congruent with the clustering of these populations observed in PCoA and DPAC analyses (Figure 4.2d, Supplementary table 3). Admixture was low within the OTU, the only exception being an

individual from Shark Bay, which displayed some admixture with OTU K (southern Australia/*H. maculosa*). SVDQuartets phylogeny exhibited strong support for a single clade comprising OTU J (100 bootstrap) (Figure 4.2a). Partial mitochondrial genes were sequenced for the Cape York (12S, COI, COIII and cytb) and Shark Bay (COI, COIII) samples. Delimitation analyses (ABGD, sPTP and GYMC) for COI were congruent with placement of both samples within the same OTU. However, the COIII gene showed greater genetic differences between individuals and each sample was placed in its own unit (Figure 4.2b).

#### 4.4.1.6. “Deep water”: OTU A

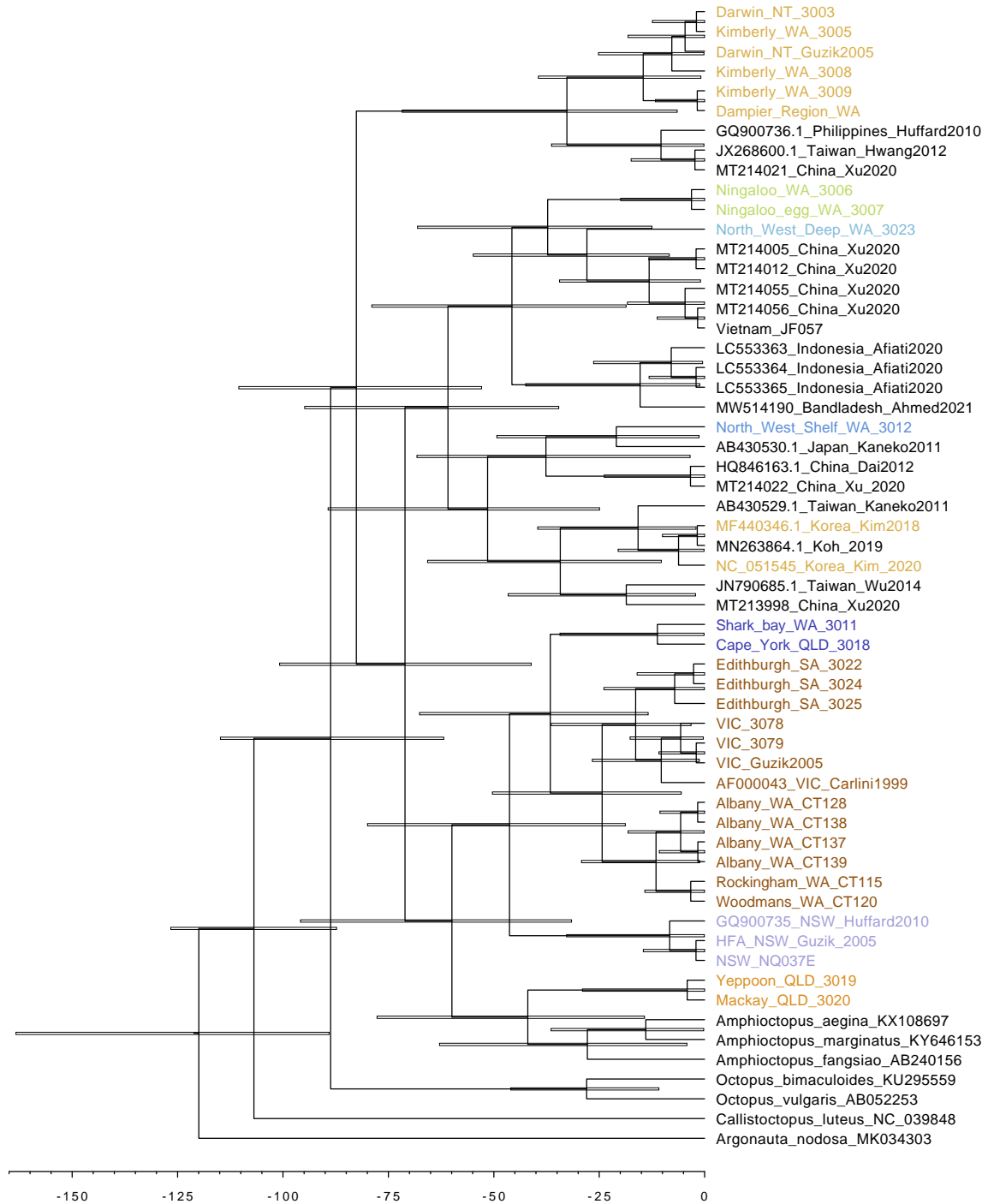
“Deep water” specimens (>100 m) collected from Yeppoon, QLD and North West (Deep) WA, formed a single OTU (A) despite their large geographic separation (i.e. over ~2400km). Specimens formed a distinct cluster in both PCoA and DAPC analyses (Figure 4.2d) with STRUCTURE showing little admixture (Figure 4.2a). No mitochondrial genes could be sequenced for “deep water” Yeppoon samples, however partial five mitochondrial genes (12S, 16S, COI and COIII and Cytb) for North West (Deep) WA were analysed. Delimitation methods (ABGD, single rate-PTP and GYMC) placed North West (Deep) WA as a unique unit distinct from all other specimens (Figure 4.2b).

#### 4.4.1.7. Singletons WA: OTU B & C

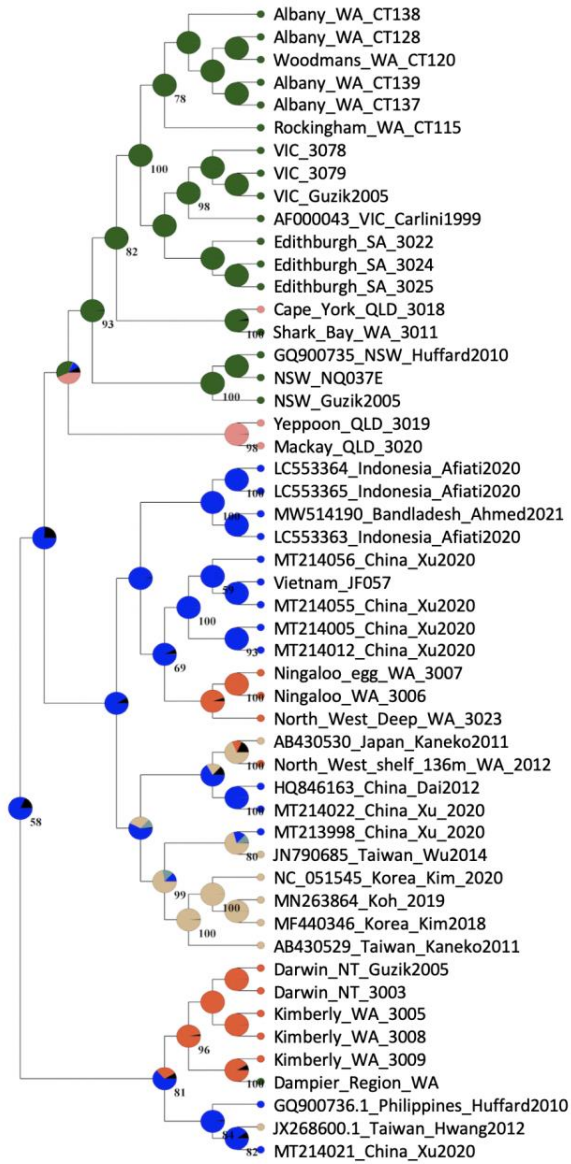
Singletons collected from North West Shelf, WA and Ningaloo, WA each formed their own distinct OTUs termed B and C respectively. OTU B, represented by a single sample (North West Shelf, WA), exhibits admixture with all adjacent populations (Figure 2a). Likewise, PCoA

and DAPC analyses place the sample in close proximity to OTUs A, C (Ningaloo), D (Taiwan-lined) and E (Taiwan-ringed) (Figure 2d, Supplementary figure 1). Low bootstrap support from the SVDQuartet phylogeny was observed for both singleton OTUs (B & C) (Figure 4.2a). Species limits estimated using fragments of five mitochondrial genes (12S, 16S, COI and COIII and Ctyb), ABGD, single rate-PTP and GYMC analyses unanimously supported the classification of both singletons as unique OTU groups (Figure 4.2b).

a)



b)



c)

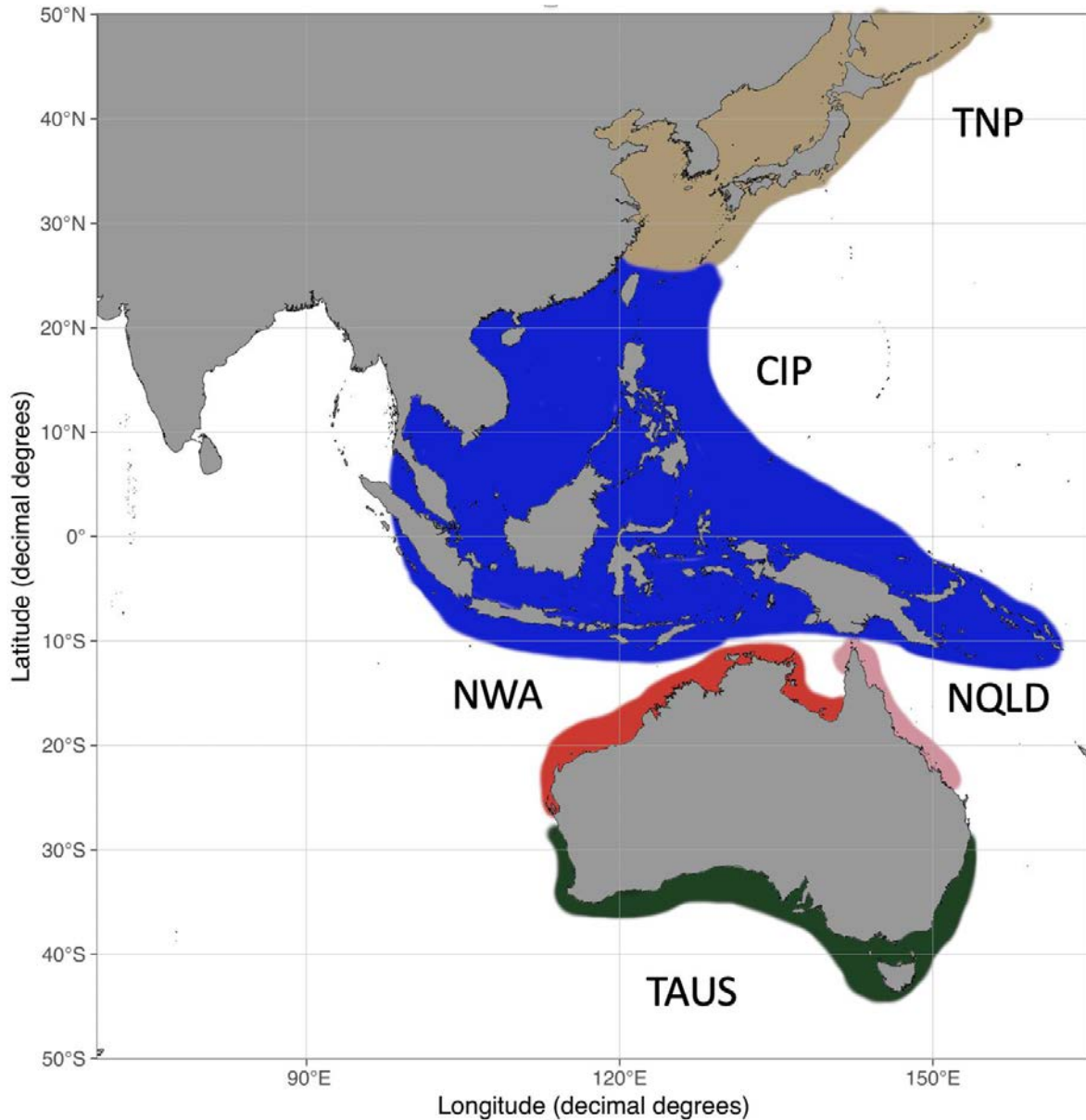


Figure 4.3. Divergence and radiation of *Haplochlana* throughout the Asia Pacific: a) Divergence time estimates of *Haplochlana* throughout the Indo Pacific generated using Bayesian methods. Bars (blue) represent standard error within a 95% confidence interval; b) Reconstructed Ancestral State Phylogeny constructed using RASP BBM. Most likely points of origin inferred for each node in the tree are shown as pie charts coloured according to major geographic regions using the colour scheme shown in c; c) Map of oceanic zones: temperate

Australasia (TAUS) = Green, North East Australia (N\_QLD) = Pink, North West Australia (NW\_AUS) = Red, Central Indo Pacific (CIP) = Blue and temperate north Pacific (TNP) = beige

#### 4.4.2. Divergence time estimation/Ancestral state reconstruction (RASP)

Divergence of the *Haplochlaena* genus is estimated to have occurred in the late Cretaceous - mid Eocene between 72.9 and 45.5mya based on four concatenated mitochondrial genes (Supplementary figure 2). A separate genealogy of the COI gene, which was able to include a greater coverage of *Haplochlaena* populations concurred with the topology of shared samples, however divergence time estimates suggest a more recent origin with the most recent common ancestor (MRCA) predicted to have emerged between 12.9 and 29.5mya (Figure 4.3a). Due to our limited sampling size and the high diversity present within the genus we are unable to identify with confidence the earliest exact location of origin for this genus. However, radiation of the genus is estimated to have occurred approximately 25mya occurring in the Central Indo-Pacific (CIP). RASP analysis indicates colonisation of northern Australia may have occurred twice with OTU G and OTUs C and B representing separate events (Figure 4.3b).

## 4.5. DISCUSSION

Using SNP and mitochondrial data from specimens across the Asia Pacific we show that the genus *Hapalochlaena* is highly diverse, encompassing between 8 and 11 genetically distinct lineages supported by a combination of Bayesian, maximum likelihood and delimitation analyses. Support was present for the current species assignments of *H. maculosa* across southern Australia and also *H. fasciata* located in NSW, Australia with each forming a single OTU. A greater diversity of OTUs than previously recognised were identified throughout the Central Indo Pacific (CIP) and Temperate North Pacific (TNP) not encapsulated in current systematics. Divergence time estimates suggest major lineages within *Hapalochlaena* diversified between ~50-25mya. These results suggest current systematics are not sufficient to describe genetic species diversity and a revision of the genus *Hapalochlaena* is required.

### 4.5.1 Species boundaries and OTUs

Greater morphological variation within the *Hapalochlaena* genus, which extends beyond the currently accepted species, has been recognised in several publications documenting differences in size, the nature of patterning of iridescent blue rings and/or lines, variability in the functionality of their ink sac and reproductive strategy (holobenthic/merobenthic) (Norman, 2000; Reid, 2016). Norman (2000) proposed a putative species for specimens found between Darwin (NT) and Cape York (QLD), termed *Hapalochlaena* sp. 1. This is corroborated by our analyses which showed a distinction between specimens collected from Darwin,



Kimberly and Exmouth (OTU G) compared with Timor Leste (OTU F) specimens sequenced in this study and believed to be *H. lunulata* (Quoy and Gaimard 1832), based on collection locality and external morphology (J. Finn pers. comm.). Low genetic similarity between OTU G and OTU J (Cape York, Torres Strait) however, suggests that the described range for *Hapalochlaena* sp. 1 may contain two distinct lineages. Norman (2000) also proposed a putative species for the Great Barrier Reef (GBR) blue-ringed octopus (*Hapalochlaena* sp. 5), which exhibits a functional ink sac and a holobenthic lifestyle distinguishing it from currently accepted species (Norman, 2000). Evidence was found for the presence of two distinct lineages (OTUs H and A) along the GBR collected between Lizard Island and Mackay, “shallow water” (0-50m) and “deep water” (> 100m). As morphological features were not examined in this study; we cannot confirm the relationship to the previously reported putative species, however *Hapalochlaena* sp. 5 may correspond to shallow water specimens used in this study due to shared location.

*Hapalochlaena* collected from depths > 100m prior have not been previously described due to scarcity and difficulty in locating specimens. “Deep water” Yeppoon specimens were genetically distinct from adjacent shallow water populations while displaying high similarity to a distant singleton collected from north west WA (OTU B) and Yeppoon samples collected from a similar depth. The impact of depth on reproductive isolation in this genus has not been investigated and due to the difficulty in obtaining specimens from depths > 100m, this also means that the range, distribution and effective population sizes of lineages from these depths have not been established.

Of the museum samples included in this study two were unique specimens with no replicates suitable for genetic analyses. Rarity in systematics is common with 13% of invertebrates identified from single specimens in American Museum of Natural History publications (Lim et al., 2012). Singletons within studies can result from difficulty in obtaining samples and/or preservation methods used (Roper, 1983). Preservation methods can greatly impact both morphological and genetic studies, this is a particular issue when examining comprehensive historic repositories for species delimitation (Appleyard et al., 2021). Formalin is a common preservative used most prominently on older samples for preservation of morphological traits (Jeon et al., 2011; Roper, 1983). However, formalin preservation inhibits genetic analyses through three main mechanisms i) facilitating binding of DNA to itself or proteins, ii) fragmentation and iii) base modification (Quach et al., 2004; Wiegand et al., 1996). With the advent of next generation sequencing in conjunction with the growing concern around the safety of formalin use an increasing number of specimens are being preserved using DNA friendly methods (ethanol/freezing) (Spencer et al., 2013). Conversely, ethanol preservation and freezing frequently results in the loss of some morphological features, particularly in octopus, which exhibit few hard body parts/diagnostic features (Lefkaditou and Bekas, 2004). As a result, many samples used in this study lack well preserved diagnostic morphological features.

Low numbers of *Haplochlæna* specimens preserved appropriately for molecular work has strong implications for delimitation using genetic information as common methods such as AGBD (Automatic Barcode Gap Discovery), PTP (Poisson Tree Process) and GYMC (General

Mixed Yule Coalescent) examine differences between intra and interspecific variation to establish species boundaries (Lim et al., 2012). In particular, ABGD generates pairwise comparisons and examines their distribution to determine the “gap” between inter and intra specific divergences (Puillandre et al., 2012). In contrast GYMC examines lineages through time and identifies the increased rate of divergence indicative of allele coalescence within a single species, assuming that a speciation event occurs at a much slower rate. The speciation event can then be inferred as occurring prior to the rapid radiation (Fonseca et al., 2021). Similar to GYMC, PTP models substitutions on a phylogenetic tree, however, it does not require time calibration (Zhang et al., 2013). These methodologies can be sensitive to sample sizes and prone to bias when applied to low and non-representative sampling (Puillandre et al., 2012). As a result, the relationship of singletons (OTU B and OTU C) could not be definitively established with each placed within its own OTU. Observed taxonomic uncertainty could be due to multiple factors including being represented by a single specimen, relatively long divergence times and incomplete sampling of close relatives (Joly et al., 2014).

Identification of *Hapalochlaena* species for use in genetic studies is often conducted based upon the iridescent blue lines and/or rings patterning the mantle and arms in conjunction with location (Kim et al., 2018). This can be misleading as it does not account for convergent evolution of morphological traits. Currently, *H. fasciata* is the only accepted species which exhibits lined patterning and has been documented from NSW, Australia. However, “lined” *Hapalochlaena* have also been documented in Japanese waters leading to the proposal of an anti-tropical distribution for this species (Norman and Kubodera, 2006). By definition

anti-tropical distributions are characterised as one contiguous range disrupted due to a range of factors (climatic shift, fluctuating sea levels etc) resulting in contraction to the extremities of the range (Briggs, 1987). While many marine taxa do exhibit true anti-tropical distributions (*Lobelia*, *Mytilus* spp., *Dasyatis*) (Hilbish et al., 2000; Kokubugata et al., 2012; Le Port et al., 2013) some species, particularly those with limited distinct morphological characteristics and incomplete systematics can be mischaracterized (Grant et al., 2005). A morphological revision of the labrid *Bidianus vulpinus* contradicted the previous anti-tropical status for the species, suggesting a complex containing a minimum of four species (Norman and Kubodera, 2006). Likewise, morphological distinctions have been observed with specimens collected from Japan given the putative species assignment of *Hapalochlaena* sp. 2 (Finn and Lu, 2015; Norman, 2000). Molecular methods can aid in the evaluation of anti-tropical taxa with limited diagnostic characters (Norman and Kubodera, 2006). Analysis of SNP data in this study indicates the morphological similarity between these groups is likely the result of convergent evolution as lined specimens from NSW and from Japan did not form a monophyletic group (Figure 4.2). Similarly, allozyme and mitochondrial gene analyses contradicted the previous anti-tropical distribution (bipolarity) hypothesis for the anchovy genus (*Engraulis*) (Grant et al., 2005). Several previous phylogenetic studies of *Hapalochlaena* included a single representation of each species and did not sample “lined” octopus from the North Pacific and NSW, Australia. As a result, this study represents the first genetic comparison of these two morphologically similar groups.

Examination of *H. maculosa* specimens across southern Australia (OTU K) using SNP data was congruent with previous work by Morse et al. (2018) suggesting a clinal species complex based on greater genetic divergence between locations at the extremities of their range (Morse et al., 2018b). Mitochondrial delimitation methods (ABGD, mPTP, bPTP & GYMC) indicate divergence between *H. maculosa* locations was insufficient to warrant species status, however when GYMC was run using a relaxed as opposed to a strict clock model, locations in the West (WA) and East (VIC and SA) were delimited as separate species.

#### 4.5.2 Dating and diversification of the genus

The *Hapalochlaena* species complex is estimated to have diverged from the lineage containing *Amphioctopus* during the early Eocene era approximately ~50mya based on divergence time estimates using partial COI sequences. This estimate is corroborated by previous divergence time estimates between octopod genomes, which indicated *H. maculosa* diverged from *O. bimaculoides* ~59mya (Tanner et al., 2017; Whitelaw et al., 2020). A notable radiation of the genus occurs post the Eocene-Oligocene extinction event (~33mya) between ~50- 25mya, which gives rise to lineages in the Central Indo Pacific (CIP) and temperate Australasia (TAUS). This suggests the point of origin for the genus is likely to have occurred in the CIP, potentially within the Indo Australian archipelago (IAA) a known hotspot for biodiversity and speciation at this time (Yasuhara et al., 2017). Composed of broad shallow water areas with complex reef systems interconnecting a large geographical area (Briggs and Bowen, 2013), the IAA currently contains the greatest species richness for coastal cephalopods globally (Rosa et al., 2019). Patterns of low endemism within the IAA coinciding with shared

taxa in TAUS and the temperate northern Pacific (TNP) indicate the IAA as a point of origin for many cephalopod species including *Haplochlæna* in the North Western Pacific. Three main hypotheses have been suggested for the patterns of species richness in the IAA i) overlapping distributions of taxa each dispersing outward from their point of origin (centre of overlap) ii) centre of diversification/radiation (centre of origin) and iii) refugia for species during climatic change (centre of accumulation/survival) (Briggs, 1999; Hughes et al., 2002; Jokieli and Martinelli, 1992; Rosa et al., 2019; Woodland, 1983). There is evidence to suggest multiple hypotheses may be contributing the patterns observed as opposed to a single driver (Mironov, 2006). Transition of the global biodiversity hotspot (hopping hotspot) from the Tethys/Arabian to the IAA during the late Eocene to early Oligocene (~33mya) extinction event was observed in patterns of four reef associated fish lineages, which showed a rapid diversification correlated with reef habitat in the IAA (Cowman and Bellwood, 2013; Gaboriau et al., 2018). Furthermore, a higher proportion of reef associated taxa indicated a greater resilience to extinction event allowing for a greater than expected retention of species diversity (Cowman and Bellwood, 2011). Heterogeneity of the IAA reefs is believed to be an important factor in retaining diversity through events of climatic upheaval providing a range of refugia compared to homogeneous habitats (Pellissier et al., 2014). Additionally, heterogeneous and complex benthic habitats may aid in facilitating sympatric and/or allopatric speciation of benthic octopods and nekto-benthic cuttlefishes and sepiolids when compared to pelagic squid (Rosa et al., 2019).

Fluctuating sea levels during the Pleistocene often coinciding with glacial periods resulted in lower sea levels leading to the exposure of the Sunda and Sahul shelves and an approximately 90% loss of shallow water coastal habitat up to 60m deep throughout the Pacific (Ludt and Rocha, 2015). Restricted water flow throughout the Indo Pacific and loss of coastal habitat during these periods greatly influenced marine species composition (Rocha and Bowen, 2008). Genetic signatures have been identified in olive sea snake (*Apiysurus laevis*) populations across Northern Australia consistent with contractions to refugia during low sea levels. Subsequent colonisation events from western to eastern populations were identified based on diversity indices indicating refugia in WA during the Pleistocene retained populations that were lost in the east (Lukoschek et al., 2008, 2007). Distributions of *Hapalochlaena* also show a distinction between east and west population across Northern Australia, which may be indicative of isolation to refugia during the Pleistocene (Ludt and Rocha, 2015).

#### *4.5.3 Mechanisms of divergence and speciation*

Dispersal ability plays an integral role on the distribution and population dynamics of marine organisms (Higgins et al., 2013). Coleoid cephalopods utilise two main reproductive strategies holobenthic and merobenthic (Doubleday et al., 2008). Holobenthic cephalopods exhibit a reduced dispersal ability relative to merobenthic species as the offspring hatch well-developed and assume a benthic lifestyle immediately. In contrast, merobenthic species have a pelagic phase allowing for dispersal on water currents post hatching before settling to a benthic lifestyle (Ibáñez et al., 2018; Villanueva et al., 2016). A merobenthic strategy has been proposed as the ancestral state for members of the family Octopodidae with the holobenthic life history

believed to be associated with colonisation of temperate and deep-water environments (Ibanez et al., 2014). Radiation of *Hapalochlaena* is predicted to have occurred in the tropical Indo Australian Archipelago (IAA) during the ~50-25mya. The subsequent colonisation of temperate zones including temperate Australasia (TAUS) are associated with the shift to a holobenthic strategy for *H. maculosa* and *H. fasciata* followed by minimal diversification relative to tropical zones. *H. lunulata* retains the predicted ancestral merobenthic strategy, occupying shallow tropical waters (Ibáñez et al., 2014). Plasticity of reproductive strategies has been observed in many octopod genera and are often associated with environmental conditions. Egg size was demonstrated to decrease from the Eastern to Western Mediterranean for the mysopsid squids (*Loligo* and *Alloteuthis*), an octopod (*Eledone cirrhosa*) and an argonaut (*Argonauta argo*). However, cuttlefish and sepiolid squids examined did not display any notable difference in egg size (Laptikhovsky et al., 2009). Cooler and less productive waters have been suggested to favour larger eggs and consequently larger yolks resulting in larger more robust offspring (Alekseev, 1981). The comparatively complex trophic structures found in tropical waters with high predator diversity may support the quantity of offspring over other adaptive strategies (Alekseev, 1981). Conversely putative *Hapalochlaena* species identified from warmer waters, such as Darwin ([*Hapalochlaena* sp. 1] and the GBR [*Hapalochlaena* sp. 5]) exhibit larger eggs, however it should be noted that the life history of these specimens have not been fully elucidated (Norman, 2000). Additionally, many of the putative species identified in this study have completely unknown life histories. Therefore, further investigation is required to fully elucidate the evolution of reproductive strategies within this genus and the resulting impact of genetic structure and speciation.



## 4.6 CONCLUSION

This study presents the first large scale genetic analyses of the *Haplochlæna* genus and demonstrates that the level of species diversity exceeds current systematics for the genus. Previously described morphological diversity within the genus is supported by the genetic analyses conducted in this study including a distinction between lined *Haplochlæna* populations within NSW, Australia and the Northern Pacific. This provides a greater understanding of species diversity within the genus and has implications for how future studies should be conducted and compared based on genetic distances observed across locations. Detailed investigation of morphological features of *Haplochlæna* specimens is required in the future to officially describe species within the genus.

## GENERAL DISCUSSION

### *5.1 Summary*

The aim of this chapter is to summarise the major findings of the thesis and place them into the broader scientific context while also highlighting future directions for research. Section 5.2 synthesises all major outcomes resulting from each chapter, which in Section 5.3, are placed into a broader context. Section 5.3 details the significance of research conducted in this thesis alongside encapsulating the current state of research and specific relevant contributions from researchers in the field. Lastly, Section 5.4 highlights areas for future research.

### *5.2 Major outcomes*

The overall aim of this thesis is to provide a greater understanding of genome evolution and speciation as it pertains to the blue-ringed octopus genus (*Hapalochlaena*), in conjunction with providing genetic tools and resources to facilitate future and ongoing research. Firstly, the genome of *H. maculosa* was annotated and compared to other available octopod genomes. Secondly, a genetic linkage map was generated for *H. maculosa* and used to aid in scaffolding of the genome. Finally, the *Hapalochlaena* species complex was examined using specimens collected from throughout its range to provide a genetic basis for re-evaluation of the systematics of the genus.

Chapter 2 of this thesis encompassed the annotation of the first blue-ringed octopus genome, the southern blue-ringed octopus (*H. maculosa*) and subsequent whole genomic and transcriptomic comparisons to non-TTX bearing octopod genomes (*Octopus bimaculoides* and *Callistoctopus minor*). Annotation of the *H. maculosa* genome identified 29,328 genes, in line with published octopod genomes *O. bimaculoides* and *C. minor* which exhibit 33,638 and 30,010 genes respectively (Albertin et al., 2015; Kim et al., 2018). Divergence times between octopod genomes place divergence of the *Hapalochlaena* lineage ~59 mya from the lineage comprising *O. bimaculoides* and ~183 mya from the lineage leading to *C. minor*. The phylogenetic relationships between the main octopod lineages observed in this study were supported by a combination of partial mitochondrial and nuclear genes (Guzik et al., 2005; Strugnell et al., 2006). Furthermore, Bayesian estimates of divergence times using a concatenated alignment of 197 genes by Tanner et al. (2017) concur with the present study, placing the divergence of *Hapalochlaena* from the *Abdopus* lineage at ~59mya.

Comparisons of gene family expansions between Molluscan lineages revealed octopod specific expansions of the neural gene families; cadherins and zinc fingers. These findings are congruent with previous analyses of the *O. bimaculoides* genome at the time of publication (Albertin et al., 2015). However, the present study utilised multiple octopod genomes and was able to determine if these expansions were lineage or octopod specific. Investigation into the impact of TTX inclusion within the venom of *H. maculosa* revealed a relatively reduced gene expansion of serine proteases in conjunction with lower expression within the posterior salivary

gland (PSG) in comparison with the two non-TTX bearing octopods. Serine proteases have been identified from the venom glands of all major coleoid cephalopod lineages (octopus, squid and cuttlefish) and are believed to play an integral role in prey capture and digestion (Fry et al., 2009; Ruder et al., 2013).

Investigation of the sodium channels ( $Na_v$ ) of octopods in this study revealed the presence of a M-T mutation shared by *Hapalochlaena lunulata* (Geffeney et al., 2019) as well as other TTX bearing taxa, a pufferfish (*Tetraodon nigroviridis*), and newt (*Taricha granulosa*) (Jost et al., 2008). This particular mutation was verified by Jost et al. (2009) to confer 10-15 fold binding resistance to TTX.

Lastly, the microbiome of *H. maculosa* was investigated to identify potential candidates for TTX production in this species. However, the PSG was found to house a diverse array of bacterial families with no dominant lineage identified. While bacteria are known producers of tetrodotoxin and common in the marine environment (Do et al., 1990), this study could not reach a consensus for the origin of TTX production within *Hapalochlaena*.

Linkage maps are highly versatile and act as a platform to conduct a plethora of studies and often act as complementary to a genome. The third chapter of this thesis details the generation of the first linkage map for any cephalopod, the southern blue-ringed octopus (*H.*

*maculosa*). Linkage map construction was achieved using 10 families and a total of 276 individuals to form a sex-averaged map with 47 linkage groups totalling 2016cM in length and composed of 5,663 loci. Currently, no octopod linkage maps are available for comparison, however karyotypes have been generated for the octopod genera *Octopus*, *Amphioctopus* and *Callistoctopus*, all of which exhibit 30n chromosomes (Wang and Zheng, 2017). The greater number of linkage groups (47) than the estimated number of chromosomes from previous studies (30) can be attributed to statistical limitations imposed by relatively small family sizes. *H. maculosa* exhibits a holobenthic life history whereby they produce a fewer number of larger eggs (6-7mm) (Tranter and Augustine, 1973) compared to the merobenthic *H. lunulata* (3.5mm) (Overath and von Boletzky, 1974). The relatively small families produced by *H. maculosa* reduce the number of informative meioses observed and as a result the resolution of linkage map can be lower.

Linkage maps often complement genome assembly allowing for reorientation of scaffolds and placement of previously disconnected scaffolds into a set of pseudo-chromosomes (Glazer et al., 2015). The *H. maculosa* genome assembly was relatively fragmented and composed of 47K scaffolds. The generated linkage map was utilized to place and orientate 1,151 scaffolds covering 34.7% of the total genome. This updated assembly allowed for the placement of three HOX genes within the same pseudo-chromosome despite being separated by large intergenic distances (~3-25Mb). This is the first documented case of HOX gene co-location for an octopod and was only possible due to the high contiguity afforded

by the linkage map. The HOX gene cluster forms a highly conserved set of genes integral for development in Metazoa (Kourakis and Martindale, 2000). Prior to the present study, the best documented set of HOX genes in an octopod (*O. bimaculoides*) places each gene within the cluster on a separate scaffold with no known order or relationship between them known. This study demonstrates the utility of a linkage map to improve fragmented non-model genome assemblies, providing a greater genetic context for gene family evolution.

Chapter 4 of this thesis aims to increase understanding of the diversity, genetic structure and species boundaries of *Hapalochlaena* throughout their range. Museum specimens were collected and genotyped from 21 locations throughout the Asia Pacific. Support was found for the current species assignment of *H. maculosa* for all locations across southern Australia (Victoria to Western Australia). However, locations at the extremities of their range showed the greatest genetic divergence. A similar pattern was observed by Morse et al. (2018) who found populations of *H. maculosa* exhibit a clinal species complex following a signature of isolation-by-distance. Similarly, *H. fasciata* observed from NSW were supported in their current species designation (Norman, 2000; Norman et al., 2016; Reid, 2016). An anti-tropical distribution of this species with purported sister taxa in the North Pacific (Japan and surrounding waters) was questioned by previous publications (Norman and Kubodera, 2006; Norman and Hochberg, 2005). Genetic analyses in this study demonstrated that any morphological similarities, primarily iridescent blue lines, are likely the result of convergence. Similarly, Norman et al. (2000) noted the North Pacific lined octopus as a potential species in its

own right designating it with the name *Hapalochlaena* sp. 2. Subsequent morphological examination has confirmed that the North Pacific lined octopus is morphologically distinct from *H. fasciata* (Finn and Lu, 2015).

Diversity within the *Hapalochlaena* species complex has previously been observed to be greater than is currently described in the systematics with three/four species accepted. Norman et al. (2000) described an additional five *Hapalochlaena* sp. which require further revision. Finn (2015) reported 12 species of blue-ringed octopus from Australia and the Indo-west Pacific, while Norman et al. (2016) reported at least 10 species across this range. In this study singletons were examined each representing a rare and difficult to collect specimen, with each forming a distinct OTU. Rarity, incomplete sampling and preservation methods for historic collections is a common issue for poorly described taxa where the full range and diversity is unknown (Lim et al., 2012). Additionally, deep water *Hapalochlaena* were also examined in this study with a minimum of two independent lineages identified. Inclusion of the rare samples in this study provide insight into the previously undocumented diversity and directs future research to address the current gaps in literature. Radiation of major *Hapalochlaena* lineages was dated to between ~50 and 25mya and estimated to have originated in the Central Indo Pacific (CIP). Currently the Indo Australian Archipelago (IAA) hosts the greatest species diversity for coastal shallow water cephalopods globally (Rosa et al., 2019). Low endemism within the IAA for cephalopods including *Hapalochlaena* indicates its location as a potential point of origin (Rosa et al., 2019). This was further supported in this study by a reconstruction of ancestral

state phylogeny, which indicated the CIP as the central point of origin for *Hapalochlaena* lineages in the North and South Pacific.

### *5.3 Significance and contributions to a greater understanding*

How did the distinctive blue-ringed octopus evolve? Understanding the evolution of an organism characterised by complex and poorly understood traits, such as TTX acquisition, requires extensive genomic resources from members of the genus and related lineages. This study investigated the species diversity of the genus, in conjunction with genome evolution between the TTX bearing *H. maculosa* genome and non TTX bearing octopods.

Members of the genus are well documented to sequester the potent neurotoxin TTX within both venom and tissues. Quantification of TTX within tissues revealed variation both within and between species (Williams and Caldwell, 2009). Williams and Caldwell (2009) examined wild caught *H. fasciata* and *H. lunulata* from the aquarium trade and found that TTX concentrations within *H. lunulata* (0-174ug) were lower overall and occurred in fewer tissues compared to *H. fasciata* (60-405ug). *Hapalochlaena fasciata* from Japanese waters were examined most recently by Yamate et al. (2021) who found large variations in TTX concentrations potentially associated with age, sex and seasonality. This suggests further research is required to fully understand the dynamics of TTX sequestration between and within *Hapalochlaena*. The results from chapter 4 provide a framework for such future work by



identifying the genetically distinct populations and their geographic ranges. In particular, OTUs identified in chapter 4 need to be taken into consideration and treated as distinct units by future studies examining TTX production and dynamics.

Understanding how TTX concentrations can vary is particularly important when it comes to venom. *Hapalochlaena* are the only documented species to include TTX within their venom (Sheumack et al., 1984). This study supports the hypothesis that acquisition of TTX as a component of venom has impacted venom evolution in the *Hapalochlaena* lineage, potentially by replacing proteinaceous octopod toxins conserved in related lineages. Comparisons of venom gland (PSG) proteomes between *H. maculosa* and the southern sand octopus (*Octopus kaurna*) demonstrated a reduction of serine protease expression in conjunction with the absence of the octopod neurotoxin tachykinin in *H. maculosa* (Whitelaw et al., 2016). *Hapalochlaena* pose a genuine risk to human health with human fatalities reported from both envenomation and accidental ingestion (Jacups and Currie, 2008; Meier and White, 2017). Clarification of systematics within this genus is crucial to prevent misidentification for both medical and academic purposes. This study provided a genetic framework to aid in the revision of systematics in conjunction with providing a taxonomic structure for future studies.

Additionally, understanding the evolution of TTX acquisition and impact on genome evolution within this genus is of importance. To date no whole genomic studies have been

conducted on members of the *Hapalochlaena* genus. The ability to perform whole genomic analyses of a cephalopod genome was only made possible in 2015 with the publication of the first cephalopod genome *Octopus bimaculoides*. The *O. bimaculoides* genome was found to be large (2.7Gb), highly repetitive and lacking evidence for whole genome duplications hypothesised to play an integral role in the evolution of the cephalopod body plan (Albertin et al., 2015). Extensive expansions of the neural gene families zinc fingers and cadherins within *O. bimaculoides* were previously believed to be exclusive to vertebrates. This study provides a greater taxonomic context to these findings demonstrating the presence of similar expansions in two other octopod genomes, suggesting that this is a shared genomic trait rather than being restricted to the genus *Octopus*.

Investigation of HOX genes in *O. bimaculoides* revealed a departure from most metazoan genomes with each gene located on a separate scaffold (Albertin et al., 2015). The HOX gene cluster plays an important role in the development of metazoans, genes are usually located in close proximity (Biscotti et al., 2014). Expression of HOX genes in *E. scolopes* revealed no deviation from collinearity, with expression along the CNS congruent with the ancestral role of axial patterning. This suggested no deviation in gene placement from related molluscan clusters (Lee et al., 2003). However, prior to this study the genomic location of HOX genes in relation to each other was unknown due to the scarcity of chromosome level cephalopod genomes. HOX genes identified in the *O. bimaculoides* genome were scattered with each located on a separate scaffold (Albertin et al., 2015). This study provided the first annotated

*Haplochlaua* genome and the first genetic linkage map of a cephalopod, which revealed the placement of three HOX genes within the same pseudo-chromosome while three additional genes from the cluster were identified on separate scaffolds. This demonstrates large intergenic distances between HOX genes are also present in *H. maculosa*, however at least three of the genes exhibit a level of co-location. It should also be noted that the order of the genes on LG 9 do not exhibit the order observed in related molluscan clusters.

TTX acquisition has occurred convergently through a plethora of distantly related taxa including gastropods, arachnids, reptiles, pufferfish and amphibians (Geffeney et al., 2005; Kaneko et al., 1997; Venkatesh et al., 2005). TTX binding resistance of sodium channels ( $Na_v$ ) is a key feature in many TTX bearing taxa allowing for the safe sequestration of the toxin. Jost et al. (2008) determined that a common M-T mutation within the third p-loop of  $Na_v$  channels in 12 taxa provided 10-15 fold binding resistance. Mutations can occur on one or more of the four  $Na_v$  p-loops and may act synergistically. The mite, *Varroa destructor*, exhibits “super resistant” channels with two mutations one on the third p-loop (M-T) and one on the fourth (D-S) conferring 1000 fold binding resistance (Du et al., 2009). *Haplochlaua maculosa* shares the M-T mutation on the third p-loop in addition to a substitution on the fourth p-loop (D-H). Located in a key binding position, the D-S substitution has also been identified in the flatworm, *Bdelloura candida* (Jeziorski et al., 1997), however TTX binding inhibition of this mutation has yet to be tested. The set of mutations observed in *H. maculosa* are shared by *H. lunulata* (Geffeney et al., 2019), however it is unknown if these mutations result from a shared ancestor

or occurred independently. Additionally, some taxa, including pufferfish, also produce TTX binding proteins within their blood or hemolymph, which may facilitate transport and/or excretion of the toxin (Matsumoto et al., 2010). Unfortunately, no homologous proteins were identified in *H. maculosa*. The source of TTX in *Hapalochlaena* is still contentious with two studies conducted to date presenting conflicting results (Chau et al., 2013; Hwang et al., 1989). Culturing of bacterial strains by Hwang et al. (1989) produced TTX producing strains from the PSG of *Hapalochlaena* collected in the Philippines. In contrast, a molecular study performed on the PSG of a *Hapalochlaena* sp. screened bacteria for three genes (polyketide synthase, non-ribosomal peptide synthase and aminotransferase) predicted to be involved in TTX biosynthesis and did not detect TTX producing bacterial strains (Chau et al., 2013). Additionally, no loci under selection detected in chapter 4 were associated with the production, transport or sequestration were identified in *Hapalochlaena*. This is due to a few key factors: i) no genes involved in the biosynthetic pathway of TTX production have been identified ii) as stated above, while a TTX binding protein has been identified in pufferfish, no genes involved in the sequestration or transport of TTX have been identified in *Hapalochlaena* iii) the large genetic divergences between locations and small sample sizes present in the study resulted in few loci under selection detected with a high level of confidence. However, if genes involved in the biosynthetic pathway of TTX could be identified the data and analysis conducted here could be applied to future work on this subject.

#### 5.4 Future directions

Cephalopod genomes remain a relatively new resource for understanding the evolution of these unique organisms and this thesis provides a strong base for myriad future work, some of which I will detail here.

The *H. maculosa* genome presented in this study could be improved by the inclusion of long read sequencing such as PacBio and/or Hi-C to generate a chromosome level assembly. Additional long read sequencing could improve contiguity and resolve scaffolds, previously unable to be joined due to the large proportion of repetitive elements present. Assembly of the additional sequencing could be aided by utilising the linkage map generated in chapter 3 to resolve ambiguous connections and correct misassemblies.

The resulting assembly could aid in the generation of related genomes within the *Hapalochlaena* genus, providing a framework from, which genus and/or species-specific features could be examined. One of the limitations present in chapter 2 is the small number of genomes examined and the large evolutionary distances between lineage, while such comparisons are useful they are limited to comparing lineages and do not provide more recent insights into the evolution of the genus. Additionally a more contiguous assembly would facilitate powerful synteny analyses with the recently published chromosome level assembly of the *Octopus sinensis* genome (Kim et al. 2020) allowing for investigation into potential genome expansions and translocations at a micro and macro scale (D'Alençon et al., 2010).

Chromosome level assemblies allow for examination of large-scale chromosomal translocation events, which have yet to be elucidated in cephalopods. Karyograms of three octopods demonstrated differences in structure, while all species contained the same number of chromosomes (30n) (Wang and Zheng, 2017). On a smaller scale the evolutionary history of key gene families such as serine proteases could be elucidated, including the number and timing of duplication and contraction events between lineages.

To further understanding of TTX acquisition in the genus several experiments are required. Firstly, the mutations identified within sodium channels in *H. maculosa* should be assayed for TTX binding inhibition. Quantification of binding inhibition is commonly tested by inducing mutation within a TTX sensitive/susceptible channel (rat example Na<sub>v</sub> 1.4) (Jost et al., 2008).

Secondly, in order to investigate the mechanism of distribution for the toxin blood should be assayed for TTX/STX binding proteins. Such proteins have been identified in the TTX bearing pufferfish (*Takifugu niphobles*) (Matsui et al., 2000), however no ortholog was present in the *H. maculosa* genome and I was unable to find a candidate using only bioinformatic means. By using a TTX/STX binding assay, proteins, which have independently evolved in octopods could be identified (Yotsu-Yamashita et al., 2001).

Thirdly, the origin of TTX in *H. maculosa* could not be resolved in this study, however if the genes involved in the biosynthetic pathway were known, then organisms could be “scanned” for bacteria containing said genes and identified as candidates for TTX production. I propose a future study examining bacteria from several genera with members able to produce TTX whereby full genomes are sequenced from a verified producer of the toxin and closely related species, which does not produce TTX. These genomes could then be compared to identify similarities between TTX producing strains independent of their relatedness. This type of study does pose some challenges, bacteria are not always culturable and as a result verification of TTX production via cultures may not be possible for some bacteria. Bacterial growth and TTX production may also be impacted by media type, temperature, salinity and pH among others (Magarlamov et al., 2017). This may be a particular issue for those strains isolated from TTX producing taxa which may facilitate production under specific conditions.

The linkage map presented here could form the basis for quantitative trait loci (QTL) analyses examining the inheritance of quantitative traits. Often QTL analyses are used in agriculture and aquaculture to understand the inheritance of desirable traits such as disease resistance (Wang et al., 2016). However, they also possess great utility in ecological studies. The ability to examine the inheritance of genes linked to specific traits throughout a geographical range opens possibilities for the investigation of local adaptation. Verhoeven et al. (2008) identified a QTL associated with flowering time which played a role in local adaptation in one

wild barley population but not another. Such studies demonstrate the applicability of QTL approaches in understanding local adaptation and selective pressures (Verhoeven et al., 2008).

Loci under selection, which could indicate adaptation to increased depth would be interesting to identify in *Haplochaena*. Morse et al. (2018) noted high genetic similarity between relatively distant populations sharing similar habitat, likewise “deep water” specimens from northern WA and Yeppoon, QLD examined in chapter 4 displayed a similar pattern. A future genotyping study with a larger sampling effort across northern Australia that includes several deep-water locations, each with an adjacent “shallow water” location could elucidate loci under selection for this trait. Alternatively, given an extensive budget whole genome sequencing (WGS) of a larger sample set than examined in chapter 4 could be conducted at low depth and assembled using the improved assembly of *H. maculosa* mentioned previously. This would provide a greater resource from which not only genes under selection but other genomic features could be examined such as transposable elements. The previous SNP based approach used in chapter 4 is limited to providing a subset of genomic locations, in comparison WGS would allow for genome wide association studies to be conducted. If loci under selection and their associated quantifiable traits could be identified than the linkage map presented here could form the basis for quantitative trait loci (QTL) analyses examining the inheritance of these quantitative traits. Often QTL analyses are used in agriculture and aquaculture to understand the inheritance of desirable traits such as disease resistance (Wang et al., 2016). However, they also possess great utility in ecological studies. The ability to examine the inheritance of genes linked to specific traits throughout a geographical range opens possibilities for the investigation of local adaptation. Verhoeven et al. (2008) identified a QTL associated with flowering time



which played a role in local adaptation in one wild barley population but not another. Such studies demonstrate the applicability of QTL approaches in understanding local adaptation and selective pressures (Verhoeven et al., 2008).

Species diversity within *Haplochlaena* examined in chapter 4 highlighted the need for a revision of current systematics for the genus. In order to describe new species extensive morphological work must be concluded including elucidation of behaviour and life history. Taxonomic descriptions from specimens across all locations are required, in addition to a larger sampling effort to increase chances of capturing potentially overlapping ranges of species and/or OTUs. Furthermore, additional sampling efforts and genetic work are required for singleton and rare specimens examined in order to properly describe the potential species and establish their distributions.

## REFERENCES

- Acosta-Jofré, M.S., Sahade, R., Laudien, J., Chiappero, M.B., 2012. A contribution to the understanding of phylogenetic relationships among species of the genus *Octopus* (Octopodidae: Cephalopoda). *Sci. Mar.* 76, 311–318.  
<https://doi.org/10.3989/scimar.03365.03B>
- Adema, C.M., Hillier, L.W., Jones, C.S., Loker, E.S., Knight, M., Minx, P., Oliveira, G., Raghavan, N., Shedlock, A., Do Amaral, L.R., 2017. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat. Commun.* 8, 15451.
- Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., Rokhsar, D.S., 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* 524, 220–224.  
<https://doi.org/10.1038/nature14668>
- Alekseev, F.E., 1981. Rass--Thorson--Marshall rule and biological structure of marine communities, in: 4th Congress of All-Union Hydrobiological Society. Theses of Reports. Part I. Naukova Dumka, Kiev. pp. 4–6.
- Allcock, A.L., Strugnell, J.M., Johnson, M.P., 2008. How useful are the recommended counts and indices in the systematics of the Octopodidae (Mollusca: Cephalopoda). *Biol. J. Linn. Soc.* 95, 205–218. <https://doi.org/10.1111/j.1095-8312.2008.01031.x>
- Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.*  
<https://doi.org/10.1186/s13059-020-1935-5>
- Appleyard, S.A., Maher, S., Pogonoski, J.J., Bent, S.J., Chua, X.Y., McGrath, A., 2021. Assessing DNA for fish identifications from reference collections: the good, bad and ugly shed light on formalin fixation and sequencing approaches. *J. Fish Biol.* 98, 1421–1432.  
<https://doi.org/10.1111/jfb.14687>

- Bartolomé, C., Maside, X., Charlesworth, B., 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol. Biol. Evol.* 19, 926–937. <https://doi.org/10.1093/oxfordjournals.molbev.a004150>
- Berger, E., 2010. Aquaculture of Octopus species: present status, problems and perspectives, *The Plymouth Student Scientist*. University of Plymouth.
- Biscotti, M.A., Canapa, A., Forconi, M., Barucca, M., 2014. Hox and parahox genes: A review on molluscs. *Genesis*. <https://doi.org/10.1002/dvg.22839>
- Boissinot, S., Entezam, A., Furano, A. V., 2001. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* 18, 926–935. <https://doi.org/10.1093/oxfordjournals.molbev.a003893>
- Bookstein, F.L., 1985. *Morphometrics in evolutionary biology: the geometry of size and shape change, with examples from fishes*. Academy of Natural Sciences of Philadelphia.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* 10, 1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Briggs, J.C., 1999. Coincident biogeographic patterns: Indo-West Pacific Ocean. *Evolution (N. Y.)* 53, 326–335. <https://doi.org/10.1111/j.1558-5646.1999.tb03769.x>
- Briggs, J.C., 1987. Antitropical distribution and evolution in the Indo-West Pacific Ocean. *Syst. Zool.* 36, 237–247.
- Briggs, J.C., Bowen, B.W., 2013. Marine shelf habitat: Biogeography and evolution. *J. Biogeogr.* <https://doi.org/10.1111/jbi.12082>
- Callaerts, P., Lee, P.N., Hartmann, B., Farfan, C., Choy, D.W.Y., Ikeo, K., Fischbach, K.F., Gehring, W.J., Gert De Couet, H., 2002. HOX genes in the sepiolid squid *Euprymna scolopes*: Implications for the evolution of complex body plans. *Proc. Natl. Acad. Sci. U. S. A.* 99,

2088–2093. <https://doi.org/10.1073/pnas.042683899>

Catchen, J., Amores, A., Bassham, S., 2020. Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3 Genes, Genomes, Genet.* 10, 4115–4128.  
<https://doi.org/10.1534/g3.120.401485>

Chakravarti, A., Lasher, L.K., Reefer, J.E., 1991. A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* 128, 175–182.  
<https://doi.org/10.1093/genetics/128.1.175>

Chau, R., Kalaitzis, J., Wood, S., Neilan, B., 2013. Diversity and biosynthetic potential of culturable microbes associated with toxic marine animals. *Mar. Drugs* 11, 2695–2712.

Chifman, J., Kubatko, L., 2014. Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics* 30, 3317–3324. <https://doi.org/10.1093/bioinformatics/btu530>

Choi, Y., Kim, S., Lee, J., 2020. Construction of an Onion (*Allium cepa* L.) Genetic Linkage Map Using Genotyping-by-Sequencing Analysis with a Reference Gene Set and Identification of QTLs Controlling Anthocyanin Synthesis and Content. *Plants* 9, 616.  
<https://doi.org/10.3390/plants9050616>

Cowman, P.F., Bellwood, D.R., 2013. Vicariance across major marine biogeographic barriers: Temporal concordance and the relative intensity of hard versus soft barriers. *Proc. R. Soc. B Biol. Sci.* <https://doi.org/10.1098/rspb.2013.1541>

Cowman, P.F., Bellwood, D.R., 2011. Coral reefs as drivers of cladogenesis: Expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. *J. Evol. Biol.* 24, 2543–2562. <https://doi.org/10.1111/j.1420-9101.2011.02391.x>

D'Alençon, E., Sezutsu, H., Legeai, F., Permal, E., Bernard-Samain, S., Gimenez, S., Gagneur, C., Cousserans, F., Shimomura, M., Brun-Barale, A., Flutre, T., Couloux, A., East, P., Gordon, K., Mita, K., Quesneville, H., Fournier, P., Feyereisen, R., 2010. Extensive synteny conservation

of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. *Proc. Natl. Acad. Sci. U. S. A.* 107, 7680–7685.  
<https://doi.org/10.1073/pnas.0910413107>

Danzmann, R.G., 2006. Linkage analysis package for outcrossed families with male or female exchange of the mapping parent, version 2.3. Univ. Guelph, Ontario.

Dawson, D.A., Burke, T., Hansson, B., Pandhal, J., Hale, M.C., Hinten, G.N., Slate, J., 2006. A predicted microsatellite map of the passerine genome based on chicken-passerine sequence similarity. *Mol. Ecol.* 15, 1299–1320. <https://doi.org/10.1111/j.1365-294X.2006.02803.x>

Do, H.K., Kogure, K., Simidu, U., 1990. Identification of deep-sea-sediment bacteria which produce tetrodotoxin. *Appl. Environ. Microbiol.* 56, 1162–1163.

Doubleday, Z.A., Pecl, G.T., Semmens, J.M., Danyushevsky, L., 2008. Stylet elemental signatures indicate population structure in a holobenthic octopus species, *Octopus pallidus*. *Mar. Ecol. Prog. Ser.* 371, 1–10. <https://doi.org/10.3354/meps07722>

Du, Y., Nomura, Y., Liu, Z., Huang, Z.Y., Dong, K., 2009. Functional expression of an arachnid sodium channel reveals residues responsible for tetrodotoxin resistance in invertebrate sodium channels. *J. Biol. Chem.* 284, 33869–33875.

Duboule, D., 2007. The rise and fall of Hox gene clusters. *Development*.  
<https://doi.org/10.1242/dev.001065>

Earl, D.A., Vonholdt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resources* 4, 259–361. <https://doi.org/10.1007/s12686-011-9548-7>

Estefanell, J., Socorro, J., Tuya, F., Izquierdo, M., Roo, J., 2011. Growth, protein retention and biochemical composition in *Octopus vulgaris* fed on different diets based on crustaceans and aquaculture by-products. *Aquaculture* 322–323, 91–98.

<https://doi.org/10.1016/j.aquaculture.2011.09.027>

Ferreira, A., da Silva, M.F., da Costa e Silva, L., Cruz, C.D., 2006. Estimating the effects of population size and type on the accuracy of genetic maps. *Genet. Mol. Biol.* 29, 187–192.  
<https://doi.org/10.1590/S1415-47572006000100033>

Fierst, J.L., 2015. Using linkage maps to correct and scaffold de novo genome assemblies: Methods, challenges, and computational tools. *Front. Genet.*  
<https://doi.org/10.3389/fgene.2015.00220>

Finn, J.K., 2015. Systematics of the blue-ringed octopuses (Octopodidae: Hapalochlaena) of Australia and the Indo-West Pacific. B. Abstr. Cephalop. Int. Advis. Council. Conf. 2015, Recent Adv. Cephalop. Sci. 236.

Finn, J.K., Lu, C.C., 2015. Assessing purported anti-tropical distributions of Australian blue-ringed octopuses (Octopodidae: Hapalochlaena). B. Abstr. Cephalop. Int. Advis. Council. Conf. 2015, Recent Adv. Cephalop. Sci. 237.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G.L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D.A., Necci, M., Nuka, G., Orengo, C.A., Park, Y., Pesseat, S., Piovesan, D., Potter, S.C., Rawlings, N.D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Xenarios, I., Yeh, L.-S., Young, S.-Y., Mitchell, A.L., 2016. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199.  
<https://doi.org/10.1093/nar/gkw1107>

Flachsenberger, W., Kerr, D.I.B., 1985. Lack of effect of tetrodotoxin and of an extract from the posterior salivary gland of the blue-ringed octopus following injection into the octopus and following application to its brachial nerve. *Toxicon* 23, 997–999.

Foll, M., Gaggiotti, O., 2008. A Genome-Scan Method to Identify Selected Loci Appropriate for

- Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 180, 977–993.  
<https://doi.org/10.1534/genetics.108.092221>
- Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates 3, 294–299.
- Fonseca, E.M., Duckett, D.J., Carstens, B.C., 2021. P2C2M.GMYC: An R package for assessing the utility of the Generalized Mixed Yule Coalescent model. *Methods Ecol. Evol.* 12, 487–493.  
<https://doi.org/10.1111/2041-210X.13541>
- Fry, B.G., Roelants, K., Norman, J.A., 2009. Tentacles of venom: toxic protein convergence in the Kingdom Animalia. *J. Mol. Evol.* 68, 311–321.
- Fu, H., Zheng, Z., Dooner, H.K., 2002. Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. *Proc. Natl. Acad. Sci. U. S. A.* 99, 1082–1087. <https://doi.org/10.1073/pnas.022635499>
- Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152.  
<https://doi.org/10.1093/bioinformatics/bts565>
- Gaboriau, T., Leprieur, F., Mouillot, D., Hubert, N., 2018. Influence of the geography of speciation on current patterns of coral reef fish biodiversity across the Indo-Pacific. *Ecography (Cop.)*. <https://doi.org/10.1111/ecog.02589>
- Gao, Y., Natsukari, Y., 1990. Karyological Studies on Seven Cephalopods. *Japanese J. Malacol.* 49, 126–145. [https://doi.org/10.18941/venusjmm.49.2\\_126](https://doi.org/10.18941/venusjmm.49.2_126)
- Geffeney, S.L., Fujimoto, E., Brodie III, E.D., Brodie Jr, E.D., Ruben, P.C., 2005. Evolutionary diversification of TTX-resistant sodium channels in a predator–prey interaction. *Nature* 434, 759.

- Geffeney, S.L., Williams, B.L., Rosenthal, J.J.C., Birk, M.A., Felkins, J., Wisell, C.M., Curry, E.R., Hanifin, C.T., 2019. Convergent and parallel evolution in a voltage-gated sodium channel underlies TTX-resistance in the Greater Blue-ringed Octopus: *Hapalochlaena lunulata*. *Toxicon* 170, 77–84.
- Glazer, A.M., Killingbeck, E.E., Mitros, T., Rokhsar, D.S., Miller, C.T., 2015. Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3 Genes, Genomes, Genet.* 5, 1463–1472.  
<https://doi.org/10.1534/g3.115.017905>
- Grant, W.S., Leslie, R.W., Bowen, B.W., 2005. Molecular genetic assessment of bipolarity in the anchovy genus *Engraulis*. *J. Fish Biol.* 67, 1242–1265. <https://doi.org/10.1111/j.1095-8649.2005.00820.x>
- Gruber, B., Unmack, P.J., Berry, O.F., Georges, A., 2018. dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol. Ecol. Resour.* 18, 691–699.
- Gueidan, C., Roux, C., Lutzoni, F., 2007. Using a multigene phylogenetic analysis to assess generic delineation and character evolution in Verrucariaceae (Verrucariales, Ascomycota). *Mycol. Res.* 111, 1145–1168. <https://doi.org/10.1016/j.mycres.2007.08.010>
- Guzik, M.T., Norman, M.D., Crozier, R.H., 2005. Molecular phylogeny of the benthic shallow-water octopuses (Cephalopoda: Octopodinae). *Mol. Phylogenet. Evol.* 37, 235–248.  
<https://doi.org/10.1016/j.ympev.2005.05.009>
- Hagen, I.J., Lien, S., Billing, A.M., Elgvin, T.O., Trier, C., Niskanen, A.K., Tarka, M., Slate, J., Sætre, G., Jensen, H., 2020. A genome-wide linkage map for the house sparrow (*Passer domesticus*) provides insights into the evolutionary history of the avian genome. *Mol. Ecol. Resour.* 20, 544–559. <https://doi.org/10.1111/1755-0998.13134>
- Han, S., Yuan, M., Clevenger, J.P., Li, C., Hagan, A., Zhang, X., Chen, C., He, G., 2018. A snp-based linkage map revealed QTLs for resistance to early and late leaf spot diseases in peanut



- (*Arachis hypogaea* L.). *Front. Plant Sci.* 9, 1012. <https://doi.org/10.3389/fpls.2018.01012>
- Hedgecock, D., Shin, G., Gracey, A.Y., van den Berg, D., Samanta, M.P., 2015. Second-generation linkage maps for the pacific oyster *Crassostrea gigas* reveal errors in assembly of genome scaffolds. *G3 Genes, Genomes, Genet.* 5, 2007–2019. <https://doi.org/10.1534/g3.115.019570>
- Higgins, K.L., Semmens, J.M., Doubleday, Z.A., Burrige, C.P., 2013. Comparison of population structuring in sympatric octopus species with and without a pelagic larval stage. *Mar. Ecol. Prog. Ser.* 486, 203–212. <https://doi.org/10.3354/meps10330>
- Hilbish, T.J., Mullinax, A., Dolven, S.I., Meyer, A., Koehn, R.K., Rawson, P.D., 2000. Origin of the antitropical distribution pattern in marine mussels (*Mytilus* spp.): Routes and timing of transequatorial migration. *Mar. Biol.* 136, 69–77. <https://doi.org/10.1007/s002270050010>
- Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42, 182–192.
- Hughes, T.P., Bellwood, D.R., Connolly, S.R., 2002. Biodiversity hotspots, centres of endemism, and the conservation of coral reefs. *Ecol. Lett.* 5, 775–784. <https://doi.org/10.1046/j.1461-0248.2002.00383.x>
- Hwang, D.F., Arakawa, O., Saito, T., Noguchi, T., Simidu, U., Tsukamoto, K., Shida, Y., Hashimoto, K., 1989. Tetrodotoxin-producing bacteria from the blue-ringed octopus *Octopus maculosus*. *Mar. Biol.* 100, 327–332.
- Ibáñez, C.M., Peña, F., Pardo-Gandarillas, M.C., Méndez, M.A., Hernández, C.E., Poulin, E., 2014. Evolution of development type in benthic octopuses: Holobenthic or pelago-benthic ancestor? *Hydrobiologia*. <https://doi.org/10.1007/s10750-013-1518-5>
- Ibáñez, C.M., Rezende, E.L., Sepúlveda, R.D., Avaria-Llautureo, J., Hernández, C.E., Sellanes, J., Poulin, E., Pardo-Gandarillas, M.C., 2018. Thorson’s rule, life-history evolution, and diversification of benthic octopuses (Cephalopoda: Octopodoidea). *Evolution (N. Y.)* 72,

1829–1839. <https://doi.org/10.1111/evo.13559>

Jacups, S.P., Currie, B.J., 2008. Blue-ringed octopuses: a brief review of their toxicology. *North. Territ. Nat.* 50–57.

Jeon, H.K., Kim, K.H., Eom, K.S., 2011. Molecular identification of *Taenia* specimens after long-term preservation in formalin. *Parasitol. Int.* 60, 203–205.  
<https://doi.org/10.1016/j.parint.2010.12.001>

Jereb, P., Roper, C., Norman, M., Finn, J., 2014. Cephalopods of the world. An annotated and illustrated catalogue of cephalopod species known to date,” in *Octopods and Vampire Squids*, Vol. 3,. Rome FAO.

Jeziorski, M.C., Greenberg, R.M., Anderson, P.A. V, 1997. Cloning of a putative voltage-gated sodium channel from the turbellarian flatworm *Bdelloura candida*. *Parasitology* 115, 289–296.

Jokiel, P., Martinelli, F.J., 1992. The Vortex Model of Coral Reef Biogeography. *J. Biogeogr.* 19, 449. <https://doi.org/10.2307/2845572>

Joly, S., Davies, T.J., Archambault, A., Bruneau, A., Derry, A., Kembel, S.W., Peres-Neto, P., Vamosi, J., Wheeler, T.A., 2014. Ecology in the age of DNA barcoding: The resource, the promise and the challenges ahead. *Mol. Ecol. Resour.* <https://doi.org/10.1111/1755-0998.12173>

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>

Jombart, T., Collins, C., 2015. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0.

Jones, D.B., Jerry, D.R., Khatkar, M.S., Raadsma, H.W., Zenger, K.R., 2013. A high-density SNP genetic linkage map for the silver-lipped pearl oyster, *Pinctada maxima*: a valuable

- resource for gene localisation and marker-assisted selection. *BMC Genomics* 14, 810.  
<https://doi.org/10.1186/1471-2164-14-810>
- Jones, O.R., Wang, J., 2010. COLONY: A program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour.* 10, 551–555. <https://doi.org/10.1111/j.1755-0998.2009.02787.x>
- Jost, M.C., Hillis, D.M., Lu, Y., Kyle, J.W., Fozzard, H.A., Zakon, H.H., 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Mol. Biol. Evol.* 25, 1016–1024.
- Kaneko, N., Kubodera, T., Iguchis, A., 2011. Taxonomic study of shallow-water octopuses (Cephalopoda: Octopodidae) in Japan and adjacent waters using mitochondrial genes with perspectives on octopus DNA barcoding. *Malacologia*.  
<https://doi.org/10.4002/040.054.0102>
- Kaneko, Y., Matsumoto, G., Hanyu, Y., 1997. TTX resistivity of Na<sup>+</sup> channel in newt retinal neuron. *Biochem. Biophys. Res. Commun.* 240, 651–656.  
<https://doi.org/10.1006/bbrc.1997.7696>
- Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., Flouri, T., 2017. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* 33, 1630–1638.  
<https://doi.org/10.1093/bioinformatics/btx025>
- Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C.F., Olason, P., Ellegren, H., 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol.* 23, 4035–4058.  
<https://doi.org/10.1111/mec.12810>
- Kim, B.-M., Kang, S., Ahn, D.-H., Jung, S.-H., Rhee, H., Yoo, J.S., Lee, J.-E., Lee, S., Han, Y.-H., Ryu, K.-B., 2018. The genome of common long-arm octopus *Octopus minor*. *Gigascience* 7,

giy119.

- Kokubugata, G., Nakamura, K., Forster, P.I., Hirayama, Y., Yokota, M., 2012. Antitropical distribution of *Lobelia* species (Campanulaceae) between the Ryukyu Archipelago of Japan and Oceania as indicated by molecular data. *Aust. J. Bot.* 60, 417–428. <https://doi.org/10.1071/BT11316>
- Kourakis, M.J., Martindale, M.Q., 2000. Combined-method phylogenetic analysis of Hox and ParaHox genes of the metazoa. *J. Exp. Zool.* 288, 175–191. [https://doi.org/10.1002/1097-010X\(20000815\)288:2<175::AID-JEZ8>3.0.CO;2-N](https://doi.org/10.1002/1097-010X(20000815)288:2<175::AID-JEZ8>3.0.CO;2-N)
- Lane, W.R., Sutherland, S., others, 1967. The ringed octopus bite: a unique medical emergency. *Med. J. Aust.* 2, 475–476.
- Laptikhovsky, V., Pereira, J., Salman, A., Arkhipov, A., Costa, A., 2009. A habitat-dependence in reproductive strategies of cephalopods and pelagophile fish in the Mediterranean Sea. *Boll. Malacol* 45, 95–102.
- Le Port, A., Pawley, M.D.M., Lavery, S.D., 2013. Speciation of two stingrays with antitropical distributions: Low levels of divergence in mitochondrial dna and morphological characters suggest recent evolution. *Aquat. Biol.* 19, 153–165. <https://doi.org/10.3354/ab00518>
- Lee, P.N., Callaerts, P., De Couet, H.G., Martindale, M.Q., 2003. Cephalopod Hox genes and the origin of morphological novelties. *Nature* 424, 1061–1065. <https://doi.org/10.1038/nature01872>
- Lefkaditou, E., Bekas, P., 2004. Analysis of beak morphometry of the horned octopus *Eledone cirrhosa* (Cephalopoda: Octopoda) in the Thracian Sea (NE Mediterranean). *Mediterr. Mar. Sci.* 5, 143–149. <https://doi.org/10.12681/mms.219>
- Leigh, J.W., Bryant, D., 2015. `popart` : full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. <https://doi.org/10.1111/2041-210X.12410>

- Leitwein, M., Guinand, B., Pouzadoux, J., Desmarais, E., Berrebi, P., Gagnaire, P.A., 2017. A dense brown trout (*Salmo trutta*) linkage map reveals recent chromosomal rearrangements in the *Salmo* genus and the impact of selection on linked neutral diversity. *G3 Genes, Genomes, Genet.* 7, 1365–1376. <https://doi.org/10.1534/g3.116.038497>
- Li, F., Bian, L., Ge, J., Han, F., Liu, Z., Li, X., Liu, Y., Lin, Z., Shi, H., Liu, C., Chang, Q., Lu, B., Zhang, S., Hu, J., Xu, D., Shao, C., Chen, S., 2020. Chromosome-level genome assembly of the East Asian common octopus (*Octopus sinensis*) using PacBio sequencing and Hi-C technology. *Mol. Ecol. Resour.* 20, 1572–1582. <https://doi.org/10.1111/1755-0998.13216>
- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Prepr.* arXiv1303.3997.
- Li, W., Godzik, A., 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>
- Lim, G.S., Balke, M., Meier, R., 2012. Determining species boundaries in a world full of rarity: Singletons, species delimitation methods. *Syst. Biol.* <https://doi.org/10.1093/sysbio/syr030>
- Lindgren, A.R., Anderson, F.E., 2018. Assessing the utility of transcriptome data for inferring phylogenetic relationships among coleoid cephalopods. *Mol. Phylogenet. Evol.* 118, 330–342.
- Ludt, W.B., Rocha, L.A., 2015. Shifting seas: The impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *J. Biogeogr.* <https://doi.org/10.1111/jbi.12416>
- Lukoschek, V., Waycott, M., Keogh, J.S., 2008. Relative information content of polymorphic microsatellites and mitochondrial DNA for inferring dispersal and population genetic structure in the olive sea snake, *Aipysurus laevis*. *Mol. Ecol.* 17, 3062–3077. <https://doi.org/10.1111/j.1365-294X.2008.03815.x>
- Lukoschek, V., Waycott, M., Marsh, H., 2007. Phylogeography of the olive sea snake, *Aipysurus*

- laevis (Hydrophiinae) indicates Pleistocene range expansion around northern Australia but low contemporary gene flow. *Mol. Ecol.* 16, 3406–3422. <https://doi.org/10.1111/j.1365-294X.2007.03392.x>
- Luu, K., Bazin, E., Blum, M.G.B., 2017. *pcadapt* : an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* 17, 67–77. <https://doi.org/10.1111/1755-0998.12592>
- Magarlamov, T.Y., Melnikova, D.I., Chernyshev, A. V., 2017. Tetrodotoxin-producing bacteria: Detection, distribution and migration of the toxin in aquatic systems. *Toxins (Basel)*. 9, 166. <https://doi.org/10.3390/toxins9050166>
- Manousaki, T., Tsakogiannis, A., Taggart, J.B., Palaiokostas, C., Tsaparis, D., Lagnel, J., Chatziplis, D., Magoulas, A., Papandroulakis, N., Mylonas, C.C., Tsigenopoulos, C.S., 2016. Exploring a nonmodel teleost genome through rad sequencing-linkage mapping in common pandora, *Pagellus erythrinus* and comparative genomic analysis. *G3 Genes, Genomes, Genet.* 6, 509–519. <https://doi.org/10.1534/g3.115.023432>
- Mäthger, L.M., Bell, G.R.R., Kuzirian, A.M., Allen, J.J., Hanlon, R.T., 2012. How does the blue-ringed octopus (*Hapalochlaena lunulata*) flash its blue rings? *J. Exp. Biol.* 215, 3752–3757.
- Matsui, T., Yamamori, K., Furukawa, K., Kono, M., 2000. Purification and some properties of a tetrodotoxin binding protein from the blood plasma of kusafugu, *Takifugu niphobles*. *Toxicon* 38, 463–468. [https://doi.org/10.1016/S0041-0101\(99\)00166-X](https://doi.org/10.1016/S0041-0101(99)00166-X)
- Matsumoto, T., Tanuma, D., Tsutsumi, K., Jeon, J.K., Ishizaki, S., Nagashima, Y., 2010. Plasma protein binding of tetrodotoxin in the marine puffer fish *Takifugu rubripes*. *Toxicon* 55, 415–420. <https://doi.org/10.1016/j.toxicon.2009.09.006>
- Meier, J., White, J., 2017. Handbook of clinical toxicology of animal venoms and poisons, *Handbook of Clinical Toxicology of Animal Venoms and Poisons*. <https://doi.org/10.1201/9780203719442>

- Mironov, A.N., 2006. Centers of marine fauna redistribution. *Entomol. Rev.* 86, S32--S44.
- Morse, P., Huffard, C.L., Meekan, M.G., McCormick, M.I., Zenger, K.R., 2018a. Mating behaviour and postcopulatory fertilization patterns in the southern blue-ringed octopus, *Haplochromis maculosa*. *Anim. Behav.* 136, 41–51.  
<https://doi.org/10.1016/j.anbehav.2017.12.004>
- Morse, P., Kjeldsen, S.R., Meekan, M.G., McCormick, M.I., Finn, J.K., Huffard, C.L., Zenger, K.R., 2018b. Genome-wide comparisons reveal a clinal species pattern within a holobenthic octopus—the Australian Southern blue-ringed octopus, *Haplochromis maculosa* (Cephalopoda: Octopodidae). *Ecol. Evol.* 8, 2253–2267.
- Nakamura, H., 1985. A Review of Molluscan Cytogenetic Information based on the CISMOCH : Computerized Index System for Molluscan Chromosomes : Bivalvia, Polyplacophora and Cephalopoda. *Venus (Japanese J. Malacol.* 44, 193–225.  
[https://doi.org/10.18941/venusjgm.44.3\\_193](https://doi.org/10.18941/venusjgm.44.3_193)
- Norman, M., Kubodera, T., 2006. Taxonomy and biogeography of an Australian subtropical octopus with Japanese affinities, in: *Proceedings of the 7th and 8th Symposia on Collection Building and Natural History Studies in Asia and the Pacific Rim.* pp. 171–189.
- Norman, M., Reid, A., 2000. *Guide to squid, cuttlefish and octopuses of Australasia.* CSIRO publishing.
- Norman, M.D., 2000. Cephalopods, a world guide: Pacific Ocean, Indian Ocean, Red Sea, Atlantic Ocean, Caribbean. *Arctic, Antarct.* 222–223.
- Norman, M.D., Finn, J.K., Hochberg, F.G., 2016. Family octopodidae. *Cephalopods world. An Annot. Illus. Cat. Cephalop. species known to date.*
- Norman, M.D., Hochberg, F.G., 2005. THE CURRENT STATE OF OCTOPUS TAXONOMY, *Phuket mar. biol. Cent. Res. Bull.*

- Overath, H., von Boletzky, S., 1974. Laboratory observations on spawning and embryonic development of a blue-ringed octopus. *Mar. Biol.* <https://doi.org/10.1007/BF00394369>
- Pellissier, L., Leprieur, F., Parravicini, V., Cowman, P.F., Kulbicki, M., Litsios, G., Olsen, S.M., Wisz, M.S., Bellwood, D.R., Mouillot, D., 2014. Quaternary coral reef refugia preserved fish diversity. *Science* (80-. ). 344, 1016–1019. <https://doi.org/10.1126/science.1249853>
- Pembleton, L.W., Cogan, N.O.I., Forster, J.W., 2013. *StAMPP* : an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* 13, 946–952. <https://doi.org/10.1111/1755-0998.12129>
- Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55, 595–609. <https://doi.org/10.1080/10635150600852011>
- Postlethwait, J.H., Johnson, S.L., Midson, C.N., Talbot, W.S., Gates, M., Ballinger, E.W., Africa, D., Andrews, R., Carl, T., Eisen, J.S., Horne, S., Kimmel, C.B., Hutchinson, M., Johnson, M., Rodriguez, A., 1994. A genetic linkage map for the zebrafish. *Science* (80-. ). 264, 699–703. <https://doi.org/10.1126/science.8171321>
- Price, N., Moyers, B.T., Lopez, L., Lasky, J.R., Grey Monroe, J., Mullen, J.L., Oakley, C.G., Lin, J., Ågren, J., Schridder, D.R., Kern, A.D., McKay, J.K., 2018. Combining population genomics and fitness QTLs to identify the genetics of local adaptation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 115, 5028–5033. <https://doi.org/10.1073/pnas.1719998115>
- Pucca, M.B., Knudsen, C., S Oliveira, I., Rimbault, C., A Cerni, F., Wen, F.H., Sachett, J., Sartim, M.A., Laustsen, A.H., Monteiro, W.M., 2020. Current Knowledge on Snake Dry Bites. *Toxins* (Basel). <https://doi.org/10.3390/toxins12110668>
- Puillandre, N., Modica, M. V., Zhang, Y., Sirovich, L., Boisselier, M.C., Cruaud, C., Holford, M., Samadi, S., 2012. Large-scale species delimitation method for hyperdiverse groups. *Mol. Ecol.* 21, 2671–2691. <https://doi.org/10.1111/j.1365-294X.2012.05559.x>



- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., others, 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
- Quach, N., Goodman, M.F., Shibata, D., 2004. In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. *BMC Clin. Pathol.* 4, 1–5.  
<https://doi.org/10.1186/1472-6890-4-1>
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rastas, P., 2017. Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33, 3726–3732.  
<https://doi.org/10.1093/bioinformatics/btx494>
- Reid, A., 2016. *Cephalopods of Australia and sub-Antarctic territories*. CSIRO Publishing.
- Richards, V.P., Henning, M., Witzell, W., Shivji, M.S., 2009. Species delineation and evolutionary history of the globally distributed spotted eagle ray (*Aetobatus narinari*). *J. Hered.* 100, 273–283. <https://doi.org/10.1093/jhered/esp005>
- Rocha, L.A., Bowen, B.W., 2008. Speciation in coral-reef fishes. *J. Fish Biol.*  
<https://doi.org/10.1111/j.1095-8649.2007.01770.x>
- Roper, C., 1983. An overview of cephalopod systematics: status, problems and recommendations. *Mem. Natl. Museum Victoria*.  
<https://doi.org/10.24199/j.mmv.1983.44.01>
- Rosa, R., Pissarra, V., Borges, F.O., Xavier, J., Gleadall, I., Golikov, A., Bello, G., Morais, L., Lishchenko, F., Roura, Á., Judkins, H., Ibáñez, C.M., Piatkowski, U., Vecchione, M., Villanueva, R., 2019. Global patterns of species richness in coastal cephalopods. *Front. Mar. Sci.* 6, 469. <https://doi.org/10.3389/fmars.2019.00469>

- Rosenberg, N.A., Burke, T., Elo, K., Feldman, M.W., Freidlin, P.J., Groenen, M.A.M., Hillel, J., Mäki-Tanila, A., Tixier-Boichard, M., Vignal, A., Wimmers, K., Weigend, S., 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics* 159, 699–713. <https://doi.org/10.1093/genetics/159.2.699>
- Rozewicki, J., Li, S., Amada, K.M., Standley, D.M., Katoh, K., 2019. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* 47, W5–W10.
- Ruder, T., Sunagar, K., Undheim, E.A.B., Ali, S.A., Wai, T.-C., Low, D.H.W., Jackson, T.N.W., King, G.F., Antunes, A., Fry, B.G., 2013. Molecular phylogeny and evolution of the proteins encoded by coleoid (cuttlefish, octopus, and squid) posterior venom glands. *J. Mol. Evol.* 76, 192–204.
- Sánchez-Molano, E., Cerna, A., Toro, M.A., Bouza, C., Hermida, M., Pardo, B.G., Cabaleiro, S., Fernández, J., Martínez, P., 2011. Detection of growth-related QTL in turbot (*Scophthalmus maximus*). *BMC Genomics* 12, 1–9. <https://doi.org/10.1186/1471-2164-12-473>
- Sansaloni, C., Petrolì, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., Kilian, A., 2011. Diversity Arrays Technology (DART) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proc.* 5, 1–2. <https://doi.org/10.1186/1753-6561-5-s7-p54>
- Schwentner, M., Timms, B. V., Richter, S., 2011. An integrative approach to species delineation incorporating different species concepts: A case study of Limnádopsis (Branchiopoda: Spinicaudata). *Biol. J. Linn. Soc.* 104, 575–599. <https://doi.org/10.1111/j.1095-8312.2011.01746.x>
- Semagn, K., Bjørnstad, A., Ndjioudjop, M.N., 2006. Principles, requirements and prospects of genetic mapping in plants. *African J. Biotechnol.* 5.
- Shaw, K.L., Lesnick, S.C., 2009. Genomic linkage of male song and female acoustic preference QTL underlying a rapid species radiation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9737–9742. <https://doi.org/10.1073/pnas.0900229106>

- Shetty, S., Griffin, D.K., Graves, J.A.M., 1999. Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosom. Res.* 7, 289–295.  
<https://doi.org/10.1023/A:1009278914829>
- Sheumack, D.D., Howden, M.E.H., Spence, I., 1984. Occurrence of a tetrodotoxin-like compound in the eggs of the venomous blue-ringed octopus (*Hapalochlaena maculosa*). *Toxicon* 22, 811–812.
- Simakov, O., Marlétaz, F., Yue, J.X., O’Connell, B., Jenkins, J., Brandt, A., Calef, R., Tung, C.H., Huang, T.K., Schmutz, J., Satoh, N., Yu, J.K., Putnam, N.H., Green, R.E., Rokhsar, D.S., 2020. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* 4, 820–830. <https://doi.org/10.1038/s41559-020-1156-z>
- Simon, C., Franke, A., Martin, A., 1991. The polymerase chain reaction: DNA extraction and amplification, in: *Molecular Techniques in Taxonomy*. Springer, pp. 329–355.
- Slater, G.S.C., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6, 1–11.
- Sohn, J. Il, Nam, J.W., 2018. The present and future of de novo whole-genome assembly. *Brief. Bioinform.* 19, 23–40. <https://doi.org/10.1093/bib/bbw096>
- Spencer, D.H., Sehn, J.K., Abel, H.J., Watson, M.A., Pfeifer, J.D., Duncavage, E.J., 2013. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *J. Mol. Diagnostics* 15, 623–633.  
<https://doi.org/10.1016/j.jmoldx.2013.05.004>
- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W., Smadja, C.M., 2017. Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Philos. Trans. R. Soc. B Biol. Sci.* 372. <https://doi.org/10.1098/rstb.2016.0455>

- Strasburg, J.L., Sherman, N.A., Wright, K.M., Moyle, L.C., Willis, J.H., Rieseberg, L.H., 2012. What can patterns of differentiation across plant genomes tell us about adaptation and speciation? *Philos. Trans. R. Soc. B Biol. Sci.* <https://doi.org/10.1098/rstb.2011.0199>
- Strugnell, J., Jackson, J., Drummond, A.J., Cooper, A., 2006. Divergence time estimates for major cephalopod groups: evidence from multiple genes. *Cladistics* 22, 89–96.
- Swofford, D.L., 2002. PAUP\*. Phylogenetic Analysis Using Parsimony (\* and Other Methods). Version 4., (Sinauer Associates Inc.: Sunderland, MA, USA.).
- Takeuchi, T., 2017. Molluscan Genomics: Implications for Biology and Aquaculture. *Curr. Mol. Biol. Reports* 3, 297–305. <https://doi.org/10.1007/s40610-017-0077-3>
- Takumiya, M., Kobayashi, M., Tsuneki, K., Furuya, H., 2005. Phylogenetic relationships among major species of Japanese coleoid cephalopods (Mollusca: Cephalopoda) using three mitochondrial DNA sequences. *Zoolog. Sci.* <https://doi.org/10.2108/zsj.22.147>
- Tanner, A.R., Fuchs, D., Winkelmann, I.E., Gilbert, M.T.P., Pankey, M.S., Ribeiro, Â.M., Kocot, K.M., Halanych, K.M., Oakley, T.H., Da Fonseca, R.R., 2017. Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. *Proc. R. Soc. B Biol. Sci.* 284, 2016–2818.
- Tiley, G.P., Burleigh, G., 2015. The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. *BMC Evol. Biol.* 15, 1–14. <https://doi.org/10.1186/s12862-015-0473-3>
- Tranter, D.J., Augustine, O., 1973. Observations on the life history of the blue-ringed octopus *Hapalochlaena maculosa*. *Mar. Biol.* 18, 115–128.
- Uribe, J.E., Zardoya, R., 2017. Revisiting the phylogeny of Cephalopoda using complete mitochondrial genomes. *J. Molluscan Stud.* 83, 133–144. <https://doi.org/10.1093/mollus/eyw052>

- Van Tuinen, M., Hedges, S.B., 2001. Calibration of avian molecular clocks. *Mol. Biol. Evol.* 18, 206–213. <https://doi.org/10.1093/oxfordjournals.molbev.a003794>
- Vaz-Pires, P., Seixas, P., Barbosa, A., 2004. Aquaculture potential of the common octopus (*Octopus vulgaris* Cuvier, 1797): A review. *Aquaculture* 238, 221–238. <https://doi.org/10.1016/j.aquaculture.2004.05.018>
- Velmurugan, J., Mollison, E., Barth, S., Marshall, D., Milne, L., Creevey, C.J., Lynch, B., Meally, H., McCabe, M., Milbourne, D., 2016. An ultra-high density genetic linkage map of perennial ryegrass (*Lolium perenne*) using genotyping by sequencing (GBS) based on a reference shotgun genome assembly. *Ann. Bot.* 118, 71–87. <https://doi.org/10.1093/aob/mcw081>
- Venkatesh, B., Lu, S.Q., Dandona, N., See, S.L., Brenner, S., Soong, T.W., 2005. Genetic basis of tetrodotoxin resistance in pufferfishes. *Curr. Biol.* 15, 2069–2072. <https://doi.org/10.1016/j.cub.2005.10.068>
- Villanueva, R., Vidal, E.A.G., Fernández-Álvarez, F., Nabhitabhata, J., 2016. Early mode of life and hatchling size in cephalopod molluscs: Influence on the species distributional ranges. *PLoS One*. <https://doi.org/10.1371/journal.pone.0165334>
- Voight, J.R., 1994. Morphological variation in shallow-water octopuses (Mollusca: Cephalopoda). *J. Zool.* 232, 491–504. <https://doi.org/10.1111/j.1469-7998.1994.tb01590.x>
- Voss, S.R., Kump, D.K., Walker, J.A., Shaffer, H.B., Voss, G.J., 2012. Thyroid hormone responsive QTL and the evolution of paedomorphic salamanders. *Heredity (Edinb)*. 109, 293–298. <https://doi.org/10.1038/hdy.2012.41>
- Wang, J., Zheng, X., 2017. Comparison of the genetic relationship between nine Cephalopod species based on cluster analysis of karyotype evolutionary distance. *Comp. Cytogenet.* 11, 477.
- Wang, L., Chua, E., Sun, F., Wan, Z.Y., Ye, B., Pang, H., Wen, Y., Yue, G.H., 2019. Mapping and Validating QTL for Fatty Acid Compositions and Growth Traits in Asian Seabass. *Mar.*

Biotechnol. 21, 643–654. <https://doi.org/10.1007/s10126-019-09909-7>

White, J., 2018. Clinical Toxicology of Blue Ringed Octopus Bites, in: Handbook of: Clinical Toxicology of Animal Venoms and Poisons. CRC Press, pp. 171–175.  
<https://doi.org/10.1201/9780203719442-14>

Whitelaw, B.L., Cooke, I.R., Finn, J., Da Fonseca, R.R., Ritschard, E.A., Gilbert, M.T.P., Simakov, O., Strugnell, J.M., 2020. Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion, and loss. *Gigascience* 9, 1–15.  
<https://doi.org/10.1093/gigascience/giaa120>

Whitelaw, B.L., Strugnell, J.M., Faou, P., Da Fonseca, R.R., Hall, N.E., Norman, M., Finn, J., Cooke, I.R., 2016. Combined Transcriptomic and Proteomic Analysis of the Posterior Salivary Gland from the Southern Blue-Ringed Octopus and the Southern Sand Octopus. *J. Proteome Res.* 15, 3284–3297. <https://doi.org/10.1021/acs.jproteome.6b00452>

Whitlock, M.C., Lotterhos, K.E., 2015. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. *Am. Nat.* 186, S24–S36.  
<https://doi.org/10.1086/682949>

Wiegand, P., Domhöver, J., Brinkmann, B., 1996. DNA-Degradation in formalinfixierten Geweben. *Pathologe* 17, 451–454. <https://doi.org/10.1007/s002920050185>

Willan, R.C., Willan, R.C., 2008. The taxonomic and nomenclatural status of the Northern Australian Greater Blue-ringed Octopus (*Hapalochlaena* sp.): a correction with potentially significant consequences. *North. Territ. Nat.* 20, 47–49.

Williams, B.L., Caldwell, R.L., 2009. Intra-organismal distribution of tetrodotoxin in two species of blue-ringed octopuses (*Hapalochlaena fasciata* and *H. lunulata*). *Toxicon* 54, 345–353.

Williams, B.L., Lovenburg, V., Huffard, C.L., Caldwell, R.L., 2011. Chemical defense in pelagic octopus paralarvae: Tetrodotoxin alone does not protect individual paralarvae of the greater blue-ringed octopus (*Hapalochlaena lunulata*) from common reef predators.

Chemoecology 21, 131.

Woodland, D.J., 1983. Zoogeography of the Siganidae (Pisces): an interpretation of distribution and richness patterns. Bull. Mar. Sci. 33, 713–717.

Wu, Y.J., Lin, C.L., Chen, C.H., Hsieh, C.H., Jen, H.C., Jian, S.J., Hwang, D.F., 2014. Toxin and species identification of toxic octopus implicated into food poisoning in Taiwan. Toxicon 91, 96–102. <https://doi.org/10.1016/j.toxicon.2014.09.009>

Yamate, Y., Takatani, T., Takegaki, T., 2021. Levels and distribution of tetrodotoxin in the blue-lined octopus *Hapalochlaena fasciata* in Japan, with special reference to within-body allocation. J. Molluscan Stud. 87, 42. <https://doi.org/10.1093/mollus/eyaa042>

Yasuhara, M., Iwatani, H., Hunt, G., Okahashi, H., Kase, T., Hayashi, H., Irizuki, T., Aguilar, Y.M., Fernando, A.G.S., Renema, W., 2017. Cenozoic dynamics of shallow-marine biodiversity in the Western Pacific. J. Biogeogr. <https://doi.org/10.1111/jbi.12880>

Yotsu-Yamashita, M., Mebs, D., Flachsenberger, W., 2007. Distribution of tetrodotoxin in the body of the blue-ringed octopus (*Hapalochlaena maculosa*). Toxicon 49, 410–412.

Yotsu-Yamashita, M., Sugimoto, A., Terakawa, T., Shoji, Y., Miyazawa, T., Yasumoto, T., 2001. Purification, characterization, and cDNA cloning of a novel soluble saxitoxin and tetrodotoxin binding protein from plasma of the puffer fish, *Fugu pardalis*. Eur. J. Biochem. 268, 5937–5946. <https://doi.org/10.1046/j.0014-2956.2001.02547.x>

Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y., Huang, L., Zheng, H., Xiang, J., 2015. Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific White Shrimp *Litopenaeus vannamei*. Sci. Rep. 5, 1–14. <https://doi.org/10.1038/srep15612>

Zarrella, I., Herten, K., Maes, G.E., Tai, S., Yang, M., Seuntjens, E., Ritschard, E.A., Zach, M., Styfhals, R., Sanges, R., 2019. The survey and reference assisted assembly of the *Octopus vulgaris* genome. Sci. data 6, 13.

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29, 2869–2876.

<https://doi.org/10.1093/BIOINFORMATICS/BTT499>

Zhang, Z.S., Hu, M.C., Zhang, J., Liu, D.J., Zheng, J., Zhang, K., Wang, W., Wan, Q., 2009.

Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Mol. Breed.* 24, 49–61.

<https://doi.org/10.1007/s11032-009-9271-1>



# APPENDICES

**SUPPLEMENTARY MATERIALS FOR CHAPTER 2: “Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion and loss”**

## **1. GENOME SEQUENCING, ASSEMBLY AND ANALYSES**

### **1.1. Library preparation**

### **1.2. Sequencing and assembly**

### **1.3. Assembly statistics**

### **1.4. Genome heterozygosity estimate**

### **1.5. Mutation rate**

### **1.6. Effective population size/Pairwise Sequentially Markovian Coalescent (PSMC) analysis**

## **2. TRANSCRIPTOME SEQUENCING AND ANALYSIS**

### **2.1. Tissues sampled, RNA preparation and sequencing**

### **2.2. Mapping reads to the genome for expression analysis**

### **2.3. *de novo* transcriptome assembly using *Trinity***

## **3. ANNOTATION OF TRANSPOSABLE ELEMENTS AND PROTEIN CODING GENES**

### **3.1. Annotation of protein coding genes**

### **3.2. Annotation completeness**

### **3.3. Transposable element annotation and expansions**

## **4. MULTI-GENE PHYLOGENY AND GENE FAMILY EXPANSION ANALYSES**

### **4.1. Multi-gene cephalopod phylogeny and dating**

### **4.2. Genome-wide gene family expansions**

## **5. ANALYSIS OF NEURAL ASSOCIATED GENE FAMILIES**

### **5.1. Zinc finger C2H2**

## **5.2. Cadherin/Protocadherin**

## **6. EVOLUTION OF THE VENOM/POSTERIOR SALIVARY GLAND IN OCTOPODS**

### **6.1. Extraction of tissue specific genes**

### **6.2. Examination of gene families and orthology**

### **6.3. Gene loss, shift**

### **6.4. Examination of selection and evolutionary rates in octopod serine proteases**

## **7. EVOLUTION OF TETRODOTOXIN RESISTANCE IN *H. MACULOSA***

### **7.1. Extraction and identification of Na<sub>v</sub> channels and p-loop regions**

### **7.2. Identification of tetrodotoxin (TTX) resistance mutation**

## **8. MICROBIOME OF THE *H. MACULOSA* POSTERIOR SALIVARY GLAND**

### **8.1. SAMSA**

## **1. GENOME SEQUENCING, ASSEMBLY AND ANALYSES**

### **1.1. Library preparation**

A single female specimen of *Hapalochlaena maculosa* was collected at Beaumaris Sea Scout Boat Shed, Beaumaris, Port Phillip bay, Australia (37°59'43.70"S 145°21.17"E) at a depth of 3m. The whole animal was stored at -80°C immediately after sampling. DNA was extracted from muscle tissue using QIAmp DNA mini kit. Illumina library preparation was conducted at the genome institute in Washington University, USA.

In order to increase coverage and improve the contiguity of contig assembly a total of 4g of arm and mantle tissue was submitted to DovetailGenomics which was then used to generate a Chicago™ library. Resulting Dovetail library concentration was 13.8nM and 4.2ng/ul with a mean size of 488nt and a total volume of 30ul.

### **1.2. Sequencing and assembly**

The Illumina library was sequenced using Illumina HiSeq 2000 at the genome institute in Washington University. Insert sizes for mate pair libraries can be found in the attached Supplementary data 2. Reads were trimmed with trimgalore ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) to remove adapters.

Several runs of Meraculous (meraculoususing) different kmer sizes were used to determine the optimal kmer (51). Dovetail sequencing was used to improve on the Illumina assembly. A HiRise™ assembly was conducted by Dovetail and compared against the original assembly (Table 2 & 3, Figure 1).

Table 1. Summary for Illumina libraries.

Read Type	Coverage at 4.3GB
All	39X
700bp insert	13X
3-8kb insert	19X

Table 2. Comparison of original Illumina and Dovetail augmented assemblies.

Estimated Chicago physical coverage (1-50 Kb pairs): 20.5 X

	Starting Assembly	Dovetail HiRise Assembly
Total Length	2,667.5 Mb	2,677.9 Mb
N50 Length	15,792 scaffolds ; min length 0.045 Mb	2,467 scaffolds ; min length 0.267 Mb
N90 Length	66,787 scaffolds ; min length 0.008 Mb	12,837 scaffolds ; min length 0.037 Mb

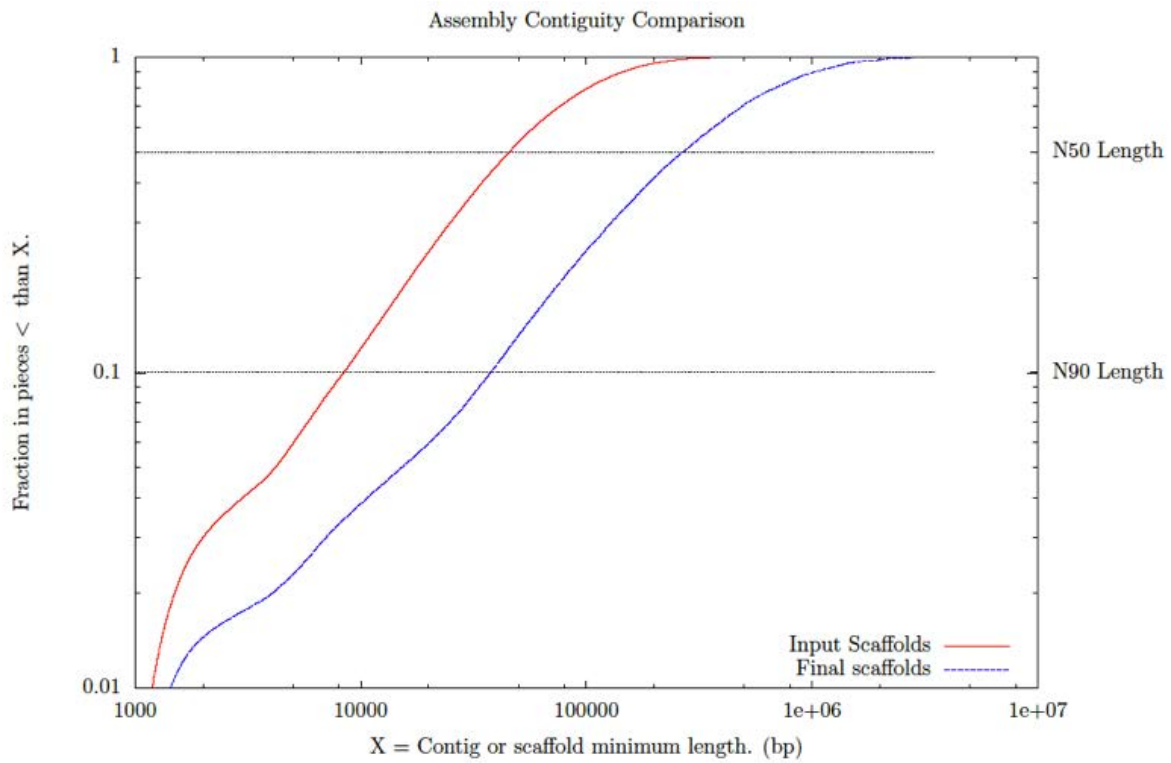


Figure 1. Comparison of assembly contiguity between original Illumina (input scaffolds) and Dovetail (Final scaffolds) augmented assemblies.

Comparative Assembly Statistics		
	Input Assembly	Dovetail HiRise Assembly
Longest scaffold	595698	3496984
Number of scaffolds	168388	64744
Number of scaffolds >1 kb	161468	57877
Contig N50	3.1 kb	3.1 kb
Number of gaps	490479	594146
Percent of genome in gaps	36.894%	37.139%

\* Note: Every join made by HiRise creates a gap.

Other Statistics	
Number of breaks made to input assembly by HiRise	172
Number of joins made by HiRise	103816
Number of gaps closed after HiRise	149
Chicago library 1 stats	365M read pairs; 2x151 bp

Table 3. Statistical comparisons between original Illumina and Dovetail augmented assemblies.

### 1.3. Assembly statistics

The *H. maculosa* genome assembly was assessed using assembly-stats

(<https://github.com/sanger-pathogens/assembly-stats>), *O. bimaculoides* and *C. minor*

assemblies were similarly examined.

Table 4. Assembly statistics for the three octopod genomes used in this study.

<b>Species</b>	<b><i>H. maculosa</i></b>	<b><i>C. minor</i></b>	<b><i>O. bimaculoides</i></b>
total_length	4009.6Mb	5090.35Mb	2371.52Mb
number	48,285	41,584	379,696
mean_length	0.083040Mb	0.122411Mb	0.006245Mb
longest	11.01Mb	3.03Mb	4.06Mb
shortest	771bp	4876bp	56bp
Gaps	4.32Mb	0	0.56Mb
N50	0.93Mb	0.29Mb	0.47Mb
N50n	1,044	3,696	1,369
N70	0.44Mb	0.2Mb	0.25Mb
N70n	2,299	6,735	2,776
N90	0.12Mb	0.09Mb	0.07Mb
N90n	5,607	11,789	6,239

#### 1.4 Genome heterozygosity estimate

*JELLYFISH* was used in conjunction with GenomeScope<sup>1</sup> to calculate heterozygosity in *H. maculosa* using a kmer frequency of 21.

Table 5. GenomeScope version 1.0 *H. maculosa* results.

Property	Min	Max
Heterozygosity	0.92%	0.97%
Genome haploid length	5,141.97 Mb	5,165.40 Mb
Genome repeat length	1,840.19 Mb	1,848.57 Mb
Genome unique length	3,301.78 Mb	3,316.82 Mb
Model fit	95.23%	99.35%
Read error rate	0.48%	0.48%

Table 6. Heterozygosity for published molluscan genomes.

Classification	Species	Heterozygosity	Publication
Cephalopod	<i>Hapalochlaena maculosa</i>	0.98%	This study
	<i>Octopus bimaculoides</i>	0.08%	<sup>2</sup>
	<i>Octopus vulgaris</i>	1.10%	<sup>3</sup>
Gastropod	<i>Elysia chlorotica</i>	3.66%	<sup>4</sup>



	<i>Pomacea canaliculata</i>	1-2%	5
Bivalve	<i>Pinctada fucata martensii</i>	2.5-3%	6
	<i>Saccostrea glomerata</i>	0.51%	7
	<i>Modiolus philippinarum</i>	2.02%	8
	<i>Bathymodiolus platifrons</i>	1.24%	8
	<i>Limnoperna fortunei</i>	2.30%	9
	<i>Chlamys farreri</i>	1.40%	10
	<i>Crassostrea gigas</i>	0.73%	11
	<i>Dreissena polymorpha</i>	2.13%	12

### 1.5 Mutation rate

Base neutral mutation rate was calculated between the *H. maculosa* and *O. bimaculoides* lineages with the assumption that the rate of mutations is equal to the rate of fixed differences between the two populations<sup>13</sup>. Orthologous genes from *O. bimaculoides* and *H. maculosa* were used. Neutrality was assumed for genes with very low expression (>10 TPM across all tissues). Neutral genes were aligned using *MAFFT*<sup>14</sup> and *codeml*<sup>15</sup> was used to calculate substitution metrics (dS). Per base neutral substitution between lineages was determined using the mean dS value divided by divergence time usually over number of generations, however *H. maculosa* is a single generation species. As octopus are diploid the rate was divided by two.

## 1.6 Effective population size (PSMC)

Historical changes in effective population size were estimated using Pairwise Sequentially Markovian Coalescent (PSMC) implemented in the software MSMC<sup>16,17</sup>. To generate inputs for MSMC we selected reads from libraries with short (500bp) insert sizes which provided 38x coverage of the genome. These were pre-processed according to GATK best practices; briefly, adapters were marked with Picard 2.2.1, reads were mapped to the *H. maculosa* genome using bwa mem (version 0.7.17) and PCR duplicates identified using Picard 2.2.1. In order to avoid inaccuracies due to poor coverage or ambiguous read mapping we masked regions where short reads would be unable to find unique matches using SNPable (<http://lh3lh3.users.sourceforge.net/snpsable.shtml>) and where coverage was more than double or less than half the genome wide average of 38x. Variant sites were called within unmasked regions and results converted to MSMC input format using msmc-tools <https://github.com/stschiff/msmc-tools>. All data for *H. maculosa* scaffolds of length greater than 1Mb was then used to generate 100 bootstrap replicates by dividing data into 500kb chunks and assembling them into 20 chromosomes with 100 chunks each. We then ran msmc2 on each bootstrap replicate and imported the resulting data into R for plotting. A mutation rate of 2.4e-9 per base per year and a generation time of 1 year were assumed in order to set a timescale in years and convert coalescence rates to effective population size.

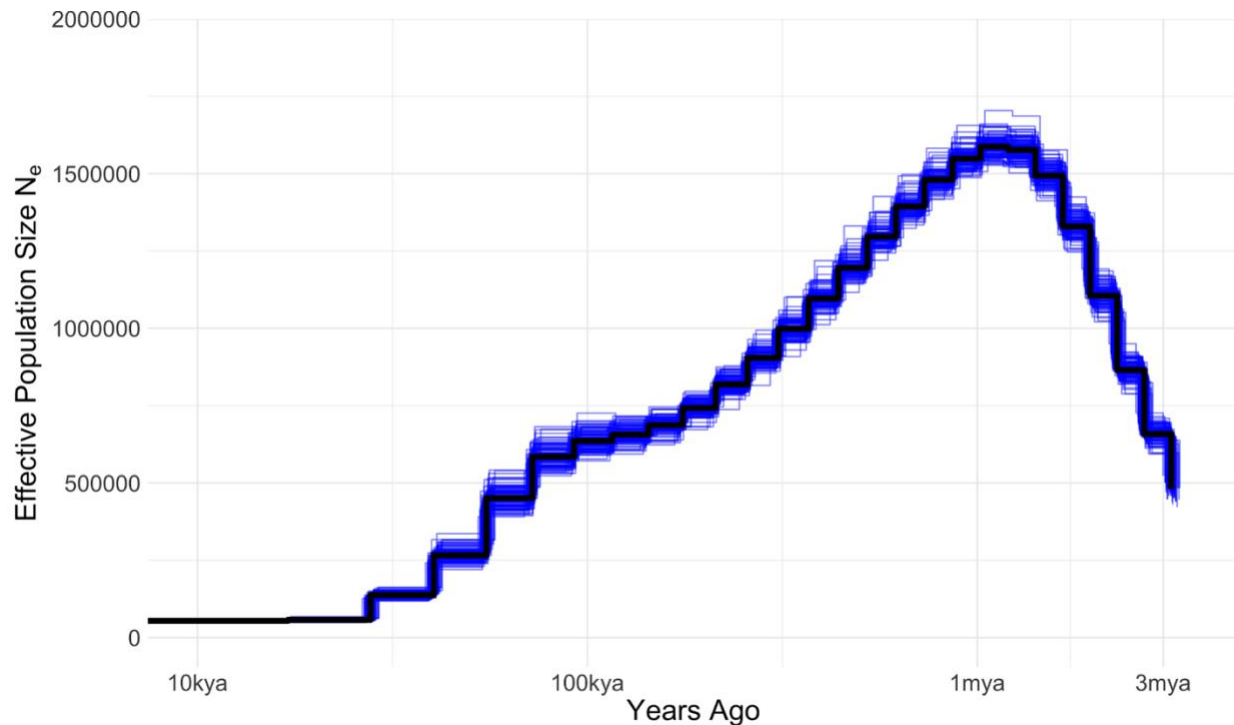


Figure 2. PSMC estimation of effective population size in *H. maculosa*.

## 2. TRANSCRIPTOME SEQUENCING AND ANALYSIS

### 2.1. Tissues sampled, RNA preparation and sequencing

Tissues were extracted from 12 tissues (brain, anterior salivary gland, digestive gland, renal, brachial heart, male reproductive tract, systemic heart, eyeballs, gills, posterior salivary gland, dorsal mantle and ventral mantle tissue) taken from a single *H. maculosa* individual collected at Beaumaris Sea Scout Boat Shed, Beaumaris, Port Phillip Bay, Australia (37°59'43.70"S 145°21.17"E). RNA was extracted using the Qiagen RNeasy mini kit according to the manufacturer's protocol. Extracted tissues were homogenised using a desktop homogeniser (IKA). Quality and quantity of samples was assessed using a Bioanalyser (nanochip)(Agilent).

Sequencing was conducted on a single lane of an Illumina HiSeq2000, with each forming 1/12th of a lane. Libraries were constructed using 3 µg of RNA at a concentration of >100 ng/µL.

## **2.2. Mapping reads to the genome for expression analysis**

Expression of gene models was determined using *kallisto*<sup>18</sup>, which performs pseudoalignments to determine the compatibility of transcripts. Transcripts used were filtered for low quality and adaptors removed prior to quantification using the *kallisto* package (<https://pachterlab.github.io/kallisto/>). Each tissue was examined separately, and abundances calculated in transcripts per million (TPM). Expression was calculated for all three octopus in this study using this method.

## **2.3. *de novo* transcriptome assembly using Trinity**

*H. maculosa* transcriptome assembly and annotation is described in a previous publication<sup>19</sup>. *De novo* assembly of the *H. maculosa* transcriptome was conducted using sequencing data from 11 tissues (brain, anterior salivary gland, digestive gland, renal, brachial heart, male reproductive tract, eyeballs, gills, posterior salivary gland, dorsal mantle and ventral mantle tissue) and Trinity (v10.11.201). Default parameters were used aside from kmer coverage, which was set to three to account for the large data volume. Protein coding sequences were identified using Trinotate<sup>15</sup> and domains assigned by Interpro<sup>20</sup>.

### 3. ANNOTATION OF TRANSPOSABLE ELEMENTS AND PROTEIN CODING GENES

#### 3.1. Annotation of protein coding genes

The first step in the annotation pipeline was identification of a training dataset suitable for training the *augustus* algorithm<sup>21</sup>. This data was derived using transcriptomic evidence via the *PASA* pipeline<sup>22</sup>. Transcripts were cleaned and filtered prior to mapping against the genome with both *BLAT*<sup>23</sup> and *GMAP*<sup>24</sup>. Validation of alignments required a percent identity of > 95% along 90% of the transcript length (default settings). ORFs were identified and extracted from the alignments using Transdecoder within the *PASA* package. ORF.gff extracted from *PASA* was converted to genbank format and split into a training subset (3,391 sequences) and a test subset (200 sequences). eTraining the *augustus* algorithm was performed using the training set of sequences and the results tested for sensitivity and specificity. The results were compared with gene prediction conducted using the gene sets for human and *Drosophila*. Optimization was then conducted using the *H. maculosa* training set for 12 rounds, after which the sensitivity and specificity were re-evaluated. Hint files were created for exons, introns and repeats to assist in gene model prediction by *augustus*<sup>21</sup>. Repeat hits were derived from the repeatmasker output, exons. Exon hints were identified by aligning transcripts against the genome and extracting alignments with a min identity of 92%, these results were filtered further using *pslCDnaFilter* to extract the best alignments (-localNearBest=0.005). Intron hints were extracted initially from *TRINITY* accepted\_hits.bam output using *bam2hints*. *Augustus* was run using these preliminary hints and an exon-exon junction database generated from the output. The pipeline has been documented in the git:[https://github.com/blwhitelaw/BRO\\_annotation](https://github.com/blwhitelaw/BRO_annotation).

Table 6. *H. maculosa* annotation statistics.

<b><i>H. maculosa</i> annotation</b>	
Primary transcripts (augustus)	22,530
Primary Transcripts (transcriptome based)	6,798
Alternative transcripts	4,031
total	33,359
<b>Primary transcript (augustus)</b>	
Average number of exons	6
Median exon length	116
Median intron length	3,671
<b>Gene model support</b>	
transcriptome support	25,257
Pfam annotation	11,578
PANTHER annotation	13,219

### 3.2 Annotation completeness (BUSCO)

Completeness of the genome was estimated using BUSCO<sup>25</sup>, which identified 87.7% complete and 7.5% fragmented genes against the metazoan database of 978 groups (Table 7).

Table 7. *H. maculosa* assembly assessed for completeness against the BUSCO Metazoan database.

Complete BUSCOs (C)	858	0.88%
Complete and single-copy BUSCOs (S)	760	0.78%
Complete and duplicated BUSCOs (D)	98	0.10%
Fragmented BUSCOs (F)	73	0.08%
Missing BUSCOs (M)	47	0.05%
Total BUSCO groups searched	978	

### 3.3 Transposable element annotation and expansions

*Repeatmodeller*<sup>26</sup> was used to create a *de novo* library of repeats for each of the three genomes (*H. maculosa*, *O. bimaculoides* and *C. minor*). Genomes were masked using the repeatmodeller library via *Repeatmasker*<sup>26</sup>.

Table 8. Summary for *H. maculosa* repeat annotation

Elements		Number of elements	Length	percentage of sequence
----------	--	--------------------	--------	------------------------

SINEs		1968743	301771845bp	7.53%
	ALUs	4	194bp	0.00%
	MIRs	118334	18379936bp	0.46%
LINEs		1099904	339826118bp	8.48%
	LINE1	37733	10860451bp	0.27%
	LINE2	159621	18598760bp	0.46%
	L3/CR1	89080	38967232bp	0.97%
LTR		162516	39307347bp	0.98%
	ERV_L	42	2749bp	0.00%
	ERV_L- MaLRs	3	176bp	0.00%
	ERV_classI	19710	1655988bp	0.04%
	ERV_classII	3265	161438bp	0.00%
DNA		2901024	400984061bp	10.00%
	hAT-Charlie	274907	40257595bp	1.00%
	TcMar- Tigger	16417	4508191bp	0.11%
Unclassified		1606419	314382396bp	7.84%



Total interspersed repeats		34.82%	1396271767bp	34.82%
Small RNA		343070	50368808bp	1.26%
Satellites		22693	5497063bp	0.14%
Simple repeats		2141863	107567262bp	2.68%
Low complexity		121842	8404129bp	0.21%

#### 4. MULTI-GENE PHYLOGENY AND GENE FAMILY EXPANSION ANALYSES

##### 4.1 Multi-gene cephalopod phylogeny and dating

The cephalopod phylogeny was constructed using the genomes (*Aplysia californica*<sup>27</sup>, *Lottia gigantea*, *Crassostrea gigas*, *Octopus bimaculoides*, *Callistoctopus minor* and *Hapalochlaena maculosa*) and transcriptomes (*Octopus kaurna*, *Octopus vulgaris*, *Sepia officinalis*<sup>28</sup> and *Idiosepius notoides*) where genomes were not available. Of the transcriptomes used, *Octopus kaurna* and *Octopus vulgaris* were sequenced and assembled in house using the same method previously described (Supplementary 2.1 & 2.3). Proteome pairs were identified using BLASTp in a mutual-best-hit approach with *O. bimaculoides* acting as the reference. We grouped together proteins from separate proteomes belonging to the same gene to form a cluster. Clusters were aligned individually with *MUSCLE*<sup>29</sup> and gapless alignments were ascertained by trimming using *Gblocks*<sup>30</sup> with default parameters. A total of 2,108 clusters were obtained. Phylogenies were constructed using *RAxML v8.0*<sup>31</sup> and divergence times estimated by *Phylobayes v4.1*<sup>32</sup>. *RAxML v8.0*<sup>31</sup> was run using the GTR+G+I model ascertained from *JmodelTest v2.1.10*. using the cAIC criterion for 100 bootstraps. *Phylobayes* estimated divergence times

under a strict clock with a mixture model of F81 + G with a burn-in of 10%. Calibrations were used as follows : divergence between *H. maculosa* and *E. scolopes* 275mya & divergence between *C. gigas* and *E. scolopes* 500mya. Two runs were performed and convergence verified using *bpcomp*, which confirmed a maximum difference of < 0.1 and *tracecomp*, which also indicated convergence with an effective sample size (EES) of > 200 for all parameters. Both programs used were from the *Phylobayes* package.

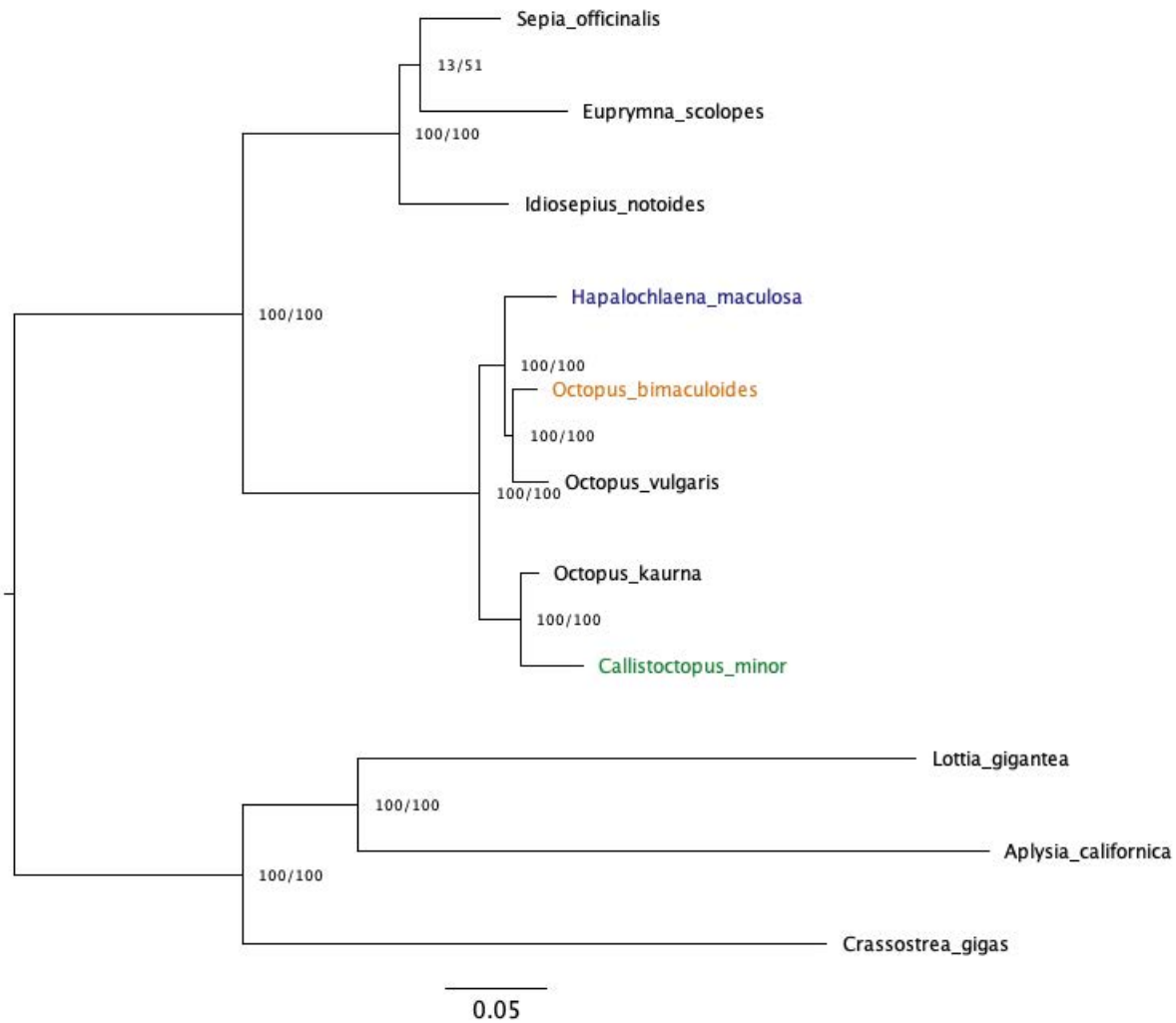


Figure 3. QI-TREE Maximum-likelihood tree.

#### 4.2 Genome-wide gene family expansions

In order to examine genome-wide gene family expansions Pfam<sup>31</sup> annotations were used to first categorize genes. Each gene was annotated using InterPro<sup>32</sup> and annotations were filtered to remove duplicates among each gene, which may contain several copies of a single domain. A total of 5565 Pfams were identified among six molluscan genomes (*Aplysia californica*, *Lottia gigantea*, *Crassostrea gigas*, *Octopus bimaculoides*, *Callistoctopus minor* and *Hapalochlaena*

*maculosa*). Significant expansions were identified by conducting an iterative Fisher's exact test in R<sup>33</sup> on a counts table. Counts of Pfams were compared between a select group (octopods) and the background (average of remaining molluscs). In addition, each of the three octopods were tested against the background to identify species specific expansions. Expansions were deemed significant if a minimum of half of the species showed a significant p-value (0.01) for the Pfam.

## 5. ANALYSIS OF NEURAL ASSOCIATED GENE FAMILIES

### 5.1 zinc finger C2H2

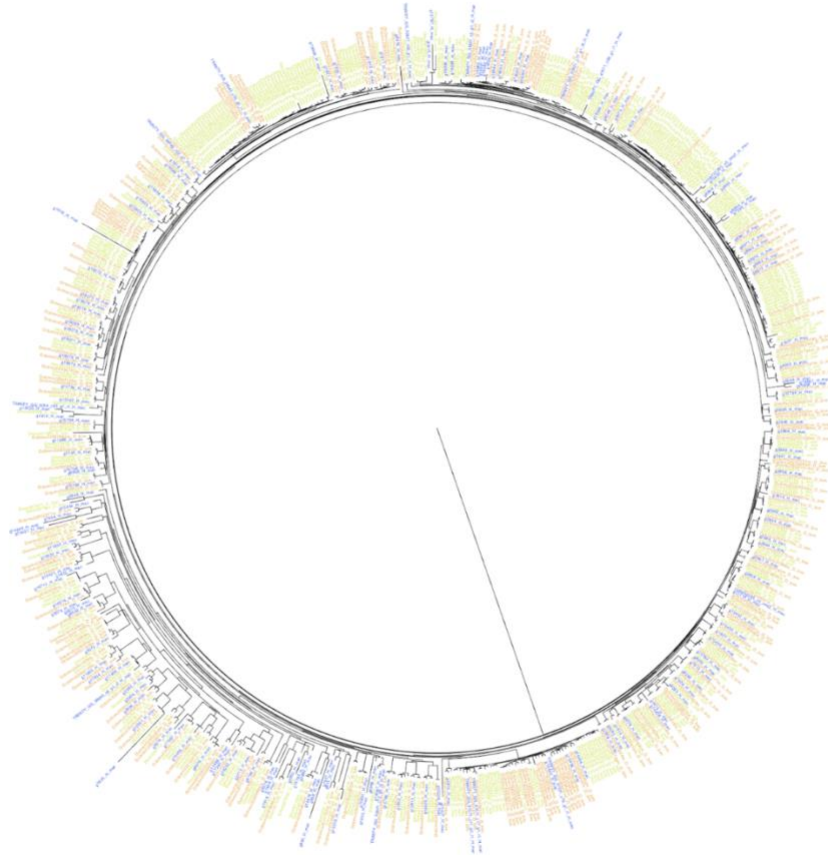
The zinc-finger gene family was examined separately due to issues with de novo annotation using a single method i.e augustus. Zinc fingers were previously dismissed as repetitive and required additional annotation methods in the previous *O. bimaculoides* genome publication<sup>1</sup>.

Members of the family were extracted from gene models annotated with the following Pfams (PF13894, PF09337, PF13912, PF00096, PF12874). However, to capture zinc fingers missed in the initial gene annotation in *H. maculosa* we examined the transcriptome for zinc finger proteins and mapped them back to the genome to identify exons. This approach could not be used in *C. minor* as all transcripts had a corresponding gene model. As an alternative, exonerate was used to identify potentially missed zinc finger genes in the *C. minor* annotation. zinc fingers identified in the previously published *O. bimaculoides* genome were aligned against the *C. minor* genome and filtered to attain matches which were > 80% identity.

## 5.2 Cadherin/Protocadherin

The cadherin family was found to be more complete than the zinc fingers and additional annotation steps were not conducted as it was not necessary. Cadherin genes were identified using the Pfams (PF08266, PF00028, PF08374). In order to identify members of the subfamily, protocadherin domains genes were examined for the number of domains. Genes with between 4-7 domains were manually classified into the protocadherin family (Figure 4). A total of 77, 164 and 116 were identified for *H. maculosa*, *C. minor* and *O. bimaculoides*, respectively.

Alignments were made for the cadherin and protocadherin families using MAFFT<sup>14</sup> and Fastree<sup>34</sup> was used to generate a phylogeny.



---

Figure 4. Phylogenetic tree of cadherins in *H. maculosa* (blue), *O. bimaculoides* (orange) and *C. minor* (green).

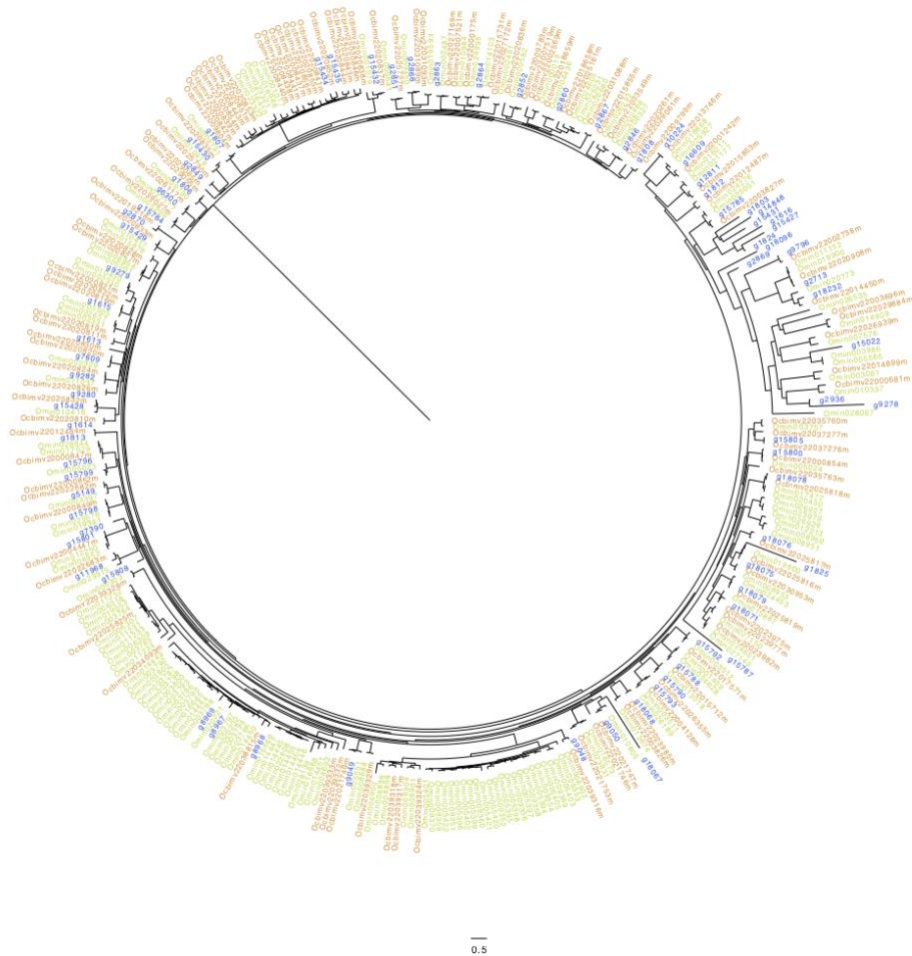


Figure 5. Phylogenetic tree of protocadherins in *H. maculosa* (blue), *O. bimaculoides* (orange) and *C. minor* (green).

## 6. EVOLUTION OF THE VENOM/POSTERIOR SALIVARY GLAND (PSG) IN OCTOPODS

### 6.1 Extraction of tissue specific genes

Specificity of gene expression was examined for all genes in *H. maculosa*, *C. minor* and *O. bimaculoides*. Expression of transcripts for each gene model within each tissue was calculated using Kallisto<sup>18</sup>. Expression was normalised prior to calculation of tau in R. A gene was deemed

to be specifically expressed in a tissue if the tau value was  $> 0.8$ . This cutoff has been previously used by <sup>35,36</sup> to classify a gene as specifically expressed or promiscuous. Overall trends shows peaks at low tau/highly generalised expression and high tau/specifically expressed within one tissue.

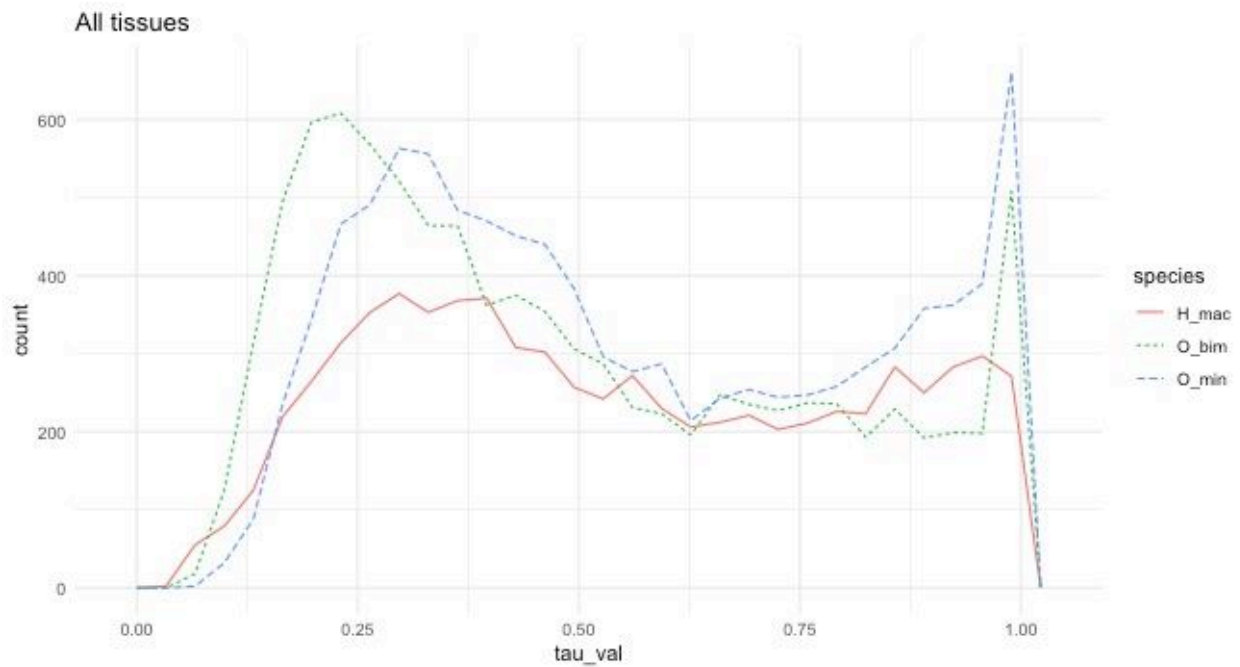


Figure 6. Distribution of tau values for genes in *H. maculosa*, *C. minor* and *O. bimaculoides*.

## 6.2 Examination of gene families and orthology

In order to identify orthologous groups between the three octopod genomes a combination of proteinortho and Orthovenn2<sup>37</sup> was used. Tau values for each gene were calculated in R to determine their level of fidelity to a tissue ( $>0.8$  = tissue specific) and Pfams used to categorise families. Specific tissues of interest shared between species (Brain and PSG) were then examined for their gene composition and expression.



### 6.3 Gene loss, shift

Patterns of gene expression, expansion and loss were examined between species. Using the defined orthologous groups, criteria were created to classify genes as shared/orthologous, specific to a species and loss of expression (expression must occur in one species and its ortholog unexpressed in another). The PSG was of particular interest in order to examine the impact of tetrodotoxin on PSG gene expression and composition. Genes with PSG specific expression in both *O. bimaculoides* and *C. minor* (non-TTX bearing) which had orthologs not specific to PSG in *H. maculosa* were examined. Additionally, the inverse of genes specific to the *H. maculosa* PSG and non-specific to the PSG of both *O. bimaculoides* and *C. minor* were compared.

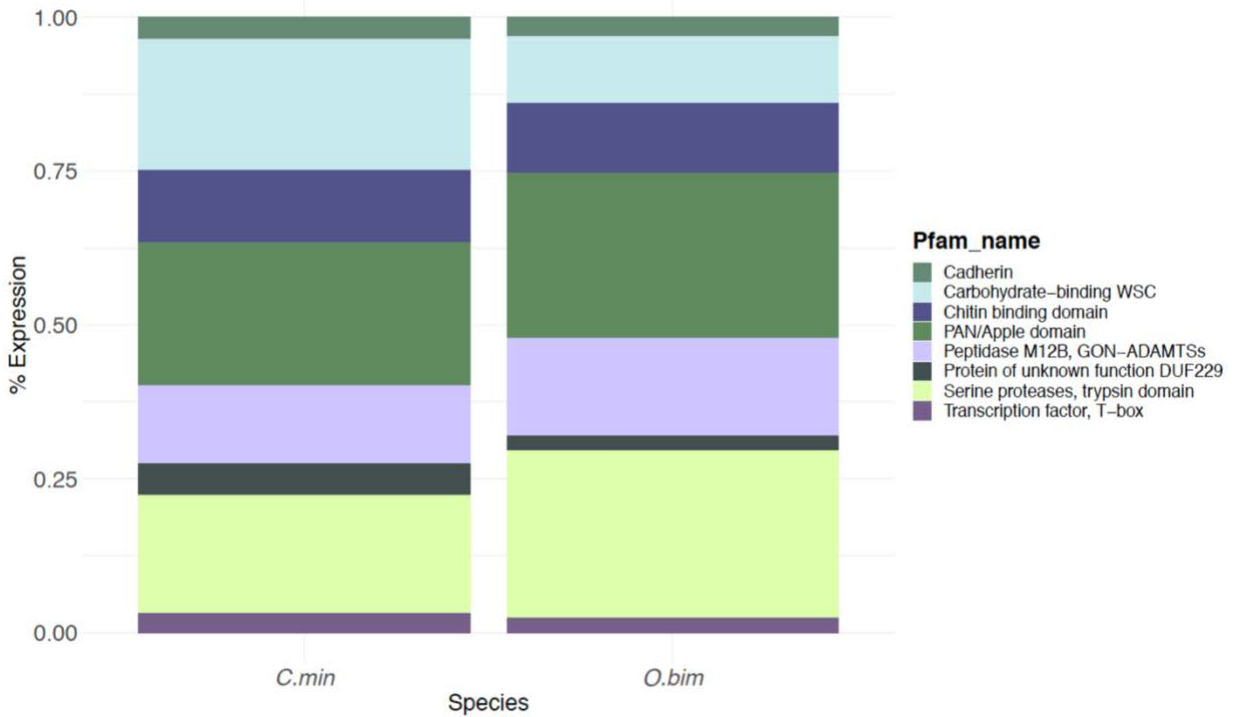


Figure 7. Orthologous genes specifically expressed in the PSG of *O. bimaculoides* and *C. minor* which have no ortholog in *H. maculosa*.

#### 6.4 Examination of selection and evolutionary rates in octopod serine proteases

Gene models (aa) from the three octopod genomes (*H. maculosa*, *O. bimaculoides* and *C. minor*) were annotated with Interproscan and serine proteases with the Pfam PF00089 extracted for examination. Gene models and their corresponding CDS sequences were imported into *Geneious v10.2.6* and selected for a single trypsin (PF00089) domain greater than 200aa/600bp long. The region containing the trypsin domain was then extracted from the nucleic acid sequences and *MAFFT v7.407* was used to align sequences using Translation align in *Geneious v10.2.6*, which interpreted the first codon as the start of the codon region and used the first translation frame. The resulting alignment was tested for an appropriate substitution

model in *jModelTest v2.2.10* and a tree was generated with *RAxML v8.0* using the GTR +G+I model and 100 bootstraps. The resulting tree and alignment were examined using *codeml* via *EasyCodeml v1.21* from the PAML package to examine non-synonymous to synonymous substitution rates for evidence of positive selection. We first used a site-based model which allows for  $\omega$  values to vary between sites along the protein. Comparison of the nested models (M1a-M2a) and (M7-M8) did not reveal any sites under positive selection ( $p > 0.05$ ). In order to access the potential for different rates of evolution within specific lineages we used a branch site model which allows for  $\omega$  values to vary between sites and branches. For the foreground a large clade of genes, majority of which were specifically expressed in the posterior salivary gland (PSG) was selected and compared to all other non-PSG specific genes. No sites among the foreground branches were significantly accelerated relative to the background. The last method implemented is similar to the branch site model, however, the rate along sites is constant and the rate between the background and foreground can differ. This also found no evidence of positive selection between the background and foreground lineages. It should be noted that serine proteases are a large and complex family and are due a more in-depth analysis in coleoid cephalopods, which could form a complete stand-alone study.

## **7. EVOLUTION OF TETRODOTOXIN RESISTANCE IN *H. MACULOSA***

### **7.1 Extraction and identification of Nav channels and p-loop regions**

Sodium channels ( $Na_v$ ) for the three octopus genomes along with all available in-house cephalopod transcriptomes were extracted manually using a series of BLAST searches against the nr database. Annotation was achieved using *interproscan*<sup>20</sup> and identification and

extraction of p-loop regions of the sodium channel alpha subunit were manually performed. Where sodium channels were incomplete alignment against related complete channels were used to extract the p-loop regions.

## 7.2 Identification of TTX resistance mutation

An assessment of current literature detailing the binding affinity of various mutations to the  $Na_v$  were used to assess TTX resistant mutations<sup>38-44</sup>. Two  $Na_v$  genes were identified in *H. maculosa* and found to contain mutations shared with the recently published *Haplochlaena lunulata*<sup>39</sup>, which have potential to inhibit TTX binding and provide resistance. Three substitutions (M1406T, D1669H and H1670S) were identified in  $Na_v1$ (Fig 8), the former of which is consistent with TTX resistance in pufferfish and garter snakes. A Met-Thr substitution in a TTX sensitive  $Na_v1.4$  rat channel was found to decreased binding affinity in pufferfish by 15-fold<sup>41</sup>. The latter two substitutions occur in the fourth p-loop at known TTX binding sites. Similar mutations at the same site in the turbellarian flatworm *Bdelloura candida* ( $BcNa_v1$ )<sup>41,45</sup> are predicted to inhibit TTX binding by preventing formation of a hydrogen bond<sup>38</sup>.

## 7.3 Patterns of resistance and expression in the Nav gene family

A total of two Nav genes were recovered for all three octopods ( $Na_v1$ ,  $Na_v2$ ), however  $Na_v2$  for *O. bimaculoides* and *C. minor* were incomplete. Function of the  $Na_v2$  gene remains unknown in cephalopods and is not believed to contribute to activation of the action potential<sup>39</sup>. However, expression data suggests it plays an important role in neural tissue with similar expression to

Nav<sub>v</sub>1 in the brain of all octopods species examined (Fig 10). Unlike the Nav<sub>v</sub>1 gene Nav<sub>v</sub>2 lacks the Met-Thr mutation believed to provide TTX resistance (Fig 9).










	<b>DI</b>	<b>DII</b>	<b>DIII</b>	<b>DIV</b>	
<b>consensus</b>	<b>DYWEN</b>	<b>EWIES</b>	<b>KGWYD</b>	<b>AGWDG</b>	
<i>Mya arenaria</i>	-----	---- <u>D</u> -	---I-	-----	
<i>Crassostrea gigas</i>	F-----	---Q-	---IE	-----	
<i>Mizuhopecten yessoensis</i>	-----	-----	--- <u>T</u> V	---S	
<i>Lottia gigantea</i>	---S-	-----	---V-	-----	
<i>Aplysia californica</i>	F--S-	-----	---I-	---SD	
<i>Loligo pealei</i>	-----	-----	---IN	-----	
<i>Euprymna scolopes</i>	-----	-----	---IN	-----	
<i>Dorytruthis pealeii</i>	-----	-----	---IN	-----	
<i>Doryteuthis opalescens</i>	-----	-----	---IN	-----	
<i>Dosidicus gigas</i>	-----	-----	---IN	-----	
<i>Grimpoteuthis</i>	-----	-----	---M-	-----	
<i>Callistoctopus minor</i>	-----	-----	-----	-----	
<i>Octopus bimaculoides</i>	-----	-----	---I-	-----	
<i>Octopus vulgaris</i>	-----	-----	---I-	-----	
<i>Hapalochlaena maculosa</i>	-----	-----	--- <u>T</u> E	---HS ☠	
<i>Hapalochlaena lunulata</i>	-----	-----	--- <u>T</u> E	---HS ☠	
<i>Taricha granulosa</i>	-----	----T	--- <u>T</u> -	---SD ☠ 	
<i>Tetraodon nigroviridis</i>	<u>C</u> -----	----N	--- <u>T</u> A	GG--Q ☠ 	
<i>Homo sapiens</i>	-----	---T	---M-	---G 	

Figure 8. Alignment of Nav1 p-loop regions

<b>Consensus</b>	<b>DI</b> <b>DYWEN</b>	<b>DII</b> <b>EWIES</b>	<b>DIII</b> <b>KGWXD</b>	<b>DIV</b> <b>AGWDG</b>
<i>Crassostrea gigas</i>	-F--D	----E	E--ME	---ND
<i>Mizuhopecten yessoensis</i>	-F--D	----P	E--ME	---ND
<i>Lottia gigantea</i>	-F--D	----P	E--ME	---ND
<i>Aplysia californica</i>	-F--D	----A	E--ME	---ND
<i>Loligo pealei</i>	----D	----P	E--ME	---ND
<i>Euprymna scolopes</i>			---ME	---ND
<i>Heterololigo bleekeri</i>	----D	----P	E--ME	---ND
<i>Grimpoteuthis</i>				---ND
<i>Callistoctopus minor</i>	----D	----P		
<i>Octopus bimaculoides</i>		----	---ME	---ND
<i>Hapalochlaena maculosa</i>	----	----	---ME	---ND

Figure 9. Alignment of Nav2 p-loop regions

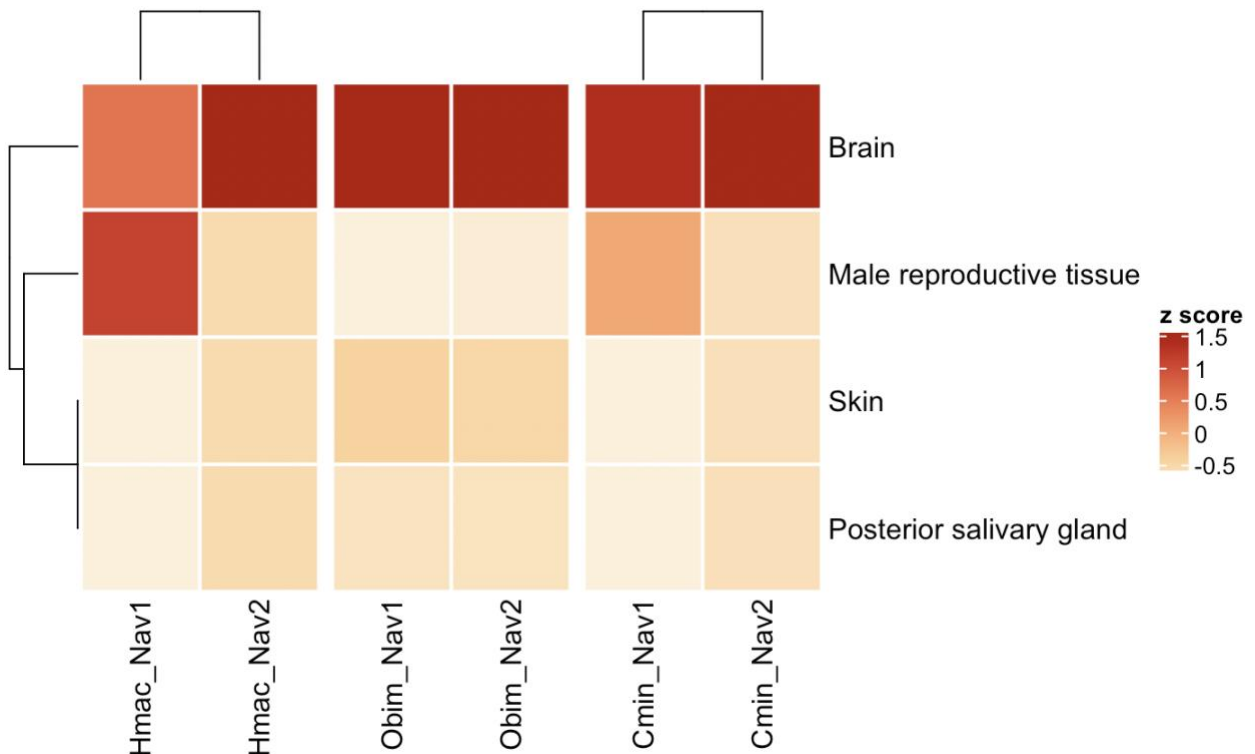


Figure 10. Expression of Nav1 and Nav2 channels across shared tissues of *H. maculosa* (Hmac), *O. bimaculoides* (Obim) and *C. minor* (Cmin).

## 8. MICROBIOME OF THE *H. MACULOSA* POSTERIOR SALIVARY GLAND

### 8.1 SAMSA

Bacterial composition in the PSG of *H. maculosa* was examined using a single ribo-depleted RNA sample. The SAMSA2 pipeline<sup>46</sup> was used to identify and quantify bacterial species. Reads were trimmed and filtered using trimmomatic<sup>47</sup> and sortMeRNA<sup>48</sup> was used to remove ribosomal DNA. As the reads were paired end PEAR<sup>49</sup> was used to merge reads prior to annotation. Annotation using DIAMOND in conjunction with the RefSeq(<ftp://ftp.ncbi.nlm.nih.gov/refseq/release/complete/>) and Subsystems(<ftp://ftp.theseed.org/subsystems/>) databases. Annotation and analyses using these databases was conducted in parallel. Analysis was limited by sample size however diversity indices were calculated in R using the SAMSA2 script “diversity\_graphs\_mod.R”.

## References

- 1 Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202-2204 (2017).
- 2 Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**, 220-224, doi:10.1038/nature14668 (2015).
- 3 Zarrella, I. *et al.* The survey and reference assisted assembly of the Octopus vulgaris genome. *Scientific data* **6**, 13 (2019).
- 4 Cai, H. *et al.* A draft genome assembly of the solar-powered sea slug Elysia chlorotica. *Scientific data* **6**, 190022 (2019).
- 5 Liu, C. *et al.* The genome of the golden apple snail Pomacea canaliculata provides insight into stress tolerance and invasive adaptation. *GigaScience* **7**, giy101 (2018).
- 6 Du, X. *et al.* The pearl oyster Pinctada fucata martensii genome and multi-omic analyses provide insights into biomineralization. *GigaScience* **6**, gix059 (2017).
- 7 Powell, D. *et al.* The genome of the oyster Saccostrea offers insight into the environmental resilience of bivalves. *DNA Research* **25**, 655-665 (2018).
- 8 Sun, J. *et al.* Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nature ecology & evolution* **1**, 0121 (2017).
- 9 Uliano-Silva, M. *et al.* A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, Limnoperna fortunei. *GigaScience* **7**, gix128 (2017).
- 10 Jiao, W. *et al.* High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA research* **21**, 85-101 (2013).
- 11 Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49 (2012).
- 12 Garton, D. W. & Haag, W. R. Heterozygosity, shell length and metabolism in the European mussel, Dreissena polymorpha, from a recently established population in Lake Erie. *Comparative Biochemistry and Physiology Part A: Physiology* **99**, 45-48 (1991).
- 13 Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624-626 (1968).
- 14 Rozewicki, J., Li, S., Amada, K. M., Standley, D. M. & Katoh, K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic acids research* **47**, W5-W10 (2019).
- 15 Adachi, J. & Hasegawa, M. *MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood.* (Institute of Statistical Mathematics Tokyo, 1996).
- 16 Li, H. & Durbin, R. Inference of human population history from whole genome sequence of a single individual. *Nature* **475**, 493 (2012).
- 17 Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nature genetics* **46**, 919 (2014).



- 18 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* **34**, 525-527 (2016).
- 19 Whitelaw, B. L. *et al.* Combined transcriptomic and proteomic analysis of the posterior salivary gland from the southern blue-ringed octopus and the southern sand octopus. *Journal of proteome research* **15**, 3284-3297 (2016).
- 20 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).
- 21 Stanke, M. *Gene prediction with a hidden Markov model*, University of Göttingen, (2004).
- 22 Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic acids research* **31**, 5654-5666 (2003).
- 23 Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome research* **12**, 656-664 (2002).
- 24 Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875 (2005).
- 25 Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution* **35**, 543-548 (2017).
- 26 Smit, A., Hubley, R & Green, P. *RepeatMasker Open-4.0.*, <<<http://www.repeatmasker.org>>. > (2013-2015).
- 27 Moroz, L. L. *et al.* The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510**, 109 (2014).
- 28 Bassaglia, Y. *et al.* ESTs library from embryonic stages reveals tubulin and reflectin diversity in *Sepia officinalis* (Mollusca—Cephalopoda). *Gene* **498**, 203-211 (2012).
- 29 Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics* **5**, 113 (2004).
- 30 Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology* **56**, 564-577 (2007).
- 31 Finn, R., Mistry, J., Tate, J., Coggill, P. & Heger, A. Pfam: the protein families database. *Nuclei. Acids Re.* (2014).
- 32 Finn, R. D. *et al.* InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Research* **45**, D190-D199, doi:10.1093/nar/gkw1107 (2016).
- 33 Team, R. C. R: A language and environment for statistical computing. (2013).
- 34 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution* **26**, 1641-1650 (2009).
- 35 Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Briefings in bioinformatics* **18**, 205-214 (2017).

- 36 Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. Tissue-specificity of gene expression diverges slowly between orthologs, and rapidly between paralogs. *PLoS computational biology* **12**, e1005274 (2016).
- 37 Xu, L. *et al.* OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic acids research* **47**, W52-W58 (2019).
- 38 Geffeney, S. L., Fujimoto, E., Brodie III, E. D., Brodie Jr, E. D. & Ruben, P. C. Evolutionary diversification of TTX-resistant sodium channels in a predator–prey interaction. *Nature* **434**, 759 (2005).
- 39 Geffeney, S. L. *et al.* Convergent and parallel evolution in a voltage-gated sodium channel underlies TTX-resistance in the Greater Blue-ringed Octopus: *Hapalochlaena lunulata*. *Toxicon* (2019).
- 40 Ulbricht, W., Wagner, H. H. & Schmidt Mayer, J. Kinetics of TTX-STX Block of Sodium Channels. *Annals of the New York Academy of Sciences* **479**, 68-83 (1986).
- 41 Jost, M. C. *et al.* Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Molecular biology and evolution* **25**, 1016-1024 (2008).
- 42 Du, Y., Nomura, Y., Liu, Z., Huang, Z. Y. & Dong, K. Functional expression of an arachnid sodium channel reveals residues responsible for tetrodotoxin resistance in invertebrate sodium channels. *Journal of Biological Chemistry* **284**, 33869-33875 (2009).
- 43 Feldman, C. R., Brodie, E. D. & Pfrender, M. E. Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. *Proceedings of the National Academy of Sciences* **109**, 4556-4561 (2012).
- 44 Frank, H. Y. & Catterall, W. A. Overview of the voltage-gated sodium channel family. *Genome biology* **4**, 207 (2003).
- 45 Jeziorski, M., Greenberg, R. & Anderson, P. Cloning of a putative voltage-gated sodium channel from the turbellarian flatworm *Bdelloura candida*. *Parasitology* **115**, 289-296 (1997).
- 46 Westreich, S. T., Korf, I., Mills, D. A. & Lemay, D. G. SAMSA: a comprehensive metatranscriptome analysis pipeline. *BMC bioinformatics* **17**, 399 (2016).
- 47 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
- 48 Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211-3217 (2012).
- 49 Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620 (2014).
- 52 Gao, F., Chen, C., Arab, D.A., Du, Z., He, Y. and Ho, S.Y., 2019. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecology and evolution*, *9*(7), pp.3891-3898.
53. Darriba, D., Taboada, G.L., Doallo, R. and Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, *9*(8), pp.772-772

**SUPPLEMENTARY MATERIALS FOR CHAPTER 3: “High density Genetic Linkage map of the southern blue-ringed octopus (*Hapalochlaena maculosa*)”**

Supplementary table 1. Recombination rates for shared loci intervals in female and male maps.

<b>interval_name</b>	<b>Rec</b>	<b>LOD</b>	<b>N</b>	<b>family_sex</b>
8777195_8793682	0	8.729	29	F131_M127_F
	0.047	4.575	21	F133_M132_M
8777838_16715831	0.033	7.126	30	F138_M133_F
	0.06	6.657	33	F144_M144_M
8778429_16717674	0	6.622	22	F145_M148_F
	0.029	8.275	34	F144_M144_M
8782658_16719272	0	2.709	9	F140_M149_F
	0	7.525	25	F140_M155_F
	0	5.418	18	F136_M145_M
8786079_8821975	0.1	4.795	30	F131_M127_F

	0	6.622	22	F145_M148_M
8788246_8819402	0.047	4.575	21	F133_M132_F
	0.333	0.811	33	F144_M144_M
8789815_8807862	0.031	7.7	32	F144_M144_F
	0.043	5.137	23	F136_M134_M
8790307_16719120	0	8.428	28	F131_M127_F
	0.03	7.987	33	F144_M144_M
8790488_8813350	0.5	0	18	F133_M132_F
	0.033	7.126	30	F144_M144_M
8791588_8818502	0	9.03	30	F138_M133_F
	0.047	4.575	21	F133_M132_M
8791778_16714510	0	7.525	25	F140_M155_F

	0.05	4.296	20	F136_M145_M
	0	6.622	22	F145_M148_M
8810238_16717194	0.095	3.453	21	F145_M148_F
	0	8.428	28	F138_M133_M
	0	6.321	21	F140_M155_M
8813074_16714430	0	6.622	22	F133_M132_F
	0.047	4.575	21	F136_M134_M
8820902_16718223	0.111	4.037	27	F138_M133_F
	0	6.622	22	F145_M148_M

---

Supplementary table 2. Significant G-values for segregation distortion test performed using LINKMFEX

Marker	G_value	N	family	sex	p_value
8797869	4.614	27	F131_M127	F	0.03171198
8777491	4.438	19	F133_M132	F	0.035147399
8779158	3.984	21	F133_M132	F	0.045934359
8780341	5.883	18	F133_M132	F	0.015287752
8782521	3.984	21	F133_M132	F	0.045934359
8782571	4.716	22	F133_M132	F	0.02988318
8788301	5.883	18	F133_M132	F	0.015287752
8813160	4.438	19	F133_M132	F	0.035147399
8814776	3.984	21	F133_M132	F	0.045934359
8819339	5.883	18	F133_M132	F	0.015287752
16716266	5.883	18	F133_M132	F	0.015287752
8802252	6.526	27	F138_M133	F	0.010630876
8819069	4.614	27	F138_M133	F	0.03171198
8781769	3.962	7	F140_M149	F	0.046538379
16716535	3.984	21	F140_M155	F	0.045934359

8790887	3.962	7	F140_M149	M	0.046538379
8801915	3.962	7	F140_M149	M	0.046538379
8779743	3.984	21	F145_M148	M	0.045934359
8798154	4.716	22	F145_M148	M	0.02988318
8801588	4.716	22	F145_M148	M	0.02988318
8819731	5.232	20	F145_M148	M	0.02217502
16715113	4.716	22	F145_M148	M	0.02988318
16716336	4.716	22	F145_M148	M	0.02988318
16716880	3.984	21	F145_M148	M	0.045934359

Supplementary table 3. HOX gene summary

Scaffold	Gene id	Hox Gene	Location	Scaffold/LG length
ScBNHFi_16706	g2744.t1	HOX1	Scaffold	360kb
ScBNHFi_31566	g9405.t1	LOX5	Scaffold	3.16Mb
ScBNHFi_22735	g5544.t1	ANTP/LOX2	Scaffold	8.53Mb
ScBNHFi_13142	g1245.t1	LOX4	LG_9	64Mb
ScBNHFi_16629	g2698.t1	POST1	LG_9	64Mb
ScBNHFi_31566	g9400.t1	SCR	LG_9	64Mb

Supplementary table 4. The top 30 Pfam domains for genes mapped within the 47 pseudo-chromosomes generated using chromonomer.

Description	Pfam ID	Number of genes
Cadherin domain	PF00028	51
WD domain, G-beta repeat	PF00400	42
Protein kinase domain	PF00069	37
Ankyrin repeats (3 copies)	PF12796	32
von Willebrand factor type A domain	PF00092	32
Zinc finger, C2H2 type	PF00096	32
Cadherin-like	PF08266	31
RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	PF00076	27
7 transmembrane receptor (rhodopsin family)	PF00001	24
Neurotransmitter-gated ion-channel ligand binding domain	PF02931	22
Helicase conserved C-terminal domain	PF00271	20
Neurotransmitter-gated ion-channel transmembrane region	PF02932	18
Ras family	PF00071	17
Major Facilitator Superfamily	PF07690	16
SAM domain (Sterile alpha motif)	PF00536	16



DEAD/DEAH box helicase	PF00270	15
Protein tyrosine kinase	PF07714	15
SAM domain (Sterile alpha motif)	PF07647	15
Homeobox domain	PF00046	14
Laminin G domain	PF02210	13
PH domain	PF00169	13
ABC transporter	PF00005	12
Leucine rich repeat	PF13855	12
BTB/POZ domain	PF00651	11
EF-hand domain pair	PF13499	11
PDZ domain (Also known as DHR or GLGF)	PF00595	11
Chitin binding Peritrophin-A domain	PF01607	10
Protein-tyrosine phosphatase	PF00102	10
Trypsin	PF00089	10

Supplementary table 5. Genome coverage estimation of sex average, female and male linkage maps.

<b>Method</b>	<b>Map</b>	<b>Ge result</b>	<b>Average between methods Ge1+Ge2/2</b>	<b>% coverage</b>
Ge1 = LG length * (marker number + 1/ marker number - 1)	Sex average	2091.75	2091.75	96.41
Ge2 = LG length +(2*ave_interval)	Sex average	2091.75		
Ge1 = LG length * (marker number + 1/ marker number - 1)	Male	1914.07	1914.06	97.22
Ge2 = LG length +(2*ave_interval)	Male	1914.06		
Ge1 = LG length * (marker number + 1/ marker number - 1)	Female	1949.44	1949.44	93.98
Ge2 = LG length +(2*ave_interval)	Female	1949.44		

**SUPPLEMENTARY MATERIALS FOR CHAPTER 4: “SNP data reveals the complex and diverse evolutionary history of the blue-ringed octopus genus (*Octopodidae: Hapalochlaena*) in the Asia-Pacific”**

Supplementary table 1. Details for specimens used in SNP analyses. Registration numbers of samples are from museum collections. Outlier test grouping indicate how samples were grouped to perform outlier analyses.

OTU_code	Sample number	Registration number	Collection location	Collection date	Location	Outlier test groupings
A	3043	NMV F236978 (SBD535496)	NE of Yeppoon, QLD	24-Nov-05	QLD	GLYE_Liz_Is_Townsville
A	3092	NMV F236978 (SBD535496)	NE of Yeppoon, QLD	24-Nov-05	QLD	GLYE_Liz_Is_Townsville
A	3023	NMV F230251	North West, WA (Deep) SS0507; Stn 017-006, WA	11-Jun-07	WA	North West, WA (Deep)
B	3012	AM C472122.001	136m shelf, WA	16-Jul-08	WA	WA_136m
C	3006	WAM S 43835	Ningaloo, WA	15-Aug-08	WA	Ningaloo
D	3089	Woolworths	Vietnam	?	Vietnam	Taiwan_Vietnam
D	3073	C.C. Lu (CL 239)	Taiwan	13-Jun-02	Taiwan	Taiwan_Vietnam
E	3066	NMV F230256	Taiwan	8-Aug-13	Taiwan	Taiwan
E	3065	NMV F230257	Taiwan	8-Aug-13	Taiwan	Taiwan
E	3064	C.C. Lu	Taiwan	5-Apr-11	Taiwan	Taiwan
F	476098.002	476098.002	off Metinaro,		Timore Leste	Timor_Leste
F	476099.002	476099.002	Off Metinaro		Timore Leste	Timor_Leste
G	DAR95	NTM DAR95	Darwin		NT	Darwin_Kimberly
G	3005	WAM S 58813	Long Reef, Kimberly, WA	21-Dec-10	WA	Darwin_Kimberly
G	3003	NMV F101643	Darwin, NT	29-Apr-95	NT	Darwin_Kimberly
G	3088	Darwin tissue	Darwin, NT		NT	Darwin_Kimberly

G	3009	WAM S 58662	Kimberly, WA	23-Oct-09	WA	Darwin_Kimberly
G	S92350	S92350	ExmouthGulf;PebbleBeach		WA	Exmouth
H	3037	QM Mo 80737	NE of Gladstone, QLD	14-Nov-05	QLD	GLYE_Liz_Is_Townsville
H	3032	QM Mo 80732	E of Yeppoon, QLD	21-Sep-04	QLD	GLYE_Liz_Is_Townsville
H	3042	NMV F236977 (SBD535415)	E of Mackay, QLD	3-Dec-05	QLD	Mackay
H	3016	NMV F230253	Lizard Is., QLD	2003	QLD	GLYE_Liz_Is_Townsville
H	3033	QM Mo 80733	NE of Gladstone, QLD	20-Sep-04	QLD	GLYE_Liz_Is_Townsville
H	3040	NMV F236975 (SBD529551)	E of Yeppoon, QLD	20-Nov-05	QLD	GLYE_Liz_Is_Townsville
H	3029	QM Mo 80725	E of Yeppoon, QLD	22-Sep-04	QLD	GLYE_Liz_Is_Townsville
H	3019	QM Mo 80731	E of Yeppoon, QLD	20-Sep-04	QLD	GLYE_Liz_Is_Townsville
H	3028	QM Mo 80724	NE of Townsville, QLD	25-Sep-03	QLD	GLYE_Liz_Is_Townsville
H	3020	QM Mo 80730	E of Mackay, QLD	17-Sep-04	QLD	Mackay
H	3030	QM Mo 80728	E of Yeppoon, QLD	22-May-04	QLD	GLYE_Liz_Is_Townsville
H	3031	QM Mo 80729	E of Mackay, QLD	28-May-04	QLD	Mackay
I	3082		CapeThreePoints		NSW	NSW_Brisbane_Redcliffe
I	3001	NMV F230252	Redcliffe, QLD		QLD	NSW_Brisbane_Redcliffe
I	483774.001	483774.001	Sydney:Manly		NSW	NSW_Brisbane_Redcliffe
I	483770.001	483770.001	Sydney,NorthHarbour		NSW	NSW_Brisbane_Redcliffe
I	483771.001	483771.001	Sydney,NorthHarbour		NSW	NSW_Brisbane_Redcliffe
I	3080	NMV F239775	Scarborough Spit, QLD	29-Jul-96	QLD	NSW_Brisbane_Redcliffe
I	483769.001	483769.001	Sydney:Manly		NSW	NSW_Brisbane_Redcliffe
I	483773.001	483773.001	Manly,DelwoodBeach		NSW	NSW_Brisbane_Redcliffe
I	483772.001	483772.001	Sydney,BotanyBay		NSW	NSW_Brisbane_Redcliffe
J	S110063	S110063	DampierArchipelago		WA	SharkBay_DampierA

J	3039	NMV F236974 (TS8017594)	Torres Strait	17-Apr-05	Torres Strait	Cape_York_TS
J	3018	QM Mo 80727	E of Cape York Peninsula, QLD	8-Feb-05	QLD	Cape_York_TS
J	3011	WAM S 15873	Shark Bay, WA	25-Sep-03	WA	SharkBay_DampierA
K	S82560	S82560	Albany, WA		WA	Albany
K	3090	Dicyemid JF030	Rye, VIC	27-Aug-96	VIC	VIC
K	SWA017	SWA017	Woodmansjetty		WA	Woodmans
K	3091	Valeria specimen	St Leonards, VIC	12-Jan-01	VIC	VIC
K	3054	NMV F221413	Woodmans Point, WA	28-Apr-15	WA	Woodmans
K	3048	NMV F160305	Albany, WA	29-Apr-07	WA	Albany
K	3055	NMV F221414	Woodmans Point, WA	28-Apr-15	WA	Woodmans
K	3079	NMV F80709	Portsea, VIC	5-Dec-96	VIC	VIC
K	3024	NMV F164728	Edithburgh, SA	26-May-97	SA	SA
K	3047	NMV F160314	Albany, WA	28-Apr-07	WA	Albany
K	3056	NMV F221415	Woodmans Point, WA	28-Apr-15	WA	Woodmans
K	3078	NMV F80708	Portsea, VIC	5-Dec-96	VIC	VIC
K	3046	NMV F160301	Woodmans Point, WA	24-Apr-07	WA	Woodmans
K	3049	NMV F160317	Albany, WA	29-Apr-07	WA	Albany
K	3045	NMV F160303	Rockingham, WA	23-Apr-07	WA	Rockmans
K	3050	NMV F160323	Albany, WA	1-May-07	WA	Albany
K	3025	NMV F164666	Edithburgh, SA	26-May-97	SA	SA
K	3022	NMV (JF MN frozen)	Edithburgh Jetty, SA	15-Dec-08	SA	SA

Supplementary table 2. Samples shared between SNP and mitochondrial data sets. Sequences obtained for a mitochondrial genes are indicated with \*. *Haplochlæna* NCBI sequence names are structured as follows: accession\_species recorded (HLU/*H. lunulata*, HFA/*H. fasciata* and HMA/*H. maculosa*)\_location\_author\_year.

Samples with corresponding SNP data	12S	16S	COI	COIII	Cytb	OTU
3023_North West, WA (Deep)_WA			*	*		A
3012_WA_136mshelf	*	*	*	*	*	B
3006_Ningaloo	*	*	*	*	*	C
3007_Ningaloo_egg	*	*	*	*	*	C
Vietnam_JF057			*			D
3003_Darwin	*		*	*		G
3005_Kimberly	*	*	*	*		G
3008_Kimberly	*		*	*		G
3009_Kimberly	*	*	*	*		G
Dampier_Region			*			G
3019_Yeppoon			*			H
3020_Mackay	*		*	*		H
3011_Shark_bay			*			J
3018_Cape_York	*		*	*	*	J

3022_Edi_SA	*	*	*	*	*	K
3024_Edi_SA	*		*			K
3025_Edi_SA	*		*			K
3078_VIC	*	*	*	*	*	K
3079_VIC	*	*	*		*	K
Albany_CT128			*			K
Albany_CT138			*			K
Albany_CT137			*			K
Albany_CT139			*			K
Rockingham_CT115			*			K
Woodmans_CT120			*			K
<b>Additional NCBI seq</b>	<b>12S</b>	<b>16S</b>	<b>COI</b>	<b>COIII</b>	<b>Cytb</b>	<b>COI NCBI accession</b>
AJ628175.1_HSP1_Darwain_Guzik2005	*		*	*	*	AJ628175
GQ900736.1_HLU_Cebu_Phillip_Huffard2010		*	*			GQ900736
JX268600.1_HLU_Pos_Taiwan_Hwang2012			*		*	JX268600
MT214021_HLU_China_Xu_2020			*			MT214021
MT214005_HLU_China_Xu_2020			*			MT214005

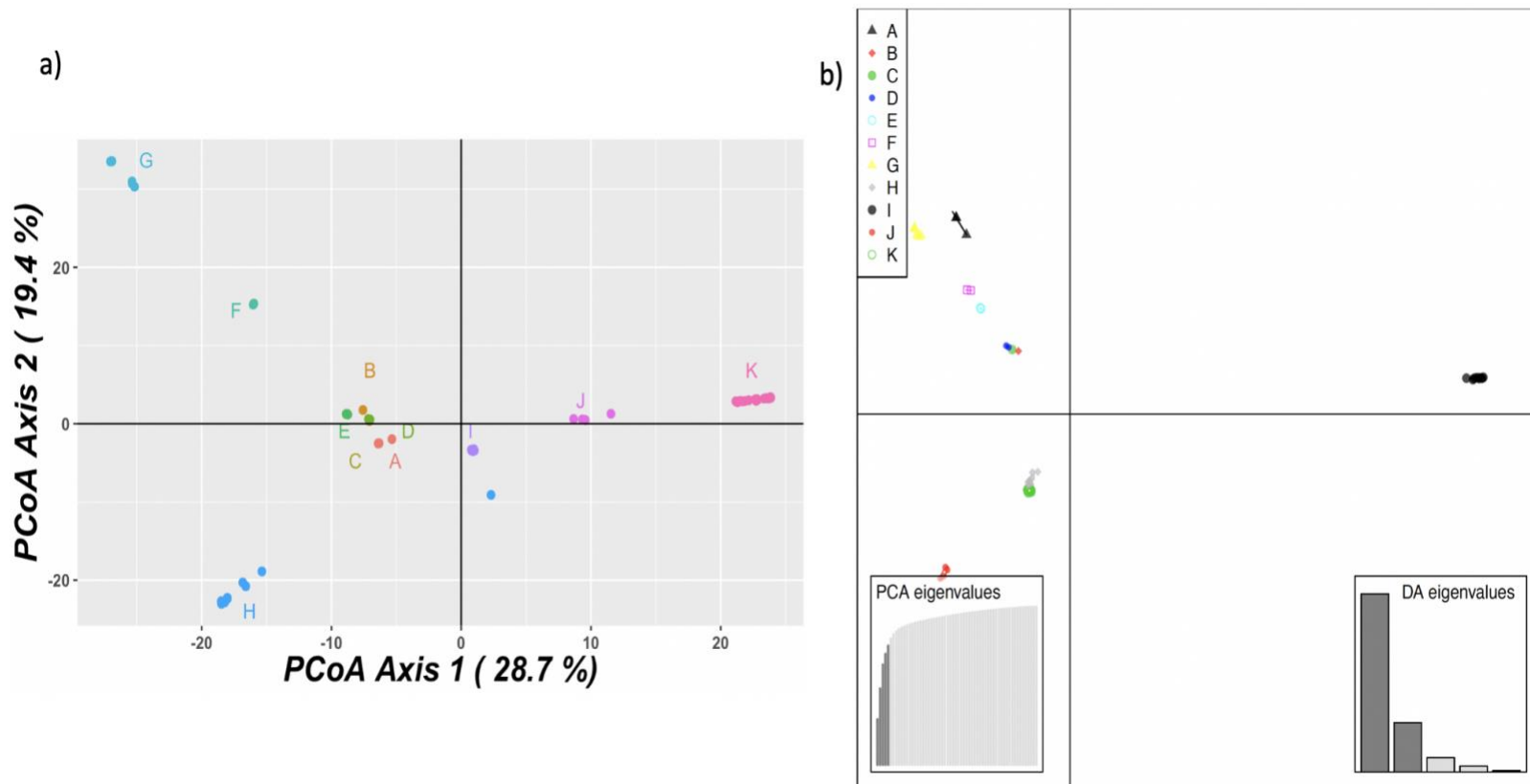
MT214012_HLU_China_Xu_2020			*			MT214012
MT214055_HLU_China_Xu_2020			*			MT214055
MT214056_HLU_China_Xu_2020			*			MT214056
LC553363_HFA_Indonesia_Afiati_2020			*			LC553363
LC553364_HFA_Indonesia_Afiati_2020			*			LC553364
LC553365_HFA_Indonesia_Afiati_2020			*			LC553365
MW514190_HFA_Bandladesh_Ahmed_at_al_2021			*			MW514190
AB430530.1_HLU_RK1_Okinawa_Kaneko2011			*	*		AB430530
HQ846163.1_HMA_Lingao_China_Dai2012		*	*			HQ846163
MT214022_HMA_China_Xu_2020			*			MT214022
AB430529.1_HFA_Taiwan_Kaneko2011			*	*		AB430529
MF440346.1_HFA_Jeju_Korea_Kim2018			*			MF440346
MN263864.1_HFA_Koh_2019			*			MN263864
NC_051545_HFA_Korea_Kim_et_al_2020			*			NC51545
JN790685.1_HFA_Taiwan_Wu2014			*			JN790685
MT213998_HMA_China_Xu_2020			*			MT213998
AJ628176.1_HMA_VIC_Guzik2005	*		*		*	AJ628176



AF000043.1_HMA_len_VIC_Carlini1999			*			AF000043
GQ900735.1_HFA_NSW_Huffard2010		*	*			GQ900735
AJ628210.1_HFA_MoretonBay_Guzik2005	*	*	*	*	*	AJ628210
Amphioctopus_aegina_KX108697	*	*	*	*	*	KX108697
Amphioctopus_marginatus_KY646153	*	*	*	*	*	KY646153
Amphioctopus_fangsiao_AB240156	*	*	*	*	*	AB240156
Octopus_bimaculoides_KU295559	*	*	*	*	*	KU295559
Octopus_vulgaris_AB052253	*	*	*	*	*	AB052253
Callistoctopus_luteus_NC_039848	*	*	*	*	*	NC39848
Argonauta_nodosa_MK034303	*	*	*	*	*	MK034303

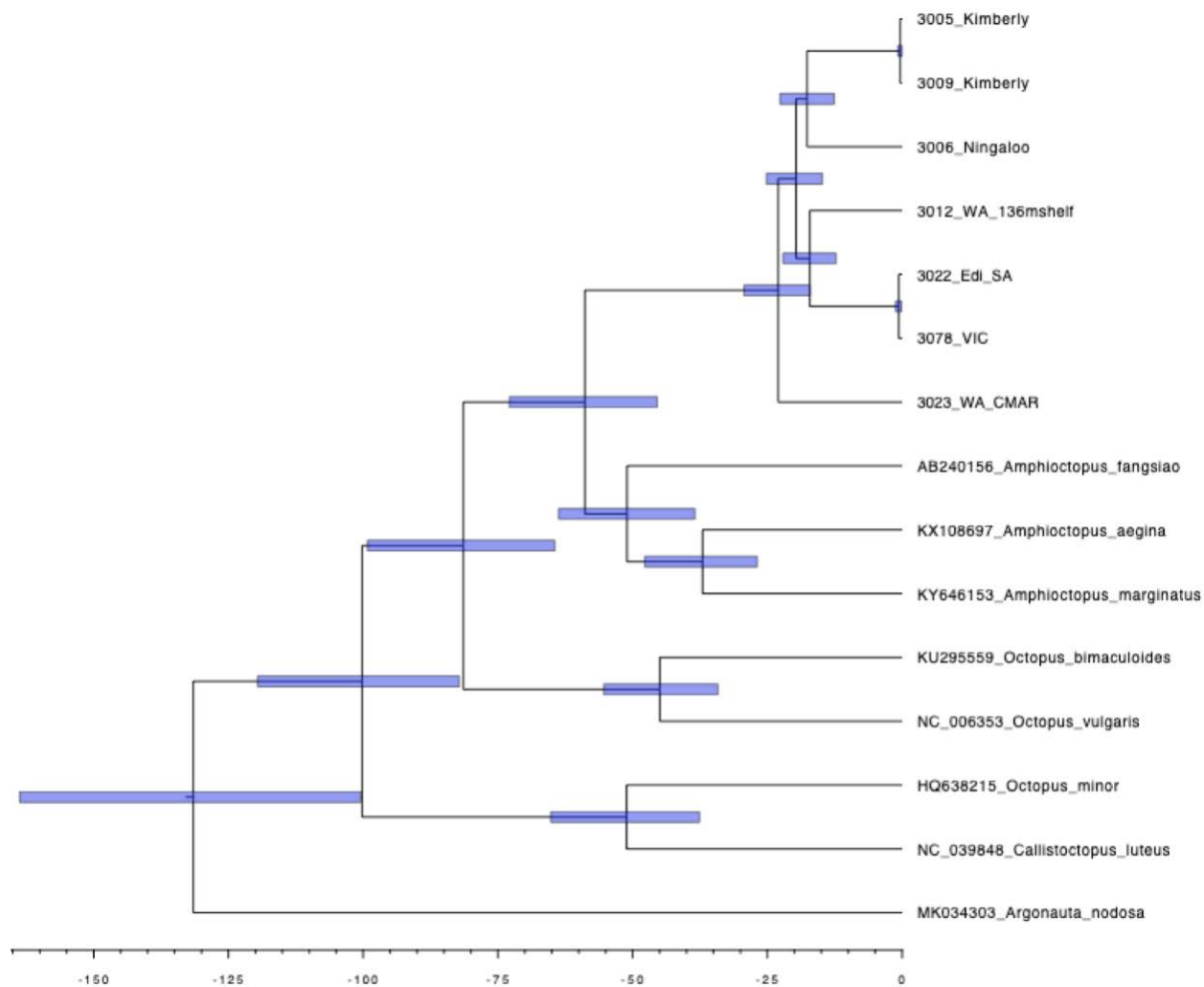
Supplementary table 3. Pairwise Fst analysis for outlier and neutral SNPs across grouped locations. Red indicates Fst values > 0.4, yellow indicates non-significant values (p>0.05)

OUTLIERS																			
	NSW_Brisbane_Redcliffe	SharkBay_DampierA	Ningaloo	GLYE_Liz_Is_Townsville_DEEP	GLYE_Liz_Is_Townsville	Albany	Cape_York_TS	VIC	Woodmans	Taiwan	Mackay	Darwin_Kimberly	SA	CMAR	Timor_Leste	WA_136m	Taiwan_Vietnam	Rockmans	Exmouth
NSW_Brisbane_Redcliffe	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SharkBay_DampierA	0.85	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Ningaloo	0.85	0.81	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GLYE_Liz_Is_Townsville_DEEP	0.89	0.88	0.90	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GLYE_Liz_Is_Townsville	0.83	0.80	0.73		0.80	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Albany	0.84	0.72	0.80		0.86		0.81	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Cape_York_TS	0.87	0.10	0.89		0.92		0.82	0.78	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
VIC	0.86	0.77	0.84		0.89		0.82	0.33	0.81	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Woodmans	0.82	0.69	0.78		0.84		0.80	0.26	0.73	0.41	NA	NA	NA	NA	NA	NA	NA	NA	NA
Taiwan	0.88	0.89	0.90		0.93		0.80	0.86	0.93	0.89	0.85	NA	NA	NA	NA	NA	NA	NA	NA
Mackay	0.78	0.60	0.38		0.64		0.13	0.68	0.65	0.69	0.66	0.68	NA	NA	NA	NA	NA	NA	NA
Darwin_Kimberly	0.90	0.88	0.85		0.90		0.85	0.88	0.90	0.90	0.87	0.88	0.78	NA	NA	NA	NA	NA	NA
SA	0.85	0.74	0.83		0.88		0.81	0.26	0.80	0.16	0.35	0.88	0.65	0.89	NA	NA	NA	NA	NA
CMAR	0.89	0.85	NA		-0.07		0.78	0.85	0.92	0.88	0.82	0.94	0.52	0.89	0.86	NA	NA	NA	NA
Timor_Leste	0.88	0.87	0.88		0.91		0.80	0.86	0.91	0.89	0.84	0.90	0.64	0.69	0.87	0.91	NA	NA	NA
WA_136m	0.85	0.76	NA		0.91		0.70	0.79	0.90	0.83	0.75	0.89	0.34	0.78	0.81	NA	0.79	NA	NA
Taiwan_Vietnam	0.86	0.84	0.48		0.90		0.74	0.82	0.91	0.86	0.80	0.90	0.56	0.85	0.85	0.90	0.87	0.78	NA
Rockmans	0.86	0.70	NA		0.92		0.79	0.27	0.83	0.48	-0.02	0.93	0.45	0.88	0.40	NA	0.89	NA	0.89
Exmouth	0.91	0.87	NA		0.94		0.83	0.88	0.93	0.90	0.86	0.93	0.61	0.89	NA	0.74	NA	0.91	NA
NEUTRAL																			
	NSW_Brisbane_Redcliffe	SharkBay_DampierA	Ningaloo	GLYE_Liz_Is_Townsville_DEEP	GLYE_Liz_Is_Townsville	Albany	Cape_York_TS	VIC	Woodmans	Taiwan	Mackay	Darwin_Kimberly	SA	CMAR	Timor_Leste	WA_136m	Taiwan_Vietnam	Rockmans	Exmouth
NSW_Brisbane_Redcliffe	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SharkBay_DampierA	0.85	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Ningaloo	0.86	0.81	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GLYE_Liz_Is_Townsville_DEEP	0.89	0.87	0.90	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GLYE_Liz_Is_Townsville	0.83	0.80	0.73		0.80	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Albany	0.84	0.72	0.81		0.86		0.82	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Cape_York_TS	0.87	0.15	0.91		0.93		0.82	0.78	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
VIC	0.86	0.76	0.85		0.89		0.83	0.28	0.82	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Woodmans	0.83	0.69	0.79		0.85		0.81	0.24	0.76	0.37	NA	NA	NA	NA	NA	NA	NA	NA	NA
Taiwan	0.88	0.89	0.90		0.93		0.79	0.86	0.93	0.89	0.85	NA	NA	NA	NA	NA	NA	NA	NA
Mackay	0.78	0.61	0.41		0.65		0.11	0.69	0.65	0.70	0.68	0.68	NA	NA	NA	NA	NA	NA	NA
Darwin_Kimberly	0.91	0.89	0.85		0.90		0.86	0.89	0.91	0.90	0.88	0.89	0.80	NA	NA	NA	NA	NA	NA
SA	0.85	0.74	0.84		0.88		0.82	0.23	0.81	0.14	0.32	0.89	0.66	0.90	NA	NA	NA	NA	NA
CMAR	0.89	0.85	NA		-0.17		0.79	0.85	0.93	0.88	0.83	0.94	0.54	0.89	0.87	NA	NA	NA	NA
Timor_Leste	0.88	0.86	0.83		0.91		0.80	0.86	0.91	0.88	0.84	0.90	0.64	0.71	0.88	0.90	NA	NA	NA
WA_136m	0.85	0.78	NA		0.89		0.70	0.79	0.90	0.83	0.76	0.87	0.34	0.83	0.82	NA	0.79	NA	NA
Taiwan_Vietnam	0.86	0.84	0.45		0.89		0.75	0.82	0.91	0.86	0.81	0.88	0.57	0.86	0.85	0.90	0.84	0.78	NA
Rockmans	0.85	0.71	NA		0.91		0.79	0.29	0.86	0.43	0.00	0.93	0.46	0.89	0.39	NA	0.88	NA	0.88
Exmouth	0.91	0.88	NA		0.94		0.84	0.88	0.94	0.91	0.87	0.94	0.64	0.89	NA	0.74	NA	0.91	NA

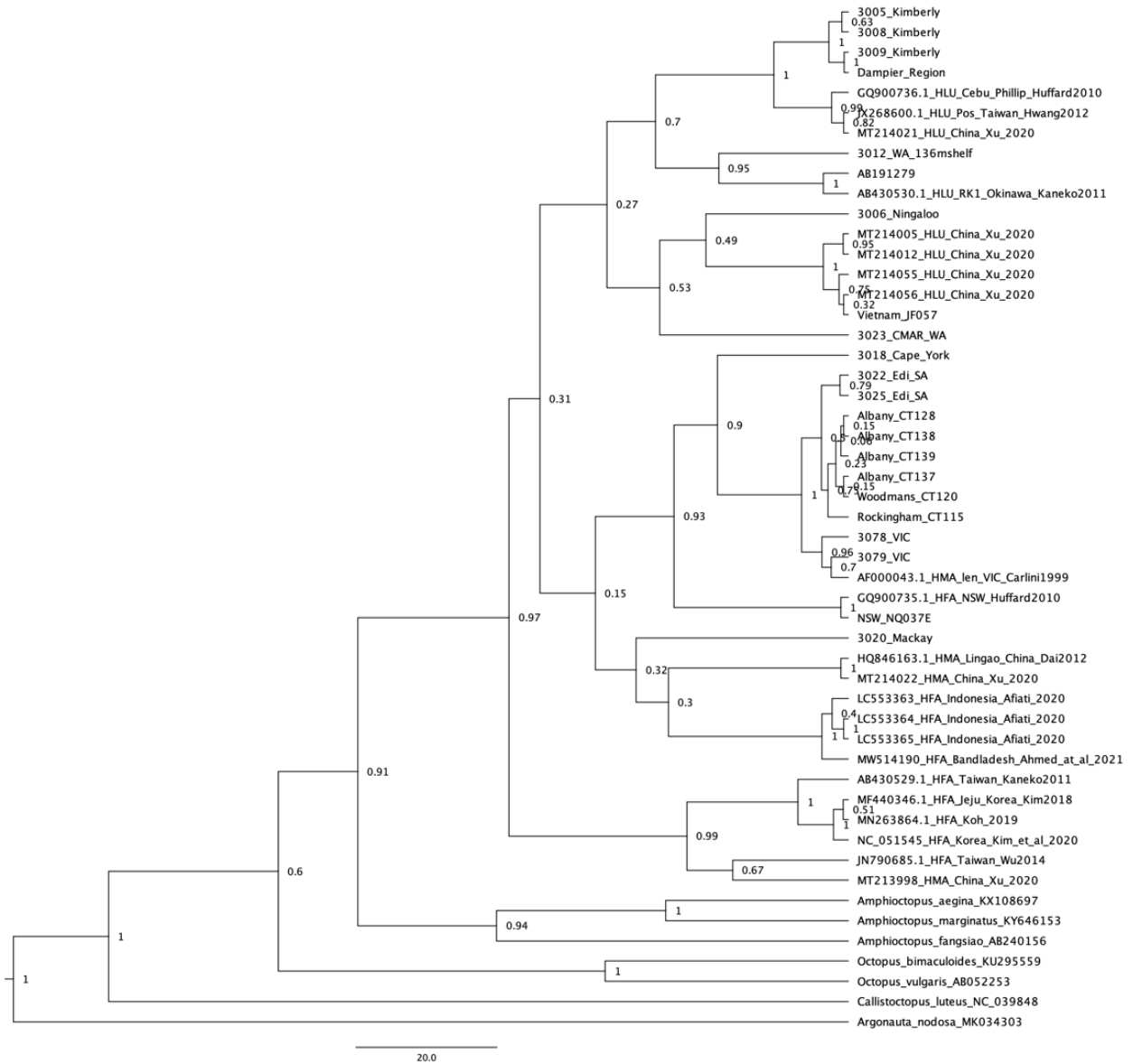


Supplementary figure 1. Delineation of *Haplochlanea* species boundaries and genetic structure throughout the Indo Pacific using 10,346 SNPs **a)** Arrangement of samples according to the first two principal components of a PCoA based on SNP data generated using the R dartR **b)** Complementary DAPC plot. Samples are coloured to represent putative taxonomic units A-K : light blue (A), dark blue (B), light green (C), light green (D), pink (E), red (F), light orange (G), dark orange (H), lilac (I), purple (J) and brown (K).





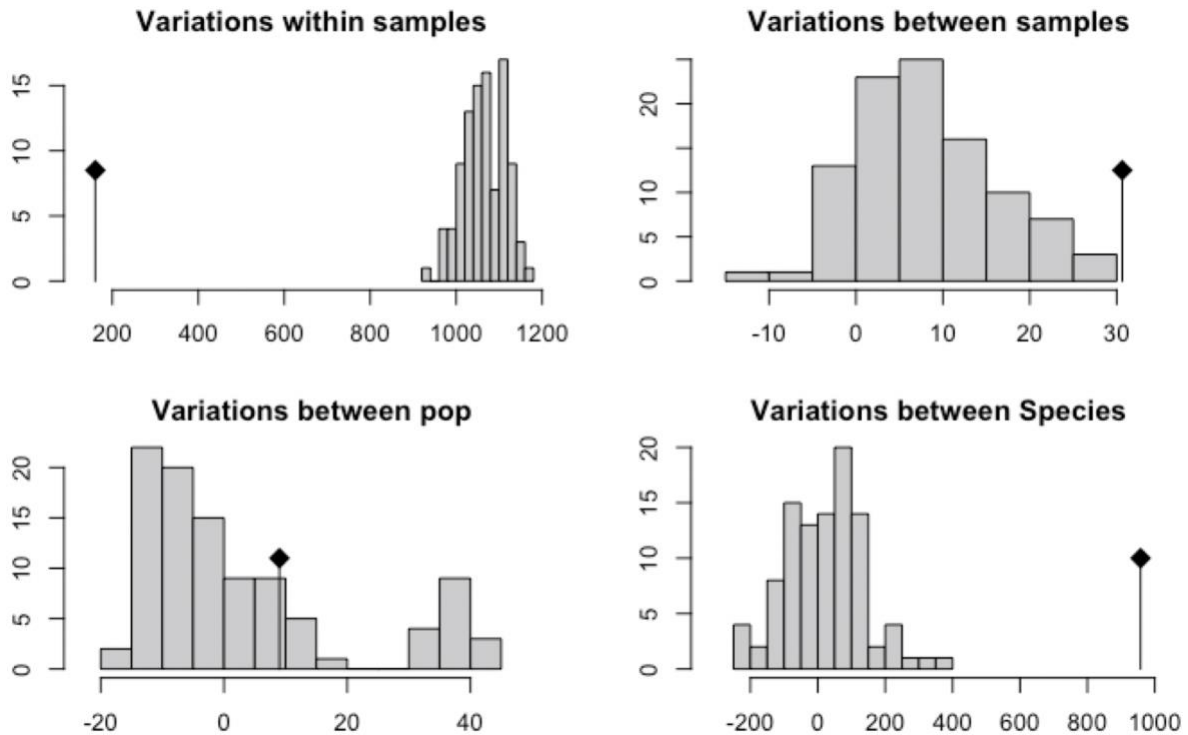
Supplementary figure 2. Bayesian phylogeny using four concatenated mitochondrial genes (12S, 16S, COI and COIII) of *Hapalochlaena* throughout the Asia Pacific generated using BEAST2. Blue node bars display standard error of divergence time estimations.



Supplementary figure 3. Bayesian COI tree of Hapalochlaena throughout the Asia Pacific generated using BEAST2. Nodes display posterior probabilities

Supplementary table 4. Diversity indices for the mitochondrial COI gene in four *Haplochlarena* organisational taxonomic units (OTU).

OTU	Diversity indices	
	Nucleotide diversity	Haplotype diversity
G	0.4	0.4
H	0	0
I	0	0
J	5	1
K	1.62	0.79



Supplementary figure 4. AMOVA of populations and predicted species using 10,346 SNPs

Supplementary table 4. Table of AMOVA of populations and predicted species using 10,346 SNPs.

<b>Results</b>	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>
Between Species	14	110411.818	7886.5584
Between pop Within Species	14	3513.348	250.9535
Between samples Within pop	33	7318.048	221.759
Within samples	62	9949.851	160.4815
Total	123	131193.066	1066.6103

<b>Components of Covariance</b>	<b>Sigma</b>	<b>%</b>
Variations Between Species	957.824858	82.72
Variations Between pop Within Species	9.015923	0.78
Variations Between samples Within pop	30.64	2.65
Variations Within samples	160.481472	13.86
Total variations	1157.96	100.00