GigaScience

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

Manuscript Number:	GIGA-D-20-00135R2
Full Title:	Adaptive venom evolution and toxicity in octopods is driven by extensive novel gene formation, expansion and loss
Article Type:	Research

| Abstract: | Background

Cephalopods represent a rich system for investigating the genetic basis underlying organismal novelties. This diverse group of specialised 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 three 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 (PSG) 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 PSG microbiome revealed a 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. |

Corresponding Author:	Brooke Lauren Whitelaw, Bachelor with Honours and current PhD candidate James Cook University College of Science and Engineering Townsville, QLD AUSTRALIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	James Cook University College of Science and Engineering
Corresponding Author's Secondary Institution:	
First Author:	Brooke Lauren Whitelaw
First Author Secondary Information:	
Order of Authors:	Brooke Lauren Whitelaw

	Ira R. Cooke
	Julian Finn
	Rute R. da Fonseca
	Elena A. Ritschard
	M. T P. Gilbert
	Oleg Simakov
	Jan M. Strugnell
Order of Authors Secondary Information:	
Response to Reviewers:	We would like to thank the reviewers for their helpful comments, which have been addressed below and the editor for considering our manuscript for publication.

Ira R. Cooke
Julian Finn
Rute R. da Fonseca
Elena A. Ritschard
M. T P. Gilbert
Oleg Simakov
Jan M. Strugnell

Response to Reviewers:

We would like to thank the reviewers for their helpful comments, which have been addressed below and the editor for considering our manuscript for publication.

Reviewer reports:
Reviewer #1: The authors have made the changes that I requested. I confess that I did not carefully read through the revision to verify that all of the minor grammatical edits I proposed were made, but I "spot checked" several and they look pretty good.

Reviewer #2: The authors have adequately revised their manuscript in response to my review and addressed the comments in my review with their letter. I have just two minor suggestions after reading the revised manuscript.

1) The authors should choose one convention for spelling the word specialized either specialised (British) or specialized (American).

We have chosen to use the British spelling and corrected the one American spelling at pg. 1, line 24.

2) In lines 372-374, the authors comment on the interactions between aspartic acid and TTX, "While it has yet to be assessed for TTX resistance, the replacement of Asp in B. candida with a neutral amino acid has been predicted to disrupt TTX binding by preventing formation of a hydrogen bond." A better reference for this statement is Shen et al. (2018). In addition, the cryo-EM structure data from this paper suggest that either a hydrogen bond or a salt bridge could form between TTX and that aspartic acid at this position of the protein.

We agree that this sentence required clarification, specifically that both hydrogen and salt bridges could be disrupted. This information has been added to the modified sentence as well as the suggested reference.

Original sentence:
"While it has yet to be assessed for TTX resistance, the replacement of Asp in B. candida with a neutral amino acid has been predicted to disrupt TTX binding by preventing formation of a hydrogen bond90"

Revised sentence: (pg. 17, lines : 396-398)
"While it has yet to be assessed for TTX resistance, the replacement of Asp in B. candida with a neutral amino acid has been predicted to disrupt TTX binding by preventing formation of a salt bridge or hydrogen bond89,91"

Additional Information:

Question	Response
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics	Yes

Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our [Minimum Standards Reporting Checklist](). Information essential to interpreting the data presented should be made available in the figure legends. Have you included all the information requested in your manuscript?	
Resources A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite [Research Resource Identifiers]() (RRIDs) for antibodies, model organisms and tools, where possible. Have you included the information requested as detailed in our [Minimum Standards Reporting Checklist]()?	Yes
Availability of data and materials All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in [publicly available repositories]() (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the "Availability of Data and Materials" section of your manuscript. Have you have met the above requirement as detailed in our [Minimum Standards Reporting Checklist]()?	Yes

1 **Adaptive venom evolution and toxicity in octopods is driven by**

2 **extensive novel gene formation, expansion and loss**

3 Brooke L. Whitelaw[1,2,*], Ira R. Cooke[3,4], Julian Finn[2], Rute R. da Fonseca[5], Elena A.

4 Ritschard[6,7], M. T P. Gilbert[8] Oleg Simakov[6], Jan M. Strugnell[1,9]

5

6 [1]Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University,

7 Townsville, Queensland, 4811, Australia

8 [2]Sciences, Museum Victoria, Carlton, Victoria 3053, Australia

9 [3]College of Public Health, Medical and Vet Sciences, James Cook University,

10 Townsville, Queensland, 4811, Australia

11 [4]La Trobe Institute of Molecular Science, La Trobe University, Melbourne, Victoria

12 3086, Australia

13 [5]Center for Macroecology, Evolution and Climate (CMEC), GLOBE Institute, University

14 of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark;

15 [6]Department of Neurosciences and Developmental Biology, University of Vienna,

16 Austria

17   [7]Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton

18   Dohrn, Italy

19   [8]Center for Evolutionary Hologenomics, GLOBE Institute, University of Copenhagen,

20   Øster Voldgade 5-7, 1350 Copenhagen, Denmark;

21   [9]Department of Ecology, Environment and Evolution, La Trobe University, Melbourne,

22   Victoria, 3086 Australia

23

24   *Brooke Whitelaw,Email: 1blwhitelaw8@gmail.com,Phone number: 0424642621

25

26   ORCIDs:

27   Brooke L. Whitelaw, 0000-0002-5555-4612;

28   Ira R. Cooke, 0000-0001-6520-1397;

29   Julian Finn, 0000-0003-1820-512X;

30   Rute R. da Fonseca, 0000-0002-2805-4698;

31   Elena A. Ritschard, 0000-0002-4956-9703;

32   M. T P. Gilbert, 0000-0002-5805-7195;

33   Oleg Simakov, 0000-0002-3585-4511;

34   Jan M. Strugnell, 0000-0003-2994-637X.

35

Abstract

Background

38   Cephalopods represent a rich system for investigating the genetic basis underlying

39   organismal novelties. This diverse group of specialised predators has evolved many

40   adaptations including proteinaceous venom. Of particular interest is the blue-ringed-

41   octopus genus (*Hapalochlaena*), which are the only octopods known to store large

42   quantities of the potent neurotoxin, tetrodotoxin, within their tissues and venom gland.

Findings

44   To reveal genomic correlates of organismal novelties, we conducted a comparative

45   study of three octopod genomes, including the Southern blue-ringed octopus

46   (*Hapalochlaena maculosa*). We present the genome of this species and reveal highly

47   dynamic evolutionary patterns at both non-coding and coding organizational levels.

48   Gene family expansions previously reported in *Octopus bimaculoides* (e.g., zinc finger

49   and cadherins, both associated with neural functions), as well as formation of novel

50   gene families, dominate the genomic landscape in all octopods. Examination of tissue-

51   specific genes in the posterior salivary gland (PSG) revealed that expression was

52   dominated by serine proteases in non- tetrodotoxin bearing octopods, while this family

53   was a minor component in *H. maculosa*. Moreover, voltage-gated sodium channels in *H.*

54   *maculosa* contain a resistance mutation found in pufferfish and garter snakes, which is

55   exclusive to the genus. Analysis of the PSG microbiome revealed a diverse array of

56   bacterial species, including genera that can produce tetrodotoxin, suggestive of a

57   possible production source.

58   ***Conclusions***

59   We present the first tetrodotoxin-bearing octopod genome *H. maculosa,* which displays

60   lineage-specific adaptations to tetrodotoxin acquisition. This genome, along with other

61   recently published cephalopod genomes, represents a valuable resource from which

62   future work could advance our understanding of the evolution of genomic novelty in

63   this family.

64

65   ***Background***

66   Reconstructing the evolution of novelties at the genomic level is becoming an

67   increasingly viable approach to understand their origin. The recent publication of

68   octopod genomes provides an opportunity to investigate the link between genomic and

69   organismal evolution in this unique lineage for which genomic resources have been

70   lacking[1]. From their emergence 275 mya[2], octopods have diversified into > 300

71   species, inhabiting tropical to polar regions, from the deep sea to shallow intertidal

72   zones[3]. As a highly diverse group, octopods show remarkable variation in body form

73   and function. They are specialised soft-bodied predators that are well adapted to their

74   environment with prehensile limbs lined with chemosensory suckers[4], the ability to

75   manipulate skin texture and colour using specialised chromatophores[5], the largest

76   invertebrate nervous systems (excluding those of other cephalopods)[6], and a

77   relatively large circumesophageal brain allowing for complex problem solving and

78   retention of information[7]. Furthermore, the cephalopods have independently evolved

79   proteinaceous venom, which is produced and stored within a specialised gland in

80   known as the posterior salivary gland (PSG). All octopods are believed to possess a

81   form of proteinaceous venom used to subdue prey[8–10]. Serine proteases are a

82   common component of cephalopod venoms and have been observed in the PSG of

83   squids, cuttlefish and octopods[10–13]. Convergent recruitment of serine proteases has

84   been observed between many vertebrate (Squamata[14–16] and Monotremata[17])

85   and invertebrate (Hymenoptera[18], Arachnida[19], Gastropoda[20], Remipedia[21]

86   and Cnidarian[22]) venomous lineages.

87   In addition to these proteinaceous venoms, the blue-ringed octopus (genus

88   *Hapalochlaena*) is the only group that also contains the potent non-proteinaceous

89   neurotoxin, tetrodotoxin (TTX)[12,23]. The mechanism of TTX resistance, which allows

90   for safe sequestration of TTX, has been attributed to several substitutions in the p-loop

91   regions of voltage-gated sodium channels(Na_v) in *H. lunulata*[24]. However, these

92   channels have yet to be examined in *H. maculosa* and *H. fasciata*. TTX resistance has

93   also been studied in a range of other genera including, pufferfish[25], newts[26,27]

94   arachnids[28], snakes[29] and gastropods[30].

95   The blue-ringed octopus is easily identified by iridescent blue rings, which

96   advertise its toxicity in an aposematic display[31–33]. Sequestration of the TTX within

97   bodily tissues is unique to this genus among cephalopods[32,34]. While other

98   unrelated TTX-bearing species primarily use TTX for defense, *Hapalochlaena* is the only

99   known taxa to utilise TTX in venom[23,35]. The impact of TTX inclusion on venom

100   composition and function has been previously investigated in the southern blue-ringed

101   octopus (*H. maculosa*)[9]. Relative to the non-TTX bearing species *Octopus kaurna, H.*

102   *maculosa* exhibited greater expression of putative dispersal factors such as

103   hyaluronidase, which serve to aid in the dispersal of toxic venom components[9].

104   Conversely, tachykinins- neurotoxins known from other octopods[36,37] were absent

105   from the *H. maculosa* PSG[9]. Further investigation into the broader impact of TTX on

106   the evolutionary trajectory of the species has yet to be addressed due to the absence of

107   a genome.

108   This study presents the genome of the southern blue-ringed octopus (*H.*

109   *maculosa,* NCBI:txid61716; marinespecies.org:taxname:342334), the first from the

110   genus *Hapalochlaena.* By using a comparative genomic approach we are able to

111   examine the emergence of octopod novelties, at a molecular level between *H. maculosa*

112   and the two non-TTX bearing octopods: the California two-spot octopus (*O.*

113   *bimaculoides*) and the long-armed octopus (*Callistoctopus minor*). We also address unique

114   features of venom evolution in octopods while also addressing the species-specific

115   evolution of tetrodotoxin acquisition and resistance in *H. maculosa.*

116

117   *Keywords:* cephalopod genome, comparative genomics, gene family expansions,

118   transposable elements, venom evolution

119

120

121   ## *Data Description*

122   ### *Genome assembly and annotation*

123   The southern blue-ringed octopus genome was sequenced using Illumina paired

124   end and Dovetail sequencing from a single female collected at Beaumaris Sea Scout

125   Boat Shed, Beaumaris, Port Phillip Bay, Victoria, Australia. The assembly was

126   composed of 48,285 scaffolds with an N50 of 0.93 Mb and total size of 4.08 GB. A total

127   of 29,328 inferred protein coding genes were predicted using a PASA[38] and an

128   Augustus[39] pipeline and supplemented with zinc finger and cadherin genes obtained

129   from aligning *H. maculosa* transcripts to *O. bimaculoides* gene models(Supplementary

130   notes 1.1-1.4). Completeness of the genome was estimated using BUSCO[40], which

131   identified 87.7% complete and 7.5% fragmented genes against the metazoan database

132   of 978 groups (Supplementary notes 3.2).

133   *H. maculosa* has a highly heterozygous genome (0.95%), similar to *O. vulgaris*

134   (1.1%)[41] but far higher than *O. bimaculoides* (0.08%)[42]. While the low

135   heterozygosity of *O. bimaculoides* is surprising, other molluscs also have highly

136   heterozygous genomes in accordance with *H. maculosa*, including the gastropods (1-

137   3.66%)[43,44] and bivalves (0.51-3%)[45–51](Supplementary table 5).

138

139   ***PSMC (Pairwise Sequentially Markovian Coalescent) and mutation rate***

140   The mutation rate for *H. maculosa* was estimated to be 2.4×10^{-9} per site per

141   generation based on analysis of synonymous differences with *O. bimaculoides*

142   (Supplementary note 1.5). The mutation rate is comparable to the average mammalian

143   mutation rate of 2.2×10^{-9} per site per generation, and *Drosophila,* 2.8×10^{-9}[52,53].

144   Due to the unavailability of a suitable closely related and comprehensive genome until

145   the publication of *O. bimaculoides* in 2015[42], this is the first genome-wide mutation

146   rate estimated for any cephalopod genome.

147   The historic effective population size (Ne) of *H. maculosa* was estimated using

148   the pairwise sequentially Markovian coalescent (PSMC) model (Supplementary Fig 2).

149   Population size was found to initially increase during the early Pleistocene, followed by

150   a steady decline which slows slightly around 100kya. Note that PSMC estimates are not

151   reliable at very recent times due to a scarcity of genomic blocks that share a recent

152   common ancestor in this highly heterozygous genome. A decline in population size

153   started during the mid-Pleistocene approximately 1mya, a time of unstable

154   environmental conditions with fluctuations in both temperature and glaciation

155   events[54–56]. Corals in the genus *Acropora* show a similar pattern of expansion and

156   contraction attributed to niche availability post mass extinction of shallow-water

157   marine organisms 2-3 mya, followed by the unstable mid-Pleistocene climate[57,58]. A

158   similar pattern of expansion and decline in effective population size has also been

159   observed in the Antarctic ice fish among other marine organisms distributed in the

160   Southern Hemisphere[59].

161

162   ***Phylogenomics***

163   A total of 2,108 (single copy/ 1-to-1) orthologous clusters were identified

164   between the molluscan genomes and transcriptomes of 11 species and used to construct

165   a time-calibrated maximum likelihood tree(Fig 1a). The phylogenetic reconstruction

166   estimated the divergence time between *H. maculosa* and its nearest relative, *O.*

167   *bimaculoides,* to be ~59 mya. *C. minor* diverged from this clade much earlier ~183

168   mya. Previous phylogenies using a combination of a small number of mitochondrial

169   and nuclear genes[60–62] and orthologs derived from transcriptomes[63] support this

170   topology. Likewise, estimates by Tanner et al.[2], using a concatenated alignment of 197

171   genes with a Bayesian approach, placed divergence of *H. maculosa* from *Abdopus*

172   *aculeatus* at ~59 mya[2].

173   Inference of "shared" phenotypic traits can be difficult to resolve with the

174   current literature. For example, false eye spots/ocelli observed in both *O. bimaculoides*

175   and *H. maculosa* are structurally very different. Each ocellus in *H. maculosa* is composed

176   of a continuous single blue ring[33], while *O. bimaculoides* has a blue ring composed of

177   multiple small rings. Morphological variations of ocelli structure and colour, in

178   conjunction with the taxonomically sporadic occurrence of this trait across species

179   within *Octopus* and *Amphioctopus,* limits our interpretation as to the evolutionary

180   history of this trait in octopods[3] . Large gaps remain in the literature between

181   phenotypic traits in cephalopods and their genomic source[1]. This study aims to

182   provide a genomic framework to enable resolution of these features by profiling

183   changes in several genomic characters: (i) gene duplications, (ii) novel gene formation,

184   and (iii) non-coding element evolution.

185

186   **Fig 1. Comparisons of molluscan genomes and gene families a)** Time-calibrated maximum likelihood

187   phylogeny of seven molluscan genomes (*Aplysia californica, Lottia gigantea, Crassostrea gigas, Euprymna*

188   *scolopes, Octopus bimaculoides, Callistoctopus minor and Hapalochlaena maculosa*) and four transcriptomes

189   (*Octopus kaurna, Octopus vulgaris, Sepia officinalis* and *Idiosepius notoides*) using 2,108 single copy

190   orthologous sequence clusters. Node labels show divergence times in millions of years (mya), blue

191   (divergence to octopods) and orange bars (decopods) represent standard error within a 95% confidence

192   interval. Octopodiformes lineages are highlighted in blue and decapod orange. Scale bar represents

193   millions of year (mya). **b)** Expansions of octopod gene families relative to molluscan genomes *Aplysia*

194   *californica* (A. cali), *Biomphalaria glabrata* (B. glab), *Crassostrea gigas* (C. gig), *Lottia gigantea* (L. gig),

195   *Euprymna scolopes* (E. scol) **c)** Lineage-specific gene expansions in the octopod genomes *Callistoctopus*

196   *minor* (C. min), *Octopus bimaculoides* (O. bim) and *Hapalochlaena maculosa* (H. mac). Domains

197   abbreviated: Chondroitin N-acetylgalactosaminyltransferase (CHGN), C2H2(Cys2-His2) zinc finger and

198   Cornifin SPRR(small proline-rich proteins).

199

200   *Organismal impact of novel genes and gene family expansions*

201   Gene family expansions between octopods (*O. bimaculoides*, *C. minor* and *H.*

202   *maculosa)* and three other molluscan genomes (*Aplysia californica*, *Lottia gigantea* and

203   *Crassostrea gigas)* were examined using Pfam annotations. A total of 5565 Pfam

204   domains were identified among six molluscan genomes. *H. maculosa* and *C. minor*

205   exhibit expansions in the cadherin gene family, characteristic of other octopod

206   genomes, including *O. bimaculoides* (Fig1b)[42,64]. *C. minor,* in particular, shows the

207   greatest expansion of this family within octopods. Expansions of protocadherins, a

208   subset of the cadherin family, have also occurred independently in squid[42], with the

209   octopod expansions occurring post divergence ~135 mya[42]. The shared ancestry of

210   octopod cadherins was also documented by Styfhals et al[64] using phylogenetic

211   inference between *O. bimaculoides* and *O. vulgaris.*Cadherins, specifically

212   protocadherins, play crucial roles in synapse formation, elimination and axon targeting

213   within mammals and are essential mediators of short-range neuronal connections[65–

214   68]. It should be noted that octopods lack a myelin sheath, as a result short-range

215   connections are integral to maintaining signal fidelity over distance[6]. The

216   independent expansions of protocadherins within chordate and cephalopod lineages are

217   believed to be associated with increased neuronal complexity[42,64]. Elevated

218   expression of protocadherins within neural tissues have been observed in *O. vulgaris*

219   and *O. bimaculoides* by both Styfhals et al[64] and Albertin et al[42] respectively. In

220   particular Styfhals et al[64] noted differential expression across neural tissues

221   including supra-esophageal mass, sub-esophageal mass, optic lobe and the stellate

222   ganglion[64]. However, functional implications of observed expression patterns remain

223   speculative without further study.

224   *H. maculosa* also shows expansions in the C2H2-type zinc finger family. Zinc

225   fingers form an ancient family of transcription factors, which among other roles serve

226   to regulate transposon splicing as well as embryonic and neural development[69,70].

227   Expansion of this type of zinc finger in *O. bimaculoides* has been associated with neural

228   tissues. It should be noted that due to the inherent difficulty in fully annotating the

229   zinc finger family, alternative methods were used to examine the number of exons in *C.*

230   *minor* with high similarity to annotated zinc finger genes in *O. bimaculoides*

231   (Supplementary notes 5.1). A total of 609 exons (not captured by published gene

232   models) from *C. minor* were found with high similarity to accepted zinc finger genes in

233   *O. bimaculoides,* suggesting this family is larger than that which the genome annotation

234   infers.

235   Examination of genes specifically expressed within neural tissues found that

236   cadherins were among the most highly expressed gene families of all octopod species.

237   Particularly in *C. minor*, relative to the other octopods, such a trend reflects the gene

238   family expansions found in this species (Fig2c). Zinc fingers were less pronounced,

239   representing 1.1% of overall expression in *C. minor* compared to cadherins at 11.3%.

240   Overall, neural tissues express a large diversity of Pfams with each species, exhibiting a

241   similar profile and proportion of orthologous to lineage-specific genes.

242

243   ***Novel patterns of gene expression***

244   High-level examination of gene dynamics (expression, loss of expression and

245   absence of expression) between octopods across different levels of orthology provides

246   insight into large-scale expression patterns and highlights lineage-specific loss of

247   expression.

248   The greatest proportion of genes in each species examined were not specific to

249   octopods or an octopus lineage (ancient genes) (Fig 2a). Expression of these genes were

250   enriched in neural tissues across all species, indicating the core conservation of neural

251   development and function. However, we also find that genes specific to each octopod

252   species also show this expression pattern. The overall elevated expression of genes

253   within neural tissues could be reflective of the extensive neural network present in

254   cephalopods, which comprises around 520 million nerve cells[71], rivalling

255   vertebrates/mammals in size[6]. Expression of many novel genes in the nervous system

256   may also indicate contribution of those genes to lineage-specific neural network

257   evolution. In contrast, genes that date back to the shared octopod ancestor show

258   highest expression in male reproductive tissues in all species.

259   Loss of expression between octopod genomes is exhibited most clearly in *H. maculosa*

260   with 11% (1993 genes) of all ancient genes having no expression, compared to 1% in

261   both *O. bimaculoides* and *C. minor*. Absence of gene expression for genes whose

262   orthologs have retained expression in one or more other species suggests a unique

263   evolutionary trajectory from other octopods. It should be noted that differences in

264   tissue sampling may in part influence these values and due to the limited sampling of

265   species, loss of expression cannot be inferred at a species level and may have occurred

266   at any point in the lineage. In order to fully understand the implications of the gene

267   family contractions and loss of expression in *H. maculosa*, relative to other octopods,

268   further investigation is required.

269

Fig2. 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 one or more 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).

277

Fig3. Dynamics of gene expression in neural and venom producing tissues of octopods. a) Tissue

specific expression of genes within the brain of *H. maculosa, O. bimaculoides* and *C. minor* (red). Venn

diagram shows numbers of shared and exclusive genes between species (Left). Bar chart of the top 5

Pfams and their contribution to overall expression in the brain (right). b) Tissue specific expression of

genes within the posterior salivary gland (PSG) of *H. maculosa, O. bimaculoides* and *C. minor* (Blue). Venn

diagram shows numbers of shared and exclusive genes between species (left). Bar chart of the top 5

Pfams and their contribution to overall expression in the PSG (right).

285

Evolution of the octopod non-coding genome

287   Similar to other cephalopod genomes, the *H. maculosa* genome has a high repeat

288   content of 37.09% (bases masked). *O bimaculoides* and *C. minor* are also highly

289   repetitive with 46.39% and 44% of their genomes composed of transposable elements

290   (TE) respectively. Of the repetitive elements, LINEs dominate the decapodiform

291   *Euprymna scolopes* genome accounting for its larger genome size[72], while SINEs are

292   expanded in all four octopod genomes. SINEs have been previously documented in *O.*

293   *bimaculoides* (7.86%)[42], comparable with *H. maculosa* (7.53%), while fewer SINEs

294   were previously reported for *C. minor* (4.7%)[73]. SINE elements also dominate the *O.*

295   *vulgaris* genome with an expansion occurring post divergence from *O. bimaculoides*[41].

296   Rolling circle (RC) elements are a prominent minor component in octopods,

297   particularly in *H. maculosa*. RC transposons have been isolated from plant (*Zea mays*)

298   and mammalian genomes. They depend greatly on proteins used in host DNA

299   replication and are the only known class of eukaryotic mobile element (transposon) to

300   have this dependence[74]. TE elements in cephalopod lineages show differing

301   expansions between most of the genomes currently available, suggesting they are

302   highly active and play a strong role in cephalopod evolution.

303 Enrichment of transposable elements associated with genes (flanking regions

304 10kb up- and downstream) was not observed compared to the whole genome for any

305 species examined. More notable were differences between species, in particular *C.*

306 *minor* shows a greater proportion of LINE to SINE elements relative to both *O.*

307 *bimaculoides* and *H. maculosa*.

308 Together, this highlights a very dynamic evolutionary composition of repeats in

309 cephalopods, that requires further study to test for any potential association with

310 changes in gene expression or genome evolution.

311

312 ***Dynamics of gene expression in the posterior salivary gland (PSG)***

313 The posterior salivary gland is the primary venom-producing gland in octopods.

314 Venom composition in the majority of octopods is primarily composed of proteinaceous

315 toxins. *Hapalochlaena* is an exception containing an additional non-proteinaceous

316 neurotoxin, TTX, within their venom. We hypothesize that the *Hapalochlaena* PSG will

317 exhibit a loss of redundant proteinaceous toxins due to the presence of TTX.

318 Examination of all PSG-specific genes from the three octopods revealed a

319 disproportionate number of genes exclusive to *H. maculosa* (Fig 3a). A total of 623

320   genes were exclusive to *H. maculosa* PSG compared to only 230 and 164 exclusive to *O.*

321   *bimaculoides* and *C. minor* PSGs, respectively. Additionally, we predict that the *H.*

322   *maculosa* PSG is functionally more diverse based on the number of Pfam families

323   detected, 532 in total. Comparatively, the PSG genes in *O. bimaculoides* and *C. minor*

324   are fewer and more specialised. Gene family expansions of serine proteases dominate

325   expression comprising over 30% of total PSG-specific expression in *C. minor* and 17-

326   20% in *O. bimaculoides* (Fig 3b). Serine proteases were also among genes whose

327   expression appears to have shifted between octopod species. Several serine proteases

328   show specific expression to the PSG of *O. bimaculoides* and *C. minor* while being

329   expressed in a non-specific pattern among brain, skin, muscle and anterior salivary

330   gland tissues in *H. maculosa* (Fig 4b). Most notable is the absence of many paralogs in

331   both *H. maculosa* and *O. bimaculoides* suggesting a lineage-specific expansion of this

332   cluster in *C. minor*. Fewer serine protease genes can also be observed in *H. maculosa*

333   (Fig 4c). Similarly, reprolysin (M12B) exhibits shifting expression in *H. maculosa*,

334   presumably from the PSG to the branchial heart, and a complete loss of paralogs from

335   the genome. While the function of this protein has not been assessed in octopus,

336 members of this protein family exhibit anticoagulant properties in snake venom[75–

337 78].

338   Serine proteases have been previously documented in cephalopod venom and

339 are prime candidates for conserved toxins in octopods. Cephalopod-specific expansions

340 have been identified with strong association to the PSG in 11 cephalopods (seven

341 octopus, two squid and two cuttlefish)[8,13]. All serine proteases identified from the

342 PSG of these species were found to belong to the cephalopod-specific clade.

343 Functionally, cephalopod venom serine proteases have yet to be assessed. However,

344 octopod venom has been observed to have strong digestive and hemolytic properties,

345 which may be in part due to this crucial protein family[79–81]. The reduced number

346 and expression of serine proteases in *H. maculosa* suggests a change in function of the

347 PSG for this species. These results support the hypothesis of toxin redundancy in the *H.*

348 *maculosa* PSG due to the incorporation of tetrodotoxin. Previous proteomic analysis of

349 the *H. maculosa* PSG revealed high expression of hyaluronidase, which often serves as a

350 dispersal factor within snake venom, facilitating the spread of toxin while not being

351 directly toxic to their prey[9,82]. While further investigation is required, the

352 incorporation of TTX within *H. maculosa* venom may have contributed to a shift in

353 function, with proteins present acting to support the spread of venom and digestion of

354 tissues.

355

356 **Fig 4. Examination of posterior salivary gland (PSG) gene expression between three octopod**

357 **genomes. a)** Heatmap of genes expressed specifically in the PSG of *H. maculosa* (tau > 0.8) and their

358 orthologs in *O. bimaculoides* and *C. minor* lacking specific expression to the PSG (tau < 0.8). Genes with

359 an ortholog lacking expression are coloured in grey while the absence of an ortholog is white. **b)**

360 Heatmap of genes expressed specifically in the (PSG) of both *O. bimaculoides* and *C. minor* (tau >0.8)

361 and their orthologs in *H. maculosa* lacking specific expression to the PSG.

362

363 *TTX resistance of the Na$_v$ channels*

364   To identify the mechanism of TTX resistance in *H. maculosa,* the voltage gated

365 sodium channel (Na$_v$) sequences were compared between susceptible (human) and

366 resistant (pufferfish, salamanders and garter snakes) species. TTX binds to the p-loop

367 regions of sodium channels, inhibiting the flow of sodium ions in neurons, resulting in

368 paralysis[83,84]. Inhibition of TTX binding has been observed in species which either

21

369   ingest TTX via prey, such as garter snakes[85], and in those which retain TTX within

370   their tissues like pufferfish[86].

371   Two Na_v genes were identified in the *H. maculosa* genome (Na_v1 and Na_v2), this

372   is congruent with the recent identification of two Na_v isoforms in *H.*

373   *lunulata*[24](Supplementary Fig 8 & 9). Among cephalopods with sequenced Na_v1

374   channels, p-loop regions are highly conserved with both DI and DII shared between all

375   species. The regions DIII and DIV closer to the C-terminal end of the protein in

376   *Hapalochlaena* sp. contain mutations, which may impact TTX binding and differ

377   between families and species as follows. Similar to the pufferfish (*Arothron,*

378   *Canthigaster, Takifugu* and *Tetraodon*)[87]and garter snake *Thamnophis couchii*[88], *H.*

379   *maculosa* Nav1 has a mutation within the third p-loop at site (DIII) from M1406T,

380   while all other cephalopods have an Ile(I) at this position (Fig 5a). The dumbo octopus

381   (*Grimpoteuthis*) is the only exception retaining the susceptible M at this site similar to

382   humans and other non-resistant mammals[83]. Additionally, the fourth p-loop (DIV) in

383   *H. maculosa* exhibits two substitutions at known TTX binding sites: D1669H and

384   H1670S. In a previous study a Met to Thr substitution into a TTX sensitive Nav1.4

385   channel decreased binding affinity to TTX by 15-fold[87]. Likewise, a 10-fold increase

386   in sensitivity was observed from a T1674M substitution in a mite (*Varroa destructor*)

387   channel VdNav1[28]. However, resistance is often a result of multiple substitutions and

388   when I1674T/D1967S occur together in VdNav1, resistance is multiplicative resulting

389   in "super resistant" channels with binding inhibition of 1000-fold. The combination of

390   M1406T/ D1669H in *H. maculosa* also occurs in the turbellarian flatworm *Bdelloura*

391   *candida*(BcNav1)[87,89]. While it has yet to be assessed for TTX resistance, the

392   replacement of Asp in *B. candida* with a neutral amino acid has been predicted to

393   disrupt TTX binding by preventing formation of a salt bridge or hydrogen bond[89,90].

394   These three substitutions (M1406T, D1669H and H1670S) in *H. maculosa*, with the

395   potential to inhibit TTX binding, have also been identified by Geffeney et al[24] in *H.*

396   *lunulata*. It has yet to be established if these mutations are derived from a shared

397   ancestor or have occurred independently.

398   While *Hapalochlaena* remains the best documented example of TTX resistance

399   among cephalopods, other species may contain some level of TTX resistance (e.g.

400   *Octopus vulgaris*)[91,92]. Saxitoxin (STX) is a similar toxin in structure and function,

401   and mutations resistant to TTX are often also STX inhibiting[93] *O. vulgaris* has been

402   observed consuming STX-contaminated bivalves with no negative impacts and as such

403   is believed to be resistant[92]. However, no mutations known to reduce TTX/STX

404   binding affinity occur in its Nav1[92,94]. The selective pressure facilitating the

405   evolution of STX/TTX resistance in these shallow water benthic octopods may be toxic

406   prey, similar to garter snakes. STX is also known as a paralytic shellfish poison (PSP).

407   Produced by photosynthetic dinoflagellates and bioaccumulated in bivalves[95], this

408   toxin contaminates a common octopus food source. Pelagic squids such as the

409   Humboldt (*D. gigas*) and longfin inshore squid (*D. pealeii*) do not appear to be TTX/STX

410   resistant; mass strandings of Humboldt squid have been associated with ingestion of

411   STX-contaminated fish[96]. Likewise, no evidence of resistance was found in the

412   sodium channel of the dumbo octopus (*Grimpoteuthis*). This species typically inhabits

413   depths of 2000-5000m and is unlikely to encounter STX-contaminated food

414   sources[97].

415

416   **Fig 5. Mechanism of tetrodotoxin resistance within the posterior salivary gland of *H. maculosa***

417   **(PSG) a)** Alignment of voltage gated sodium channel alpha subunits (DI, DII,DIII & DIV) p-loop regions.

418   Mutations conferring resistance are coloured in green (pufferfish), orange (salamander), purple (clam)

419   and blue (octopus). Susceptible mutations at the same site are Black and bolded. Sites which may be

420   involved with resistance are in bold. **b)** Schematic of voltage-gated sodium channel (Na_v) alpha subunits

421   (DI, DII, DIII and DIV). Each unit is composed of six subunits 1-4 (blue) and 5-6 (yellow). Alternating

422   extra and intercellular loops are shown in black with the p-loops between subunits 5 and 6 highlighted

423   in red. Mutations conferring resistance are shown within black circles on p-loops.

424

425   ***Microbiome of the PSG***

426   TTX is produced through a wide variety of bacteria, which are common in

427   marine   sediments and have been isolated from organisms such as

428   pufferfish[25,98,99]. Sequestration of TTX is not exclusive to the blue-ringed octopus

429   among molluscs. Gastropods such as *Pleurobranchaea maculata*[100] and *Niotha*

430   *clathrata*[30], as well as some bivalves, are also capable of sequestering TTX[95]. The

431   commonly held hypothesis for TTX acquisition within *Hapalochlaena* is that it is

432   bacterial in origin, and is either ingested or endosymbiotic[100,101]. Analysis of a

433   ribo-depleted RNA sample from the PSG of *H. maculosa* revealed a highly diverse

434   composition of bacterial genera with Simpson's and Shannon's diversity indices of 4.77

435   and 0.94, respectively. The dominant phyla were Proteobacteria and Firmicutes,

436   composing respectively 41% and 22% of overall bacterial species detected (Fig 5a-b).

437   To date, 151 strains of TTX-producing bacteria have been identified from 31 genera. Of

438   these, 104 are members of Proteobacteria[102]. The genera *Pseudomonas* and *Bacillus*

439   belonging to the phyla Proteobacteria and Firmicutes, respectively, have been

440   previously identified in the PSG of *Hapalochlaena* sp *(Octopus maculosus)*[101].

441   Examination of these bacterial strains revealed TTX production, and extracts injected

442   into mice proved to be lethal[101]. A more recent study on the bacterial composition

443   of *H. maculosa* PSG did not identify TTX-producing strains[100]. However, only a small

444   subset of the many strains identified were tested. Congruent with our findings the

445   diversity of bacterial genera was high and this may complicate identification of species

446   responsible for TTX production. The biosynthetic pathway of TTX has yet to be

447   elucidated, and as a result, only culturable bacterial species can be tested for TTX

448   production.

449

450   **Fig 6. Assessment of bacteria within the posterior salivary gland of *H. maculosa* (PSG). a)** Bacterial composition at

451   the phylum level of a *H. maculosa* posterior salivary/venom gland. **b)** Composition of the largest Phylum,

452   Protobacteria of a *H. maculosa* posterior salivary/venom gland.

453

454   ***Conclusions***

455   This work describes the genome of a unique TTX bearing mollusc, the southern

456   blue-ringed octopus (*Hapalochlaena maculosa*). Much of cephalopod evolution is barely

457   understood due to sparseness of genomic data. Our analysis provides the first glimpse

458   into genomic changes underlying genome evolution of closely related octopod species.

459   While the size, heterozygosity and repetitiveness of the blue ring genome is congruent

460   with previously published octopod genomes, we find similar yet independent

461   expansions of key neuronal gene families across all three species and show evidence for

462   the involvement of gene novelty in the evolution of key neuronal, reproductive, and

463   sensory tissues. The evolution of venom in octopods also differs between species, with

464   *H. maculos*a showing a reduction in the number and expression of serine proteases in

465   their venom gland relative to the other octopods in this study. Inclusion of TTX in *H.*

466   *maculosa* distinguishes this species from related octopods and is believed to impact

467   toxin recruitment and retention, as the highly potent TTX is sufficient to subdue

468   common octopod prey without additional toxins.

469

470   ***Methods***

471   ***Genome sequencing and assembly***

472 DNA was extracted from a single *H. maculosa* female collected at Port Phillip Bay,

473 Victoria, Australia. Two types of Illumina libraries were constructed, standard paired

474 end and Illumina mate pairs (Supplementary data 2). Dovetail sequencing, Chicago

475 libraries improved upon original sequencing resulting in an overall coverage of 71X.

476 Assembly-stats[103] was used to ascertain the quality of the assembly and relevant

477 metrics (Supplementary notes 1).

478

479 ***Transcriptome sequencing***

480 The *H. maculosa* transcriptome was generated using 12 tissues (brain, anterior salivary

481 gland, digestive gland, renal, brachial heart, male reproductive tract, systemic heart,

482 eyeballs, gills, posterior salivary gland, dorsal mantle and ventral mantle tissue). RNA

483 was extracted using the Qiagen RNeasy kit. Construction of cDNA libraries was

484 outsourced to AGRF (Australian Genome Research Facility), Melbourne and conducted

485 using their TruSeq mRNA Library Prep with polyA selection and unique dual indexing

486 method. Libraries were constructed using 3 μg of RNA at a concentration of >100

487 ng/μL. Each tissue was sequenced on 1/12th of an Illumina HiSeq2000 lane with one

488 lane used in total.

489

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)[104]. 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 (Trinotate, RRID:SCR_018930) [105] and domains

assigned by Interpro v72.0 (InterPro, RRID:SCR_006695) [106].

497

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 the git[107]. Additional genes were predicted by mapping

506 raw expressed reads against the genome. Functional annotation of gene models was

507 achieved using InterPro v72.0 (InterPro, RRID:SCR_006695)[106]. Completeness of

508 genes was assessed using BUSCO v3 Metazoan database (BUSCO,

509 RRID:SCR_015008)[40].

510

511 *Heterozygosity*

512 JELLYFISH v2.2.1 (Jellyfish, RRID:SCR_005491) was used in conjunction with

513 GenomeScope (GenomeScope, RRID:SCR_017014)[108] to calculate heterozygosity in

514 *H. maculosa* using a kmer frequency of 21 (Supplementary table 5).

515

516 *Repetitive and transposable elements*

517 Repetitive and transposable elements were annotated using RepeatModeler v1.0.9

518 (RepeatScout) (RepeatModeler, RRID:SCR_015027) and masking performed with

519 RepeatMasker v4.0.8 (RepeatMasker, RRID:SCR_012954)[109](Supplementary notes

520 3.3). Analysis of gene associated TE was conducted by extracting TE within flanking

521 regions 10K upstream and downstream of genes using Bedtools v2.27.1 (BEDTools,

522 RRID:SCR_006646)[110].

523

Calibration of sequence divergence with respect to time

525   Divergence times between the molluscan genomes (*Crassostrea gigas, Lottia gigantea,*

526 *Aplysia californica, Euprymna scolopes, Octopus bimaculoides, Callistoctopus minor* and

527 *Hapalochlaena maculosa*) and transcriptomes (*Sepia officinalis, idiosepius notoides,*

528 *Octopus kaurna* and *Octopus vulgaris*) was obtained using a mutual best hit (MBH)

529 approach. Bioprojects for each genome used are as follows: *Crassostera gigas*

530 (PRJNA629593 & PRJEB3535), *Lottia gigantea* (PRJNA259762 & PRJNA175706),

531 *Aplysia californica* (PRJNA629593 & PRJNA13635) and (*Euprymna scolopes*

532 PRJNA47095). *Octopus bimaculoides* was obtained from this link [111]. The *, Idiosepius*

533 *notoides (*BioProject: PRJNA302677) transcriptome was sequenced and assembled using

534 the same method previously described for the *H. maculosa* transcriptome. Whole

535 genomes and transcriptomes were BLASTed against *Octopus bimaculoides*. The resulting

536 hits were filtered, and alignments shared between all species extracted. A maximum

537 likelihood phylogeny was generated using RAxML v8.0 (RAxML,

538 RRID:SCR_006086)[112]. Phylobayes v3.3 (PhyloBayes, RRID:SCR_006402)[113] was

539 used to calculate divergence times (Supplementary 4.1).

540

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 38x coverage

of reads from libraries with short (500bp) 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. 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[117]

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[118] . 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 assembled imported the

557   resulting data into R for plotting. A mutation rate of 2.4e-9 per base per year and a

558   generation time of 1 year were assumed in order to set a timescale in years and convert

559   coalescence rates to effective population size.

560

561   ***Mutation rate***

562   Mutation rate was calculated by extracting orthologous genes from *O. bimaculoides* and

563   *H. maculosa.* Neutrality was assumed for genes with very low expression (>10 TMP

564   across all tissues). Neutral genes were aligned using MAFFT v7.407[119] and codeml

565   (PAML, RRID:SCR_014932)[120] was used to calculate substitution metrics (dS). Per

566   base neutral substitution between lineages was determined using the mean dS value

567   divided by divergence time (refer to *Calibration of sequence divergence with respect to*

568   *time*) over the number of generations. As octopus are diploid the rate was divided by

569   two. Divergence between species was calculated using Phylobayes v3.3 (PhyloBayes,

570   RRID:SCR_006402)[113].

571

572   ***Quantifying gene expression/ specificity***

573   Gene expression within individual tissues was calculated using Kallisto (kallisto,

574   RRID:SCR_016582)[121] for the transcriptomic data sets of *H. maculosa, O.*

575   *bimaculoides* and *C. minor*. Defaults were used and counts were calculated as TPM.

576   Gene specificity was defined as any gene with a tau value > 0.80.

577

578   ***Gene model expression dynamics***

579   Patterns of gene expression and loss were assessed across octopod genomes at differing

580   taxonomic/organismal levels. Gene models were classified as lineage-specific, octopod

581   specific or non-specific (orthologous to a gene outside of octopods). Expression at each

582   level was determined using whole transcriptomes from all tissues of each species. Genes

583   with expression within one or more tissues were determined to be expressed, loss of

584   expression was classified as a gene with a single ortholog in each species, which is

585   expressed in one or more species and not expressed in the remaining species.

586

587   ***Dynamics of PSG gene expression***

588   In order to identify patterns of PSG specific gene expression (losses and shifts) between

589   the three available octopod genomes, genes with expression specific to the PSG of each

590   species were examined separately. Specific gene expression was defined as a tau value

591   > 0.8. Orthologous groups were identified between species using Orthovenn2[122]

592   and sequences which were identified as lineage-specific were confirmed using BLAST.

593   Types of expressions were categorized as follows: A loss of expression requires a gene

594   to be present in all three octopods and expressed in one or more species while having

595   no detectable expression in at least one species. A shift in expression occurs when an

596   ortholog present in all species is expressed in different tissues.

597

598   ***The role of the Nav in TTX resistance***

599   Sodium channels for the three octopus genomes along with all available in-house

600   cephalopod transcriptomes were extracted manually using a series of BLAST searches

601   against the nr database. Annotation was achieved using Interpro v72.0 (InterPro,

602   RRID:SCR_006695)[106] and identification and extraction of p-loop regions of the

603   sodium channel alpha subunit were manually performed. Where sodium channels were

604   incomplete alignment against related complete channels were used to extract the p-

605   loop regions. Individual mutations with potential to confer resistance were identified

606   manually in Geneious v10.1[123].

607

608   *Microbiome of PSG*

609   A single ribo-depleted RNA sample of *H. maculosa* PSG was examined using the

610   SAMSA2 pipeline[124] to identify the bacterial composition and corresponding

611   molecular functions. Two databases were used Subsys and NCBI RefBac. The Krona

612   package[125] was used to produce visualizations of each dataset.

613

614   **Availability of source code and requirements**

615   Project name: BRO_annotation

616   Project home page: https://github.com/blwhitelaw/BRO_annotation

617   Operating system(s): linux

618   Programming language: Unix/Bash

619   Other requirements: HPC

620   License: GPL-2.0 License

621   Any restrictions to use by non-academics: none

622   RRID: SCR_019072

623

624 **Availability of supporting data and materials**

625 Genomic and transcriptomic data produced and used in this paper have been made

626 available in the NCBI BioProject: PRJNA602771 under the following accession

627 numbers: raw transcriptome (SAMN13930963 - SAMN13930975), genome assembly

628 (SAMN13906985), raw genome reads (SAMN13906958), gene models

629 (SAMN13942395). Voucher specimen for the transcriptome is stored at Melbourne

630 museum. All supporting data and materials are available in the *GigaScience* GigaDB

631 database [126]. This includes expression data for the transcriptome, raw

632 transcriptomes reads, gene models, gene annotation gff and assembled genome, as well

633 as files used in figure generation (i.e. trees, heatmaps).

634

635 **Supplementary Information**

636 Supplementary Notes 1-8, Supplementary Tables 1-8, Supplementary Figs 1-10

637 Supplementary Data 2: Table of genomic Illumina library insert sizes

638

639 **Abbreviations**

640 TTX: Tetrodotoxin, STX: Saxitoxin, PSG: Posterior Salivary Gland, CHGN:Chondroitin

641 N-acetylgalactosaminyltransferase, C2H2(Cys2-His2) zinc finger, Cornifin SPRR:Small

642 Proline-Rich Proteins, LINE: Long Interspersed Nuclear Element, SINE: Short

643 Interspersed Nuclear Element, Mya: Million Years Ago, BUSCO: Benchmarking

644 Universal Single-Copy Orthologs, PSMC: *Pairwise Sequentially Markovian Coalescent* and

645 MSMC:multiple sequentially Markovian coalescent

646

647 **Ethics declaration**

648 Animal Ethics Approval

649 Field collection of fishes, cephalopods (nautiluses, squids, cuttlefishes and octopuses}

650 and decapod crustaceans (crabs, lobsters, crayfishes and their allies) in for Museum

651 Victoria" (Animal Ethics Committee: Museums Victoria; AEC Approval Number: 10006)

652

653 **Competing interests**

654 Authors have no conflicts/competing interests to declare.

655

References

663
664

665 1.   Ritschard EA, Whitelaw B, Albertin CB, Cooke IR, Strugnell JM, Simakov O.
666   Coupled Genomic Evolutionary Histories as Signatures of Organismal Innovations
667   in Cephalopods. BioEssays. 2019;

668 2.   Tanner AR, Fuchs D, Winkelmann IE, Gilbert MTP, Pankey MS, Ribeiro ÂM, et al.
669   Molecular clocks indicate turnover and diversification of modern coleoid
670   cephalopods during the Mesozoic Marine Revolution. Proc R Soc B Biol Sci.
671   2017;284(1850):2016–818.

672 3.   Jereb Roper, C. F. E., Norman, M. D., and Finn, J. K. P. Cephalopods of the
673   world. An annotated and illustrated catalogue of cephalopod species known to
674   date," in Octopods and Vampire Squids, Vol. 3,. Rome FAO. 2014;

675 4.   Graziadei PPC, Gagne HT. Sensory innervation in the rim of the octopus sucker. J
676   Morphol. 1976;150(3):639–79.

677 5.   Froesch D. Projection of chromatophore nerves on the body surface of Octopus
678   vulgaris. Mar Biol. 1973;19(2):153–5.

679 6.   Budelmann BU. The cephalopod nervous system: what evolution has made of the
680   molluscan design. In: The nervous systems of invertebrates: An evolutionary and
681   comparative approach. Springer; 1995. p. 115–38.

682   7.   Gray EG, Young JZ. Electron microscopy of synaptic structure of Octopus brain. J
683   Cell Biol. 1964;21(1):87–103.

684   8.   Fingerhut LCHW, Strugnell JM, Faou P, Labiaga ÁR, Zhang J, Cooke IR. Shotgun
685   proteomics analysis of saliva and salivary gland tissue from the common octopus
686   Octopus vulgaris. J Proteome Res. 2018;17(11):3866–76.

687   9.   Whitelaw BL, Cooke IR, Finn J, Zenger K, Strugnell JM. The evolution and origin
688   of tetrodotoxin acquisition in the blue-ringed octopus (genus Hapalochlaena).
689   Aquat Toxicol. 2019;206:114–22.

690   10.   Cooke IR, Whitelaw B, Norman M, Caruana N, Strugnell JM. Toxicity in
691   cephalopods. Evol Venom Anim Their Toxins. 2015;1–15.

692   11.   Whitelaw BL, Strugnell JM, Faou P, Da Fonseca RR, Hall NE, Norman M, et al.
693   Combined Transcriptomic and Proteomic Analysis of the Posterior Salivary Gland
694   from the Southern Blue-Ringed Octopus and the Southern Sand Octopus. J
695   Proteome Res. 2016;15(9):3284–97.

696   12.   Fry BG, Roelants K, Norman JA. Tentacles of venom: toxic protein convergence in
697   the Kingdom Animalia. J Mol Evol. 2009;68(4):311–21.

698   13.   Ruder T, Sunagar K, Undheim EAB, Ali SA, Wai T-C, Low DHW, et al. Molecular
699   phylogeny and evolution of the proteins encoded by coleoid (cuttlefish, octopus,
700   and squid) posterior venom glands. J Mol Evol. 2013;76(4):192–204.

701   14.   Fry BG, Winter K, Norman JA, Roelants K, Nabuurs RJA, Van Osch MJP, et al.
702   Functional and structural diversification of the anguimorpha lizard venom
703   system. Mol Cell Proteomics. 2010;9(11):2369–90.

704   15.   Hendon RA, Tu AT. Biochemical Characterization of the Lizard Toxin Gilatoxin.
705   Biochemistry. 1981;20(12):3517–22.

706   16.   Matsui T, Fujimura Y, Titani K. Snake venom proteases affecting hemostasis and
707   thrombosis. Biochimica et Biophysica Acta - Protein Structure and Molecular
708   Enzymology. 2000. p. 146–56.

709   17.   Mitreva M, Papenfuss AT, Whittington CM, Locke DP, Mardis E, Wilson RK, et al.
710   Novel venom gene discovery in the platypus. Genome Biol. 2010;11:1–13.

711   18.   Choo YM, Lee KS, Yoon HJ, Qiu Y, Wan H, Sohn MR, et al. Antifibrinolytic role
712   of a bee venom serine protease inhibitor that acts as a plasmin inhibitor. PLoS
713   One. 2012;7(2):e32269.

714   19.   Veiga SS, Da Silveira RB, Dreyfuss JL, Haoach J, Pereira AM, Mangili OC, et al.

715   Identification of high molecular weight serine-proteases in Loxosceles intermedia
716   (brown spider) venom. Toxicon. 2000;38(6):825–39.

717 20. Modica MV, Lombardo F, Franchini P, Oliverio M. The venomous cocktail of the
718   vampire snail Colubraria reticulata (Mollusca, Gastropoda). BMC Genomics.
719   2015;16(1):1–21.

720 21. von Reumont BM, Undheim EAB, Jauss RT, Jenner RA. Venomics of remipede
721   crustaceans reveals novel peptide diversity and illuminates the venom's
722   biological role. Toxins (Basel). 2017;9(8):234.

723 22. Jaimes-Becerra A, Chung R, Morandini AC, Weston AJ, Padilla G, Gacesa R, et al.
724   Comparative proteomics reveals recruitment patterns of some protein families in
725   the venoms of Cnidaria. Toxicon. 2017;137:19–26.

726 23. Crone HD, Leake B, Jarvis MW, Freeman SE. On the nature of "Maculotoxin", a
727   toxin from the blue-ringed octopus (Hapalochlaena maculosa). Toxicon.
728   1976;14(6):423–6.

729 24. Geffeney SL, Williams BL, Rosenthal JJC, Birk MA, Felkins J, Wisell CM, et al.
730   Convergent and parallel evolution in a voltage-gated sodium channel underlies
731   TTX-resistance in the Greater Blue-ringed Octopus: Hapalochlaena lunulata.
732   Toxicon. 2019;170:77–84.

733 25. Wu Z, Yang Y, Xie L, Xia G, Hu J, Wang S, et al. Toxicity and distribution of
734   tetrodotoxin-producing bacteria in puffer fish Fugu rubripes collected from the
735   Bohai Sea of China. Toxicon. 2005;46(4):471–6.

736 26. Tsuruda K, Arakawa O, Kawatsu K, Hamano Y, Takatani T, Noguchi T. Secretory
737   glands of tetrodotoxin in the skin of the Japanese newt Cynops pyrrhogaster.
738   Toxicon. 2002;40(2):131–6.

739 27. Wakely JF, Fuhrman GJ, Fuhrman FA, Fischer HG, Mosher HS. The occurrence of
740   tetrodotoxin (tarichatoxin) in Amphibia and the distribution of the toxin in the
741   organs of newts (Taricha). Toxicon. 1966;3(3):195–203.

742 28. Du Y, Nomura Y, Liu Z, Huang ZY, Dong K. Functional expression of an arachnid
743   sodium channel reveals residues responsible for tetrodotoxin resistance in
744   invertebrate sodium channels. J Biol Chem. 2009;284(49):33869–75.

745 29. McGlothlin JW, Chuckalovcak JP, Janes DE, Edwards S V, Feldman CR, Brodie Jr
746   ED, et al. Parallel evolution of tetrodotoxin resistance in three voltage-gated
747   sodium channel genes in the garter snake Thamnophis sirtalis. Mol Biol Evol.

748   2014;31(11):2836–46.

749 30. Cheng CA, Hwang DF, Tsai YH, Chen HC, Jeng SS, Noguchi T, et al. Microflora

750   and tetrodotoxin-producing bacteria in a gastropod, Niotha clathrata. Food Chem

751   Toxicol. 1995;33(11):929–34.

752 31. Williams BL, Caldwell RL. Intra-organismal distribution of tetrodotoxin in two

753   species of blue-ringed octopuses (Hapalochlaena fasciata and H. lunulata).

754   Toxicon. 2009;54(3):345–53.

755 32. Yotsu-Yamashita M, Mebs D, Flachsenberger W. Distribution of tetrodotoxin in

756   the body of the blue-ringed octopus (Hapalochlaena maculosa). Toxicon.

757   2007;49(3):410–2.

758 33. Mäthger LM, Bell GRR, Kuzirian AM, Allen JJ, Hanlon RT. How does the blue-

759   ringed octopus (Hapalochlaena lunulata) flash its blue rings? J Exp Biol.

760   2012;215(21):3752–7.

761 34. Williams BL, Stark MR, Caldwell RL. Microdistribution of tetrodotoxin in two

762   species of blue-ringed octopuses (Hapalochlaena lunulata and Hapalochlaena

763   fasciata) detected by fluorescent immunolabeling. Toxicon. 2012;60(7):1307–13.

764 35. Savage IVE, Howden MEH. Hapalotoxin, a second lethal toxin from the octopus

765   Hapalochlaena maculosa. Toxicon. 1977;15(5):463–6.

766 36. Kanda A, Iwakoshi-Ukena E, Takuwa-Kuroda K, Minakata H. Isolation and

767   characterization of novel tachykinins from the posterior salivary gland of the

768   common octopus Octopus vulgaris. Peptides. 2003;24(1):35–43.

769 37. Ruder T, Ali SA, Ormerod K, Brust A, Roymanchadi M-L, Ventura S, et al.

770   Functional characterization on invertebrate and vertebrate tissues of tachykinin

771   peptides from octopus venoms. Peptides. 2013;47:71–6.

772 38. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, et al.

773   Improving the Arabidopsis genome annotation using maximal transcript

774   alignment assemblies. Nucleic Acids Res. 2003;31(19):5654–66.

775 39. Stanke M. Gene prediction with a hidden Markov model. University of Göttingen;

776   2004.

777 40. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et

778   al. BUSCO applications from quality assessments to gene prediction and

779   phylogenomics. Mol Biol Evol. 2017;35(3):543–8.

780 41. Zarrella I, Herten K, Maes GE, Tai S, Yang M, Seuntjens E, et al. The survey and

781   reference assisted assembly of the Octopus vulgaris genome. Sci data.

782   2019;6(1):13.

783   42.   Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, et

784   al. The octopus genome and the evolution of cephalopod neural and

785   morphological novelties. Nature [Internet]. 2015;524(7564):220–4. Available

786   from: https://dx.doi.org/10.1038/nature14668

787   43.   Cai H, Li Q, Fang X, Li J, Curtis NE, Altenburger A, et al. A draft genome

788   assembly of the solar-powered sea slug Elysia chlorotica. Sci data.

789   2019;6:190022.

790   44.   Liu C, Zhang Y, Ren Y, Wang H, Li S, Jiang F, et al. The genome of the golden

791   apple snail Pomacea canaliculata provides insight into stress tolerance and

792   invasive adaptation. Gigascience. 2018;7(9):giy101.

793   45.   Du X, Fan G, Jiao Y, Zhang H, Guo X, Huang R, et al. The pearl oyster Pinctada

794   fucata martensii genome and multi-omic analyses provide insights into

795   biomineralization. Gigascience. 2017;6(8):gix059.

796   46.   Powell D, Subramanian S, Suwansa-ard S, Zhao M, O'Connor W, Raftos D, et al.

797   The genome of the oyster Saccostrea offers insight into the environmental

798   resilience of bivalves. DNA Res. 2018;25(6):655–65.

799   47.   Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, et al. Adaptation to deep-sea

800   chemosynthetic environments as revealed by mussel genomes. Nat Ecol Evol.

801   2017;1(5):121.

802   48.   Uliano-Silva M, Dondero F, Dan Otto T, Costa I, Lima NCB, Americo JA, et al. A

803   hybrid-hierarchical genome assembly strategy to sequence the invasive golden

804   mussel, Limnoperna fortunei. Gigascience. 2017;7(2):gix128.

805   49.   Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress

806   adaptation and complexity of shell formation. Nature. 2012;490(7418):49.

807   50.   Jiao W, Fu X, Dou J, Li H, Su H, Mao J, et al. High-resolution linkage and

808   quantitative trait locus mapping aided by genome survey sequencing: building up

809   an integrative genomic framework for a bivalve mollusc. DNA Res.

810   2013;21(1):85–101.

811   51.   Garton DW, Haag WR. Heterozygosity, shell length and metabolism in the

812   European mussel, Dreissena polymorpha, from a recently established population

813   in Lake Erie. Comp Biochem Physiol Part A Physiol. 1991;99(1–2):45–8.

814   52.   Simmons MJ, Crow JF. Mutations affecting fitness in Drosophila populations.
815   Annu Rev Genet. 1977;11(1):49–78.

816   53.   Kumar S, Subramanian S. Mutation rates in mammalian genomes. Proc Natl Acad
817   Sci. 2002;99(2):803–8.

818   54.   Hayward BW, Kawagata S, Grenfell HR, Sabaa AT, O'Neill T. Last global
819   extinction in the deep sea during the mid-Pleistocene climate transition.
820   Paleoceanogr Paleoclimatology. 2007;22(3).

821   55.   Hofreiter M, Stewart J. Ecological change, range fluctuations and population
822   dynamics during the Pleistocene. Curr Biol. 2009;19(14):R584–94.

823   56.   Reynolds T V, Matthee CA, Von Der Heyden S. The influence of Pleistocene
824   climatic changes and ocean currents on the phylogeography of the southern
825   African barnacle, Tetraclita serrata (Thoracica; Cirripedia). PLoS One.
826   2014;9(7):e102115.

827   57.   Huber CD, DeGiorgio M, Hellmann I, Nielsen R. Detecting recent selective sweeps
828   while controlling for mutation rate and background selection. Mol Ecol.
829   2016;25(1):142–56.

830   58.   Mao Y, Economo EP, Satoh N. The roles of introgression and climate change in
831   the rise to dominance of Acropora Corals. Curr Biol. 2018;28(21):3373-3382. e5.

832   59.   Kim B-M, Amores A, Kang S, Ahn D-H, Kim J-H, Kim I-C, et al. Antarctic blackfin
833   icefish genome reveals adaptations to extreme environments. Nat Ecol Evol.
834   2019;3(3):469.

835   60.   Guzik MT, Norman MD, Crozier RH. Molecular phylogeny of the benthic shallow-
836   water octopuses (Cephalopoda: Octopodinae). Mol Phylogenet Evol [Internet].
837   2005;37(1):235–48. Available from:
838   https://dx.doi.org/10.1016/j.ympev.2005.05.009

839   61.   Strugnell J, Jackson J, Drummond AJ, Cooper A. Divergence time estimates for
840   major cephalopod groups: evidence from multiple genes. Cladistics.
841   2006;22(1):89–96.

842   62.   Strugnell JM, Rogers AD, Prodöhl PA, Collins MA, Allcock AL. The thermohaline
843   expressway: the Southern Ocean as a centre of origin for deep-sea octopuses.
844   Cladistics. 2008;24(6):853–60.

845   63.   Lindgren AR, Anderson FE. Assessing the utility of transcriptome data for
846   inferring phylogenetic relationships among coleoid cephalopods. Mol Phylogenet

847   Evol. 2018;118:330–42.

848   64. Styfhals R, Seuntjens E, Simakov O, Sanges R, Fiorito G. In silico Identification
849   and Expression of Protocadherin Gene Family in Octopus vulgaris. Frontiers in
850   Physiology. 2019. p. 1905.

851   65. Zipursky SL, Sanes JR. Chemoaffinity revisited: dscams, protocadherins, and
852   neural circuit assembly. Cell. 2010;143(3):343–53.

853   66. Tepass U, Truong K, Godt D, Ikura M, Peifer M. Cadherins in embryonic and
854   neural morphogenesis. Nat Rev Mol Cell Biol. 2000;1(2):91.

855   67. Peek SL, Mah KM, Weiner JA. Regulation of neural circuit formation by
856   protocadherins. Cellular and Molecular Life Sciences. 2017. p. 4133–57.

857   68. De Wit J, Ghosh A. Specification of synaptic connectivity by cell surface
858   interactions. Nat Rev Neurosci. 2016;1:4.

859   69. Layden MJ, Meyer NP, Pang K, Seaver EC, Martindale MQ. Expression and
860   phylogenetic analysis of the zic gene family in the evolution and development of
861   metazoans. Evodevo. 2010;1(1):12.

862   70. Liu H, Chang L-H, Sun Y, Lu X, Stubbs L. Deep vertebrate roots for mammalian
863   zinc finger transcription factor subfamilies. Genome Biol Evol. 2014;6(3):510–25.

864   71. Young JZ. The number and sizes of nerve cells in Octopus. In: Proceedings of the
865   Zoological Society of London. Wiley Online Library; 1963. p. 229–54.

866   72. Belcaid M, Casaburi G, McAnulty SJ, Schmidbaur H, Suria AM, Moriano-
867   Gutierrez S, et al. Symbiotic organs shaped by distinct modes of genome
868   evolution in cephalopods. Proc Natl Acad Sci [Internet]. 2019;116(8):3030–5.
869   Available from: https://dx.doi.org/10.1073/pnas.1817322116

870   73. Kim B-M, Kang S, Ahn D-H, Jung S-H, Rhee H, Yoo JS, et al. The genome of
871   common long-arm octopus Octopus minor. Gigascience. 2018;7(11):giy119.

872   74. Kapitonov V V, Jurka J. Rolling-circle transposons in eukaryotes. Proc Natl Acad
873   Sci. 2001;98(15):8714–9.

874   75. Oulion B, Dobson JS, Zdenek CN, Arbuckle K, Lister C, Coimbra FCP, et al. Factor
875   X activating Atractaspis snake venoms and the relative coagulotoxicity
876   neutralising efficacy of African antivenoms. Toxicol Lett. 2018;15:119–28.

877   76. Rogalski A, Soerensen C, op den Brouw B, Lister C, Dashevsky D, Arbuckle K, et
878   al. Differential procoagulant effects of saw-scaled viper (Serpentes: Viperidae:
879   Echis) snake venoms on human plasma and the narrow taxonomic ranges of

880   antivenom efficacies. Toxicol Lett. 2017;280:159–70.

881   http://www.sciencedirect.com/science/article/pii/S0378427417312675

882   77.   Sousa LF, Zdenek CN, Dobson JS, den Brouw B op, Coimbra FCP, Gillett A, et al.

883   Coagulotoxicity of Bothrops (lancehead pit-vipers) venoms from Brazil:

884   Differential biochemistry and antivenom efficacy resulting from prey-driven

885   venom variation. Toxins (Basel). 2018;

886   78.   Fox JW, Serrano SMT. Structural considerations of the snake venom

887   metalloproteinases, key members of the M12 reprolysin family of

888   metalloproteinases. Toxicon. 2005;45(8):969–85.

889   79.   Grisley MS, Boyle PR. Bioassay and proteolytic activity of digestive enzymes from

890   octopus saliva. Comp Biochem Physiol Part B Comp Biochem. 1987;88(4):1117–

891   23.

892   80.   Key LN, Boyle PR, Jaspars M. Novel activities of saliva from the octopus Eledone

893   cirrhosa (Mollusca; Cephalopoda). Toxicon. 2002;40(6):677–83.

894   81.   McDonald NM, Cottrell GA. Purification and mode of action of toxin from

895   Eledone cirrosa. Comp Gen Pharmacol. 1972;3(10):243–8.

896   82.   Bordon KCF, Perino MG, Giglio JR, Arantes EC. Isolation, enzymatic

897   characterization and antiedematogenic activity of the first reported rattlesnake

898   hyaluronidase from Crotalus durissus terrificus venom. Biochimie.

899   2012;94(12):2740–8.

900   83.   Frank HY, Catterall WA. Overview of the voltage-gated sodium channel family.

901   Genome Biol. 2003;4(3):207.

902   84.   How C-K, Chern C-H, Huang Y-C, Wang L-M, Lee C-H. Tetrodotoxin poisoning.

903   Am J Emerg Med. 2003;21(1):51–4.

904   85.   Brodie III ED, Brodie Jr ED. Tetrodotoxin resistance in garter snakes: an

905   evolutionary response of predators to dangerous prey. Evolution (N Y).

906   1990;44(3):651–9.

907   86.   Nakamura M, Yasumoto T. Tetrodotoxin derivatives in puffer fish. Toxicon.

908   1985;23(2):271–6.

909   87.   Jost MC, Hillis DM, Lu Y, Kyle JW, Fozzard HA, Zakon HH. Toxin-resistant

910   sodium channels: parallel adaptive evolution across a complete gene family. Mol

911   Biol Evol. 2008;25(6):1016–24.

912   88.   Feldman CR, Brodie ED, Pfrender ME. Constraint shapes convergence in

913   tetrodotoxin-resistant sodium channels of snakes. Proc Natl Acad Sci.
914   2012;109(12):4556–61.

915 89. Jeziorski MC, Greenberg RM, Anderson PA V. Cloning of a putative voltage-gated
916   sodium channel from the turbellarian flatworm Bdelloura candida. Parasitology.
917   1997;115(3):289–96.

918 90. Shen H, Li Z, Jiang Y, Pan X, Wu J, Cristofori-Armstrong B, et al. Structural basis
919   for the modulation of voltage-gated sodium channels by animal toxins. Science
920   (80-). 2018;362:6412.

921 91. Flachsenberger W, Kerr DIB. Lack of effect of tetrodotoxin and of an extract from
922   the posterior salivary gland of the blue-ringed octopus following injection into
923   the octopus and following application to its brachial nerve. Toxicon.
924   1985;23(6):997–9.

925 92. Lopes VM, Baptista M, Repolho T, Rosa R, Costa PR. Uptake, transfer and
926   elimination kinetics of paralytic shellfish toxins in common octopus (Octopus
927   vulgaris). Aquat Toxicol. 2014;146:205–11.

928 93. Ulbricht W, Wagner H, Schmidtmayer J. Kinetics of TTX-STX Block of Sodium
929   Channels. Ann N Y Acad Sci. 1986;479(1):68–83.

930 94. Monteiro A, Costa PR. Distribution and selective elimination of paralytic shellfish
931   toxins in different tissues of Octopus vulgaris. Harmful Algae. 2011;10(6):732–7.

932 95. Li S-C, Wang W-X, Hsieh D. Feeding and absorption of the toxic dinoflagellate
933   Alexandriumtamarense by two marine bivalves from the South China Sea. Mar
934   Biol. 2001;139(4):617–24.

935 96. Braid HE, Deeds J, DeGrasse SL, Wilson JJ, Osborne J, Hanner RH. Preying on
936   commercial fisheries and accumulating paralytic shellfish toxins: a dietary
937   analysis of invasive Dosidicus gigas (Cephalopoda Ommastrephidae) stranded in
938   Pacific Canada. Mar Biol. 2012;159(1):25–31.

939 97. Shea EK, Ziegler A, Faber C, Shank TM. Dumbo octopod hatchling provides
940   insight into early cirrate life cycle. Curr Biol. 2018;28(4):R144–5.

941 98. Lee M-J, Jeong D-Y, Kim W-S, Kim H-D, Kim C-H, Park W-W, et al. A
942   tetrodotoxin-producing Vibrio strain, LM-1, from the puffer fish Fugu
943   vermicularis radiatus. Appl Environ Microbiol. 2000;66(4):1698–701.

944 99. Yotsu M, Yamazaki T, Meguro Y, Endo A, Murata M, Naoki H, et al. Production
945   of tetrodotoxin and its derivatives by Pseudomonas sp. isolated from the skin of a

946   pufferfish. Toxicon. 1987;25(2):225–8.

947 100. Chau R, Kalaitzis J, Wood S, Neilan B. Diversity and biosynthetic potential of
948   culturable microbes associated with toxic marine animals. Mar Drugs.
949   2013;11(8):2695–712.

950 101. Hwang DF, Arakawa O, Saito T, Noguchi T, Simidu U, Tsukamoto K, et al.
951   Tetrodotoxin-producing bacteria from the blue-ringed octopus Octopus
952   maculosus. Mar Biol. 1989;100(3):327–32.

953 102. Magarlamov T, Melnikova D, Chernyshev A. Tetrodotoxin-producing bacteria:
954   Detection, distribution and migration of the toxin in aquatic systems. Toxins
955   (Basel). 2017;9(5):166.

956 103. assembly-stats. https://github.com/sanger-pathogens/assembly-stats.Accessed 29
957   Sep 2020

958 104. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-
959   length transcriptome assembly from RNA-Seq data without a reference genome.
960   Nat Biotechnol. 2011;29(7):644–52.

961 105. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, et
962   al. A tissue-mapped axolotl de novo transcriptome enables identification of limb
963   regeneration factors. Cell Rep. 2017;18(3):762–76.

964 106. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5:
965   genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–
966   40.

967 107. BRO_annotation. https://github.com/blwhitelaw/BRO_annotation. Accessed 29
968   Sep 2020

969 108. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et
970   al. GenomeScope: fast reference-free genome profiling from short reads.
971   Bioinformatics. 2017;33(14):2202–4.

972 109. Smit Hubley, R & Green, P. AFA. RepeatMasker Open-4.0.

973 110. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing
974   genomic features. Bioinformatics. 2010;26(6):841–2.

975 111. Octopus bimaculoides genome files.
976   http://octopus.unit.oist.jp/OCTDATA/BASIC/Metazome/Obimaculoides_280.fa.g
977   z. Accessed 29 Sep 2020

978 112. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses

979   with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90.

980   113. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large

981   phylogenies by maximum likelihood. Syst Biol. 2003;52(5):696–704.

982   114. Li H, Durbin R. Inference of human population history from whole genome

983   sequence of a single individual. Nature. 2012;475(7357):493.

984   115. Schiffels S, Durbin R. Inferring human population size and separation history

985   from multiple genome sequences. Nat Genet. 2014;46(8):919.

986   116. Li H, Durbin R. Fast and accurate short read alignment with Burrows--Wheeler

987   transform. bioinformatics. 2009;25(14):1754–60.

988   117. SNPable. http://lh3lh3.users.sourceforge.net/snpable.shtml. Accessed 29 Sep

989   2020

990   118. msmc-tools. https://github.com/stschiff/msmc-tools. Accessed 29 Sep 2020

991   119. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated

992   protein sequence and structural alignment. Nucleic Acids Res. 2019;47(W1):W5–

993   10.

994   120. Adachi J, Hasegawa M. MOLPHY version 2.3: programs for molecular

995   phylogenetics based on maximum likelihood. Institute of Statistical Mathematics

996   Tokyo; 1996.

997   121. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq

998   quantification. Nat Biotechnol. 2016;34(5):525–7.

999   122. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, et al. OrthoVenn2: a web server for

1000   whole-genome comparison and annotation of orthologous clusters across multiple

1001   species. Nucleic Acids Res. 2019;47(W1):W52–8.

1002 123. Geneious v10.1.

1003   https://assets.geneious.com/documentation/geneious/release_notes.html#v10.1

1004   Accessed 29 Sep 2020

1005 124. Westreich ST, Korf I, Mills DA, Lemay DG. SAMSA: a comprehensive

1006   metatranscriptome analysis pipeline. BMC Bioinformatics. 2016;17(1):399.

1007 125. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in

1008   a Web browser. BMC Bioinformatics. 2011;12(1):385.

1009 126. Whitelaw BL; Cooke IR; Finn J; da Fonseca RR; Ritschard EA; Gilbert MTP;

1010   Simakov O; Strugnell JM. Supporting data for "Adaptive venom evolution and

1011   toxicity in octopods is driven by extensive novel gene formation, expansion and
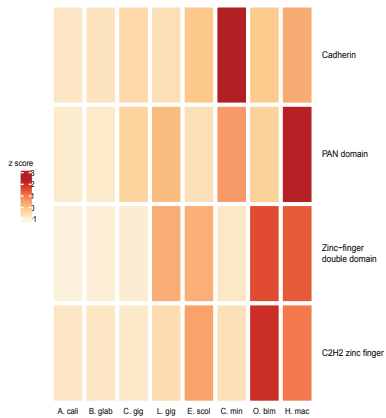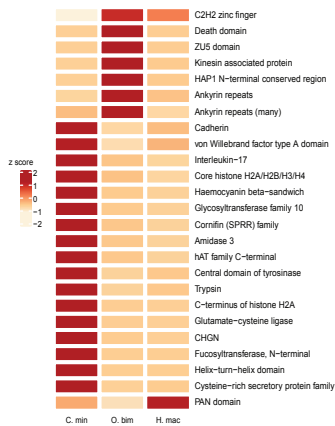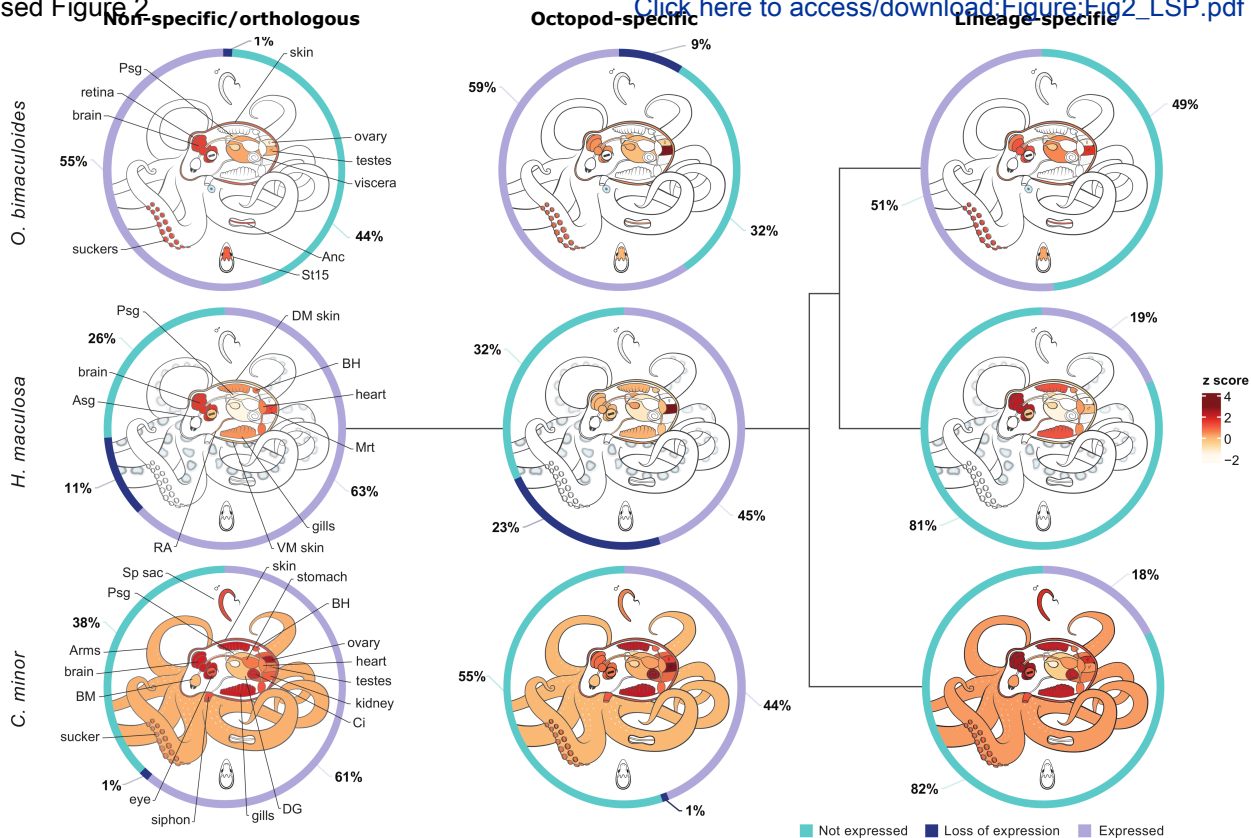
1012   loss" *GigaScience* Database 2020. http://dx.doi.org/10.5524/100793

1013

1014

Figure1

a)

Callistoctopus sp.

Octopus kaurna

Octopus vulgaris

Octopus bimaculoides

Hapalochlaena maculosa

90.91

183.56

29.88

59.45

281.53

Octopods

Sepia officinalis

Idiosepius notoides

Euprymna scolopes

102.75

201.7

Decapods

549.5

Crassostrea gigas

Lottia gigantea

Aplysia californica

533.89

259.6

60.0

b)

z score

Cadherin

PAN domain

Zinc-finger
double domain

C2H2 zinc finger

A. cali B. glab C. gig L. gig E. scol C. min O. bim H. mac

c)

z score

C2H2 zinc finger
Death domain
ZU5 domain
Kinesin associated protein
HAP1 N-terminal conserved region
Ankyrin repeats
Ankyrin repeats (many)
Cadherin
von Willebrand factor type A domain
Interleukin-17
Core histone H2A/H2B/H3/H4
Haemocyanin beta-sandwich
Glycosyltransferase family 10
Cornifin (SPRR) family
Amidase 3
hAT family C-terminal
Central domain of tyrosinase
Trypsin
C-terminus of histone H2A
Glutamate-cysteine ligase
CHGN
Fucosyltransferase, N-terminal
Helix-turn-helix domain
Cysteine-rich secretory protein family
PAN domain

C. min O. bim H. mac

Non-specific/orthologous   Octopod-specific   Lineage-specific

Figure3

a)

b)

Click here to access/download;Figure;Fig3_V2.pdf ⬇

Figure4

Revised Figure 5

a)

consensus	DI DYWEN	DII EWIES	DIII KGWxD	DIV AGWDG
Mya arenaria	- - - - -	- - -D-	- - -I-	- - - - -
Crassostrea gigas	F- - - -	- - -Q-	- - -IE	- - - - -
Mizuhopecten yessoensis	- - - - -	- - - - -	- - -TV	- - - -S
Lottia gigantea	- - -S-	- - - - -	- - -V-	- - - - -
Aplysia californica	F- -S-	- - - - -	- - -I-	- - -SD
Euprymna scolopes	- - - - -	- - - - -	- - -IN	- - - - -
Doryteuthis pealeii	- - - - -	- - - - -	- - -IN	- - - - -
Doryteuthis opalescens	- - - - -	- - - - -	- - -IN	- - - - -
Dosidicus gigas	- - - - -	- - - - -	- - -IN	- - - - -
Grimpoteuthis	- - - - -	- - - - -	- - -M-	- - - - -
Callistoctopus minor	- - - - -	- - - - -	- - - - -	- - - - -
Octopus bimaculoides	- - - - -	- - - - -	- - -I-	- - - - -
Octopus vulgaris	- - - - -	- - - - -	- - -I-	- - - - -
Hapalochlaena maculosa	- - - - -	- - - - -	- - -TE	- - -HS ☠
Hapalochlaena lunulata	- - - - -	- - - - -	- - -TE	- - -HS ☠
Taricha granulosa	- - - - -	- - - -T	- - -T-	- - -SD ☠
Tetraodon nigroviridis	C- - - -	- - - -N	- - -TA	GG- -Q ☠
Homo sapiens	- - - - -	- - - -T	- - -M-	- - - -G

b)

Extracellular

Intracellular

$NH_{2}+$ CHOOH-

P-loop
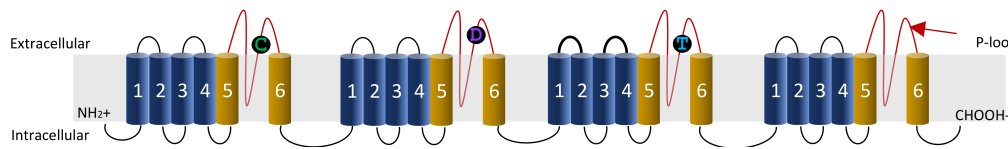
Figure6

Click here to access/download
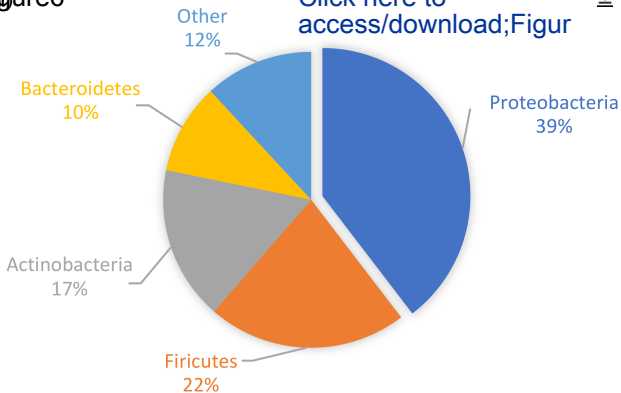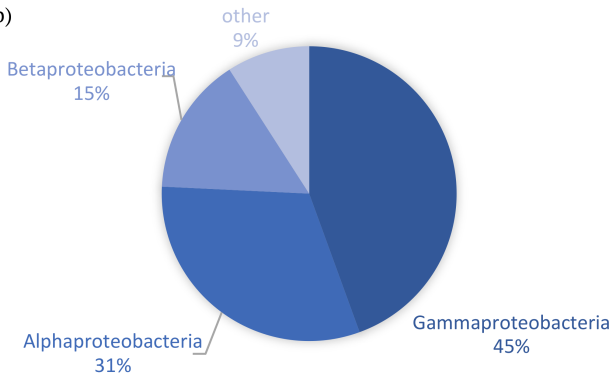
Supplementary Material

SUPPLEMENTARY MATERIALS_GS2_5_8_2020.docx

Supplementary Material(insert sizes)

Click here to access/download
Supplementary Material
BRO Sequencing Insert Sizes.xlsx

JAMES COOK
UNIVERSITY
AUSTRALIA
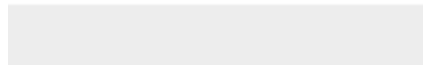
Dear Dr. Goodman

I am pleased to submit an original research piece titled "**Adaptive venom evolution and
toxicity in octopods is driven by extensive novel gene formation, expansion and loss**
" for consideration to be published in *GigaScience*.

Much of cephalopod evolution remains unknown due to sparseness of their genomic
sampling. Cephalopod genomes are some of the largest and most repetitive animal genomes
and exhibit drastically different evolutionary trajectories relative to other better
documented lineages. A more focused genomic study to reveal how individual genomic
changes are associated with the evolution of novel organs, tissues, or adaptations, within a
single group of cephalopods has been missing so far. We present such a study, focussing on
adaptations in the toxic blue-ringed octopus the *Hapalochlaena maculosa,* for which we
provide a high quality genome assembly based on multiple technologies. Members of the
genus *Hapalochlaena* are the only octopods to contain the lethal neurotoxin, tetrodotoxin
(TTX), within their venom and tissues and are a prime example of the origin of
evolutionary novelties within octopods.

Using global comparative genomics approaches and focused study on TTX evolution, we
report key findings:
- Gene family expansions crucial for the development of complex neural networks are
  present in cephalopods, yet are differentially expanding in all three octopod species
- Novel gene formation at different phylogenetic levels can be associated with
  evolution in a specific set of cephalopod tissues
- Changes in Posterior Salivary Gland composition (PSG) between TTX bearing and
  non-TTX bearing species
- Convergently evolved mutations consistent with TTX resistance detected in *H.
  maculosa*

We firmly believe that our manuscript is suited for publication by *GigaScience* as one of the
first to explore the evolutionary genomic basis for novelties within octopods and
cephalopods in general. Our whole genomic comparisons provide insight into the defining
structure/features of octopod genomes at the species-specific level. Additionally, we

examine the impact of TTX on the evolution of venom in *H. maculosa* relative to non-TTX bearing octopods.

Yours sincerely,

Brooke L. Whitelaw
James Cook University,
Townsville, QLD 4810
Mobile: 0424642621
brooke.whitelaw@my.jcu.edu.au

Assoc. Prof. Jan Strugnell
James Cook University,
Townsville, QLD 4810
jan.strugnell.jcu.edu.au

and

Prof. Oleg Simakov
University of Vienna
Universitätsring 1, 1010 Wien, Austria
oleg.simakov@univie.ac.at