GigaScience

Genome sequencing of deep-sea hydrothermal vent snails reveals adaptations to extreme environments

--Manuscript Draft--

Manuscript Number:	GIGA-D-20-00187R1
Full Title:	Genome sequencing of deep-sea hydrothermal vent snails reveals adaptations to extreme environments
Article Type:	Data Note
Funding Information:	National Key R&D Programme of China (No. 2018YFC0310702) — Dr. Xiang Zeng
Abstract:	Background The scaly-foot snail (Chrysomallon squamiferum) is highly adapted to deep-sea hydrothermal vents and drew people's interest once it was found. However, the limited information on its genome impeded related research and understanding of its adaptation to deep-sea hydrothermal vents. Findings Here, we report the whole-genome sequencing and assembly of the scaly-foot snail and another snail (Gigantopelta aegis), which inhabits similar environments. Using ONT, 10X genomic, and Hi-C technologies, we obtained a chromosome-level genome of C. squamiferum with an N50 size of 20.71 Mb. By constructing a phylogenetic tree, we found that these two deep-sea snails were independent of other snails, and their divergence from each other occurred approximately 66.3 million years ago. Comparative genomic analysis showed that different snails have diverse genome sizes and repeat contents. Deep-sea snails have more DNA transposons and LTRs, but fewer LINEs, than other snails. Gene family analysis revealed that deep-sea snails experienced stronger selective pressures than freshwater snails, and the nervous system, immune system, metabolism, DNA stability, antioxidation and biomineralization-related gene families were significantly expanded in scaly-foot snails. We also found 251 class II histocompatibility antigen H2-Aal, which uniquely exist in the Gigantopelta aegis genome, which is important for investigating the evolution of MHC genes. Conclusion Our study provides new insights into deep-sea snail genomes and valuable resources for further studies.
Corresponding Author:	Yaolei Zhang BGI Qingdao, CHINA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	BGI
Corresponding Author's Secondary Institution:	
First Author:	Yaolei Zhang
First Author Secondary Information:	
Order of Authors:	Yaolei Zhang
	Xiang Zeng
	Lingfeng Meng
	Guangyi Fan

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

	Jie Bai
	Jianwei Chen
	Yue Song
	Inge Seim
	Congyan Wang
	Zenghua Shao
	Nanxi Liu
	Haorong Lu
	Xiaoteng Fu
	Liping Wang
	Xin Liu
	Shanshan Liu
	Zongze Shao
Order of Authors Secondary Information:	
Response to Reviewers:	Response to editors and reviewers: Dear editor and reviewers: On behalf of all the coauthors, we would like to thank you very much and all the reviewers for the time spent to assess our manuscript (GIGA-D-20-00187) and for your relevant remarks and suggestions that allow to improve the quality of this manuscript. We have checked the manuscript carefully and revised it according to the comments. Sincerely, Xiang Zeng, Yaolei Zhang Reviewer reports: Reviewer #1: The Data Note by Zeng et al. reported two genome assemblies of deep sea gastropods, Chrysomallon squamiferum and Gigantopelta aegis. I would ask the authors for additional informations about assembling process and references of data sources in order to guarantee the quality of the data and analyses. I also found there are many ambiguous expressions in the present manuscript, making it unclear how the genome resources can contribute to understand biology of these animals. Please find my specific comments and concerns below, which need to be addressed. Response: Thank you very much for your thoughtful and helpful suggestions. We have revised our manuscript according to your comments. Please see that. Background The authors should mention the fact that the Chrysomallon squamiferum genome has been published by Sun et al. (Nat Commun 11, 1657, 2020) somewhere in the Background section. I would suggest the authors to explain that they analyzed the genome of "white scaly foot individual" while Sun et al. sequenced "black" one, to emphasize the uniqueness of this study. Response: Thank you very much for this helpful suggestion. We have added descriptions as you suggested. Please see lines 91-93 “And the whole genome of black scaly-foot snail was reported recently, which highlighted its evolutionary mechanisms of biomineralised armour [9]” and 109-110: “In this study, we sequenced and assembled genomes of the white scaly-foot snail (Figure 1a), which is different from the published black individual”. Line 93 Remove "sp. nov.". This abbreviation is used when new species is named. Response: Thank you. We removed this abbreviation. See lines 94-96: “The genus includes two species, Gigantopelta chessoia from East Scotia Ridge and Gigantopelta

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

aegis from the Southwest Indian Ridge [6]. Both Chrysomallon and Gigantopelta are members of the family Peltospiridae."

Data description
Line 110
Was the insert size 350bp (main text) or 300bp (Table S2)?
Response: Thank you very much. We modified it as "350 bp" in Table S2.

Line 121-122
As mentioned above, the C. squamiferum genome has been published. Therefore this sentence needs to be removed.
Response: Thank you. We delete this sentence.

Line 156-160
These sentences do not make sense to me. Why despite "precise functions of these repeats have not been studied," the authors can infer the composition of repeat elements "may be closely associated with adaptation to extreme environment"? Please describe more specifically by mentioning some references that support this idea.
Response: Thank you and sorry for confusing you. We added more description and relevant references about the importance of transposons/repeats. Please see lines: 156-161 "Although most of the precise functions of these repeats have not been studied in depth, repeats have been thought to have a regulatory function in related genes that play an important role in the life cycle and can introduce great genome flexibility [18]. And in the mammalian genome, transposons were described to be redundant enhancers that regulate their target genes which are higher or tissue specially expressed, indicating the importance of transposons"

Lines 173-175
It seems the authors assumed the split of C. squamiferum and G. aegis was related to the mass extinction event around 66 MYA. However, estimated divergence time is considerably ambiguous (42.4-100 MYA, Fig 1a), making the idea less reliable.
Response: Thank you. We agree with you and delete this description.

Deleted sentences: "This time is consistent with the most recent 'mass extinction', at the end of the Cretaceous geological period ~66 MYA, where ~76% of species became extinct" for your reference.

Lines 178-180
This sentence is difficult to understand. Speciation and demographic histories of each species are different topics.
Response: Thank you. We also delete this sentence.

Deleted sentences: "As the speciation of the two deep-sea snails may be related to geological events (see above)" for your reference.

Lines 190-191
This sentence is not clear. Please describe what "major geological events" affected the population sizes, with references describing the geological events.

An: Thank you. We added one example of geological events and relevant reference. Please see lines 194-197: "For example, the Cretaceous/Paleogene (K/Pg) extinction event caused extinction of three-quarters of species on earth and affected population dynamics approximately 66 million years ago, when an asteroid impact caused global environmental devastation [21, 22]"

Lines 191-193
The recent decreased population size was reported by ref[8] and was not related to this study (Fig2b). Then, this sentence may be put in the Background section.
An: Thank you. We deleted this sentence.

Deleted sentences: " Unfortunately, the C. squamiferum population size has dramatically decreased recently due to deep-sea mining [8], which has made this species endangered" just for your reference

Lines 196-210
Comparisons of Ks and Ka values among snail species should be tested statistically. In the figures 2C and S3 it is not clear whether these values are significantly different.
Response: Thank you. We have added statistical test (Mann-Whitney U test) for Ka/Ks comparisons. Please see lines 204-213: "We found that the Ka values of the two deep-sea snails (average: 0.37 and 0.41) were higher (Mann-Whitney U test, p-value<0.001) than that of the shallow-water limpet (0.35) but similar to those of two freshwater snails (0.39 and 0.41), which suggests that the genes of deep-sea and freshwater snails both evolved faster after their divergence from shallow-water limpet . The Ks values of the deep-sea (3.34 and 3.09) and freshwater (3.19 and 3.24) snails were also similar and lower (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (3.72). Additionally, the Ka/Ks values of the deep-sea snails (average: 0.13 and 0.15) were approximately ~20% and ~40% higher (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (0.11)".

Discussion
Lines 309-310
I have no idea what the "infamous Cambrian Explosion" means. Please explain the authors' idea more in detail.
Response: Thank you. It was modified to "Cambrian Explosion". It was a clerical error. See line 319.

Lines 328-329
Please describe what are "adaptive needs" and "region-specific features" specifically.
Response: Thank you. We deleted this sentence.

Materials and Methods
Line 370
350bp or 300bp?
Response: It is "350 bp" actually for BGISEQ.

Lines 403-405
Based on the description, the 10X Chromium reads were used only for polishing, not for scaffolding. On the other hand, there are stats of scaffolds before Hi-C scaffolding in Table S3.
Response: Yes, the 10X Chromium reads were used only for polishing, not for scaffolding.

My questions are;
i) Were the scaffolds in Table S3 generated using 10X Chromium reads?
ii) If so, the scaffolds were improved very little (sequences are reduced from 6449 to 6444), indicating there was problem in 10X Chromium sequencing. How the authors interpreted the results?
Response: Thank you. The 6,444 scaffolds (also 6,444 contigs) were generated using Oxford Nanopore reads. We then used 10X Chromium reads to polish (error correction) 6,444 scaffolds with software Pilon. Pilon introduced 5 bp gaps so the contig number became 6,449 while the scaffold number was still 6,444.

Line 436
"Lottia"
Response: Thank you. We corrected this word.

Line 454
Describe a reference for GLEAN.
Response: Thank you. We added a reference for GLEAN. Please see line 465.

Lines 464-467
Describe sources or references for these genomic data.
Response: Thank you. We added sources for these genomic data. Please see lines 474-479: " all the protein sequences from selected 10 representative species (8 species including Aplysia californica (GCF_000002075.1), Octopus bimaculoides (GCF_001194135.1), Biomphalaria glabrata (GCF_000457365.1), Crassostrea gigas (GCF_000297895.1), Lottia gigantea (GCF_000327385.1), Pomacea canaliculate (GCF_003073045.1), Pinctada fucata (GCA_002216045.1), Helobdella robusta

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

(GCF_000326865.1) from NCBI database, C. squamiferum and G. aegis from this research) were compared using blastp with the E-value threshold set as 1e-7."

Lines 484-487
Describe references of these fossil records rather than summary database (Timetree.org) so that readers can refer the original data.
Response: Thank you very much and we can't agree with you more. However, each of the time point between two species refers a lot of references and TimeTree database summarized all of these references to estimate one divergence time with a confidence interval. For example, the divergence time between Aplysia californica and Crassostrea gigas was estimated to be 537 MYA with a confidence interval of 516.3-558.3 MYA based on 11 references. We used this confidence interval time to calibrate our estimation. This is a common method used frequently in nowadays genome research. So here we cite TimeTree database (Timetree.org) for reference which includes many references. However, if you think list all the references is a must, we are pleased to do this.

Methods of SNP identification and PSMC (lines 177-193) were not described.
Response: Thank you. We have added methods of SNP identification and PSMC. Please see lines 503-516: "SNP calling and estimation of history population sizes About 50X clean WGS reads were mapped to genomes of C. squamiferum and G. aegis using BWA mem (v0.7.12-r1039) [73] with default parameters respecitvely. Then SAMtools (v0.1.19-44428cd) [74] and "SortSam.jar" in the picard package (v1.54) was used to convert and sort BAM files. Local realignment was again carried out using RealignerTargetCreator and IndelRealigner in GATK (v3.6) [75] with default parameters. SNPs were identified using HaplotypeCaller and filtered using VariantFiltration with parameter "-filter-expression "QD < 2.0 || MQ < 40.0 || ReadPosRankSum < -8.0 || FS > 60.0" --filter-name LowQualFilter --genotype-filter-expression "DP < 5.0" --genotype-filter-name lt_5". Estimatation of history population sizes were carried out using PSMC (v0.6.5-r67) [76]. Firstly, diploid genome references were constructed using samtools and bcftools call with "samtools mpileup -C50" and "vcfutils.pl vcf2fq -d 20 -D 100". Secondly, the demographic history was inferred using PSMC with parameters '-N25 -t15 -r5 -p 4+25*2+4+6' [77]."

Fig 1a
Add the size of the scale for C. squamiferum. No scale is indicated for G.aegis.
Response: Thank you. Scale was added in Fig 1a. Also, we added "Scale bar = 1cm" in Fig 1 legend.
Fig 2a
This cartoon is too ambiguous and not suitable for scientific paper. The molecular phylogeny should be clearly shown by solid lines.
Response: Thank you. We have updated Fig. 2a. Please see that.

Fig.3d
This figure is not very informative for readers. The authors may want to draw molecular phylogeny trees for BTBD6 and HTR4.
Response: Thank you. We also deleted Fig.3d.

Reviewer #2: The manuscript entitled "Genome sequencing of deep-sea hydrothermal vent snails reveals adaptations to extreme environments" presents a nice description of a good genome assembly (16 chromosomes representing ~80% of the genome) of the scaly foot snail (Chrysomallon squamiferum) and compare it to genomes of other molluscan species. Overall the paper is well written and presents a nice view of some unique adaptations by this deep-sea mollusc. One concern that I had is throughout the manuscript (starting at line 164 and onward) the authors describe comparing two mussels, two freshwater snails and two shallow-water snails to their genomes. However, these other molluscan species include C. gigas and P. fucata…which are both oysters and not mussels, and while two of the other molluscs included in the tree are in Gastropoda and considered snails, Lottia gigantea is a limpet and Aplysia californica is a sea slug. I would encourage the authors to describe all of these species more accurately, i.e., as limpet and sea slug, because these are very different from what people commonly think of when they hear "snail", represented by the more traditional Pomacea and Biomphalaria. Referring to all the "snails" as gastropods would be a more suitable term that captures the true diversity of this large group. But

when discussing individual species, I would prefer to see the more accurate descriptions because limpets and sea slugs are different from traditional snails, and will have unique adaptations of their own related to their unique characteristics. Overall, the authors give a good general description of the results and present a reasonable discussion about some of the potential adaptations that they observed in the genome. One minor point - thioredoxins are much more likely play a role in repairing proteins that have been altered by oxidation (Lines 255-256), so to limit this expansion to innate immunity leaves out a lot of other possibilities. My other question was regarding the source of the genomic DNA. The authors describe using muscle samples for isolating DNA, but it is not clear if DNA from one individual was used for all sequencing or if pooling occurred?
Response: Thank you very much for your approval and your thoughtful advices. We have updated our descriptions based on your suggestions. Please see below response.

1) About the scientific name (line 164 and onward), we modified this part as "we compared them with two shallow-water bivalves (P.fucata and C.gigas) and four shallow-water gastropods, including two fresh-water snails (B. glabrata and P. canaliculate), one limpet (B. glabrata) and one sea slug (A. californica). The California two-spot octopus (O. bimaculoides) and the freshwater leech (H.robusta) were used as the outgroup (Figure 2a) (lines 167-171)."
2) About thioredoxins, we added the description of thioredoxin as redox proteins and references. Please lines 260-265: "For example, increased expression of thioredoxin 1 (Txn1; 22 copies in C. squamiferum) was identified. Thioredoxin 1 (Txn1), a redox protein, is important in regulation of cellular redox homeostasis and anti-apoptotic functions. Txn1 stimulates cell proliferation and cell cycle progression, induces hypoxia-inducible factor-1α (HIF-1α) and angiogenesis, and alters the balance between the matrix metalloproteinases and their tissue inhibitors [29, 30]"
3) About isolating DNA, all DNA was isolated from one individual and we add clear description about this in lines 374-375: "DNA was extracted from muscle sample of one individual using the cetyl trimethylammonium bromide (CTAB) method and a DNeasy blood & tissue kit (QIAGEN)."

Reviewer #3: The manuscript of Zeng et al seems to describe a well-put together genome for one species of deep-sea snail, with an additional 'draft' genome for another species. It is clear and well-written, with most of the methods described sufficiently. My main criticism is that I found some of the discussions regarding the adaptative significance and/or putative "function" of various TE content and gene-family expansion results quite speculative, given that the comparative results are often observational with no hypothesis testing or statistical framework. That may well be beyond the remit of the paper, but the language could be more careful in places to reflect the putative nature of any hypothesised effects. Nonetheless I have no doubt that the genomes themselves will be useful additions to the community for future work on mollusc and animal evolution.
Response: Thank you very much for your approval of our manuscript and your helpful criticism. Yes, you are quite right. Here we did not show any experiments results to valid our hypothesis or speculation because this is a data description paper, mainly focusing on observational data. More investigations including both in-depth analysis and experiments based on these two genomes and these findings will be carried out in the future to verify function of important genes or TEs or conserved no-coding regions. These are important and interesting issues and must be done then. Thank you again.

Minor comments:
- Typo line 38: "impedes"
Response: Thank you. This word was modified

- Line 158: is there a reference or two for this? I would assume that most TEs are simply selfish genetic elements that do not serve a "function" per se but exist only for their own purpose, i.e. to copy themselves independently of the host genome
- Line 158-160: but most TE content differences are probably driven by stochastic forces (i.e. drift) rather than deterministic forces such as adaptation, and here we have only 2 data points. The language used for this statement is careful, but I wonder if it is

too far to extrapolate that some differences in TE content may be adaptive
Response: Thank you very much for your helpful thoughts. We added more descriptions and references to support “TEs are functional” to make it clear. Please see lines 156-161: “Although most of the precise functions of these repeats have not been studied in depth, repeats have been thought to have a regulatory function in related genes that play an important role in the life cycle and can introduce great genome flexibility[18]. And in the mammalian genome, transposons were described to be redundant enhancers that regulate their target genes which are higher or tissue specially expressed, indicating the importance of transposon”.

- Line 196: I don't know what the authors mean by this statement
Response: Thank you. We deleted the sentence.

Deleted sentence “The evolution and expression of single-copy orthologous genes are unique features of organisms.” for your reference.

- Section on Ka/Ks values: there is no impression given about the statistical significance of the differences observed between Ka/Ks in any given lineage, or what the distribution of error looks like for these point estimates. Perhaps a more refined PAML analysis could resolve this? It is also not written how Ka/Ks values were calculated
Response: Thank you. We have added statistical test (Mann-Whitney U test) for Ka/Ks comparisons. Please see lines 204-213: “We found that the Ka values of the two deep-sea snails (average: 0.37 and 0.41) were higher (Mann-Whitney U test, p-value<0.001) than that of the shallow-water limpet (0.35) but similar to those of two freshwater snails (0.39 and 0.41), which suggests that the genes of deep-sea and freshwater snails both evolved faster after their divergence from shallow-water limpet . The Ks values of the deep-sea (3.34 and 3.09) and fresh-water (3.19 and 3.24) snails were also similar and lower (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (3.72). Additionally, the Ka/Ks values of the deep-sea snails (average: 0.13 and 0.15) were approximately ~20% and ~40% higher (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (0.11).

And the Ka/Ks values were calculated actually using codeml in PAML package. We have added this information to lines 200-204: “To explore the evolutionary rate of single-copy orthologous genes, we calculated the synonymous substitution rate (Ka) and nonsynonymous substitution rate (Ks) values of 1,324 single-copy orthologous genes shared by the two deep-sea snails, one shallow-water limpet (L. gigantea), and two freshwater snails (B. glabrata and P. canaliculate) using Codeml in PAML package[23].”

- Typo line 214: CAFE not CAFÉ
Response: Thank you. It was modified.
- Line 329: "region-specific feature shared between lineages" - not sure what is meant by this?
Response: Thank you. We deleted this confusing sentence.
- Line 350: it seems speculative - surely both immune response and biomineralization are "vital" for all snails, not particularly deep-sea ones?
Response: Thank you. We modified it and specified it in C. squamiferum which is a chemosynthetic snail species depending on endosymbionts. See lines 360-362 “In particular, we found that DMBT1 gene families that encode multiple SRCR domains expanded significantly in C. squamiferum. These genes play important roles in immune response and biomineralization, both of which are vital for deep sea chemosynthetic snail”.

- Line 454: reference for GLEAN is missing
Response: Thank you. We added this reference (65). See line 465

- Line 469: references for Solar and Hcluster are missing, and what is a H-score?
Response: Thank you. Solar, Hcluster_sg and H-score are tools and concept of TreeFam tools. We added reference for Solar and Hcluster_sg. H-score means hcluster score. Details can be found from TreeFam tools.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

	- Figure 2a: it's a weird looking tree that, in fact, looks a bit like a snail itself! Are the widths of the blobs representative of the error around the divergence times or topological support? Response: Thank you and sorry for confusing you. We updated Fig.2a. Please see that.
Additional Information:	
Question	**Response**
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics 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?	Yes
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	Yes

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

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?	

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Manuscript Click here to access/download;Manuscript;Manuscript_Revison_2-2.docx

1 **Genome sequencing of deep-sea hydrothermal vent snails reveals adaptions to**
2 **extreme environments**

3 Xiang Zeng[1†], Yaolei Zhang[2,3,4†], Lingfeng Meng[2†], Guangyi Fan[2,3,6], Jie Bai[3], Jianwei
4 Chen[2], Yue Song[2] , Inge Seim[7,8], Congyan Wang[2], Zenghua Shao[2], Nanxi Liu[3], Haorong
5 Lu[3], Xiaoteng Fu[1], Liping Wang[1],Xin Liu[2,3,5], Shanshan Liu[2*], Zongze Shao[1*]
6

7 [1]Key Laboratory of Marine Biogenetic Resources, Third Institute of Oceanography,
8 Ministry of Natural Resources, No.178 Daxue Road, Xiamen 361005, China.
9 [2]BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China
10 [3]BGI-Shenzhen, Shenzhen, 518083, China
11 [4]Department of Biotechnology and Biomedicine, Technical University of Denmark,
12 Lyngby, 2800, Denmark
13 [5]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China
14 [6]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China.
15 [7]Integrative Biology Laboratory, College of Life Sciences, Nanjing Normal University,
16 Nanjing, 210046, China
17 [8]Comparative and Endocrine Biology Laboratory, Translational Research Institute-
18 Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland
19 University of Technology, Woolloongabba, 4102, Australia
20

21 [†] These authors contributed equally.
22 *Correspondence: shaozz@163.com (Z.S.); liushanshan@genomics.cn (S.L.)
23

24 **ORCIDs:**
25 Yaolei Zhang: 0000-0003-2131-0254
26 Xiang Zeng: 0000-0001-6662-4551
27 Lingfeng Meng: 0000-0001-9051-0992
28 Guangyi Fan: 0000-0001-7365-1590
29 Jie Bai: 0000-0003-3839-5450
30 Jianwei Chen: 0000-0003-2332-4609
31 Yue Song: 0000-0002-2974-6442
32 Inge Seim: 0000-0001-8594-7217

33 Congyan Wang: 0000-0002-0734-5461
34 Zenghua Shao: 0000-0001-5000-3216
35 Nanxi Liu: 0000-0002-9524-1952
36 Haorong Lu: 0000-0003-0132-8166
37 Xiaoteng Fu: 0000-0001-7937-6980
38 Liping Wang: 0000-0001-7660-7288
39 Xin Liu: 0000-0003-3256-2940
40 Shanshan Liu: 0000-0002-5756-1728
41 Zongze Shao: 0000-0002-4784-090X
42
43
44
45
46
47
48 **Abstract**

49 **Background**

50 The scaly-foot snail (*Chrysomallon squamiferum*) is highly adapted to deep-sea
51 hydrothermal vents and has drawn much interest since its discovery. However, the limited
52 information on its genome has impeded further related research and understanding of its
53 adaptation to deep-sea hydrothermal vents.

54 **Findings**

55 Here, we report the whole-genome sequencing and assembly of the scaly-foot snail and
56 another snail (*Gigantopelta aegis*), which inhabits similar environments. Using Oxford
57 Nanopore Technology, 10X Genomic, and Hi-C technologies, we obtained a chromosome-
58 level genome of *C. squamiferum* with an N50 size of 20.71 Mb. By constructing a
59 phylogenetic tree, we found that these two deep-sea snails are independent of other snails.
60 Their divergence from each other occurred approximately 66.3 million years ago.
61 Comparative genomic analysis showed that different snails have diverse genome sizes and
62 repeat contents. Deep-sea snails have more DNA transposons and LTRs, but fewer LINEs,
63 than other snails. Gene family analysis revealed that deep-sea snails experienced stronger

64 selective pressures than freshwater snails, and the nervous system, immune system,
65 metabolism, DNA stability, antioxidation and biomineralization-related gene families were
66 significantly expanded in scaly-foot snails. We also found 251 H-2 class II
67 histocompatibility antigen, A-U alpha chain-like (*H2-Aal*) genes, which exist uniquely in
68 the *Gigantopelta aegis* genome. This finding is important for investigating the evolution
69 of major histocompatibility complex (MHC) genes.

70 **Conclusion**

71 Our study provides new insights into deep-sea snail genomes and valuable resources for
72 further studies.

73

74 **Keywords:** Deep-sea snails; Genome assembly; Comparative genomics;
75 Biomineralization;

76

77

78

79 **Background**

80 The discovery of deep-sea hydrothermal vents in the late 1970s expanded our knowledge
81 of the extent of life on Earth [1]. Deep-sea macrobenthos, which are animals that inhabit
82 deep-sea hydrothermal vents, face high hydrostatic pressure, variable temperatures and pH,
83 and high levels of hydrogen sulphide, methane, and heavy metals [2]. To date, the literature
84 contains a limited number of studies on the genetics of macrobenthos. A recent report on
85 the genome of deep-sea hydrothermal vent/cold seep mussels (*Bathymodiolus platifrons*)
86 showed that, while most of the genes present in a related shallow-water mussel (*Modiolus*
87 *philippinarum*) have been retained, many gene families have expanded in the *B. platifrons*
88 genome. These families include those that are associated with stabilising protein structures,
89 removing toxic substances from cells, and the immune response to symbionts [3].

90 Gastropods represent the largest class of the phylum Mollusca, with different estimates
91 of diversity varying from 80,000 to 150,000 species [4]. More than 218 gastropod (i.e. snail
92 and slug) species have been described from chemosynthetic ecosystems (i.e. solely rely on
93 endosymbiotic bacteria for sustenance), of which more than 138 are believed to be endemic
94 to these ecosystems [5]. Gastropods are an important component of the fauna in
95 hydrothermal vents in terms of abundance and biomass [6]. Due to the lack of samples and
96 fossil evidence, studies on the evolution and adaptation of deep sea chemosynthetic
97 gastropods are very limited. The scaly-foot snail: *Chrysomallon squamiferum* (*C.*
98 *squamiferum*) is only found in hydrothermal vents at a depth of ~3,000-metres in the Indian
99 Ocean. There are two types of varieties: black (due to greigite, which is an iron sulphide
100 mineral that covers its exterior) scaly-foot individuals from the Kairei field on the central
101 Indian ridge and Longqi field on the Southwest Indian ridge, and white scaly-foot
102 individuals from the Solitaire field on the Central Indian Ridge [7] and Wocan field on the
103 Carlsberg ridge of Northwest Indian ocean (this study). In particular, *C. squamiferum* has
104 been included in the International Union for Conservation of Nature (IUCN) Red List of
105 Endangered Species on July 18, 2019 [8]. Furthermore, the recently reported whole
106 genome of the black scaly-foot snail was reported recently, which highlighted its evolved
107 defence mechanisms of biomineralised armour [9]. *Gigantopelta* spp. is a major
108 megafaunal gastropod genus found in some hydrothermal fields. The genus includes two
109 species, *Gigantopelta chessoia* from East Scotia Ridge and *Gigantopelta aegis* from the

110 Southwest Indian Ridge [6]. Both *Chrysomallon* and *Gigantopelta* are members of the
111 family Peltospiridae. They live in high-density aggregations and share several features,
112 such as a large body size (up to > 45 mm, compared to typical sizes in other taxa of 10-15
113 mm, a 10-50 fold increase in body volume) and an enlarged oesophageal gland [10].
114 In this study, we sequenced and assembled genomes of the white scaly-foot snail
115 *Chrysomallon squamiferum* (NCBI: txid216257; marinespecies.org:taxname:736932)
116 (**Figure 1a**), which differ from the published genomes of the black varieties, from the
117 Wocan field on the Carlsberg ridge of Northwest Indian ocean and *Gigantopelta aegis*
118 (NCBI: txid1735272; marinespecies.org:taxname:853164) *(G. aegis*, **Figure 1a**) from the
119 Southwest Indian Ridge. We gained insights into the evolution, gene family expansions,
120 and adaptations of these extremophile gastropods.
121
122 **Data Description**
123 **Genome assembly and annotation**
124 The *C. squamiferum* genome was sequenced using a combination of sequencing libraries
125 – 10X Genomics, Oxford Nanopore Technology (ONT), and Hi-C – to generate ~369.03
126 Gb of raw data (**Table S1**). Due to the limited sample material, *G. aegis* was sequenced
127 from whole genome shotgun libraries (with 350 bp to 10 kb inserts on the BGISEQ-500,
128 RRID:SCR_017979) to generate 910.08 Gb of raw data (**Table S2**). The genome of *C.*
129 *squamiferum* was assembled with long ONT reads by using Canu v1.7 (Canu, RRID:SCR
130 015880) [11] and WTDBG (WTDBG, RRID:SCR_017225) [12]. After polishing the
131 genome with 10X Genomics sequencing data, a 454.58 Mb assembly (a little smaller than
132 the estimated genome size: 495 Mb, **Figure S1**) with 6,449 contigs and an N50 of 541.32
133 kb was generated (**Table S3**). Next, Hi-C data were used to anchor the assembly, yielding
134 a 16-chromosome assembly (**Figure 1b**). This effort increased the N50 size to ~20.71 Mb
135 (**Table 1**). The 16 chromosomes cover ~80% of the whole genome, and the average length,
136 maximal length, and minimal length of the 16 chromosomes were 22.67, 46.78, and 10.64
137 Mb, respectively, (**Table S4**). A Benchmarking Universal Single-Copy Orthologs
138 (BUSCO, RRID:SCR_015008) completeness score of 94.8% for this genome suggested
139 that it was of good quality (**Table S5**). An approximately 1.29 Gb (a little smaller than the
140 estimated genome size:1.50 Gb, **Figure S1**) genome assembly of *G. aegis* with a scaffold

141 N50 of 120.96 kb (**Table S6**) and a BUSCO completeness score of 92.4% (**Table S7**) was
142 obtained using Platanus (Platanus, RRID:SCR_015531) [13]. After masking repeat
143 elements, we employed homologous and *de novo* prediction methods to construct gene
144 models for the two genomes, obtaining 28,781 *C. squamiferum* genes and 25,601 *G. aegis*
145 genes (**Tables S8** and **S9**). The gene sets were functionally annotated using KEGG (KEGG,
146 RRID:SCR_012773), Swiss-Prot (UniProt, RRID:SCR_002380), InterPro (InterPro,
147 RRID:SCR_006695), and TrEMBL (TrEMBL, RRID:SCR_002380) (**Tables S10** and
148 **S11**).
149
150 **Genome sizes and repeat contents.**
151 The genome assembly sizes of *C. squamiferum* (~455.36 Mb) and *G. aegis* (~1.29 Gb)
152 differed from those of freshwater snails (~916 Mb (*Biomphalaria glabrata*) [14] and ~440
153 Mb (*Pomacea canaliculate*) [15]), which suggests that there is significant genome size
154 diversity within snails (**Figure 1c**). In the absence of ploidy effects [16, 17], differences in
155 genome size often stem from the accumulation of various repetitive elements. A
156 comparison of the repeat elements (**Figure 1c** and **Table S12**) supported this trend. The
157 genomes of *C. squamiferum* and *P. canaliculate* (smaller genome sizes) contained fewer
158 repeats than *B. glabrata* and *G. aegis*, whereas *G. aegis* had more repeats than *B. glabrata*
159 (**Figure 1d**). This finding suggests that snail genome sizes correlate with repeat content.
160 Despite the similar genome sizes of *C. squamiferum* and *P. canaliculata,* their genome
161 landscapes were distinct. For example, ~10.17% of the *C. squamiferum* genome consisted
162 of tandem repeats compared to ~2.89% in *P. canaliculata* (**Table S12**). DNA transposons
163 and LTRs comprise ~17.73% and ~5.99% of the *C. squamiferum* genome, respectively,
164 but only ~6.84% and ~3.53% in *P. canaliculata*. LINEs made up ~8.63% of the *P.*
165 *canaliculata* genome compared to ~5.65% in *C. squamiferum*. Similarly, although the
166 larger *G. aegis* and *B. glabrata* genomes have similar proportions of tandem repeats, *G.*
167 *aegis* had a higher percentage of DNA transposons (~32.15% versus ~20.20%) and LTRs
168 (~13.32% versus ~3.75%). LINEs made up ~23.93% of the *B. glabrata* genome compared
169 to ~11.51% in *G. aegis*. Taken together, these data suggest that deep-sea hydrothermal vent
170 snail genomes have more DNA transposons and LTRs and fewer LINEs than their
171 freshwater counterparts. In particular, DNA/CMC-EnSpm, DNA/TcMar−Tc1, and

172 DNA/DNA were the main factors that caused the differences in DNA transposon content
173 in the four snail genomes (**Figure 1d**). We found that LINE/L2, LINE/RTE-BovB,
174 LINE/LINE, and LINE/CR1 were much higher in fresh-water snail genomes than in deep-
175 sea snails. Although most of the precise functions of these repeats have not been studied in
176 depth, repeats have been thought to have a regulatory function in related genes that play an
177 important role in the life cycle and can introduce great genome flexibility [18]. Also, in the
178 mammalian genome, transposons have been described as redundant enhancers that regulate
179 their target genes, which are higher or specific tissue expressed, indicating the importance
180 of transposons[19]. Thus, we might infer that the expansion of DNA transposons and LTRs,
181 as well as the absence of some LINEs, may be closely associated with important genes that
182 help these deep-sea snails adapt to extreme environments.

183 **Construction of phylogenetic relationships for deep-sea snails**

184 To determine the phylogenetic relationships between deep-sea snails and other molluscs,
185 we compared their genomes with those from two shallow-water bivalves (*P.fucata* and
186 *C.gigas*) and four shallow-water gastropods, including two fresh-water snails (*B. glabrata*
187 and *P. canaliculate*), one limpet (*B. glabrata*) and one sea slug (*A. californica*). The
188 genomes of the California two-spot octopus (*O. bimaculoides*) and the freshwater leech
189 (*H.robusta*) were used as the outgroup (**Figure 2a**). We identified 26,668 gene families in
190 the ten species examined (**Table S13**). Phylogenetic trees were constructed from 406
191 shared single-copy orthologs. Both maximum likelihood (ML) and Bayesian methods
192 revealed the same topology (**Figure 2a** and **Figure S2**), which is consistent with a recent
193 study [15]. In the tree, bivalves and gastropods are clearly separated and the two deep-sea
194 snails are located on the same branch and are independent of other snails (although their
195 genome sizes are quite different). We estimated that *C. squamiferum* and *G. aegis* diverged
196 from a common ancestor approximately 66.3 million years ago (MYA).
197

198 **Demographic histories of the deep-sea snails**

199 Based on these two assembled genomes, we estimated their historical effective
200 population size (N_e) using whole-genome genetic variation. We identified ~3.51 and ~3.19
201 million heterozygous SNPs with nucleotide diversities of 0.0077 and 0.0025 for *C.*
202 *squamiferum* and *G. aegis*, respectively. We estimated changes in N_e using the pairwise

203 sequential Markovian coalescent (PSMC, RRID:SCR_017229) method, which can infer
204 demography from approximately 20,000 to 1 MYA [20]. The effective population sizes of
205 *C. squamiferum* and *G. aegis* – species derived from different geographical locations in the
206 Indian Ocean – are distinct (**Figure 2b**). In the demographic history of *G. aegis* decreased
207 until ~250 thousand years ago, followed by an N_e increase, from ~50,000 to 450,000
208 individuals, 20,000 years ago. Several cycles of increasing and decreasing N_e have been
209 observed for *C. squamiferum*, with the effective population size recovering and stabilising
210 at 35,000 individuals approximately 70 thousand years ago. Thus, although deep-sea
211 habitats are inhabited, deep-sea snail populations are sensitive to habitat disturbances. It
212 was reported that vent organisms are exquisitely sensitive to nuances in fluid flux, such as
213 chemical compositions, temperature, geological setting and biological interactions[21].Our
214 results revealed that the demographic histories of these two snails differd because their
215 habitat conditions are markedly different.
216
217 **Evolution of single-copy orthologous genes**
218 To explore the evolutionary rate of single-copy orthologous genes, we calculated the
219 synonymous substitution rate (*Ka*) and nonsynonymous substitution rate (*Ks*) values of
220 1,324 single-copy orthologous genes shared by the two deep-sea snails, one shallow-water
221 limpet (*L. gigantea*), and two freshwater snails (*B. glabrata* and *P. canaliculate*) using
222 Codeml in the PAML package (PAML, RRID:SCR_014932) [22] (**Figure 2c**, **Figure S3,**
223 and **Table S15**). We found that the *Ka* values of the two deep-sea snails (average: 0.37 and
224 0.41) were higher (Mann-Whitney U test, p-value<0.001) than that of the shallow-water
225 limpet (0.35) but similar to those of two freshwater snails (0.39 and 0.41), which suggests
226 that the genes of deep-sea and freshwater snails both evolved faster after their divergence
227 from the shallow-water limpet. The *Ks* values of the deep-sea (3.34 and 3.09) and fresh-
228 water (3.19 and 3.24) snails were also similar to and lower (Mann-Whitney U test, p-
229 value<0.001) than those of the shallow-water limpet (3.72). Additionally, the *Ka*/*Ks* values
230 of the deep-sea snails (average: 0.13 and 0.15) were approximately ~20% and ~40% higher
231 (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (0.11). From
232 these findings, we could infer that deep-sea snails have experienced stronger selective

233 pressures than the shallow and fresh water species discussed here, possibly to allow
234 adaptation to life in hydrothermal vents.
235
236 **Expanded gene families in deep-sea snail genomes**
237 ***Nervous system***
238 Using CAFE (CAFE, RRID:SCR_005983) [23] (*see details in Methods*), we identified
239 two significantly (p-value < 0.01) expanded gene families in the two deep-sea snail
240 genomes compared to the freshwater snails and shallow-water limpet. The BTB/POZ
241 domain-containing protein 6 (*BTBD6*) had 56 copies in *C. squamiferum* and 35 copies in
242 *G. aegis*, while fewer than 5 copies were found in the four other snail species examined
243 (**Figure 3a**). We found 17 *BTBD6* genes on chromosome 16 of *C. squamiferum,* and these
244 genes showed traces of tandem duplications (**Figure 3b**). In *G. aegis*, we also found several
245 tandem gene clusters (**Figure 3b**). *HTR4* (5-hydroxytryptamine receptor 4) had 12 copies
246 in *C. squamiferum* and 18 copies in *G. aegis*, while only one copy was found in the other
247 snail species (**Figure 3c**). The expansions of these gene families also displayed tandem
248 duplications (**Figure S4**). Both of these genes have roles in neuroregulation; *BTBD6*, which
249 is an adaptor of the Cul3 ubiquitin ligase complex and is essential for neural differentiation
250 [24], while *HTR4* modulates the release of various neurotransmitters [25]. A previous study
251 revealed that a large unganglionated nervous system exists in *C. squamiferum* [7]. We
252 speculate that the expansions of *BTBD6* and *HTR4* contribute to this system by sustaining
253 life in a deep-sea environment.
254
255 ***Metabolism related genes***
256 *C. squamiferum* houses abundant endosymbionts in its greatly enlarged oesophageal
257 gland and these endosymbionts supply nutrition for its host. KEGG enrichment analysis on
258 the 183 expanded gene families of *C. squamiferum* revealed significant enrichment for
259 metabolic pathways (q-value < 0.0001, **Table S16**). Among these genes, nine gene families
260 encoded enzymes in the glycolysis pathway and citrate cycle (TCA cycle). For example,
261 the genes for isocitrate dehydrogenase (IDH), which catalyses the oxidative
262 decarboxylation of isocitrate to produce α-ketoglutarate and CO2, expanded significantly
263 ($p < 0.01$). The α-ketoglutarate dehydrogenase complex (OGDC) consists of three

264 components: oxoglutarate dehydrogenase (OGDH), dihydrolipoyl succinyltransferase
265 (DLST), and dihydrolipoyl dehydrogenase (DLD), among which the genes for OGDH were
266 expanded ($p < 0.01$, **Figure 4a**). IDH and OGDC are two rate-limiting enzymes in the TCA
267 cycle, and the related biochemical reactions are irreversible (**Figure 4b**).
268
269 ***Defence mechanisms***
270 Endosymbiotic bacteria are critical for snail life in deep-sea hydrothermal vent
271 ecosystems [26]. These bacterial taxa are largely restricted to chemosynthetic
272 environments, with some being exclusive to vents [27]. The divergent evolution of the *C.*
273 *squamiferum* and *G. aegis* genomes may have generated diverse defence mechanisms.
274 A total of 183 expanded gene families were identified in the *C. squamiferum* genome.
275 As expected, many of these families have roles in the immune system. However, unlike
276 the freshwater snail *B. glabrata* [14] and deep-sea mussels [3], we did not detect an
277 expansion of the Toll-like receptor 13 (*TLR13*) gene family, but identified other expanded
278 gene families (**Figure 4a**). For example, increased expression of thioredoxin 1 (*Txn1*; 22
279 copies in *C. squamiferum*) was identified. Thioredoxin 1 (*Txn1*), a redox protein, is
280 important for regulating of cellular redox homeostasis and anti-apoptotic functions. Txn1
281 stimulates cell proliferation and cell cycle progression, induces hypoxia-inducible factor-
282 1α (HIF-1α) and angiogenesis, and alters the balance between the matrix
283 metalloproteinases and their tissue inhibitors [28, 29]. Txn1 also plays a pivotal role in T
284 cell activation in mice [30]. Although T-cell related adaptive immunity only appears in
285 vertebrates, the existence and expansion of this gene may assist the innate immune system
286 of *C. squamiferum*. Glutamine-fructose-6-phosphate transaminase (*GAFT*; 21 copies in *C.*
287 *squamiferum*) promotes the biosynthesis of chitin [31, 32], which is one of the stable
288 components of the crustacean shell and provides protection against predation and infection.
289 We identified expanded gene families that maintain the stability of nucleic acids and
290 proteins, such as heat shock protein 90 (Hsp90; 13 copies in *C. squamiferum***, Figure 4a**),
291 which protects proteins against heat stress [33]; the single-stranded DNA-binding proteins,
292 encoded by SSB genes (19 copies in *C. squamiferum,* and 1 copy in other species, **Figure**
293 **4a)**, which are required for DNA replication, recombination, and repair processes [34]; and
294 catalase (*CAT*, 6 copies *C. squamiferum*; **Figure 4c**), which is critical in the response

295 against oxidative stress [35]. The elevated levels of heavy metals and sulphide and high
296 temperatures in hydrothermal vents are likely to greatly increase the risk of DNA damage
297 and misfolded proteins. Thus, these expanded gene families may help these snails resist
298 environmental stress.

299 We also found a special gene family, deleted in malignant brain tumours 1 (*DMBT1*),
300 expanded (70 copies, **Figure 4a**) in the *C. squamiferum* genome. *DMBT1* can encode three
301 glycoproteins (DMBT1 (deleted in malignant brain tumours 1 protein), SAG (salivary
302 agglutinin), and GP340 (lung glycoprotein-340)) and belongs to the scavenger receptor
303 cysteine-rich (SRCR) protein superfamily of the immune system [36]. This gene consists
304 of the SRCR, CUB, and zona pellucida domains, and all 70 copies of this gene in *C.*
305 *squamiferum* contain the SRCR domain, which can bind a broad range of pathogens,
306 including cariogenic *streptococci*, *Helicobacter pylori,* and HIV [37]. However, previous
307 studies have shown that SRCR domains that contain proteins are commonly expressed in
308 the shell martrix [38] and have been proven to be potentially linked to biomineralization
309 [39], which would be associated with the foot scales of *C. squamiferum*. Nonetheless, the
310 expansion of this gene family will either strengthen the immune ability or help construct
311 the scale armour of these snails.

312 Correspondingly, we identified the expansion of 198 gene families (containing 4,515
313 genes) in the *G. aegis* genome. These families were enriched in 58 KEGG pathways
314 (*q*value < 0.05) (**Table S17**). The majority of these pathways were associated with the
315 immune and disease response, and included terms such as ‘infection’, ‘NOD-like receptor
316 signalling’, ‘Tumour necrosis factor (TNF) signalling pathway’, and ‘Antigen processing
317 and presentation’ (**Figure S5**). Surprisingly, we found 251 copies of the H-2 class II
318 histocompatibility antigen, A-U alpha chain-like (H2-Aal) genes, which is one of the major
319 histocompatibility complex (MHC) genes in vertebrates [40]. The existence and super
320 expansion of this gene family in the invertebrate positions in *G. aegis* is useful for the study
321 of immune system evolution.

322

323 Discussion

324 Molluscs are a highly diverse group, and their high biodiversity makes them an excellent
325 model to address topics such as biogeography, adaptability, and evolutionary processes

326 [41]. Members of the family Peltospiridae in the gastropod clade Neomphalina are
327 restricted to chemosynthetic ecosystems and, so far, are only known from hot vents [6].
328 Based on the chromosome-scale genome assembly analyses of the scaly-foot snail (*C.*
329 *squamiferum*) and deep-sea snail (*G. aegis*), which both belong to the Peltospiridae family
330 from chemosynthetic ecosystems, our results provide insight into the possible evolution
331 and adaptation mechanisms of hydrothermal vent animals.
332 By constructing a phylogenetic tree, we found that snails diverged from other molluscs
333 approximately 555.2 MYA (**Figure 2a**). These two deep-sea snails were found to be
334 independent of other shallow-water gastropods around 536.6 MYA. At the end of the
335 Cretaceous geological period, approximately 66.3 MYA, *C. squamiferum* and *G. aegis*
336 diverged from each other and later had different *Ne* (**Figure 2b**). This finding indicated
337 that they faced different environmental factors and selected pressures. This evolutionary
338 time frame implies that the last common ancestor of all molluscs (LCAM) already lived
339 before the Cambrian Explosion (530-540 MYA), which was also speculated by the
340 palaeobiological hypothesis [42]. It also elucidated that deep-sea gastropod lineages
341 originated at least around 540 MYA and diverged from other gastropods in the same age
342 of the oldest molluscs taxons, Aculifera and Conchifera [43, 44]. The deep-sea gastropod
343 lineages were also confirmed by the phylogenetic analysis of mitogenomes [45]. Further
344 conceived by the evolutionary rate of single-copy orthologous genes, deep-sea gastropod
345 lineages have experienced stronger selective pressures than shallow-water gastropods
346 (**Figure 2c**).
347 Transposable elements (TEs) play multiple roles in driving genome evolution in
348 eukaryotes[46]. The genome sizes of four representative snails were quite divergent (440
349 Mb-1.29 Gb). The deep-sea snail *G. aegis* had the largest genome (1.29 Gb), with the
350 highest percentage of DNA transposons (32.15%). Deep-sea snails (*C. squamiferum* and
351 *G. aegis*) had more DNA transposons and LTRs than other snails, but fewer LINEs. LTR
352 class has been identified as the main contributor to open chromatin regions and
353 transcription factor binding sites [47, 48]. LINEs may be associated with the duplicability
354 of genomic regions, which are always shared between related lineages[49]. Thus, the
355 higher portions of DNA transposons and LTRs may be the results of genome evolution due

356 to environmental changes and associated with the ability of deep-sea snails to adapt to
357 extreme environments.

358 Specifically, we analysed expanded gene families in deep-sea snail genomes (**Figure
359 4a**). They both significantly expanded the nervous system, especially *BTBD6* and *HTR4*,
360 which are involved in the neuroregulation of activities, such as movement, predation, and
361 resistance to environmental change. As for the chemosynthetic snails, they both had
362 expanded immune system-related gene families. In the *C. squamiferum* genome, the
363 expansions of *Txn1* and *GAFT* were found. In the *G. aegis* genome, different immune and
364 disease response genes families were expanded, such as *H2-Aal* genes. These expanded
365 gene families were different from those found in fresh water snails and deep-sea mussels.

366 Interestingly, in the scaly-foot snail (*Chrysomallon squamiferum*) genome, genes
367 involved in the main metabolic pathways were significantly enriched, including the
368 glycolysis pathway and the citrate cycle (TCA cycle) . Other enriched gene families
369 included the single-stranded DNA-binding protein (*SSB)* family, which stabilise ssDNA;
370 heat shock protein 90 (*Hsp90*) family, which keep proteins folded properly; and the
371 catalase (*CAT*) family, which prevents the generation of free radicals due to exposure to
372 peroxides. The expansions of these gene families may have provided deep-sea snails with
373 better immune reactions with symbionts, rapid nerve signal conduction, stronger
374 metabolism, and effective resistance while adapting to their hydrothermal vent habitat.

375 In particular, we found that *DMBT1* gene families that encode multiple SRCR domains
376 were expanded significantly in *C. squamiferum*. These genes play important roles in
377 immune response and biomineralisation, both of which are vital for deep-sea
378 chemosynthetic snail.

379 In conclusion, the genome analysis of deep-sea snails (*C. squamiferum* and *G. aegis*)
380 from hydrothermal vents revealed mechanisms of their evolution and molecular
381 adaptations to extreme environments, and will be a valuable resource for studying the
382 evolution of invertebrates.

383

384 **Materials and Methods**

385 **Sample collection and DNA isolation**

386 *C. squamiferum* samples were obtained from the Wocan vent field (60.5°E 6.4°N, 2919m
387 depth) on the Carlsberg Ridge, northwest Indian Ocean, in March 2017 during the Chinese
388 DY38th cruise. *G. aegis* samples were obtained from the Longqi vent field (37.5°S, 49.4°E,
389 2,780 m) on the southwest Indian ridge in March 2015 during the Chinese DY35th cruise.
390 DNA was extracted from the muscle sample of one individual using the cetyl
391 trimethylammonium bromide (CTAB) method and a DNeasy blood & tissue
392 kit (QIAGEN). DNA quality and quantity were checked using pulsed field gel
393 electrophoresis and a Qubit Fluorometer (Thermo Scientific).

394

395 **Libraries preparation and sequencing**

396 ***Whole Genome Shotgun Sequencing***

397 Four WGS libraries were prepared for sequencing: one short insert size library (350 bp)
398 and three mate-pair large insert size libraries (2 kb, 5 kb, and 10 kb). Libraries were
399 constructed using an MGI Easy FS DNA Library Prep Set kit (MGI, China). Paired-end
400 reads (100 bp) and mate-pair reads (50 bp) were obtained from the BGISEQ-500 platform.

401

402 ***10X Genomics sequencing***

403 To prepare the Chromium library, 1 ng of high quality DNA was denatured, spiked into
404 the reaction mix, and mixed with gel beads and emulsification oil to generate droplets
405 within a Chromium Genome chip. Then, the rest of the steps were completed following the
406 standard protocols for performing PCR. After PCR, the standard circularisation step for
407 BGISEQ-500 was carried out, and DNA nanoballs (DNBs) were prepared [50]. Paired-end
408 reads with a length of 150 bp were generated on the BGISEQ-500 platform [51].

409

410 ***Oxford Nanopore Technologies***

411 DNA for long-read sequencing was isolated from the muscle tissues of our samples.
412 Using 5 flow cells and the ONT chemistry for the GridION X5 sequencer (GridION,
413 RRID:SCR_017986).

414 following manufacturer's protocols, we generated 39.61 Gbp of raw genome sequencing
415 data.
416
417 ***Hi-C* library and sequencing**
418 The Hi-C library was prepared following the standard *in situ* Hi-C [52] protocol for muscle
419 samples, using *DpnII* (NEB, Ipswich, America) as the restriction enzyme. After that, a
420 standard circularization step was carried out, followed by DNA nanoballs (DNB)
421 preparation following the standard protocols of the BGISEQ-500 sequencing platform as
422 previously described [50]. Paired-end reads with a length of 100 bp were generated on the
423 BGISEQ-500 platform[51].
424
425 **Genome assembly**
426 For the genome assembly of *Chrysomallon squamiferum*, Canu v1.7 was first used to
427 perform corrections of ONT reads with the parameters "correctedErrorRate=0.105
428 corMinCoverage=0 minReadLength=1000 minOverlapLength=800". Then, wtdbg (v1.2.8)
429 was used to assemble the genome with the parameters "--tidy-reads 3000 -k 0 -p 21 -S 4 -
430 -rescue-low-cov-edges" using corrected reads generated by Canu. Next, we made use of
431 the sequencing reads from the 10X Genomic library to carry out genome polishing using
432 Pilon v1.22 (Pilon, RRID:SCR_014731) with its default parameters. Quality control of Hi-
433 C sequencing reads was first performed using the HiC-Pro pipeline (HiC-Pro,
434 RRID:SCR_017643) [53] with the parameters "[BOWTIE2_GLOBAL_OPTIONS = --
435 very-sensitive -L 30 --score-min L,-0.6,-0.2 --end-to-end –reorder;
436 BOWTIE2_LOCAL_OPTIONS = --very-sensitive -L 20 --score-min L,-0.6,-0.2 --end-to-
437 end –reorder; IGATION_SITE = GATC; MIN_FRAG_SIZE = 100; MAX_FRAG_SIZE
438 = 100000; MIN_INSERT_SIZE = 50; MAX_INSERT_SIZE = 1500]". In total,
439 23,646,810 pairs of valid reads were obtained. Next, the valid Hi-C data was used to anchor
440 the nanopore contigs onto chromosomes separately by applying the 3D-DNA pipeline [54].
441 The contact maps were then generated by the Juicer pipeline [55], and the boundaries for
442 each chromosome were manually rectified by visualising the inter.hic file in Juicebox [56].
443 16 chromosomes were identified by combining the linkage information from the agp file.

444 For the genome assembly of *G. aegis*, we obtained only WGS sequencing reads because
445 of limited DNA and tissue samples. Platanus v1.2.4 [13] was used to assemble the genome
446 with WGS clean data with the parameters "assemble –k 29 –u 0.2, scaffold -l 3 -u 0.2 -v
447 32 -s 32 and gap_close –s 34 –k 32 –d 5000". BUSCO v2 were used to evaluate genome
448 assemblies with the metazoan_odb9 database.
449
450 **Genome annotation**
451 ***Repeat annotation***
452 Homolog-based and *de novo* prediction methods were used to detect repeat contents. In
453 particular, RepeatMasker v4.0.5 (RepeatMasker, RRID:SCR_012954) [57] and
454 RepeatProteinMask v4.0.5 (RepeatProteinMask, RRID:SCR_012954) were used to detect
455 transposable elements against the Repbase database[58] at the nuclear and protein levels,
456 respectively. RepeatMasker was used again to detect species-specific transposable
457 elements against databases generated by RepeatModeler v1.0.8 (RepeatModeler,
458 RRID:SCR_015027) and LTR-FINDER v1.0.6 (LTR-FINDER, RRID:SCR 015247) [59].
459 Moreover, Tandem Repeat Finder v4.0.7 [60] was utilised to predict tandem repeats.
460
461 ***Gene annotation***
462 We combined homology-based and *de novo* evidence to predict protein-coding genes in
463 two genomes. For the homology-based method, we used six relative gene sets of *Aplysia*
464 *californica*, *Bathymodiolus platifrons*, *Biomphalaria glabrata*, *Lottia gigantea*, *Modiolus*
465 *philippinarum*, and *Pomacea canaliculata*. First, these homologous protein sequences were
466 aligned onto each assembled genome using TBLASTN (TBLASTN, RRID_SCR 011822),
467 with an *E*-value cut-off of 1×10^{-5}, and the alignment hits were linked to candidate gene
468 loci by GenBlastA [61]. Second, we extracted genomic sequences of candidate gene
469 regions, including 2 kb flanking sequences, and then used GeneWise v2.2.0 (GeneWise,
470 RRID:SCR_015054) [62] to determine gene models.
471
472 In the *de novo* method, we used Augustus (Augustus, RRID:SCR_008417)[63] to predict
473 the gene models on repeat-masked genome sequences. We selected high-quality genes with
474 intact open reading frames (ORFs) and the highest GeneWise [62] score from a homology-

475 based gene set to train Augustus with default parameters before prediction. Gene models
476 with incomplete ORFs and small genes with protein-coding lengths less than 150 bp were
477 filtered out. Finally, a BLASTP (BLASTP, RRID:SCR_001010) search of predicted genes
478 was performed against the Swiss-Prot database [64]. Genes with matches to Swiss-Prot
479 proteins containing any one of the following keywords were filtered: transpose, transposon,
480 retrotransposon, retrovirus, retrotransposon, reverse transcriptase, transposase, and
481 retroviral. Finally, the results of the homology- and *de novo*-based gene sets were merged
482 using GLEAN (GLEAN, RRID:SCR_002890)[65] to yield a nonredundant reference gene
483 set.
484
485 ***Gene function annotation***
486 We annotated the protein-coding genes by searching against the following public
487 databases: Swiss-Prot [66], the Kyoto Encyclopedia of Genes and Genomes [67],
488 InterPro [68], and TrEMBL [66].
489
490 **Phylogenetic tree reconstruction and divergence time estimation**
491 The TreeFam tool (Tree families database, RRID:SCR 013401) [69] was used to identify
492 gene families as follows: first, all the protein sequences from a selection of 10
493 representative species (8 species including *Aplysia californica* (GCF_000002075.1),
494 *Octopus bimaculoides* (GCF_001194135.1), *Biomphalaria glabrata* (GCF_000457365.1),
495 *Crassostrea gigas* (GCF_000297895.1), *Lottia gigantea* (GCF_000327385.1), *Pomacea*
496 *canaliculate* (GCF_003073045.1), *Pinctada fucata* (GCA_002216045.1), *Helobdella*
497 *robusta* (GCF_000326865.1) from the NCBI database, *C. squamiferum* and *G. aegis* from
498 this research) were compared using blastp with the *E*-value threshold set as 1e-7. Then,
499 alignment segments of each protein pair were concatenated using the in-house software
500 Solar v0.9.6 [69]. H-scores were computed based on Bit-scores and were used to evaluate
501 the similarity among proteins. Finally, gene families were obtained by clustering
502 homologous gene sequences using Hcluster_sg v0.5.0 [69].
503
504 We obtained 406 one-to-one single-copy orthology gene families based on gene family
505 classification. Then, these gene families were extracted and aligned using guidance from

506 amino-acid alignments created using the default parameters of the MUSCLE (MUSCLE,
507 RRID:SCR_011812) [70] programme. All sequence alignments were then concatenated to
508 construct 1 super-matrix and then a phylogenetic tree was constructed under a
509 GTR+gamma model for nucleotide sequences using ML and Bayesian methods. The same
510 set of codon sequences were used for phylogenetic tree construction and estimation of
511 divergence time. The PAML mcmctree programme [71, 72] was used to determine
512 divergence times with the approximate likelihood calculation method, and the correlated
513 molecular clock and REV substitution model. The concatenated coding sequences of one-
514 to-one orthologous genes and the phylogenomics topology were used as inputs. We used
515 five calibration time points based on fossil records: *A. californica* - *C. gigas* (~516.3 - 558.3
516 MYA), *A. californica* - *P. canaliculata* (~310 – 496 MYA), *A. californica* - *Octopus*
517 *bimaculoides* (~551 – 628 MYA), *C. gigas* - *H. robusta* (~585 – 790 MYA), and *C. gigas*–
518 *P. fucata* (394 MYA) (http://www.timetree.org), were used as constraints in the
519 MCMCTree estimation.
520

521 **SNP calling and estimation of history population sizes**

522 About 50X clean WGS reads were mapped to genomes of *C. squamiferum* and *G. aegis*
523 using BWA-MEM (v0.7.12-r1039) (BWA, RRID:SCR_010910) [73] with default
524 parameters respecitvely. Then, SAMtools (v0.1.19-44428cd) (Samtools,
525 RRID:SCR_002105) [74] and "SortSam.jar" in the Picard package (v1.54) was used to
526 converte and sort BAM files. Local realignment was again carried out using
527 RealignerTargetCreator and IndelRealigner in GATK v3.6 (GATK, RRID:SCR_001876)
528 [75] with default parameters. SNPs were identified using HaplotypeCaller and filtered
529 using VariantFiltration with parameter "-filter-expression "QD < 2.0 || MQ < 40.0 ||
530 ReadPosRankSum < -8.0 || FS > 60.0" --filter-name LowQualFilter --genotype-filter-
531 expression "DP < 5.0" --genotype-filter-name lt_5". Estimatation of the historical effective
532 population sizes were carried out using PSMC v0.6.5-r67 [76]. Firstly, diploid genome
533 references were constructed using SAMtools and BCFtools call with "samtools mpileup -
534 C50" and "vcfutils.pl vcf2fq -d 20 -D 100". Secondly, the demographic history was
535 inferred using PSMC with parameters '-N25 -t15 -r5 -p 4+25*2+4+6' [77].

536

537 **Expansion and contraction of gene families**

538 We used CAFE (Computational Analysis of gene Family Evolution) v2.1 [23] to analyse
539 gene family expansion and contraction under the maximum likelihood framework. The
540 gene family results from the TreeFam pipeline and the estimated divergence time between
541 species were used as inputs. We used the parameters "-p 0.01, -r 10000, -s" to search for
542 the birth and death parameter (λ) of gene families, calculated the probability of each gene
543 family with observed sizes using 10,000 Monte Carlo random samplings, and reported birth
544 and death parameters in gene families with probabilities less than 0.01.

545

546

547 **Figure legends**

548 **Figure 1. Genome characteristics of *C. squamiferum* and *G. aegis*. a)** Photos of two
549 species. Left: *C. squamiferum*; right: *G. aegis.* Scale bar = 1cm. **b**) Heat map of chromatin
550 interaction relationships at a 125 kb resolution of 16 chromosomes. **c**) Genome sizes and
551 transposable elements in *C. squamiferum, G. aegis,* and two representative freshwater snail
552 genomes. **d**) Distribution of repeat sub-types of four species.

553

554 **Figure 2. Phylogenetic tree, estimated N_e, and evolution of single copy orthologous**
555 **genes of deep-sea snails**. **a)** Phylogenetic tree of ten representative molluscs. Expanded
556 and contracted gene families were identified using CAFE. Divergence time was estimated
557 using MCMCtree. Species names in red represent two deep-sea snails. Red dots represent
558 calibration time from TimeTree database. The timescale also refers to the TimeTree
559 database. **b)** Estimated demographic histories of two deep-sea snails. The generation time
560 set to "3" refers to the land snail [78]. The μ values are calculated in **Table S15**. **c)**. Box
561 plot of *K*a/*K*s values for five species.

562

563 **Figure 3. Expansion of nervous system-related genes a**) Phylogentic tree of *BTBD6*
564 genes in the examined species. The grey ellipses mark different clusters of genes. **b**)
565 Expansion pattern of *BTBD6* genes in two deep-sea snails. Grey lines represent scaffold
566 sequences. Coloured rectangles represent *BTBD6* genes. Symbols "//" represent other
567 genes along the scaffolds. The blue numbers: "1" represent only one gene between the
568 tandem duplicated genes. **c**) Expansion of *HTR4* genes. The species legend in the middle
569 was used for **a** and **c**. Gene trees of **a** and **c** were constructed using MUSCLE (v3.8.31)[70]
570 and FastTree (v2.1.10)[79].

571

572 **Figure 4**. **Expansion of immune, metabolism, DNA stability, and antioxidation genes**.
573 **a**) Gene numbers of four defence-related genes (*DMBT1*, *GAFT*, *Hsp90,* and *Txn1*), three
574 metabolism-related genes (*OGDHE1*, *OGDHE2,* and *IDH*), and the *SSB* gene. **b**) TCA
575 cycle signal pathway. The brown ellipses represent important enzymes and the expansion
576 of these genes (*OGDHE1*, *OGDHE2,* and *IDH*). **c**) Expansion of the catalase (*CAT*) gene
577 family in selected species.

578

579 **Table 1. Genome assembly and annotation of *Chrysomallon squamiferum* and**
580 ***Gigantopelta aegis*.**

Species	*Chrysomallon squamiferum*	*Gigantopelta aegis*
Genome size	455.36 Mb	1.29 GB
Scaffold N50	20.7 Mb	120.96 kb
Contig N50	541.32 kb	6.96 kb
Number of genes	28,781	25,601
Repeat content	30.56%	64.17%
GC content	34.48%	37.45%
Complete BUSCO	94.80%	92.40%

581

582 **Data and code availability**

583 The genome assemblies of these two genomes have been deposited in GenBank under the
584 accession number CNP0000854. The raw sequencing reads were also uploaded to the
585 SRA database under accession number CNP0000854. All supporting data are available in
586 the *GigaScience* GigaDB database [80].

587 **Additional Files**

588 Additional File 1: Supplementary Figures and Tables.docx

589

590 **Abbreviations**

591 ONT: Oxford Nanopore Technology; Hi-C: chromosome conformation capture; DNA:
592 Deoxyribonucleic acid; LTR: long terminal repeat; LINEs: Long interspersed nuclear
593 elements; MHC: major histocompatibility complex; IUCN: International Union for
594 Conservation of Nature; Gb: gigabase pairs; Mb: megabase pairs; kb: kilobase pairs; bp:
595 base pairs; BUSCO: Benchmarking Universal Single-Copy Orthologs; KEGG: Kyoto
596 Encyclopedia of Genes and Genomes; CR1: chicken repeat 1; ML: Maximum
597 Likelihood; MYA: million years ago; PSMC: pairwise sequential Markovian coalescent;

598 SNP: Single Nucleotide Polymorphism; K/Pg: Cretaceous/Paleogene; PAML:
599 Phylogenetic Analysis by Maximum Likelihood; CAFE: Computational Analysis of gene
600 Family Evolution; BTBD6: BTB/POZ domain-containing protein 6; HTR4: 5-
601 hydroxytryptamine receptor 4; IDH: isocitrate dehydrogenase; TCA: tricarboxylic acid;
602 OGDC: α-ketoglutarate dehydrogenase complex; DLD: dihydrolipoyl dehydrogenase;
603 OGDH: oxoglutarate dehydrogenase; DLST: dihydrolipoyl succinyltransferase; CO2:
604 carbon dioxide; TLR13: Toll-like receptor 13; Txn1: Thioredoxin 1; HIF-1α: hypoxia-
605 inducible factor-1α; GAFT: glutamine-fructose-6-phosphate transaminase; CAT:
606 catalase; DMBT1: deleted in malignant brain tumours 1; GP340; glycoprotein-340; SAG:
607 salivary agglutinin; SRCR: scavenger receptor cysteine-rich; HIV: human
608 immunodeficiency virus; TNF: Tumour necrosis factor; LCAM: last common ancestor of
609 all molluscs; TEs: The transposable elements; Hsp90: heat shock protein 90; PCR:
610 Polymerase chain reaction; DNBs: DNA nanoballs; WGS: whole genome sequence;
611 BWA: Burrows-Wheeler-Alignment; GATK: Genome Analysis Toolkit; NCBI: National
612 Center for Biotechnology Information; SRA: Sequence Read Archive.

613 **Author contributions**

614 Z.S., S.L., G.F., and X.L. conceived and managed this project and amended the
615 manuscript. X.Z., Y.Z., L.M., and I.S. performed the evolutionary analysis and wrote the
616 manuscript. L.M., J.C., and Y.S. performed genome assembly and annotation. J.B., S.L.,
617 X.F., C.W., Z.S., H.L., N.L., and L.W. were responsible for sample collection, DNA
618 extraction, and library construction.
619

620 **Acknowledgments**

621 We thank Dr. Yadong Zhou for providing the animal photographs. We are very grateful to
622 the RV "Xiangyanghong 9" crew and HOV "Jiaolong" operation teams for helping us

623 collect samples from the deep-sea hydrothermal vent field. This work was supported by
624 the National Key R&D Programme of China (No. 2018YFC0310702).
625

626 **Supplemental information**

627 Supplemental Information can be found online.

628

629 **Declaration of interests**

630 The authors declare that they have no competing interests.

631

632 **References**

633 1. Corliss JB, Dymond J, Gordon LI, Edmond JM, Von Herzen RP, Ballard RD, et al.
634 Submarine Thermal Springs on the Galápagos Rift. Science. 1979;203 4385:1073-
635 83.
636 2. VAN DOVER CL. The ecology of deep-sea hydrothermal vents. Princeton:
637 Princeton University Press; 2000.
638 3. Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y, et al. Adaptation to deep-sea
639 chemosynthetic environments as revealed by mussel genomes. Nature ecology &
640 evolution. 2017;1 5:0121.
641 4. Parkhaev PY. The Cambrian 'basement'of gastropod evolution. Geological Society,
642 London, Special Publications. 2007;286 1:415-21.
643 5. Sasaki T, Warén A, Kano Y, Okutani T and Fujikura K. Gastropods from Recent
644 Hot Vents and Cold Seeps: Systematics, Diversity and Life Strategies. 2010.
645 6. Chong C, Linse K, Copley TJ and Rogers DA. The 'scaly-foot gastropod': a new
646 genus and species of hydrothermal vent-endemic gastropod (Neomphalina:
647 Peltospiridae) from the Indian Ocean. 2015;81 3:322-34.
648 7. Chen C, Copley JT, Linse K and Rogers AD. Low connectivity between 'scaly-foot
649 gastropod' (Mollusca: Peltospiridae) populations at hydrothermal vents on the
650 Southwest Indian Ridge and the Central Indian Ridge. Organisms Diversity &
651 Evolution. 2015;15 4:663-70.
652 8. Sigwart JD, Chen C, Thomas EA, Allcock AL, Bohm M and Seddon M. Red Listing
653 can protect deep-sea biodiversity. Nature Ecology and Evolution. 2019;3 8:1134-.
654 9. Sun J, Chen C, Miyamoto N, Li R, Sigwart JD, Xu T, et al. The Scaly-foot Snail
655 genome and implications for the origins of biomineralised armour. Nature
656 communications. 2020;11 1:1-12.
657 10. Chen C, Uematsu K, Linse K and Sigwart JD. By more ways than one: Rapid
658 convergence at hydrothermal vents shown by 3D anatomical reconstruction of
659 Gigantopelta (Mollusca: Neomphalina). BMC Evolutionary Biology. 2017;17 1:62.
660 11. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH and Phillippy AM. Canu:
661 scalable and accurate long-read assembly via adaptive k-mer weighting and repeat

662 separation. Genome research. 2017;27 5:722-36.
663 12. Ruan J and Li H. Fast and accurate long-read assembly with wtdbg2. Nature
664 Methods. 2019:1-4.
665 13. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient
666 de novo assembly of highly heterozygous genomes from whole-genome shotgun
667 short reads. Genome Res. 2014;24 8:1384-95. doi:10.1101/gr.170720.113.
668 14. Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P, et al. Whole
669 genome analysis of a schistosomiasis-transmitting freshwater snail. Nature
670 communications. 2017;8:15451.
671 15. Liu C, Zhang Y, Ren Y, Wang H, Li S, Jiang F, et al. The genome of the golden
672 apple snail Pomacea canaliculata provides insight into stress tolerance and invasive
673 adaptation. GigaScience. 2018;7 9:giy101.
674 16. Biemont C. Genome size evolution: within-species variation in genome size.
675 Nature Publishing Group, 2008.
676 17. Dehal P and Boore JL. Two rounds of whole genome duplication in the ancestral
677 vertebrate. PLoS biology. 2005;3 10.
678 18. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al.
679 Ten things you should know about transposable elements. Genome biology.
680 2018;19 1:1-12.
681 19. Barth NK, Li L and Taher L. Independent transposon exaptation is a widespread
682 mechanism of redundant enhancer evolution in the mammalian genome. Genome
683 biology and evolution. 2020;12 3:1-17.
684 20. Li H and Durbin R. Inference of human population history from individual whole-
685 genome sequences. Nature. 2011;475 7357:493-6.
686 21. Van Dover CL, Arnaud-Haond S, Gianni M, Helmreich S, Huber JA, Jaeckel A, et
687 al. Scientific rationale and international obligations for protection of active
688 hydrothermal vent ecosystems from deep-sea mining. Marine Policy. 2018;90:20-
689 8.
690 22. Yang Z. PAML: a program package for phylogenetic analysis by maximum
691 likelihood. CABIOS 13, 555–556. Supplementary material The following
692 supplementary material is available for this article online: Fig S. 1997;1.
693 23. De Bie T, Cristianini N, Demuth JP and Hahn MW. CAFE: a computational tool for
694 the study of gene family evolution. Bioinformatics. 2006;22 10:1269-71.
695 24. Sobieszczuk DF, Poliakov A, Xu Q and Wilkinson DG. A feedback loop mediated
696 by degradation of an inhibitor is required to initiate neuronal differentiation. Genes
697 & development. 2010;24 2:206-18.
698 25. Conductier G, Dusticier N, Lucas G, Côté F, Debonnel G, Daszuta A, et al. Adaptive
699 changes in serotonin neurons of the raphe nuclei in 5-HT4 receptor knock-out
700 mouse. European Journal of Neuroscience. 2006;24 4:1053-62.
701 26. Goffredi SK, Warén A, Orphan VJ, Van Dover CL and Vrijenhoek RC. Novel forms
702 of structural integration between microbes and a hydrothermal vent gastropod from
703 the Indian Ocean. Appl Environ Microbiol. 2004;70 5:3082-90.
704 27. Wolff T. Composition and endemism of the deep-sea hydrothermal vent fauna.
705 CBM-Cahiers de Biologie Marine. 2005;46 2:97-104.
706 28. Farina AR, Tacconelli A, Cappabianca L, Masciulli MP, Holmgren A, Beckett GJ,
707 et al. Thioredoxin alters the matrix metalloproteinase/tissue inhibitors of

708 metalloproteinase balance and stimulates human SK-N-SH neuroblastoma cell
709 invasion. European Journal of Biochemistry. 2001;268 2:405-13.
710 29. Mukherjee A and Martin S. The thioredoxin system: a key target in tumour and
711 endothelial cells. The British journal of radiology. 2008;81 special_issue_1:S57-
712 S68.
713 30. Muri J, Heer S, Matsushita M, Pohlmeier L, Tortola L, Fuhrer T, et al. The
714 thioredoxin-1 system is essential for fueling DNA synthesis during T-cell metabolic
715 reprogramming and proliferation. Nature communications. 2018;9 1:1851.
716 31. Kato N, Dasgupta R, Smartt C and Christensen B. Glucosamine: fructose-6-
717 phosphate aminotransferase: gene characterization, chitin biosynthesis and
718 peritrophic matrix formation in Aedes aegypti. Insect molecular biology. 2002;11
719 3:207-16.
720 32. Lagorce A, Le Berre-Anton V, Aguilar-Uscanga B, Martin-Yken H,
721 Dagkessamanskaia A and François J. Involvement of GFA1, which encodes
722 glutamine–fructose-6-phosphate amidotransferase, in the activation of the chitin
723 synthesis pathway in response to cell-wall defects in Saccharomyces cerevisiae.
724 European journal of biochemistry. 2002;269 6:1697-707.
725 33. Csermely P, Schnaider T, So C, Prohászka Z and Nardai G. The 90-kDa molecular
726 chaperone family: structure, function, and clinical applications. A comprehensive
727 review. Pharmacology & therapeutics. 1998;79 2:129-68.
728 34. Marceau AH. Functions of single-strand DNA-binding proteins in DNA replication,
729 recombination, and repair. Single-Stranded DNA Binding Proteins. Springer; 2012.
730 p. 1-21.
731 35. Nazıroğlu M. Molecular role of catalase on oxidative stress-induced Ca2+ signaling
732 and TRP cation channel activation in nervous system. Journal of Receptors and
733 Signal Transduction. 2012;32 3:134-41.
734 36. Hohenester E, Sasaki T and Timpl R. Crystal structure of a scavenger receptor
735 cysteine-rich domain sheds light on an ancient superfamily. Nature Structural &
736 Molecular Biology. 1999;6 3:228.
737 37. Ligtenberg AJ, Karlsson NG and Veerman EC. Deleted in malignant brain tumors-
738 1 protein (DMBT1): a pattern recognition receptor with multiple binding sites.
739 International journal of molecular sciences. 2010;11 12:5212-33.
740 38. Aguilera F, McDougall C and Degnan BM. Co-option and de novo gene evolution
741 underlie molluscan shell diversity. Molecular biology and evolution. 2017;34
742 4:779-92.
743 39. Mann K, Edsinger-Gonzales E and Mann M. In-depth proteomic analysis of a
744 mollusc shell: acid-soluble and acid-insoluble matrix of the limpet Lottia gigantea.
745 Proteome science. 2012;10 1:28.
746 40. Benoist CO, Mathis DJ, Kanter MR, Williams II VE and McDevitt HO. Regions of
747 allelic hypervariability in the murine Aα immune response gene. Cell. 1983;34
748 1:169-77.
749 41. Lindberg DR PW, Haszprunar G. The Mollusca: Relationships and Patterns
750 fromTheir First Half-Billion Years. Oxford: Oxford University Press; 2004.
751 42. Wanninger A and Wollesen T. The evolution of molluscs. Biological Reviews.

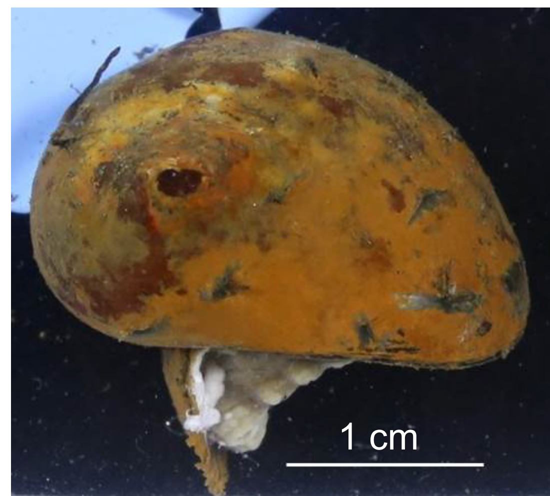
752 2019;94 1:102-15.
753 43. Vinther J. A molecular palaeobiological perspective on aculiferan evolution.
754 Journal of natural history. 2014;48 45-48:2805-23.
755 44. Vinther J. The origins of molluscs. Palaeontology. 2015;58 1:19-34.
756 45. Lee H, Chen W, Puillandre N, Aznarcormano L, Tsai M and Samadi S.
757 Incorporation of deep-sea and small-sized species provides new insights into
758 gastropods phylogeny. Molecular Phylogenetics and Evolution. 2019;135:136-47.
759 46. Kazazian HH. Mobile elements: drivers of genome evolution. science. 2004;303
760 5664:1626-32.
761 47. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, et al. Widespread contribution
762 of transposable elements to the innovation of gene regulatory networks. Genome
763 research. 2014;24 12:1963-76.
764 48. Jacques P-E, Jeyakani J and Bourque G. The majority of primate-specific
765 regulatory sequences are derived from transposable elements. PLoS genetics.
766 2013;9 5.
767 49. Janoušek V, Laukaitis CM, Yanchukov A and Karn RC. The role of retrotransposons
768 in gene family expansions in the human and mouse genomes. Genome biology and
769 evolution. 2016;8 9:2632-50.
770 50. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al.
771 Human genome sequencing using unchained base reads on self-assembling DNA
772 nanoarrays. Science. 2010;327 5961:78-81. doi:10.1126/science.1181498.
773 51. Huang J, Liang X, Xuan Y, Geng C, Li Y, Lu H, Qu S, Mei X, Chen H, Yu T,
774 Sun N, Rao J, Wang J, Zhang W, Chen Y, Liao S, Jiang H, Liu X, Yang Z, Mu F,
775 Gao S. 2018. BGISEQ-500 Sequencing. protocols.io
776 https://dx.doi.org/10.17504/protocols.io.pq7dmzn
777 52. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et
778 al. A 3D map of the human genome at kilobase resolution reveals principles of
779 chromatin looping. Cell. 2014;159 7:1665-80. doi:10.1016/j.cell.2014.11.021.
780 53. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, et al. HiC-Pro: an
781 optimized and flexible pipeline for Hi-C data processing. Genome Biol.
782 2015;16:259. doi:10.1186/s13059-015-0831-x.
783 54. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De
784 novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length
785 scaffolds. Science. 2017;356 6333:92-5. doi:10.1126/science.aal3327.
786 55. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer
787 Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments.
788 Cell Syst. 2016;3 1:95-8. doi:10.1016/j.cels.2016.07.002.
789 56. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al.
790 Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited
791 Zoom. Cell Syst. 2016;3 1:99-101. doi:10.1016/j.cels.2015.07.012.
792 57. Tarailo-Graovac M and Chen N. Using RepeatMasker to identify repetitive
793 elements in genomic sequences. Current Protocols in Bioinformatics. 2009:4.10. 1-
794 4.. 4.
795 58. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O and Walichiewicz J.
796 Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and
797 genome research. 2005;110 1-4:462-7.

798 59. Xu Z and Wang H. LTR_FINDER: an efficient tool for the prediction of full-length
799 LTR retrotransposons. Nucleic acids research. 2007;35 suppl 2:W265-W8.
800 60. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic
801 acids research. 1999;27 2:573.
802 61. She R, Chu JS, Wang K, Pei J and Chen N. GenBlastA: enabling BLAST to identify
803 homologous gene sequences. Genome Res. 2009;19 1:143-9.
804 doi:10.1101/gr.082081.108.
805 62. Birney E, Clamp M and Durbin R. GeneWise and Genomewise. Genome Res.
806 2004;14 5:988-95. doi:10.1101/gr.1865504.
807 63. Keller O, Kollmar M, Stanke M and Waack S. A novel hybrid gene prediction
808 method employing protein multiple sequence alignments. Bioinformatics. 2011;27
809 6:757-63. doi:10.1093/bioinformatics/btr010.
810 64. UniProt C. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res.
811 2019;47 D1:D506-D15. doi:10.1093/nar/gky1049.
812 65. Elsik CG, Mackey AJ, Reese JT, Milshina NV, Roos DS and Weinstock GM.
813 Creating a honey bee consensus gene set. Genome biology. 2007;8 1:1.
814 66. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, et
815 al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.
816 Nucleic acids research. 2003;31 1:365-70.
817 67. Kanehisa M and Goto S. KEGG: kyoto encyclopedia of genes and genomes.
818 Nucleic acids research. 2000;28 1:27-30.
819 68. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, et al. The
820 InterPro database, an integrated documentation resource for protein families,
821 domains and functional sites. Nucleic acids research. 2001;29 1:37-40.
822 69. Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L, et al. TreeFam: a
823 curated database of phylogenetic trees of animal gene families. Nucleic Acids Res.
824 2006;34 Database issue:D572-80. doi:10.1093/nar/gkj118.
825 70. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
826 throughput. Nucleic Acids Res. 2004;32 5:1792-7. doi:10.1093/nar/gkh340.
827 71. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol.
828 2007;24 8:1586-91. doi:10.1093/molbev/msm088.
829 72. Yang Z and Rannala B. Bayesian estimation of species divergence times under a
830 molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol.
831 2006;23 1:212-26. doi:10.1093/molbev/msj024.
832 73. Li H and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler
833 transform. Bioinformatics. 2010;26 5:589-95. doi:10.1093/bioinformatics/btp698.
834 74. Guindon S, Delsuc F, Dufayard J-F and Gascuel O. Estimating maximum
835 likelihood phylogenies with PhyML. Bioinformatics for DNA sequence analysis.
836 2009:113-37.
837 75. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-
838 Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome
839 Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11 0
840 1-33. doi:10.1002/0471250953.bi1110s43.
841 76. Li H and Durbin R. Inference of human population history from individual whole-
842 genome sequences. Nature. 2011;475 7357:493-6. doi:10.1038/nature10231.
843 77. Warren WC, Kuderna L, Alexander A, Catchen J, Perez-Silva JG, Lopez-Otin C, et

844 al. The Novel Evolution of the Sperm Whale Genome. Genome Biol Evol. 2017;9
845 12:3260-4. doi:10.1093/gbe/evx187.
846 78. Schilthuizen M. Rapid, habitat-related evolution of land snail colour morphs on
847 reclaimed land. Heredity. 2013;110 3:247-52.
848 79. Price MN, Dehal PS and Arkin AP. FastTree 2–approximately maximum-likelihood
849 trees for large alignments. PloS one. 2010;5 3:e9490.
850 80. Zeng X; Zhang Y; Meng L; Fan G; Bai J; Chen J; Song Y; Seim I; Wang C; Shao
851 Z; Liu N; Lu H; Fu X; Wang L; Liu X; Liu S; Shao Z. Supporting data for "Genome
852 sequencing of deep-sea hydrothermal vent snails reveals adaptions to extreme
853 environments" GigaScience Database 2020. http://dx.doi.org/10.5524/100817.
854

Click here to access/download;Figure;Figure 1.pdf

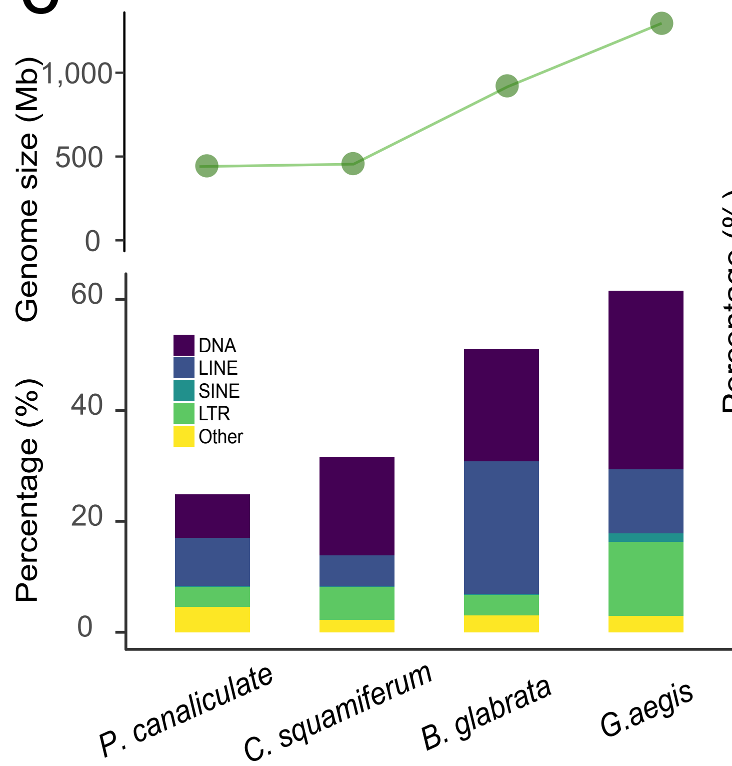
a



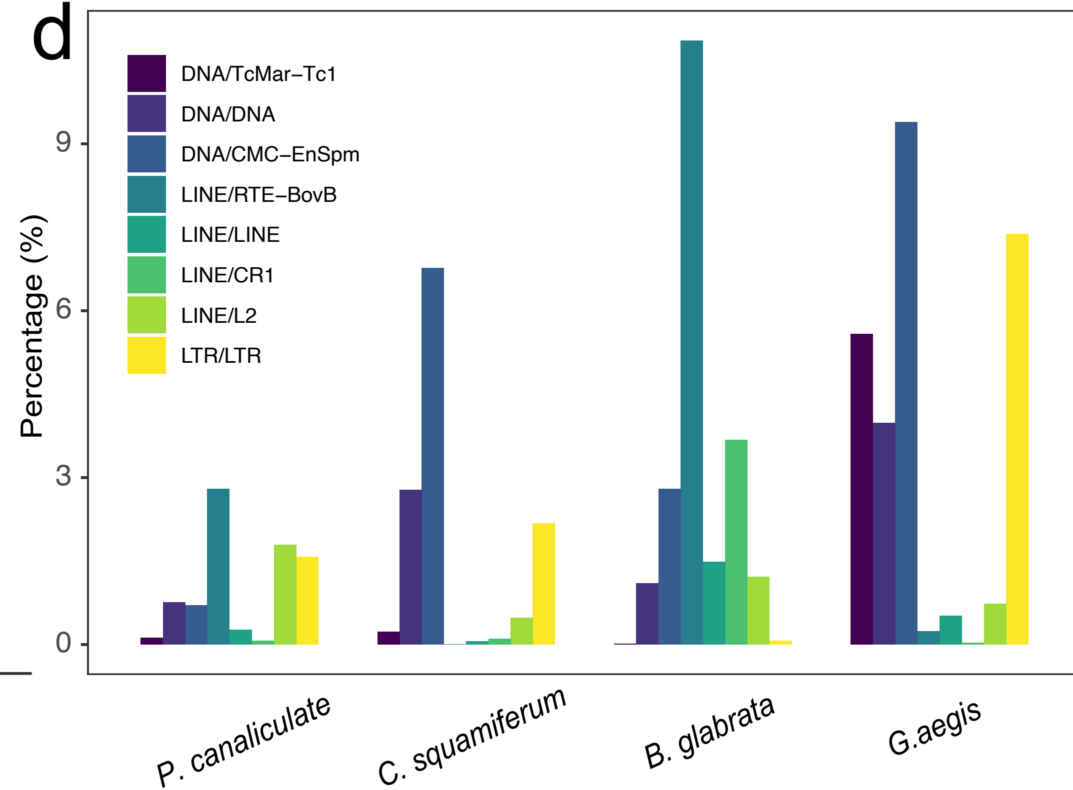
C. squamiferum *G.aegis*

b

c



d



Click here to access/download;Figure;Figure 2.pdf

Fig 2

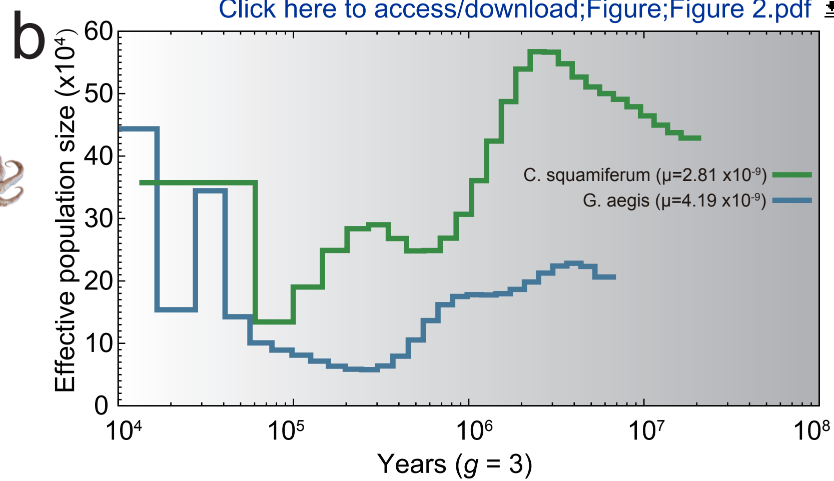
a

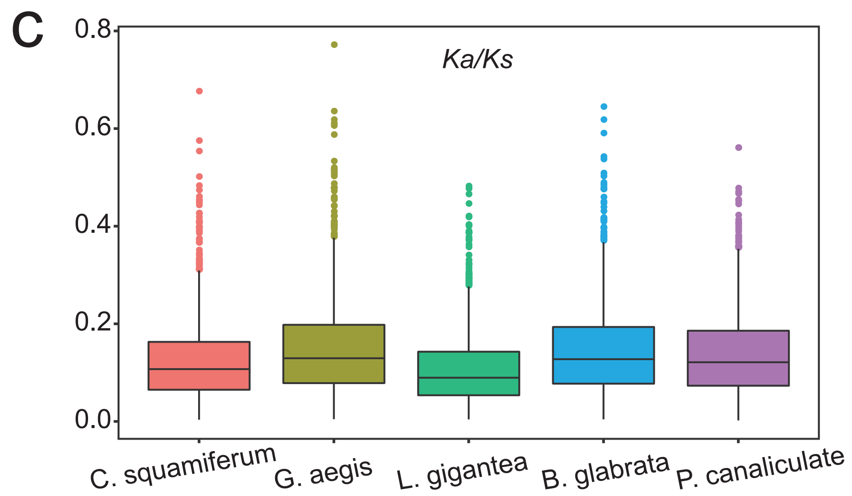
H. robusta
+1677/-19359
575.0(560.6-584.0)
O. bimaculoides
+873/-18124
569.9(557.5-576.9)
C. gigas
+2713/-3182
252.3(184.3-306.7)
P. fucata
+3777/-3480
555.2(543.7-561.2)
L. gigantea
+2160/-14397
551.0(539.4-557.3)
G. aegis
+2301/-6786
66.3(42.4-100.0)
C. squamiferum
+3088/-3857
536.6(525.1-543.3)
P. canaliculate
+1585/-4825
374.3(305.5-468.6)
B. glabrata
+2383/-1739
250.5(201.3-329.1)
A. californica
+948/-3786

Divergence time
Expanded gene families
Contracted gene families

Neo-Proterozoic
Paleozoic
Mesozoic
Cenozoic
Proterozoic
Phanerozoic

688 600 400 200 0 MYA





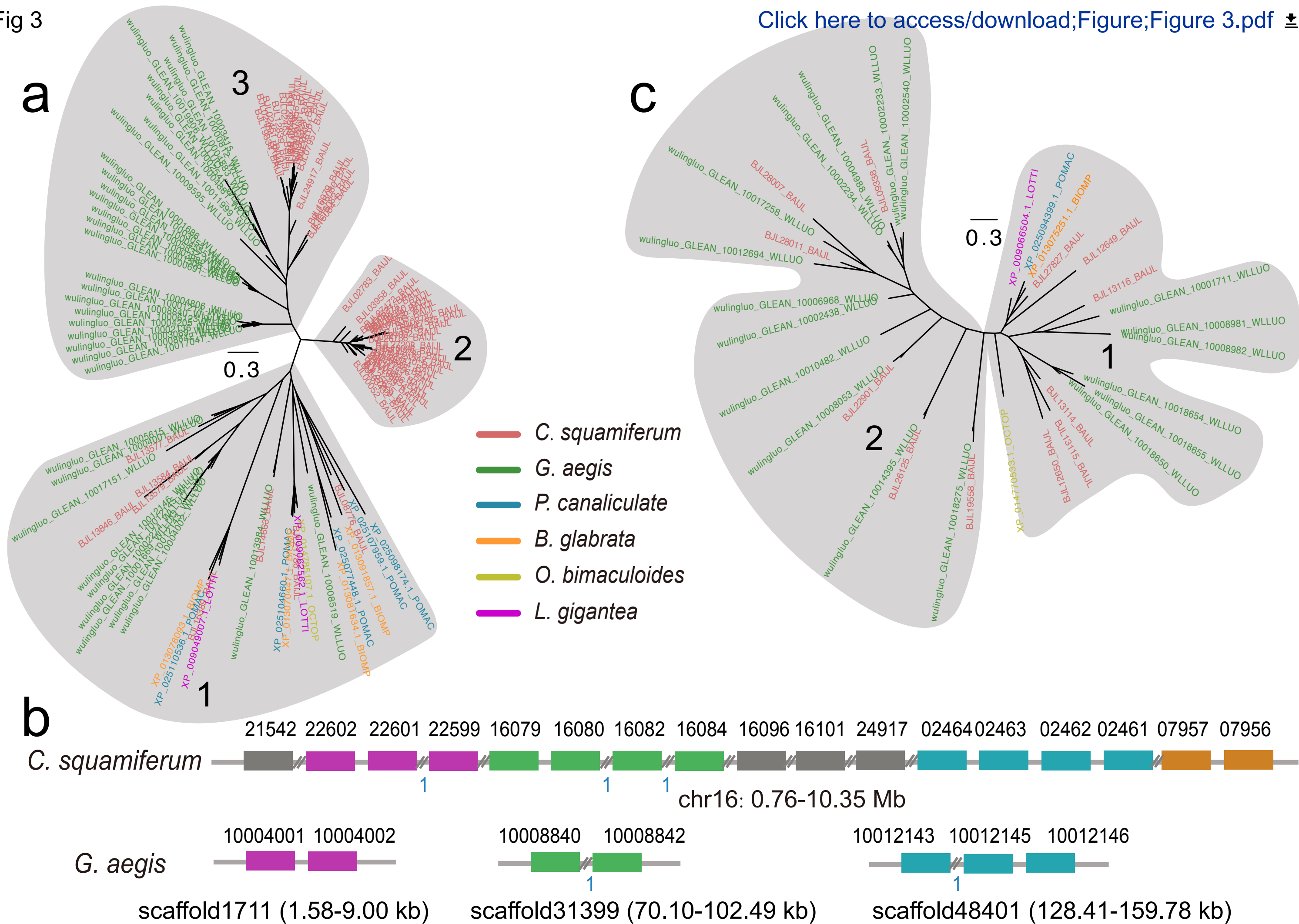
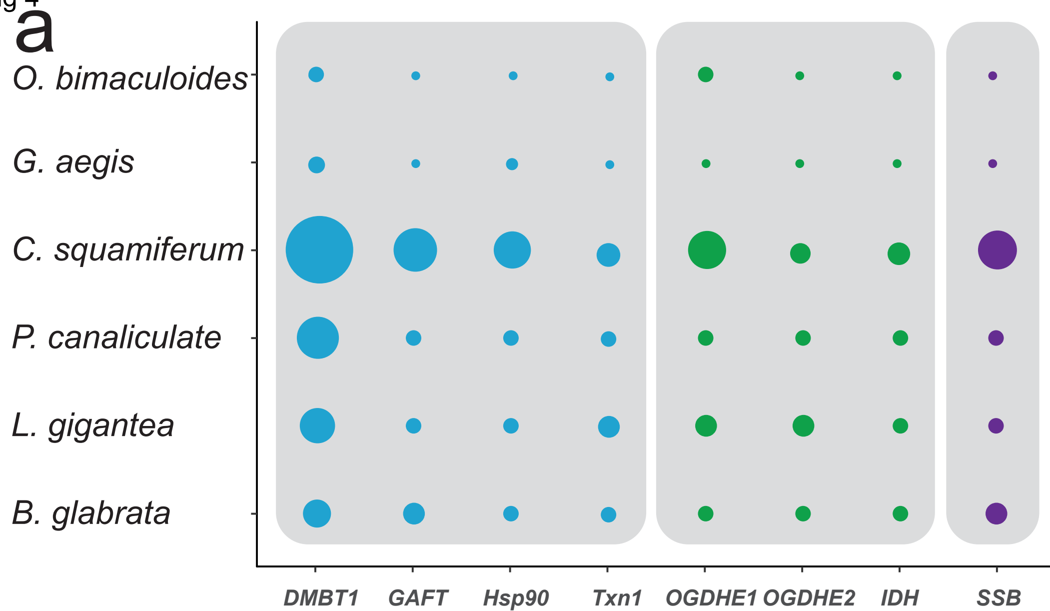
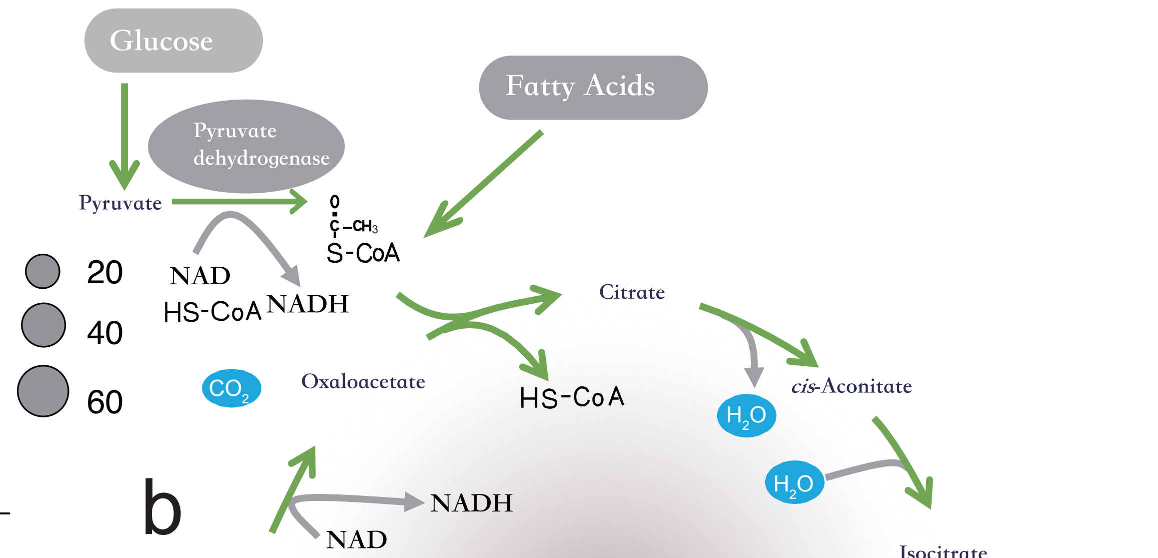


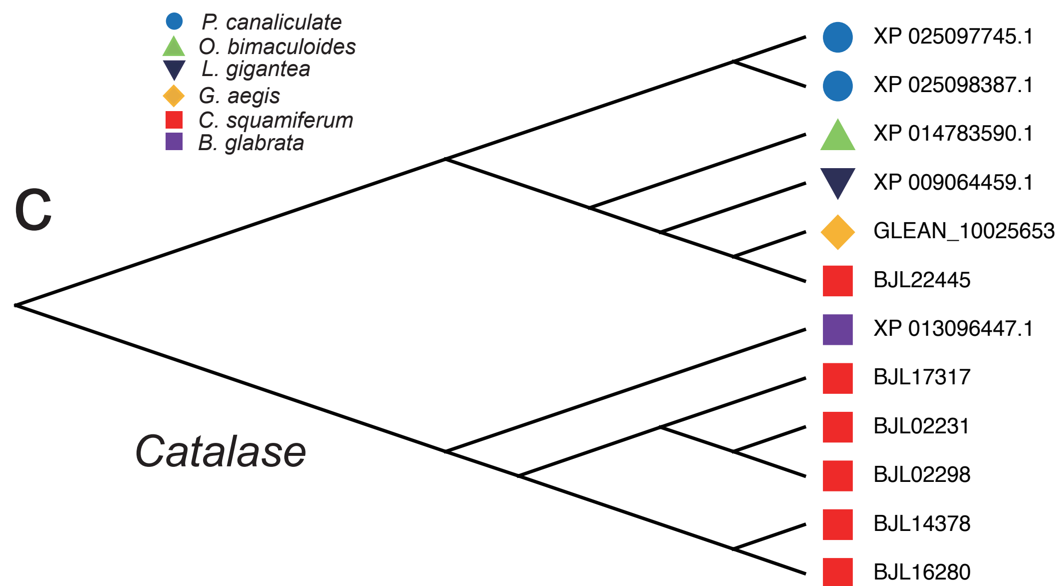
Fig 3
Click here to access/download;Figure;Figure 3.pdf
a
3
2
1
0.3
c
0.3
1
2
C. squamiferum
G. aegis
P. canaliculate
B. glabrata
O. bimaculoides
L. gigantea
b
21542 22602 22601 22599 16079 16080 16082 16084 16096 16101 24917 02464 02463 02462 02461 07957 07956
C. squamiferum
1
1
1
chr16: 0.76-10.35 Mb
10004001 10004002
10008840 10008842
10012143 10012145 10012146
G. aegis
1
1
scaffold1711 (1.58-9.00 kb)
scaffold31399 (70.10-102.49 kb)
scaffold48401 (128.41-159.78 kb)

Fig 4

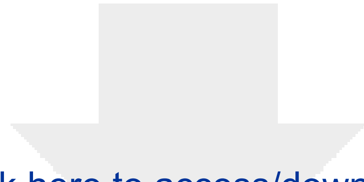
Click here to access/download;Figure;Figure 4.pdf



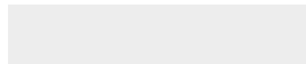


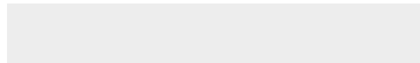


Suppl figs and table



Click here to access/download
Supplementary Material
Supplementary Figures and Tables-revision_2.docx





Point to point response letter
Click here to access/download;Personal Cover;PointToPointResponseLetter.docx

Response to editors and reviewers:

Dear editor and reviewers:
On behalf of all the coauthors, we would like to thank you very much and all the reviewers for the time spent to assess our manuscript (GIGA-D-20-00187) and for your relevant remarks and suggestions that allow to improve the quality of this manuscript.
We have checked the manuscript carefully and revised it according to the comments.

Sincerely,
Xiang Zeng, Yaolei Zhang

Reviewer reports:
Reviewer #1: The Data Note by Zeng et al. reported two genome assemblies of deep sea gastropods, Chrysomallon squamiferum and Gigantopelta aegis.
I would ask the authors for additional informations about assembling process and references of data sources in order to guarantee the quality of the data and analyses. I also found there are many ambiguous expressions in the present manuscript, making it unclear how the genome resources can contribute to understand biology of these animals.
Please find my specific comments and concerns below, which need to be addressed.
Response: Thank you very much for your thoughtful and helpful suggestions. We have revised our manuscript according to your comments. Please see that.

Background
The authors should mention the fact that the Chrysomallon squamiferum genome has been published by Sun et al. (Nat Commun 11, 1657, 2020) somewhere in the Background section.
I would suggest the authors to explain that they analyzed the genome of "white scaly foot individual" while Sun et al. sequenced "black" one, to emphasize the uniqueness of this study.
Response: Thank you very much for this helpful suggestion. We have added descriptions as you suggested. Please see **lines 91-93** "And the whole genome of black scaly-foot snail was reported recently, which highlighted its evolutionary mechanisms of biomineralised armour [9]" and 109-110: "In this study, we sequenced and assembled genomes of the white scaly-foot snail (Figure 1a), which is different from the published black individual".

Line 93
Remove "sp. nov.". This abbreviation is used when new species is named.
Response: Thank you. We removed this abbreviation. See **lines 94-96**: "The genus includes two species, *Gigantopelta chessoia* from East Scotia Ridge and *Gigantopelta aegis* from the Southwest Indian Ridge [6]. Both *Chrysomallon* and *Gigantopelta* are members of the family Peltospiridae."

Data description

Line 110

Was the insert size 350bp (main text) or 300bp (Table S2)?

Response: Thank you very much. We modified it as “350 bp” in Table S2.

Line 121-122

As mentioned above, the *C. squamiferum* genome has been published. Therefore this sentence needs to be removed.

Response: Thank you. We delete this sentence.

Line 156-160

These sentences do not make sense to me. Why despite "precise functions of these repeats have not been studied," the authors can infer the composition of repeat elements "may be closely associated with adaptation to extreme environment"? Please describe more specifically by mentioning some references that support this idea.

Response: Thank you and sorry for confusing you. We added more description and relevant references about the importance of transposons/repeats. Please see **lines: 156-161** “Although most of the precise functions of these repeats have not been studied in depth, repeats have been thought to have a regulatory function in related genes that play an important role in the life cycle and can introduce great genome flexibility [18]. And in the mammalian genome, transposons were described to be redundant enhancers that regulate their target genes which are higher or tissue specially expressed, indicating the importance of transposons”

Lines 173-175

It seems the authors assumed the split of C. squamiferum and G. aegis was related to the mass extinction event around 66 MYA. However, estimated divergence time is considerably ambiguous (42.4-100 MYA, Fig 1a), making the idea less reliable.

Response: Thank you. We agree with you and delete this description.

Deleted sentences: “This time is consistent with the most recent ‘mass extinction’, at the end of the Cretaceous geological period ~66 MYA, where ~76% of species became extinct” for your reference.

Lines 178-180

This sentence is difficult to understand. Speciation and demographic histories of each species are different topics.

Response: Thank you. We also delete this sentence.

Deleted sentences: “As the speciation of the two deep-sea snails may be related to geological events (see above)” for your reference.

Lines 190-191
This sentence is not clear. Please describe what "major geological events" affected the population sizes, with references describing the geological events.

An: Thank you. We added one example of geological events and relevant reference. Please see **lines 194-197**: “For example, the Cretaceous/Paleogene (K/Pg) extinction event caused extinction of three-quarters of species on earth and affected population dynamics approximately 66 million years ago, when an asteroid impact caused global environmental devastation [21, 22]”

Lines 191-193
The recent decreased population size was reported by ref[8] and was not related to this study (Fig2b). Then, this sentence may be put in the Background section.
An: Thank you. We deleted this sentence.

Deleted sentences: ” Unfortunately, the C. squamiferum population size has dramatically decreased recently due to deep-sea mining [8], which has made this species endangered” just for your reference

Lines 196-210
Comparisons of Ks and Ka values among snail species should be tested statistically. In the figures 2C and S3 it is not clear whether these values are significantly different.
Response: Thank you. We have added statistical test (Mann-Whitney U test) for Ka/Ks comparisons. Please see **lines 204-213**: “We found that the Ka values of the two deep-sea snails (average: 0.37 and 0.41) were higher (Mann-Whitney U test, p-value<0.001) than that of the shallow-water limpet (0.35) but similar to those of two freshwater snails (0.39 and 0.41), which suggests that the genes of deep-sea and freshwater snails both evolved faster after their divergence from shallow-water limpet . The Ks values of the deep-sea (3.34 and 3.09) and freshwater (3.19 and 3.24) snails were also similar and lower (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (3.72). Additionally, the Ka/Ks values of the deep-sea snails (average: 0.13 and 0.15) were approximately ~20% and ~40% higher (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (0.11)”.

Discussion
Lines 309-310
I have no idea what the "infamous Cambrian Explosion" means. Please explain the authors' idea more in detail.
Response: Thank you. It was modified to “Cambrian Explosion”. It was a clerical error. See **line 319**.

Lines 328-329
Please describe what are "adaptive needs" and "region-specific features" specifically.

Response: Thank you. We deleted this sentence.

Materials and Methods
Line 370
350bp or 300bp?
Response: It is “350 bp” actually for BGISEQ.

Lines 403-405
Based on the description, the 10X Chromium reads were used only for polishing, not for scaffolding. On the other hand, there are stats of scaffolds before Hi-C scaffolding in Table S3.
Response: Yes, the 10X Chromium reads were used only for polishing, not for scaffolding.

My questions are;
i) Were the scaffolds in Table S3 generated using 10X Chromium reads?
ii) If so, the scaffolds were improved very little (sequences are reduced from 6449 to 6444), indicating there was problem in 10X Chromium sequencing. How the authors interpreted the results?
Response: Thank you. The 6,444 scaffolds (also 6,444 contigs) were generated using Oxford Nanopore reads. We then used 10X Chromium reads to polish (error correction) 6,444 scaffolds with software Pilon. Pilon introduced 5 bp gaps so the contig number became 6,449 while the scaffold number was still 6,444.

Line 436
"Lottia"
Response: Thank you. We corrected this word.

Line 454
Describe a reference for GLEAN.
Response: Thank you. We added a reference for GLEAN. Please see **line 465**.

Lines 464-467
Describe sources or references for these genomic data.
Response: Thank you. We added sources for these genomic data. Please see **lines 474-479**: “ all the protein sequences from selected 10 representative species (8 species including Aplysia californica (GCF_000002075.1), Octopus bimaculoides (GCF_001194135.1), Biomphalaria glabrata (GCF_000457365.1), Crassostrea gigas (GCF_000297895.1), Lottia gigantea (GCF_000327385.1), Pomacea canaliculate (GCF_003073045.1), Pinctada fucata (GCA_002216045.1), Helobdella robusta (GCF_000326865.1) from NCBI database, C. squamiferum and G. aegis from this research) were compared using blastp with the E-value threshold set as 1e-7.”

Lines 484-487

Describe references of these fossil records rather than summary database (Timetree.org) so that readers can refer the original data.

Response: Thank you very much and we can’t agree with you more. However, each of the time point between two species refers a lot of references and TimeTree database summarized all of these references to estimate one divergence time with a confidence interval. For example, the divergence time between *Aplysia californica* and *Crassostrea gigas* was estimated to be 537 MYA with a confidence interval of 516.3-558.3 MYA based on 11 references. We used this confidence interval time to calibrate our estimation. This is a common method used frequently in nowadays genome research. So here we cite TimeTree database (Timetree.org) for reference which includes many references. However, if you think list all the references is a must, we are pleased to do this.

Methods of SNP identification and PSMC (lines 177-193) were not described.

Response: Thank you. We have added methods of SNP identification and PSMC. Please see **lines 503-516**: “SNP calling and estimation of history population sizes

About 50X clean WGS reads were mapped to genomes of C. squamiferum and G. aegis using BWA mem (v0.7.12-r1039) [73] with default parameters respecitvely. Then SAMtools (v0.1.19-44428cd) [74] and “SortSam.jar” in the picard package (v1.54) was used to convert and sort BAM files. Local realignment was again carried out using RealignerTargetCreator and IndelRealigner in GATK (v3.6) [75] with default parameters. SNPs were identified using HaplotypeCaller and filtered using VariantFiltration with parameter “-filter-expression "QD < 2.0 || MQ < 40.0 || ReadPosRankSum < -8.0 || FS > 60.0" --filter-name LowQualFilter --genotype-filter-expression "DP < 5.0" --genotype-filter-name lt_5”. Estimatation of history population sizes were carried out using PSMC (v0.6.5-r67) [76]. Firstly, diploid genome references were constructed using samtools and bcftools call with “samtools mpileup -C50” and “vcfutils.pl vcf2fq -d 20 -D 100”. Secondly, the demographic history was inferred using PSMC with parameters ‘-N25 -t15 -r5 -p 4+25*2+4+6’ [77].”

Fig 1a

Add the size of the scale for C. squamiferum. No scale is indicated for G.aegis.

Response: Thank you. Scale was added in **Fig 1a**. Also, we added “Scale bar = 1cm” in **Fig 1** legend.

Fig 2a

This cartoon is too ambiguous and not suitable for scientific paper. The molecular phylogeny should be clearly shown by solid lines.

Response: Thank you. We have updated **Fig. 2a**. Please see that.

Fig.3d

This figure is not very informative for readers. The authors may want to draw molecular phylogeny trees for BTBD6 and HTR4.

Response: Thank you. We also deleted **Fig.3d**.

Reviewer #2: The manuscript entitled "Genome sequencing of deep-sea hydrothermal vent snails reveals adaptations to extreme environments" presents a nice description of a good genome assembly (16 chromosomes representing ~80% of the genome) of the scaly foot snail (Chrysomallon squamiferum) and compare it to genomes of other molluscan species. Overall the paper is well written and presents a nice view of some unique adaptations by this deep-sea mollusc. One concern that I had is throughout the manuscript (starting at line 164 and onward) the authors describe comparing two mussels, two freshwater snails and two shallow-water snails to their genomes. However, these other molluscan species include C. gigas and P. fucata…which are both oysters and not mussels, and while two of the other molluscs included in the tree are in Gastropoda and considered snails, Lottia gigantea is a limpet and Aplysia californica is a sea slug. I would encourage the authors to describe all of these species more accurately, i.e., as limpet and sea slug, because these are very different from what people commonly think of when they hear "snail", represented by the more traditional Pomacea and Biomphalaria. Referring to all the "snails" as gastropods would be a more suitable term that captures the true diversity of this large group. But when discussing individual species, I would prefer to see the more accurate descriptions because limpets and sea slugs are different from traditional snails, and will have unique adaptations of their own related to their unique characteristics. Overall, the authors give a good general description of the results and present a reasonable discussion about some of the potential adaptations that they observed in the genome. One minor point - thioredoxins are much more likely play a role in repairing proteins that have been altered by oxidation (Lines 255-256), so to limit this expansion to innate immunity leaves out a lot of other possibilities. My other question was regarding the source of the genomic DNA. The authors describe using muscle samples for isolating DNA, but it is not clear if DNA from one individual was used for all sequencing or if pooling occurred?

Response: Thank you very much for your approval and your thoughtful advices. We have updated our descriptions based on your suggestions. Please see below response.

1) About the scientific name (line 164 and onward), we modified this part as "we compared them with two shallow-water bivalves (P.fucata and C.gigas) and four shallow-water gastropods, including two fresh-water snails (B. glabrata and P. canaliculate), one limpet (B. glabrata) and one sea slug (A. californica). The California two-spot octopus (O. bimaculoides) and the freshwater leech (H.robusta) were used as the outgroup (**Figure 2a**) (**lines 167-171**)."

2) About thioredoxins, we added the description of thioredoxin as redox proteins and references. Please **lines 260-265**: "For example, increased expression of thioredoxin 1 (Txn1; 22 copies in C. squamiferum) was identified. Thioredoxin 1 (Txn1), a redox protein, is important in regulation of cellular redox homeostasis and anti-apoptotic functions. Txn1 stimulates cell proliferation and cell cycle progression, induces

hypoxia-inducible factor-1α (HIF-1α) and angiogenesis, and alters the balance between the matrix metalloproteinases and their tissue inhibitors [29, 30]"
3) About isolating DNA, all DNA was isolated from one individual and we add clear description about this in **lines 374-375**: "DNA was extracted from muscle sample of one individual using the cetyl trimethylammonium bromide (CTAB) method and a DNeasy blood & tissue kit (QIAGEN)."

Reviewer #3: The manuscript of Zeng et al seems to describe a well-put together genome for one species of deep-sea snail, with an additional 'draft' genome for another species. It is clear and well-written, with most of the methods described sufficiently. My main criticism is that I found some of the discussions regarding the adaptative significance and/or putative "function" of various TE content and gene-family expansion results quite speculative, given that the comparative results are often observational with no hypothesis testing or statistical framework. That may well be beyond the remit of the paper, but the language could be more careful in places to reflect the putative nature of any hypothesised effects. Nonetheless I have no doubt that the genomes themselves will be useful additions to the community for future work on mollusc and animal evolution.
Response: Thank you very much for your approval of our manuscript and your helpful criticism. Yes, you are quite right. Here we did not show any experiments results to valid our hypothesis or speculation because this is a data description paper, mainly focusing on observational data. More investigations including both in-depth analysis and experiments based on these two genomes and these findings will be carried out in the future to verify function of important genes or TEs or conserved no-coding regions. These are important and interesting issues and must be done then. Thank you again.

Minor comments:
- Typo line 38: "impedes"
Response: Thank you. This word was modified

- Line 158: is there a reference or two for this? I would assume that most TEs are simply selfish genetic elements that do not serve a "function" per se but exist only for their own purpose, i.e. to copy themselves independently of the host genome
- Line 158-160: but most TE content differences are probably driven by stochastic forces (i.e. drift) rather than deterministic forces such as adaptation, and here we have only 2 data points. The language used for this statement is careful, but I wonder if it is too far to extrapolate that some differences in TE content may be adaptive
Response: Thank you very much for your helpful thoughts. We added more descriptions and references to support "TEs are functional" to make it clear. Please see **lines 156-161**: "Although most of the precise functions of these repeats have not

been studied in depth, repeats have been thought to have a regulatory function in related genes that play an important role in the life cycle and can introduce great genome flexibility[18]. And in the mammalian genome, transposons were described to be redundant enhancers that regulate their target genes which are higher or tissue specially expressed, indicating the importance of transposon".

- Line 196: I don't know what the authors mean by this statement

Response: Thank you. We deleted the sentence.

Deleted sentence "The evolution and expression of single-copy orthologous genes are unique features of organisms." for your reference.

- Section on Ka/Ks values: there is no impression given about the statistical significance of the differences observed between Ka/Ks in any given lineage, or what the distribution of error looks like for these point estimates. Perhaps a more refined PAML analysis could resolve this? It is also not written how Ka/Ks values were calculated

Response: Thank you. We have added statistical test (Mann-Whitney U test) for Ka/Ks comparisons. Please see **lines 204-213**: "We found that the Ka values of the two deep-sea snails (average: 0.37 and 0.41) were higher (Mann-Whitney U test, p-value<0.001) than that of the shallow-water limpet (0.35) but similar to those of two freshwater snails (0.39 and 0.41), which suggests that the genes of deep-sea and freshwater snails both evolved faster after their divergence from shallow-water limpet . The Ks values of the deep-sea (3.34 and 3.09) and fresh-water (3.19 and 3.24) snails were also similar and lower (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (3.72). Additionally, the Ka/Ks values of the deep-sea snails (average: 0.13 and 0.15) were approximately ~20% and ~40% higher (Mann-Whitney U test, p-value<0.001) than those of the shallow-water limpet (0.11).

And the Ka/Ks values were calculated actually using codeml in PAML package. We have added this information to **lines 200-204**: "To explore the evolutionary rate of single-copy orthologous genes, we calculated the synonymous substitution rate (Ka) and nonsynonymous substitution rate (Ks) values of 1,324 single-copy orthologous genes shared by the two deep-sea snails, one shallow-water limpet (L. gigantea), and two freshwater snails (B. glabrata and P. canaliculate) using Codeml in PAML package[23]."

- Typo line 214: CAFE not CAFÉ

Response: Thank you. It was modified.

- Line 329: "region-specific feature shared between lineages" - not sure what is meant by this?

Response: Thank you. We deleted this confusing sentence.

- Line 350: it seems speculative - surely both immune response and biomineralization are "vital" for all snails, not particularly deep-sea ones?

Response: Thank you. We modified it and specified it in *C. squamiferum* which is a chemosynthetic snail species depending on endosymbionts. See **lines 360-362** “In particular, we found that DMBT1 gene families that encode multiple SRCR domains expanded significantly in C. squamiferum. These genes play important roles in immune response and biomineralization, both of which are vital for deep sea chemosynthetic snail”.

- Line 454: reference for GLEAN is missing

Response: Thank you. We added this reference (65). See **line 465**

- Line 469: references for Solar and Hcluster are missing, and what is a H-score?

Response: Thank you. Solar, Hcluster_sg and H-score are tools and concept of TreeFam tools. We added reference for Solar and Hcluster_sg. H-score means hcluster score. Details can be found from TreeFam tools.

- Figure 2a: it's a weird looking tree that, in fact, looks a bit like a snail itself! Are the widths of the blobs representative of the error around the divergence times or topological support?

Response: Thank you and sorry for confusing you. We updated **Fig.2a**. Please see that.