1 TITLE: Giant African snail genomes provide insights into molluscan
2 whole-genome duplication and aquatic-terrestrial transition

3 Conghui Liu⁺, Yuwei Ren⁺, Zaiyuan Li, Qi Hu, Lijuan Yin, Xi Qiao, Yan Zhang,
 4 Longsheng Xing, Yu Xi, Fan Jiang, Sen Wang, Cong Huang, Bo Liu, Hengchao Wang,
 5 Hangwei Liu, Fanghao Wan, Wanqiang Qian*, Wei Fan*

6 Guangdong Laboratory of Lingnan Modern Agriculture, Shenzhen ; Genome
 7 Analysis Laboratory of the Ministry of Agriculture and Rural Affairs; Agricultural
 8 Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences,
 9 Shenzhen, 518120, China.

10 The authors with symbol “⁺” contributed equally: Conghui Liu, Yuwei Ren; the
 11 authors with symbol “*” represented corresponding authors: Wanqiang Qian, Wei
 12 Fan.

13 Email: qianwanqiang@caas.cn; fanwei@caas.cn.

14

15

16

17

18

19

20

21

22 Abstract

23 Whole-genome duplication (WGD) has been observed across a wide variety
 24 of eukaryotic groups, contributing to evolutionary diversity and environmental
 25 adaptability. Mollusks are the second largest group of animals, and are among
 26 the organisms that have successfully adapted to the nonmarine realm through
 27 aquatic-terrestrial (A-T) transition, and no comprehensive research on WGD has
 28 been reported in this group. To explore WGD and the A-T transition in Mollusca,
 29 we assembled a chromosome-level reference genome for the giant African snail
 30 *Achatina immaculata*, a global invasive species, and compared the genomes of
 31 two giant African snails (*A. immaculata* and *Achatina fulica*) to the other
 32 available mollusk genomes. The chromosome-level macrosynteny, colinearity
 33 blocks, Ks peak and Hox gene clusters collectively suggested the occurrence of a
 34 WGD event shared by *A. immaculata* and *A. fulica*. The estimated timing of this
 35 WGD event (~70 MYA) was close to the speciation age of the
 36 Sigmurethra-Orthurethra (within Stylommatophora) lineage and the
 37 Cretaceous-Tertiary (K-T) mass extinction, indicating that the WGD reported
 38 herein may have been a common event shared by all Sigmurethra-Orthurethra
 39 species and could have conferred ecological adaptability and genomic plasticity
 40 allowing the survival of the K-T extinction. Based on macrosynteny, we deduced
 41 an ancestral karyotype containing 8 conserved clusters for the
 42 Gastropoda-Bivalvia lineage. To reveal the mechanism of WGD in shaping
 43 adaptability to terrestrial ecosystems, we investigated gene families related to the

44 **respiration, aestivation and immune defense of giant African snails. Several**
 45 **mucus-related gene families expanded early in the Stylommatophora lineage,**
 46 **functioning in water retention, immune defense and wound healing. The**
 47 **hemocyanins, PCK and FBP families were doubled and retained after WGD,**
 48 **enhancing the capacity for gas exchange and glucose homeostasis in aestivation.**
 49 **After the WGD, zinc metalloproteinase genes were highly tandemly duplicated to**
 50 **protect tissue against ROS damage. This evidence collectively suggests that**
 51 **although the WGD may not have been the direct driver of the A-T transition, it**
 52 **provided an important legacy for the terrestrial adaptation of the giant African**
 53 **snail.**

54 **Introduction**

55 Whole-genome duplication (WGD) is proposed to be a key evolutionary event
 56 driving phenotypic complexity, functional novelty, and ecological adaptation ^{1,2}. With
 57 the progress of high-throughput sequencing technologies, almost all fundamental
 58 lineages of land plants have been characterized as having undergone WGD, while far
 59 fewer WGDs among animal lineages have been reported, especially in invertebrate ^{3,4}.
 60 Based on chromosome counts and genome size, WGD events have been speculated to
 61 have occurred in mollusks, the second largest animal group, which have survived
 62 major extinction events and become some of the most successful conquerors of
 63 marine, freshwater and terrestrial environments ^{5,6}. Giant African snails, a group of
 64 species within Achatinidae possessing unusually large body sizes, have evolved from

65 aquatic ancestors and achieved a pulmonate terrestriality through an aquatic-terrestrial
66 (A-T) transition ⁷. Because of their remarkable ecological adaptability, some members
67 of the giant African snails are considered global invasive species, causing serious
68 damage to agriculture and households ⁸. The success of their colonization of terrestrial
69 ecosystems and adaptation to diverse environments suggests that giant African snails
70 possess advanced genetic plasticity and evolutionary novelties, although the key
71 drivers and underlying mechanisms remain poorly understood.

72 In contrast to plants, only a few ancient WGDs have been well documented in
73 animals, because polyploidy is usually an evolutionary dead end, resulting from
74 associated meiotic difficulties ^{9,10}. In vertebrates, several ancient WGD events that
75 occurred at the origin of vertebrates and teleost fishes have been proposed to have
76 shaped genetic diversity and adaptive radiation ^{11,12}. In animals other than vertebrates,
77 less conclusive evidence of ancient WGD events has been reported as there is a
78 deficiency of continuous high-quality genome data. Nevertheless, there are several
79 known cases of large-scale genome duplication in groups, such as rotifers,
80 chelicerates and insects. The genome of bdelloid rotifers is tetraploid, and its scaffolds
81 are rearranged during asexual reproduction ¹³. A possible ancestral WGD in
82 chelicerates has been identified in horseshoe crab and was dated to at least 135
83 million years ago (MYA) through the phylogenetic analysis of Hox genes ¹⁴, and
84 subsequent surveys in chelicerates revealed that the spider and scorpion lineages
85 underwent another separate WGD before 430 MYA ¹⁵. Because of the fragmental
86 assembly and uncertain accuracy of traditional analysis methods, the validation of

87 WGD in invertebrates remains controversial. Surprisingly, Li et al. reported 18
88 ancient WGDs during the evolution of insects in a recent paper, while another
89 macrosyteny analysis in lepidopterans showed that the gene-tree-based and Ks-based
90 detection of WGD adopted by Li are unreliable, and suggested that WGD events
91 should be verified by chromosome-scale genome assembly and macrosyteny
92 analysis ^{16,17}.

93 Regarding the A-T transition, the conquest of land by organisms that evolved
94 from aquatic ancestors represents one of the most astonishing events in the history of
95 life on Earth ¹⁸. The dramatic environmental changes from homogeneous water
96 habitat to the heterogenous land environment brought about a significant revolution in
97 species evolution, leading to radiation diversity and life complexity ^{19,20}. This step
98 was achieved in multiple phyla independently via a set of adaptations such as water
99 balance, air breathing, nitrogen excretion, neural-immune system interactions, and
100 certain behaviors ²¹. Within Mollusca, the clade Pulmonata includes several lineages
101 that invaded the terrestrial zone and non-aquatic ecosystems, especially the
102 Stylommatophora ²². The giant African snail within Stylommatophora is a
103 representative of the land snails that has also developed fundamental machinery and
104 behaviors such as a pulmonate lung, complex immune system, mucus production, and
105 aestivation, making it highly adapted to the terrestrial environment ²²⁻²⁴. Several
106 reports on amphibious species have shown that the expansion and positive selection of
107 genes related to innate immunity, nitrogen excretion, hormonal regulation and
108 pulmonary surfactants were closely connected to the A-T transition ²⁵, although the

109 genomic features and evolutionary characteristics of terrestrial invertebrates are still
110 poorly described. Studies of the giant African snail, a promising model for terrestrial
111 mollusks, will facilitate the elucidation of invertebrate A-T transition.

112 Mollusca is the second largest animal phylum constituting 7% of living animals
113 on Earth ²⁶, and comprises numerous species of evolutionary and economic
114 importance ^{27,28}. However, the available genomic resources for Mollusca are still quite
115 insufficient in comparison with those for other large phyla such as Arthropoda and
116 Nematoda ^{29,30}. The genomes of aquatic mollusks such as California sea hare ³¹,
117 Pacific oyster ³², pearl oyster ^{33,34}, owl limpet ³⁵, California two-spot octopus ³⁶,
118 golden mussel ³⁷, *Biomphalaria glabrata* ³⁸, and golden apple snail ³⁹ have been
119 sequenced, whereas few terrestrial species in Mollusca have well-documented
120 genomic information. Recently, the genomic data of the invasive land snail *A. fulica*
121 were released, but without in-depth studies of related biological issues ⁴⁰. To address
122 the genetic and evolutionary characteristics of terrestrial mollusks, we report the
123 genome of another giant African snail with a larger body size and greater invasiveness
124 (Fig. 1a, b and c) ⁴¹, and provide insights into the molluscan WGD, A-T transition,
125 and invasion mechanisms, through comparative genomic analysis between the two
126 giant African snails and other closely related mollusks.

127

128 **Results**

129 **The *A. immaculata* genome provides a better assembly and annotation for giant** 130 **African snails**

131 We generated 200 Gb (121 ×) of PacBio SMRT raw reads with an average read
132 length of 14 kb, and 145 Gb (82 ×) of Illumina HiSeq paired-end reads using DNA
133 extracted from a single adult of *A. immaculata* (Supplementary Table 1). After quality
134 filtering, 199 Gb (120 ×) of clean PacBio SMRT reads were corrected with Canu ⁴²,
135 assembled with wtdbg2 ⁴³, and polished with pilon ⁴⁴, resulting in an assembly of 563
136 raw contigs with a total length of 1,653 Mb and an N50 length of 3.80 Mb. Based on
137 the Hi-C data, 1,648 Mb (99.7%) of final contigs were anchored and arranged into 31
138 linkage groups, each corresponding to a natural chromosome (Fig. 1d, Supplementary
139 Fig. 1), where the longest was 111.19 Mb and the shortest 34.32 Mb. According to an
140 estimated genome size of 1.75 Gb based on the distribution of the k-mer frequency ³²
141 (Supplementary Fig. 2), ~95% of the *A. immaculata* genome was assembled. To
142 further confirm the accuracy and completeness of the assembly, we mapped the
143 Illumina reads to the assembled reference genome. Significantly, 99.11% of the
144 genome-derived reads could be aligned to the reference genome with a coverage rate
145 of 96.50%, suggesting no obvious bias in sequencing and assembly. The high-quality
146 reference genome provides a good foundation for gene annotation.

147 Protein-coding genes were predicted in the reference genome by using EVM,
148 integrating *de novo* prediction, transcriptome and homology data (Supplementary

Table 2). In total, 28,702 gene models were predicted as the reference gene set, with coding regions spanning ~39.1 Mb (2.37%) of the genome (Supplementary Table 3 and 4). The distribution of CDS length in *A. immaculata* was similar to that in closely related species (Supplementary Fig. 3 and 4). Overall, 87.5% of the reference genes were supported by transcriptome data, and 96.27% of the eukaryotic core genes from OrthoDB (<http://www.orthodb.org/>) were identified in the reference gene set by BUSCO. These results were comparable to those obtained from other published molluscan genomes (Supplementary Table 4). In the functional annotation, a total of 26,616 (92.73%) reference genes were annotated according to at least one functional database (Supplementary Fig. 5).

The quality of this assembly is much better than that of other molluscan genomes published thus far (Supplementary Table 4). In particular, the coverage rate and sequence continuity were greatly improved compared with the most recent published *A. fulica* genome. The estimated genome size was 2.12 Gb, and the *A. fulica* genome was assembled into 1.86 Gb. The coverage rate of the *A. immaculata* genome (95%) was 8% higher than that of *A. fulica* (87%). The N50 and N90 lengths of the *A. immaculata* contigs were increased by 5.3 times and 4.9 times, respectively. With better assembly quality, ~5000 additional gene models were annotated in *A. immaculata*, and the BUSCO rate was improved from 93% (*A. fulica*) to 96% (*A. immaculata*). The high quality of the assembly and annotation of *A. immaculata* provide a better resource for research on giant African snails.

170 **Signs of adaptive evolution in giant African snails**

171 To gain insights into the evolution of giant African snails (*A. immaculata* and *A.*
172 *fulica*), a total of 292,034 reference genes from ten molluscan genomes (Fig. 2) were
173 clustered into 17,949 orthologous groups (OGs) containing at least two genes each.
174 Then, a phylogenetic tree was built based on 229 high-confidence single-copy
175 orthologous genes with RAxML⁴⁵ and the divergence time was estimated using R8S
176 ⁴⁶. The results showed that *A. immaculata* diverged from *A. fulica* 21 million years
177 ago (MYA), from *B. glabrata* (Panpulmonata) 174 MYA, from *A. californica*
178 (Euopisthobranchia) 205 MYA, and from *P. canaliculata* (Caenogastropoda) 416
179 MYA (Fig. 2a). Through CAFE analysis, we identified 2225 expanded OGs in giant
180 African snails (both *A. immaculata* and *A. fulica*). The functions of these OGs are
181 mainly related to signal transduction; the endocrine, immune and nervous systems;
182 longevity regulation and reproduction (Supplementary Fig. 6). Additionally, 836 OGs
183 were found exclusively in the lineage of giant African snails, which mainly functioned
184 in neurohormonal regulation and mucus synthesis and included such as acetylcholine
185 receptor, pannexin, tenascin, adrenocorticotrophic hormone receptor, neuropeptide
186 receptor, 5-hydroxytryptamine receptor, mucin, heparan sulfate glucosamine
187 3-O-sulfotransferase and proteophosphoglycan genes (Fig. 2b, Source Data file). This
188 evidence suggests that the expanded and lineage-specific OGs may play important
189 roles in the adaptation and invasion of giant African snails.

190 This high-quality genome assembly enables a comprehensive analysis of
191 transposable elements (TEs), which play multiple roles in driving genome evolution

192 in eukaryotes⁴⁷. In total, we identified 954 Mb of repetitive sequences in the
 193 assembled *A. immaculata* genome and 1,366 Mb in *A. fulica* (Fig. 2c). Next, we
 194 analyzed the divergence rate of each class of TE among the available molluscan
 195 genomes. Notably, the TE class of DNA transposons showed a specific peak at a
 196 divergence rate of 4~6% for four invasive species, *A. immaculata*, *A. fulica*, *P.*
 197 *canaliculata* and *C. gigas* (Fig. 2d), indicating a recent explosion of DNA transposons.
 198 We identified 9,291 genes in region that contained DNA transposons distributed
 199 within the specific divergence peak. Based on KEGG annotation, these genes were
 200 mainly enriched in signal transduction; the endocrine, immune and nervous systems
 201 and reproduction (Supplementary Fig. 7). TEs are powerful facilitators of evolution
 202 that generate the “evolutionary potential” to introduce small adaptive changes within
 203 a lineage, and the importance of TEs in stress responses and adaptation has been
 204 reported in numerous studies^{48,49}. The recent explosion of DNA TEs in giant African
 205 snails could have played important roles in promoting their potential plasticity in
 206 stress adaptation.

207 **Whole genome duplication in the Sigmurethra-Orthurethra branch**

208 Whole-genome duplication (WGD) has rarely been reported in animals,
 209 especially in invertebrates, although there is growing suspicions of paleopolyploidy in
 210 mollusks based largely on genome sizes and chromosome counts⁶. With
 211 chromosome-level assemblies, we searched for macrosynteny based on homologous
 212 gene pairs among four molluscan genomes, from *A. immaculata*, *A. fulica*, *P.*

213 *canaliculata* and *P. fucata*. Our chromosome-level macrosynteny revealed a WGD
 214 event shared by two giant African snails, *A. immaculata* and *A. fulica*. The 31
 215 chromosomes of *A. immaculata* could be divided into 14 groups with the preservation
 216 of correspondence (Fig. 3a, Supplementary Fig. 8), and the same situation was found
 217 in *A. fulica* (Fig. 3b, Supplementary Fig. 8). In *A. immaculata*, 2092 homologous gene
 218 pairs with mutual best BLASTP hits were located on the corresponding chromosomes,
 219 and this number was 2364 for *A. fulica* (Supplementary Table 5). The comparison of
 220 the giant African snails, *P. canaliculata* (n = 14) and *P. fucata* (n = 14) revealed that
 221 the karyotype doubled in the lineage leading to giant African snails (Fig. 3c, d and e).
 222 The chromosomes of *A. immaculata* and *A. fulica* showed a 1 to 1 corresponding
 223 relationship (Fig. 3c, Supplementary Fig. 8), while most chromosomes from *P.*
 224 *canaliculata* (Fig. 3d, Supplementary Fig. 8) and *P. fucata* (Fig. 3d, Supplementary
 225 Fig. 8) shared macrosynteny with 2 corresponding chromosomes from giant African
 226 snails. The colinearity blocks identified by MCScanX based on BLASTP hits in
 227 corresponding chromosome groups, also suggested that WGD events had occurred in
 228 two giant African snails (Supplementary Fig. 9 and 10). In the gene age distribution
 229 plot of homologous gene pairs, a specific Ks peak shared by *A. immaculata* and *A.*
 230 *fulica* was observed, which was consistent with WGD. The duplication of Hox gene
 231 clusters was powerful clue leading to the discovery of ancient WGDs in vertebrates⁵⁰,
 232 so we further compared the giant African snails' genomes to those with a single Hox
 233 cluster and no evidence of WGD. Duplicated Hox gene clusters with specific
 234 rearrangements were found in two giant African snail genomes. In conclusion, the

235 macrosyteny, colinearity blocks, Ks peak and Hox gene clusters collectively
236 suggested that a WGD event occurred before the speciation of giant African snails.

237 The timing of WGD events has been reported to show a significant correlation
238 with specific geological and global climatic change ⁵¹. According to the gene age
239 distribution of homologous gene pairs and the divergence time of *B. glabrata* and giant
240 African snails (~174 MYA), we deduced that the timing of the WGD event was ~70
241 MYA. An earlier estimation suggested the occurrence of a WGD event at the
242 Sigmurethra-Orthurethra branch within Stylommatophora by comparison of
243 chromosome numbers among closely related mollusks ⁶. The speciation age of the
244 Sigmurethra-Orthurethra branch was reported to be 65 MYA, which was largely in
245 consistent with our deduced timing, indicating that the WGD reported here was a
246 common event shared by all Sigmurethra and Orthurethra species. The timing of the
247 WGD of the Sigmurethra-Orthurethra branch was also close to Cretaceous-Tertiary
248 (K-T) mass extinction (~66 MYA), in which global climate change caused the
249 extinction of 60-70% of all plant and animal life, including most mollusks ⁵². In plants,
250 the K-T mass extinction is considered a shared common causal factor in the
251 genome-wide doubling of diverse angiosperm lineages ⁵³. It has also been previously
252 suggested that polyploidization in animals is correlated with periods of unstable
253 environments ¹. The WGD of the Sigmurethra-Orthurethra branch is expected to have
254 provided ecological adaptability and genome plasticity and allowed these taxa to
255 survive the K-T mass extinction.

256 As the terrestrial area expanded during the K-T mass extinction due to

257 Maastrichtian sea-level regression ⁵⁴, the WGD is expected to have promoted the
258 adaptability of land snails to terrestrial ecosystems and speciation diversity. The
259 functions of WGD-derived homologous gene pairs were significantly enriched in
260 biological regulation, signal transduction, energy generation and the response to
261 stimulus (Supplementary Fig. 11). These functions are closely related to terrestrial
262 living, indicating that the retained WGD-associated genes increased the terrestrial
263 ecological tolerance of giant African snails.

264 The chromosome-level assembly of *A. immaculata*, *A. fulica*, *P. canaliculata* and
265 *P. fucata*, allowed us to infer karyotype evolution within the clade of
266 Gastropoda-Bivalvia based on macrosynteny. Our results indicated a monoploid
267 chromosome number of n=8 for the root of the Gastropoda-Bivalvia lineage. This
268 deduced chromosome number was consistent with that of Patellogastropoda (n=8~9)
269 ⁵⁵, which has been reported to exhibit monophyly in the Gastropoda-Bivalvia lineage
270 and to present preserved ancestral characters ⁵⁶. During subsequent speciation, 6
271 breaks in the lineages of Gastropoda (n=14) and *P. fucata* (n=14) were observed at
272 different locations. In the Gastropoda clade, Stylommatophora possessed 17
273 chromosomes, with fusion of chr4 and chr13 of Gastropoda, and showed 4 breaks in
274 chr5 ,chr6 ,chr7 and chr8 of Gastropoda, while *P. canaliculata* retained the monoploid
275 chromosome karyotype of Gastropoda (n=14). The WGD in the lineage of the
276 Sigmurethra-Orthurethra branch doubled the chromosome count (n=17+17). After
277 WGD, three fusions (chr6-chr15, chr8-chr16 and chr10-chr17 of Stylommatophora)
278 resulted in the 31 chromosomes of the Sigmurethra-Orthurethra branch (n=17+14)

(Fig. 4). In the lineage of *A. immaculata* and *A. fulica*, a chromosome number $n = 31$ has remained since their speciation in 21 MYA. The demonstration of chromosome duplications, fusions and breaks provides insights into the ancestral karyotype and the mechanism of karyotype evolution in Gastropoda-Bivalvia.

Expansion of hemocyanins and zinc metalloproteinases improves terrestrial respiratory function

The innovation of respiratory gas exchange is the signature of the aquatic-to-terrestrial transition, which was one of the most conspicuous evolutionary events to have occurred on Earth⁵⁷. The evolution of the oxygen transportation system allowed land snails to utilize O₂ from air far more efficiently than aquatic mollusks⁵⁸. Additionally, an advanced system is needed to eliminate the accompanying oxidative stress to maintain O₂ homeostasis. However, the underlying mechanisms of O₂ transport and antioxidation are less well understood.

O₂ transport in most members of Gastropoda and Cephalopoda is dependent on hemocyanin⁵⁹. Based on orthologous and phylogenetic analysis, hemocyanin genes were observed in six species of Gastropoda and Cephalopoda, which was consistent with previous reports^{59,60}. Notably, there are 4 hemocyanin genes in both *A. immaculata* and *A. fulica*, which is twice as many as in any other Gastropoda species (Fig. 5a). Moreover, the four homologous genes of *A. immaculata* and *A. fulica* are located on two chromosomes that derived from WGD, whereas those of other species without WGD were only located in one chromosome or scaffold (Fig. 5b,

Supplementary Fig. 12 and 13). Thus, the doubling of the hemocyanin gene number through WGD may have increased the O₂ transport ability of giant African snails to adapt to land living.

Reactive oxygen species (ROS) are generated during O₂ metabolism, and excessive ROS cause oxidative stress and trigger inflammation⁶¹. However, antioxidant enzymes protect the host from excess oxidative damage. Superoxide dismutase (SOD), acid phosphatase (AP) and glutathione S transferase (GST) genes were identified in giant African snails, with gene numbers comparable to those of other molluscan species (Supplementary Table 6). In addition to antioxidant enzymes, metalloproteinases have the ability to hydrolyze fibronectin to reduce the injuries caused by ROS^{62,63}. The number of zinc metalloproteinase genes in both giant African snails was largely expanded, to 11 in *A.immaculata* and 9 in *A. fulica*, compared to only 1 or 2 in other species (Fig. 5c, Supplementary Fig. 14 and 15). Importantly, the zinc metalloproteinase genes of the two species were located in two syntenic clusters on homologous chromosomes, while no homologous genes were found on the corresponding duplicated chromosomes resulting from WGD. The finding that the same location pattern was shared by the two species indicates that these genes were tandemly duplicated after the WGD event but before the speciation of *A.immaculata* and *A. fulica*. Furthermore, the expression levels of all zinc metalloproteinases in the hepatopancreas and blood were higher than those in other tissues, which was consistent with the importance of the antioxidative function of these two tissues. Therefore, the expansion of zinc metalloproteinases after WGD

might have played important roles in the defense of giant African snails against the damage resulting from oxidative stress and inflammation.

Glucose homeostasis and ureagenesis benefit survival in aestivation

Aestivation in land animals is a special long-term torpid state that occurs in response to the extreme conditions of summer, such as desiccation, high temperature and starvation⁶⁴. Terrestrial mollusks originating from aquatic ancestors, developed aestivation as a strategy for surviving drastic environmental changes on land. During aestivation, giant African snails seal their shell aperture with epiphragma, and their body remains within the solid shell for several months²⁴. During this long period, these snails are isolated from feeding and excretion, but their blood sugar level and toxic nitrogenous waste are maintained within a normal range. However, how the snails regulate blood glucose homeostasis and eliminate toxins during aestivation is still poorly understood.

Without sugar intake during aestivation, blood glucose homeostasis is mainly achieved via the exploitation of endogenous resources and the reduction of energy expenditure⁶⁵. Regarding endogenous resources, glycogen is used initially, after which some portion of the carbohydrate skeleton associated with protein metabolism is primarily employed to produce blood glucose through gluconeogenesis⁶⁵. In this study, we analyzed the expression level changes of two rate-limiting gluconeogenic enzymes phosphoenolpyruvate carboxykinase (PCK, EC:4.1.1.32) and fructose-1,6 - bisphosphatase (FBP, EC:3.1.3.11). Both enzymes are encoded by two homologous

genes derived from WGD, and the expression level of one homologous gene is significantly increased in aestivation (negative binomial generalized log-linear model, $p < 0.01$; Supplementary Table 7 and 8), while the other displays relatively constant expression. On the other hand, the consumption of glucose through the tricarboxylic acid (TCA) cycle is minimized in aestivation⁶⁶. Several important enzymes are involved in the TCA, such as citrate synthase (EC:2.3.3.1), and malate dehydrogenase (EC:1.1.1.37) (Fig. 6a)⁶⁷. Citrate synthase is encoded by two homologous genes derived from WGD, where one of the homologous genes is significantly downregulated (negative binomial generalized log-linear model, $p < 0.01$; Supplementary Table 7 and 8), while the other gene remains stable. The gene encoding malate dehydrogenase also presented significantly downregulated expression. This evidence indicates that blood glucose homeostasis during aestivation is achieved via both the upregulation of gluconeogenesis and downregulation of the TCA cycle.

The oxidative deamination of amino acids generates toxic ammonium (NH_3 , NH_4^+), which can cause immunosuppression and neurotoxic effects^{68,69}. In a normal state, giant African snails mainly convert ammonium into uric acid and excrete it out of their body through the urine, while during aestivation, they convert ammonium into urea and store it in their body^{24,70}. The accumulation of nontoxic urea during aestivation has been proposed to play a role in the detoxification of nitrogenous substances, the reduction of evaporative water loss, and the reutilization of nitrogen resources⁶⁵. In this study, we analyzed the expression profiles of three important

enzymes involved in the urea cycle, including carbamoyl phosphate synthetase (CPS, EC:6.3.5.5), argininosuccinate synthetase (ASS, EC:6.3.4.5), and arginase (EC:3.5.3.1)⁷¹. All of the genes encoding these enzymes were significantly upregulated in the aestivation group compared with the normal group (Fig. 6b), providing insight into the mechanism of the transformation of ammonia into urea during the aestivation of giant African snails.

Expanded mucus-related gene families reinforce immune defense

Molluscan immune defense has drawn increasing attention because of its economic and evolutionary importance^{5,72}. As one of the most successful colonizers of terrestrial environments within Mollusca, giant African snails are remarkably adaptive and may possess advanced molecular mechanisms for host-defense against biotic and abiotic stresses⁷³. Although giant African snails lack the canonical vertebrate immunoglobulin, they have developed diverse repertoires of receptors, regulators, and effectors²³. To investigate the genes and pathways involved in immune and stress responses, we characterized the immune system on the basis of the genome of *A. immaculata*, including pattern recognition receptors, soluble factors, and mucus-related gene families (Fig. 7a, Supplementary Table 9). The transcriptomes of the hemocytes after different stimuli (lipopolysaccharide/LPS, peptidoglycan/PGN, poly (I:C)/IC and β -glucan/GLU) were also analyzed to address the potential roles of these genes (Supplementary Fig. 16).

The mucus of giant African snails is mainly composed of glycosaminoglycans

and glycoproteins, and its antimicrobial properties were first reported in the 1980s⁷⁴. In the *A. immaculata* genome, we identified 11 heparan sulfate glucosamine 3-O-sulfotransferases (OSTs) functioning in glycosaminoglycan synthesis and 99 mucins that encode the major glycoprotein in mucus. OST genes were found to be expanded in Panpulmonata (10 in *A. immaculata*, 11 in *A. fulica*, 10 in *B. glabrata*) and Cephalopoda (13 in *O. bimaculoides*), in which they were two times more abundant than in other mollusks (Supplementary Table 10). The expansion of OSTs in Panpulmonata occurred in subfamilies OST-1 and OST-3B (Fig. 7b). Acharan sulfate, a product of OST, is a primary constituent of mucus, and the expansion of the OSTs was therefore closely related to the abundant surface mucus. The increased expression levels of OSTs observed after LPS stimulus suggest a possible immune role in the response to Gram-negative bacteria. The mucin genes were also found to be expanded on the Stylommatophora branch (99 for *A. immaculate*, 71 for *A. fulica*), showing a number approximately three times greater than in other mollusks, indicating that the mucus of Stylommatophora contains more mucin proteins to defend against microorganism infection (Supplementary Fig. 17, Supplementary Table 10). Increased expression patterns of mucins were mainly detected in the LPS, IC and GLU groups, implying their immune roles against Gram-negative bacteria, viruses and fungi, respectively.

A total of 34 mucins from 19 OGs and 3 OSTs from 2 OGs were found exclusively on the Stylommatophora branch (Source Data file). Their homologous pseudogenes were located within the corresponding duplicated chromosomes,

408 indicating that these genes were generated before WGD, doubled in number through
409 WGD, and transformed into pseudogenes after WGD (Supplementary Table 11 and
410 12). In addition to their roles in immunity, functions of the mucins and OSTs in
411 wound healing, locomotion and other terrestrial-related processes were observed,
412 indicating that their expansions before WGD may have played important roles in the
413 A-T transition.

414 **Discussion**

415 WGD provides evolutionary novelties that increase environmental adaptation and
416 species diversity, although this process is regarded as rare in animals because it is
417 hampered by sex determination ⁷⁵. As one of the best-known simultaneous
418 hermaphrodites ⁷⁶, the giant African snail possesses the ability to undergo
419 autofecundation under certain circumstances, thereby overcoming the evolutionary
420 dead end resulting from WGD. Based on two high-quality assemblies from giant
421 African snails, we report a high-credibility WGD on the Sigmurethra-Orthurethra
422 branch (Order: Stylommatophora) with collective evidence including macrosynteny
423 data, colinearity blocks, the gene-age distribution and Hox gene clusters. In particular,
424 chromosome-level macrosynteny was employed for WGD identification, which was
425 highly recommended in a recent paper to avoid possible false-positive results ¹⁷. To
426 the best of our knowledge, this WGD is the first to be reported in Mollusca and the
427 first to be reported in invertebrates based on chromosome-level macrosynteny. In
428 contrast to the two WGDs found in chelicerates (~135 and ~430 MYA), this

molluscan WGD was a relatively more recent event with an estimated timing of ~70 MYA. In contrast to ancient WGDs, the identification of this recent event will help to reveal the adaptive mechanisms and consequences of WGD, especially regarding subgenome divergence, providing more traceable genomic clues⁷⁷. In most cases, WGD occurs only under remarkably uncommon circumstances and provides a driving force during subsequent evolution. As an example, the WGDs identified in a cluster of angiosperm plants share a common causal factor corresponding with the K-T mass extinction⁷⁸. The timing of the WGD that occurred in chelicerates at ~430 MYA and that in *Sigmurethra-Orthurethra* at ~70 MYA is close to the Ordovician-Silurian (O-S) extinction and the K-T mass extinction, respectively, indicating that the invertebrate WGD was also connected to mass extinction events. Furthermore, it has been reported that WGD is followed by a substantial increase in morphological complexity and species numbers². Therefore, the species richness and wide ranging adaptations of invertebrates imply that more undiscovered WGDs exist in this clade and might be revealed as the available genomic resources increases.

The A-T transition from water-living to land-living was a milestone in the evolutionary history on Earth²⁰. The *Stylommatophora* lineage, which originated from a marine ancestor, and breaths through permanently air-filled lungs successfully completed terrestrial adaptation approximately 100~150 MYA⁷⁹. With all the associated anatomical and physiological changes required by terrestrial living⁸⁰, *Stylommatophora* is considered to provide good study material for the investigation of the A-T transition. The WGD event and sea-level regression that

occurred during the K-T boundary are expected to have resulted in greater adaptability to terrestrial ecosystems⁵⁴. To reveal the mechanism underlying the A-T transition and the influence of WGD, we investigated gene families related to respiration, aestivation and immune defense from the giant African snail. The terrestrial adaptation of Stylommatophora was initiated before WGD (~70 MYA), and we found that several mucus-related gene families expanded early in the Stylommatophora lineage (100~150 MYA); these genes included the mucin and OST families, functioning in water retention, immune defense and wound healing. WGD has been proposed to provide functional redundancy and mutational robustness to increase the rates of evolution and adaptation⁸¹. The genes encoding hemocyanins, involved in O₂ transport, and PCK and FBP, involved in gluconeogenesis, were doubled and retained after WGD, enhancing the capacity for gas exchange and glucose homeostasis in aestivation. The extra chromosome copy resulting from genome duplication might limit the risk of genetic variation in one copy of the chromosome. In the post-WGD period, zinc metalloproteinase genes were highly tandemly duplicated, resulting in the protection of tissue against ROS damage arising from respiration. This evidence indicates that although the A-T transition of the giant African snail was not initially driven by WGD, WGD could have facilitated its terrestrial adaptation by providing additional genomic resources, thus increasing the survival rate in the drastic transition from water to land.

Biological invasion has become an increasing serious problem worldwide, as the international flow of people and goods has increased over the years⁸². With its rich

species diversity and strong environmental adaptability, Mollusca is among the phyla with the greatest numbers of invasive species, such as the golden apple snail and giant African snail, which are listed among the top 100 global invasive species, although the invasion mechanism of mollusks is not yet clear. Under genetic bottlenecks, the invasive species can still adapt to new habitats and expand their populations, which is referred to as the ‘genetic paradox of invasive species’⁸³. Previous reports have shown that TEs are powerful facilitators of rapid adaptation that generate “evolutionary potential” by introducing stress-induced changes in invasive species⁸⁴. In this study, recent TE explosions were observed in all 4 invasive mollusks but were absent in the other noninvasive mollusks. Genes located near recently emerged TEs were enriched in the function of stress responses, which is consistent with our previous findings in the golden apple snail³⁹, indicating that the recent TE explosion might be a common genetic force contributing to biological invasion. In addition, WGD has been proposed to provide robustness of genetic variation and redundant gene resources for the rapid evolution of novel functions, driving phenotypic complexity and ecological adaptation. Therefore, WGD may be another explanation for the genetic paradox of biological invasion, and species exhibiting WGD may present greater invasiveness.

In summary, we have revealed a WGD occurring on the Sigmurethra-Orthurethra branch within Stylommatophora providing genomic evidence for the A-T transition in Mollusca, and we propose WGD as a potential mechanism contributing to biological invasion. In addition, the obtained genome sequences of giant African snails will

enable us to develop more environmentally friendly and efficient control measures using species-specific gene targets, benefiting the protection of agricultural crops and ecological environment as well as the prevention of human disease caused by zoonotic parasites. On the other hand, giant African snails are considered as a high-protein food source in some parts of the world, especially in Africa and Asia. The invasive characteristics of giant African snails, such as their rapid growth, high production rate, and ability to survive harsh conditions, endow them with the potential to be cultivated as an economic species. Thus, the genome sequence of giant African snails provides a powerful platform for the genetic breeding of this species, turning “waste” into wealth.

References

- 1 Mable, B., Alexandrou, M. & Taylor, M. Genome duplication in amphibians and fish: an extended synthesis. *Journal of Zoology* **284**, 151-182 (2011).
- 2 Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725 (2009).
- 3 Clark, J. W. & Donoghue, P. C. J. Whole-genome duplication and plant macroevolution. *Trends. Plant. Sci.* **23**, 933-945 (2018).
- 4 Orr, H. A. " Why polyploidy is rarer in animals than in plants" revisited. *The American Naturalist* **136**, 759-770 (1990).
- 5 Guo, X., He, Y., Zhang, L., Lelong, C. & Jouaux, A. Immune and stress responses in oysters with insights on adaptation. *Fish Shellfish Immunol.* **46**,

- 107-119 (2015).
- 6 Hallinan, N. M. & Lindberg, D. R. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. *Genome. Biol. Evol.* **3**, 1150-1163 (2011).
- 7 Jörger, K. M. *et al.* On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia. *BMC Evol. Biol.* **10**, 323 (2010).
- 8 Cowie, R. H., Dillon, R. T., Robinson, D. G. & Smith, J. W. Alien non-marine snails and slugs of priority quarantine importance in the United States: A preliminary risk assessment. *Am. Malacol. Bull.* **27**, 113-133 (2009).
- 9 Comai, L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836 (2005).
- 10 Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411 (2017).
- 11 Furlong, R. F. & Holland, P. W. Were vertebrates octoploid? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **357**, 531-544 (2002).
- 12 Hoegg, S., Brinkmann, H., Taylor, J. S. & Meyer, A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* **59**, 190-203 (2004).
- 13 Flot, J.-F. *et al.* Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* **500**, 453 (2013).
- 14 Kenny, N. J. *et al.* Ancestral whole-genome duplication in the marine

- 538 chelicerate horseshoe crabs. *Heredity* **116**, 190 (2016).
- 539 15 Schwager, E. E. *et al.* The house spider genome reveals an ancient
540 whole-genome duplication during arachnid evolution. *BMC Biol.* **15**, 62
541 (2017).
- 542 16 Li, Z. *et al.* Multiple large-scale gene and genome duplications during the
543 evolution of hexapods. *PNAS* **115**, 4713-4718 (2018).
- 544 17 Nakatani, Y. & McLysaght, A. Macrosynteny analysis shows the absence of
545 ancient whole-genome duplication in lepidopteran insects. *PNAS* **116**,
546 1816-1818 (2019).
- 547 18 Lillywhite, H. B., Albert, J. S., Sheehy III, C. M. & Seymour, R. S. Gravity
548 and the evolution of cardiopulmonary morphology in snakes. *Comp. Biochem.*
549 *Physiol. A Mol. Integr. Physiol.* **161**, 230-242 (2012).
- 550 19 Vermeij, G. J. & Dudley, R. Why are there so few evolutionary transitions
551 between aquatic and terrestrial ecosystems? *Biol. J. Linn. Soc.* **70**, 541-554
552 (2000).
- 553 20 Volkmann, D. & Baluška, F. Gravity: one of the driving forces for evolution.
554 *Protoplasma* **229**, 143-148 (2006).
- 555 21 Little, C. *The terrestrial invasion: an ecophysiological approach to the origins*
556 *of land animals.* (CUP Archive, 1990).
- 557 22 Romero, P. E., Pfenninger, M., Kano, Y. & Klussmann-Kolb, A. Molecular
558 phylogeny of the Ellobiidae (Gastropoda: Panpulmonata) supports
559 independent terrestrial invasions. *Mol. Phylogenet. Evol.* **97**, 43-54 (2016).

- 560 23 Mukherjee, S., Sarkar, S., Munshi, C. & Bhattacharya, S. *The Uniqueness of*
561 *Achatina fulica in its Evolutionary Success.* (InTech, 2017).
- 562 24 Hiong, K. C., Loong, A. M., Chew, S. F. & Ip, Y. K. Increases in urea
563 synthesis and the ornithine–urea cycle capacity in the Giant African Snail,
564 *Achatina fulica*, during fasting or aestivation, or after the Injection with
565 ammonium chloride. *J. Exp. Zool. A Comp. Exp. Biol.* **303**, 1040-1053 (2005).
- 566 25 Biscotti, M. A. *et al.* The Lungfish Transcriptome: A Glimpse into Molecular
567 Evolution Events at the Transition from Water to Land. *Sci. Rep.* **6**, 21571,
568 doi:10.1038/srep21571 (2016).
- 569 26 Bouchet, P. & Duarte, C. M. *The exploration of marine biodiversity: scientific*
570 *and technological challenges.* (Fundación BBVA, 2006).
- 571 27 Thiengo, S. C., Faracas, F. A., Salgado, N. C., Cowie, R. H. & Fernandez, M.
572 A. Rapid spread of an invasive snail in South America: the giant African snail,
573 *Achatina fulica*, in Brasil. *Biol. Invasions* **9**, 693-702 (2007).
- 574 28 Ferreira, F. S., Albuquerque, U. P., Coutinho, H. D., Almeida Wde, O. & Alves,
575 R. R. The trade in medicinal animals in northeastern Brazil. *Evid. Based*
576 *Complement. Alternat. Med.* **2012**, 126938, doi:10.1155/2012/126938 (2012).
- 577 29 Serb, J. M. & Lydeard, C. Complete mtDNA sequence of the North American
578 freshwater mussel, *Lampsilis ornata* (Unionidae): an examination of the
579 evolution and phylogenetic utility of mitochondrial genome organization in
580 Bivalvia (Mollusca). *Mol. Biol. Evol.* **20**, 1854-1866 (2003).
- 581 30 Faure, E. & Casanova, J. P. Comparison of chaetognath mitochondrial

- 582 genomes and phylogenetical implications. *Mitochondrion* **6**, 258-262 (2006).
- 583 31 Aplysia Genome Project. Broad Institute. Vertebrate Biology Group.
584 doi:<https://www.broadinstitute.org/aplysia/aplysia-genome-project> (2009).
- 585 32 Liu, B. *et al.* Estimation of genomic characteristics by analyzing k-mer
586 frequency in de novo genome projects. *arXiv* (2013). URL:
587 <https://arxiv.org/abs/1308.2012>.
- 588 33 Du, X. *et al.* The pearl oyster *Pinctada fucata martensii* genome and
589 multi-omic analyses provide insights into biomineralization. *Gigascience* **6**,
590 1-12 (2017).
- 591 34 Takeuchi, T. *et al.* Draft genome of the pearl oyster *Pinctada fucata*: a platform
592 for understanding bivalve biology. *DNA Res.* **19**, 117-130 (2012).
- 593 35 Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian
594 genomes. *Nature* **493**, 526-531 (2013).
- 595 36 Albertin, C. B. *et al.* The octopus genome and the evolution of cephalopod
596 neural and morphological novelties. *Nature* **524**, 220-224 (2015).
- 597 37 Uliano-Silva, M. *et al.* A hybrid-hierarchical genome assembly strategy to
598 sequence the invasive golden mussel, *Limnoperna fortunei*. *Gigascience* **7**,
599 doi:10.1093/gigascience/gix128 (2018).
- 600 38 Adema, C. M. *et al.* Whole genome analysis of a schistosomiasis-transmitting
601 freshwater snail. *Nat. Commun.* **8**, 15451, doi:10.1038/ncomms15451 (2017).
- 602 39 Liu, C. *et al.* The genome of the golden apple snail *Pomacea canaliculata*
603 provides insight into stress tolerance and invasive adaptation. *Gigascience* **7**,

doi:10.1093/gigascience/giy101 (2018).

40 Guo, Y. *et al.* A chromosomal-level genome assembly for the giant African
snail *Achatina fulica*. *Gigascience* **8**, doi:10.1093/gigascience/giz124 (2019).

41 Tan, S. K. & Low, M. E. Y. An update to the West African *Limicolaria*
flammea (Müller, 1774) in Singapore, and its distinction from the confamilial
Achatina fulica Bowdich, 1822 (Mollusca: Gastropoda: Achatinidae). *Nature*
in Singapore **4**, 311-317 (2011).

42 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive
k-mer weighting and repeat separation. *Genome Res.* **27**, 722-736 (2017).

43 Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *BioRxiv*,
(2019). URL: <https://www.biorxiv.org/content/10.1101/530972v1.full>.
DOI: <https://doi.org/10.1101/530972>.

44 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial
variant detection and genome assembly improvement. *PloS one* **9**, e112963
(2014).

45 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and
post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

46 Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and
divergence times in the absence of a molecular clock. *Bioinformatics* **19**,
301-302 (2003).

47 Feschotte, C. & Wessler, S. R. Mariner-like transposases are widespread and
diverse in flowering plants. *Proceedings of the National Academy of Sciences*

626 **99**, 280-285 (2002).

627 48 Hua-Van, A., Le Rouzic, A., Boutin, T. S., Filée, J. & Capi, P. The struggle for
628 life of the genome's selfish architects. *Biology direct* **6**, 19 (2011).

629 49 Werren, J. H. Selfish genetic elements, genetic conflict, and evolutionary
630 innovation. *PNAS* **108**, 10863-10870 (2011).

631 50 Amores, A. *et al.* Zebrafish hox clusters and vertebrate genome evolution.
632 *Science* **282**, 1711-1714 (1998).

633 51 Sessa, E. B. Polyploidy as a mechanism for surviving global change. *New*
634 *Phytol.* **221**, 5-6 (2019).

635 52 Petersen, S. V., Dutton, A. & Lohmann, K. C. End-Cretaceous extinction in
636 Antarctica linked to both Deccan volcanism and meteorite impact via climate
637 change. *Nat. Commun.* **7**, 12079 (2016).

638 53 Fawcett, J. A., Maere, S. & Van De Peer, Y. Plants with double genomes might
639 have had a better chance to survive the Cretaceous–Tertiary extinction event.
640 *PNAS* **106**, 5737-5742 (2009).

641 54 Marshall, C. R. & Ward, P. D. Sudden and gradual molluscan extinctions in
642 the latest Cretaceous of western European Tethys. *Science* **274**, 1360-1363
643 (1996).

644 55 Thiriou-Quiévreux, C. Advances in chromosomal studies of gastropod
645 molluscs. *J. Mollus. Stud.* **69**, 187-202 (2003).

646 56 Aktipis, S. W. & Giribet, G. A phylogeny of Vetigastropoda and other
647 “archaeogastropods”: re-organizing old gastropod clades. *Invertebr. Biol.* **129**,

648 220-240 (2010).

649 57 Ashley-Ross, M. A., Hsieh, S. T., Gibb, A. C. & Blob, R. W. Vertebrate land
650 invasions-past, present, and future: an introduction to the symposium. *Integr.*
651 *Comp. Biol.* **53**, 192-196 (2013).

652 58 Hsia, C. C., Schmitz, A., Lambertz, M., Perry, S. F. & Maina, J. N. Evolution
653 of air breathing: oxygen homeostasis and the transitions from water to land
654 and sky. *Compr. Physiol.* **3**, 849-915 (2013).

655 59 Markl, J. Evolution of molluscan hemocyanin structures. *Biochim. Biophys.*
656 *Acta* **1834**, 1840-1852 (2013).

657 60 Kato, S., Matsui, T., Gatsogiannis, C. & Tanaka, Y. Molluscan hemocyanin:
658 structure, evolution, and physiology. *Biophys. Rev.* **10**, 191-202 (2018).

659 61 Shen, S. *et al.* ABCG2 reduces ROS-mediated toxicity and inflammation: a
660 potential role in Alzheimer's disease. *J. Neurochem.* **114**, 1590-1604 (2010).

661 62 Kruse, M. N. *et al.* Human meprin α and β homo-oligomers: cleavage of
662 basement membrane proteins and sensitivity to metalloprotease inhibitors.
663 *Biochem. J.* **378**, 383-389 (2004).

664 63 Faa, G. *et al.* Zinc in gastrointestinal and liver disease. *Coordin. Chem. Rev.*
665 **252**, 1257-1269 (2008).

666 64 Storey, K. B. Life in the slow lane: Molecular mechanisms of estivation. *Comp.*
667 *Biochem. Physiol. A Mol. Integr. Physiol.* **133**, 733-754 (2002).

668 65 Ip, Y. K. & Chew, S. F. *Nitrogen Metabolism and Excretion During*
669 *Aestivation*. (Springer Heidelberg Dordrecht London New York, 2010).

- 670 66 Bell, R. A., Dawson, N. J. & Storey, K. B. Insights into the in vivo regulation
671 of glutamate dehydrogenase from the foot muscle of an estivating land snail.
672 *Enzyme Res.* **2012**, 317314, doi:10.1155/2012/317314 (2012).
- 673 67 Bulutoglu, B., Garcia, K. E., Wu, F., Minter, S. D. & Banta, S. Direct
674 Evidence for Metabolon Formation and Substrate Channeling in Recombinant
675 TCA Cycle Enzymes. *ACS Chem. Biol.* **11**, 2847-2853 (2016).
- 676 68 Oja, S. S., Saransaari, P. & Korpi, E.R. Neurotoxicity of Ammonia.
677 *Neurochem. Res.* **42**, 713 (2017).
- 678 69 Hermenegildo, C., Monfort, P. & Felipo, V. Activation of
679 N-Methyl-D-Aspartate Receptors in Rat Brain In Vivo Following Acute
680 Ammonia Intoxication: Characterization by In Vivo Brain Microdialysis.
681 *Hepatology* **31**, 709-715 (2000).
- 682 70 Onadeko, A. B. & Egonmwan, R. I. Water Relations, Weight Loss and
683 Nitrogenous Waste Products in the Giant African Land Snail *Archachatina*
684 *Marginata* (Swainson) – Pulmonata: Achatinidae. *J. Sci. Res. Dev.* **17**, 40-47
685 (2017).
- 686 71 Diez-Fernandez, C., Gallego, J., Haberle, J., Cervera, J. & Rubio, V. The
687 Study of Carbamoyl Phosphate Synthetase 1 Deficiency Sheds Light on the
688 Mechanism for Switching On/Off the Urea Cycle. *J Genet Genomics* **42**,
689 249-260 (2015).
- 690 72 Wang, L., Qiu, L., Zhou, Z. & Song, L. Research progress on the mollusc
691 immunity in China. *Dev. Comp. Immunol.* **39**, 2-10 (2013).

- 692 73 Begum, N., Matsumoto, M., Tsuji, S., Toyoshima, K. & Seya, T. The primary
693 host defense system across humans, flies and plants. *Curr. Trends Immunol.* **3**,
694 59-74 (2000).
- 695 74 Iguchi, S. M., Aikawa, T. & Matsumoto, J. J. Antibacterial activity of snail
696 mucus mucin. *Comp. Biochem. Physiol. A Physiol.* **72**, 571-574 (1982).
- 697 75 Wertheim, B., Beukeboom, L. & Van de Zande, L. Polyploidy in animals:
698 effects of gene expression on sex determination, evolution and ecology.
699 *Cytogenet. Genome Res.* **140**, 256-269 (2013).
- 700 76 Barrows, E. M. *Animal behavior desk reference: a dictionary of animal*
701 *behavior, ecology, and evolution.* (CRC press, 2000).
- 702 77 Soltis, D. E. *et al.* Polyploidy and angiosperm diversification. *Am. J. Bot.* **96**,
703 336-348 (2009).
- 704 78 Vanneste, K., Baele, G., Maere, S. & Van de Peer, Y. Analysis of 41 plant
705 genomes supports a wave of successful genome duplications in association
706 with the Cretaceous–Paleogene boundary. *Genome Res.* **24**, 1334-1347 (2014).
- 707 79 Kameda, Y. & Kato, M. Terrestrial invasion of pomatiopsid gastropods in the
708 heavy-snow region of the Japanese Archipelago. *BMC Evol. Biol.* **11**, 118
709 (2011).
- 710 80 Barker, G. M. *Gastropods on land: phylogeny, diversity and adaptive*
711 *morphology.* 1-146 (CAB International, 2001).
- 712 81 Crow, K. D. & Wagner, G. P. What is the role of genome duplication in the
713 evolution of complexity and diversity? *Mol. Biol. Evol.* **23**, 887-892 (2006).

714 82 Wan, F., Jiang, M. & Zhan, A. *Biological invasions and its management in*
715 *china*. Vol. 1 (Springer, 2017).

716 83 Frankham, R. Resolving the genetic paradox in invasive species. *Heredity* **94**,
717 385-385 (2005).

718 84 Stapley, J., Santure, A. W. & Dennis, S. R. Transposable elements as agents of
719 rapid adaptation may explain the genetic paradox of invasive species. *Mol.*
720 *Ecol.* **24**, 2241-2252 (2015).

721

722

723

724 **Figure legends**

725 **Figure 1. The general characteristics of *A.immaculata*.** (a) The morphological
726 difference: *A. immaculata* has relatively longer shell with pink or purple columella,
727 while *A. fulica* has shorter shell with white columella. (b) Global invasion regions for
728 *A. immaculata* (top) and *A. fulica* (bottom). Currently, *A. fulica* is more widely
729 distributed, but *A. immaculata* has more advantage with a faster invasion speed. (c)
730 The sketch map showed the major physiological specialty of *A. immaculata*, including
731 immune innovation, eupulmonates lung, estivation and mucus. (d) Genomic features
732 showed by Circos plot. Track n: 31 linkage groups of the genome; Track d-m:
733 expression profile of brain, egg, eye, hemocytes, hepatopancreas, kidney, lungs,
734 muscle, ovary and testis tissues; Track c: distribution of transposon elements; Track b:

735 distribution of gene density; Track a: distribution of GC content.

736 **Figure 2. Evolutionary analysis of *A.immaculata*.** (a) Phylogenetic placement of *A.*
737 *immaculata* within the dated tree of mollusks. The estimated divergence time is
738 shown at each branching point, and each lineage is shown in different color blocks. *A.*
739 *immaculata* is highlighted in red. (b) Categories of orthologous and paralogous genes
740 in various phyla and species. (c) Classification and contents of repetitive sequences in
741 *A. immaculata* and other mollusks. (d) Distribution of divergence rate for the class of
742 DNA transposons in mollusks genomes. The divergence rate was calculated by
743 comparing all TE sequences identified in the genome to the corresponding consensus
744 sequence in each TE subfamily. The arrow indicates that both *A. immaculata* and *A.*
745 *fulica* had a recent explosion of TEs at divergence rate of ~5%.

746 **Figure 3. Whole genome duplication shared by *A. immaculata* and *A. fulica*.**
747 Circos plots showed the homologous gene pairs in chromosome-level microsynteny
748 among the *A. immaculata* (a) and *A. fulica* (b) individually, as well as chromosomes
749 between *A. immaculata* and *A. fulica* (c), *A. immaculata* and *P. canaliculata* (d), *A.*
750 *immaculata* and *P. fucata* (e). (f) Gene age distribution of Ks values calculated from
751 orthologous gene pairs, among *A. immaculata*, *A. fulica*, and *P. canaliculata*
752 individually, as well as between *A. immaculata* and *A. fulica*, *A. immaculata* and *B.*
753 *glabrata*. (E) Comparison of the catalog of Hox genes. *A. immaculata* and *A. fulica*
754 have more Hox genes than the other molloscus, because of the retained genes on the
755 duplicated chromosome. These evidences collectively support a whole genome

756 duplication event shared by *A. immaculata* and *A. fulica*, which is not found in other
757 mollusks.

758 **Figure 4. Chromosome relationship and karyotype inferring.** Clustering of
759 chromosomes for *A. immaculata* **(a)** and *A. fulica* **(b)**. Each group corresponding to an
760 orthologous chromosome pair derived from WGD, with the links between
761 chromosome pairs representing the mutual best-hit gene pairs. **(c)** Evolution of
762 karyotype in mollusks. The common ancestor of Gastropoda and Bivalvia is estimated
763 to have 8 monoploid chromosomes, while that of Gastropoda has 14. Chromosome
764 breaks, fusions, as well as chromosome number doubling by WGD, formed the
765 current karyotype of living mollusks.

766 **Figure 5. Oxygen transport and anti-oxidization in the respiratory.** **(a)** The
767 phylogenetic relationship of hemocyanin genes among mollusks. The protein IDs
768 covered with different colors meant different species. **(b)** The phylogenetic
769 relationship of zinc metalloproteinases among gastropoda, with same style to that of
770 **(a)**. **(c)** Comparison of hemocyanin locations in the chromosomes of *A. immaculata*, *A.*
771 *fulica* and *P. canaliculata*. Hemocyanin genes of *A. immaculata* and *A. fulica* were
772 distributed in two chromosomes, whereas *P. canaliculata* was only in one contig. **(d)**
773 The evolution process of zinc metalloproteinase genes in *A. immaculata* and *A. fulica*.
774 The shared ancient ancestor might have only one zinc metalloproteinase gene located
775 on one chromosome, WGD (~70 MYA) event doubles the chromosome and gene
776 number, followed by gene expansion on one chromosome and deletion on the
777 corresponding chromosome, resulting in a tandem-cluster of 11 genes and 9 genes in

778 *A. immaculata* and *A. fulica*, respectively.

779 **Figure 6. Flowchart of glucose homeostasis and urea-cycle in aestivation of *A.***

780 *immaculata*. (a) The key processes and related gene expression change of
781 gluconeogenesis and TCA cycle in aestivation compared with normal group. The left
782 part represented the major genes participated in gluconeogenesis, while the right part
783 represented the major genes participated in TCA cycle. Two types of arrows were
784 used, double-line arrow meant a number of steps, and single-line arrow meant one
785 step. Two shapes were used, rectangle meant production, and ellipse meant enzymes,
786 with color red and blue representing up-regulated and down-regulated respectively.
787 Ellipse with symbol “*” meant this enzyme had two genes derived from WGD. (b)
788 The key processes and related gene expression change of ornithine-urea cycle in
789 aestivation compared with normal groups, the style is same to that of (a).

790 **Figure 7. The immune repertoire of *A. immaculata* and the expansion of OST. (a)**

791 The immunity of *A. immaculata* is presented in mucus, pattern recognition receptors,
792 and soluble effectors. The mucus system covered epidermis, mainly including
793 archacin, mucin and acharan sulfate secreted through the NDST pathway, which were
794 presented with green ellipse. The pattern recognition system on the cytomembrane
795 (orange color) contained PGRP, lectin, TFNR, TLR, FREP, GGBP, SR, TEP, and
796 secreted soluble effector (red ellipse) in the downstream immune cascading. (b) The
797 expansion of OST genes. Phylogenetic analyses were performed using MEGA7
798 through maximum likelihood. The OST genes from different species were indicated in
799 circus filled with different colors.

800 **Methods**

801 **Samples collection and genome sequencing**

802 Adults of *A. immaculata* were collected from a local farm in Yangjiang,
803 Guangdong province, China, and maintained at 25 ± 2 °C for a week before
804 processing. Genomic DNA was extracted from the foot muscles of a single snail for
805 constructing PCR-free Illumina 350-bp insert libraries and PacBio 20-kb insert library,
806 and sequenced on Illumina HiSeq-X and PacBio SMRT platforms, respectively. The
807 Hi-C library was prepared using the muscle tissue of another single snail by the
808 following methods: Nuclear DNA was cross-linked in situ, extracted, and then
809 digested with a restriction enzyme. The sticky ends of the digested fragments were
810 biotinylated, diluted, and then ligated to each other randomly. Biotinylated DNA
811 fragments were enriched and sheared again for preparing the sequencing library,
812 which was then sequenced on a HiSeq-X platform (Illumina).

813 **RNA sample preparation and transcriptome sequencing**

814 Ten tissues including brains, eggs (2 days post fertilization), eyes, hemocytes,
815 hepatopancreas, kidneys, lungs, muscles, ovaries and testes from six animals were
816 collected as replicates. Eighty snails were employed for the immune elicitor challenge
817 experiment, and they were equally divided into control group, LPS
818 (lipopolysaccharide) group, PGN (peptidoglycan) group, GLU (β -glucan) group and
819 IC (poly I:C) group. The five groups of snails received injections of 100 μ L

phosphatic buffer solution (PBS, 0.14 M NaCl, 3 mM KCl, 8 mM NaH₂PO₄·12H₂O, 1.5 mM K₂HPO₄, pH7.4), LPS from *Escherichia coli* 0111:B4 (Sigma-Aldrich, 0.5 mg·ml⁻¹ in PBS), PGN from *Staphylococcus aureus* (Sigma-Aldrich, 0.8 mg·ml⁻¹ in PBS), GLU from *Saccharomyces cerevisiae* (Sigma-Aldrich, 1.0 mg·ml⁻¹ in PBS), and poly I:C (Sigma-Aldrich, 1.0 mg·ml⁻¹ in PBS), respectively. These treated snails were maintained after injection, and fifteen individuals from each group were randomly sampled at 12 h post-injection. Non-aestivation snails were feeding with enough food and water, whereas aestivation snails were fasting and treated with high temperature, and hemolymph of these two groups were collected around 10 days after the epiphragma formation. Hemolymph samples of all the treated group were collected from five individuals were pooled into one sample. There were three replicates for each sample. In final, total RNAs were extracted from the stored tissues, and then mRNAs were pulled out by beads with poly-T for constructing cDNA libraries (insert 350-bp), and sequenced on an Illumina HiSeq-X sequencer.

Genome assembly

The Illumina raw reads were filtered by trimming the adapter sequence and low-quality regions (https://github.com/fanagislab/assembly_2ndGeneration/tree/master/clean_illumina), resulting in clean and high-quality reads with an average error rate < 0.001. For the PacBio raw data, the short subreads (< 2 kb) and low-quality (error rate > 0.2) subreads were filtered out, and only one representative subread was retained for each

PacBio read. The clean PacBio reads were corrected by Canu 1.8⁴² and then assembled by Wtdbg2⁴³. The PacBio reads were employed to polish the raw contigs by a module within Wtdbg2, after which Illumina reads were aligned to the contigs by BWA-MEM, and single base errors in the contigs were corrected by Pilon 2.10⁴⁴ with the parameters “-fix bases, -nonpf, -minqual 20”. Next, Hi-C sequencing data were aligned to the haploid reference contigs by HiC-Pro 2.11.1⁸⁵, and then these contigs were clustered, ordered, and orientated into chromosomes with LACH-ESIS⁸⁶.

Genome annotation

A de novo repeat library for *A. immaculata* was constructed by RepeatModeler (v1.0.4; <http://www.repeatmasker.org/RepeatModeler.html>). TEs in the *A. immaculata* genome were identified by RepeatMasker (v4.0.6; <http://www.repeatmasker.org/>) using both the Repbase library and the de novo library. Tandem repeats in the *A. immaculata* genome were predicted using Tandem Repeats Finder v4.07b⁸⁷. The divergence rates of TEs were calculated between the identified TE elements in the genome and their consensus sequence at the TE family level.

The gene models in the *A. immaculata* genome were predicted by EVM v1.1.1⁸⁸ integrating evidences from ab initio predictions, homology-based searches and RNA-seq alignments. Then, these gene models were annotated by RNA-seq data, UniProt database and InterProScan software⁸⁹. Finally, the gene models were retained if they had at least one piece of supporting evidence from the UniProt database, InterProScan domain and RNA-seq data. Gene functional annotation was performed

by aligning the protein sequences to the NCBI NR, UniProt, COG and KEGG databases with BLASTP v2.3.0+ under an E-value cutoff of 10^{-5} and choosing the best hit. Pathway analysis and functional classification were conducted based on the KEGG database⁹⁰. InterProScan was used to assign IPR domains and GO terms to the gene models.

Evolutionary analysis

Orthologous and paralogous groups were assigned from 11 species (*A. immaculata*, *A. fulica*, *Aplysia californica*, *Biomphalaria glabrata*, *Pomacea canaliculata*, *Lottia gigantea*, *Crassostrea gigas*, *Pintada fucata*, *Limnoperna fortunei*, *Octopus bimaculoides* and *Lingula anatina*) by OrthoFinder⁹¹ with default parameters. OGs that contained only one gene for each species were selected to construct the phylogenetic tree. The protein sequences of each gene family were independently aligned by muscle v3.8.31⁹² and then concatenated into one super-sequence. The phylogenetic tree was constructed by maximum likelihood (ML) using RAxML 8.2.12⁴⁵, with the best-fit model (LG+IGF) estimated by ProtTest3⁹³. The absolute rates of molecular evolution and divergence times were estimated by r8s⁴⁶. The tree was calibrated with the following time frames to constrain the age of the nodes between the species: minimum = 260 Ma and maximum = 290 MYA for *P. fucata* and *C. gigas*⁹⁴; minimum = 450 MYA and maximum = 480 MYA for *A. californica* (or *B. glabrata*) and *L. gigantea*⁹⁵. The calibration time (fossil record time) interval (550-610 MYA) of *O. bimaculoides* was adopted from previous results⁹⁶.

883 **Transcriptome data analysis**

884 Transcriptome reads were trimmed with the same method for genomic reads
 885 (https://github.com/fanagislab/assembly_2ndGeneration/tree/master/clean_illumina),
 886 and then mapped to the reference genome of *A. immaculata* using TopHat (v. 2.1.0)
 887 with default settings. The expression level of each reference gene in terms of FPKM
 888 was computed by cufflinks v2.2.1. A gene was considered to be expressed if the
 889 FPKM > 0. Differential gene expression analysis was conducted using cuffdiff v2.2.1.

890 **WGD verification**

891 The analyses incorporating chromosome-level macrosynteny analysis, colinearity
 892 blocks, Ks peak and Hox gene clusters were employed to verify the WGD in giant
 893 African snails. Based on 4 chromosome-level molluscan genomes, the macrosynteny
 894 was identified using homologous gene sets. Genes from different species would be
 895 identified as homologous gene pairs when they had mutual best BLASTP hits with
 896 each other. The conserved macrosynteny between species with chromosome-level
 897 assemblies was displayed in dot plot. Each dot in the dot plot comparison represents a
 898 one-to-one homologous gene pair mentioned above. Based on the dot plot, we
 899 inferred the circus plot and dual synteny plot. Then, MCScanX was also used with
 900 default parameters to identify the colinearity blocks in *A. immaculata* and *A. fulica*⁹⁷.
 901 Ks distribution of gene pairs in colinearity blocks was calculated by ParaAT⁹⁸ and
 902 KaKs_calculator 2.0⁹⁹. The homeobox genes were identified in the giant African
 903 snails using BLAST ($E < e^{-5}$) against all homeodomain sequences from the

904 HomeoDB database (<http://homeodb.zoo.ox.ac.uk/>), and were further confirmed by
905 comparing to the Conserved Domains Database (<http://www.ncbi.nlm.nih.gov/cdd>).

906 **Gene family analysis**

907 To identify gene families involved in pathways of respiration, aestivation and
908 immune functions, we performed manual curation for identification of homologous
909 genes by three steps. Initially, we aligned known genes of other close species to the *A.*
910 *immaculata* genome by BLASTP with best hits ($E < e^{-5}$), and followed by the
911 analysis of paralogous genes performed by OrthoFinder. Then, the obtained genes
912 were used to perform phylogeny analysis by maximum likelihood (ML) method with
913 MEGA7¹⁰⁰, to further validate the accuracy and reveal the phylogenetic relationship
914 of these genes.

915 **Data availability**

916 Data relating with the findings of this work are available within the paper and the
917 Supplementary Information files. A reporting summary for this Article is available as
918 a Supplementary Information file. Source data are provided as a Source Data file. All
919 the raw sequencing data generated during this study have been deposited at NCBI as a
920 BioProject under accession PRJNA561271. Genomic and transcriptome sequence
921 reads have been deposited in the SRA database with BioSample: SAMN12612888.
922 The Whole Genome Shotgun project of *A. immaculata* has been deposited at
923 DDBJ/ENA/GenBank under the accession WNKJ000000000. The version described in

924 this paper is version WNKJ00000000. The genome assemblies and annotation files
925 are available at the website
926 ftp://ftp.agis.org.cn/~fanwei/Achatina_immaculata_genome/.

927 **Code availability**

928 The in-house software clean_adapter [[https://github.com/fanagislab/](https://github.com/fanagislab/assembly_2ndGeneration/blob/master/clean_illumina/clean_adapter)
929 [assembly_2ndGeneration/blob/master/clean_illumina/clean_adapter](https://github.com/fanagislab/assembly_2ndGeneration/blob/master/clean_illumina/clean_adapter)] and
930 clean_lowqual
931 [[https://github.com/fanagislab/assembly_2ndGeneration/blob/master/clean_illumina/](https://github.com/fanagislab/assembly_2ndGeneration/blob/master/clean_illumina/clean_lowqual)
932 [clean_lowqual](https://github.com/fanagislab/assembly_2ndGeneration/blob/master/clean_illumina/clean_lowqual)] are used to filter the adapter and low-quality sequence.

933 **Methods references**

- 934 85 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data
935 processing. *Genome Biol.* **16**, 259 (2015).
- 936 86 Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome
937 assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119 (2013).
- 938 87 Benson, G. Tandem repeats finder: a program to analyze DNA sequences.
939 *Nucleic Acids Res.* **27**, 573-580 (1999).
- 940 88 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using
941 EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome*
942 *Biol.* **9**, R7, doi:10.1186/gb-2008-9-1-r7 (2008).
- 943 89 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.

944 *Bioinformatics* **30**, 1236-1240 (2014).

945 90 Aoki, K. F. & Kanehisa, M. Using the KEGG database resource. *Current*

946 *protocols in bioinformatics* **11**, 1.12. 11-11.12. 54 (2005).

947 91 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole

948 genome comparisons dramatically improves orthogroup inference accuracy.

949 *Genome Biol.* **16**, 157 (2015).

950 92 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and

951 high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).

952 93 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection

953 of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165 (2011).

954 94 Sun, J. *et al.* Adaptation to deep-sea chemosynthetic environments as revealed

955 by mussel genomes. *Nat. Ecol. Evol.* **1**, 0121 (2017).

956 95 Benton, M., Donoghue, P. & Asher, R. *The Timetree of Life, chapter*

957 *Calibrating and constraining molecular clocks.* (Oxford University Press,

958 2009).

959 96 Zapata, F. *et al.* Phylogenomic analyses of deep gastropod relationships reject

960 Orthogastropoda. *P. Roy. Soc. B Biol. Sci* **281**, 20141739 (2014).

961 97 Wang, Y. *et al.* MCSScanX: a toolkit for detection and evolutionary analysis of

962 gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49-e49 (2012).

963 98 Zhang, Z. *et al.* ParaAT: a parallel tool for constructing multiple

964 protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **419**,

965 779-781 (2012).

966 99 Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: a
967 toolkit incorporating gamma-series methods and sliding window strategies.
968 *Genom. Proteom. Bioinf.* **8** 77-80 (2010).
969 100 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary
970 Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**,
971 1870-1874 (2016).
972

973 **Acknowledgements**

974 The work was funded by the National key research and development program of
975 China (2016YFC1200600), the Agricultural Science and Technology Innovation
976 Program && The Elite Young Scientists Program of CAAS, Fundamental Research
977 Funds for Central Non-profit Scientific Institution (No.Y2017JC01), the Agricultural
978 Science and Technology Innovation Program Cooperation and Innovation Mission
979 (CAAS-XTCX2016), Fund of Key Laboratory of Shenzhen
980 (ZDSYS20141118170111640), as well as National Natural Science Foundation of
981 China (Grant Nos. 31801804).

982 **Author contributions**

983 W. F. and W. Q. Q. conceived and led the project; C. H. L. and Y. W. R. prepared
984 DNA and RNA for sequencing; C. H. L. performed genome assembly, annotation,
985 evolution, whole genome duplication, and immune analysis; Y.W.R. performed

986 respiration and aestivation analysis; W. F., W. Q. Q., C. H. L. and Y. W. R. wrote and
987 revised the manuscript.

988 **Competing interests**

989 The authors declare no competing interests.

990

a

b

c

d

e

f

g