



OPEN


Pulmonate slug evolution is reflected in the de novo genome of *Arion vulgaris* Moquin-Tandon, 1855

Zeyuan Chen^{1,2}, Özgül Doğan³, Nadège Guiguelmoni⁴, Anne Guichard^{5,6} & Michael Schrödl^{1,2,7}

Stylommatophoran pulmonate land slugs and snails successfully completed the water-to-land transition from an aquatic ancestor and flourished on land. Of the 30,000 estimated species, very few genomes have so far been published. Here, we assembled and characterized a chromosome-level genome of the “Spanish” slug, *Arion vulgaris* Moquin-Tandon, 1855, a notorious pest land slug in Europe. Using this reference genome, we conclude that a whole-genome duplication event occurred approximately 93–109 Mya at the base of Stylommatophora and might have promoted land invasion and adaptive radiation. Comparative genomic analyses reveal that genes related to the development of kidney, blood vessels, muscle, and nervous systems had expanded in the last common ancestor of land pulmonates, likely an evolutionary response to the terrestrial challenges of gravity and water loss. Analyses of *A. vulgaris* gene families and positively selected genes show the slug has evolved a stronger ability to counteract the greater threats of external damage, radiation, and water loss lacking a protective shell. Furthermore, a recent burst of long interspersed elements in the genome of *A. vulgaris* might affect gene regulation and contribute to rapid phenotype changes in *A. vulgaris*, which might be conducive to its rapid adaptation and invasiveness.

Land slugs and snails (Mollusca: Gastropoda), which are often abundant in gardens, forests, fields, and orchards, are, for the most part, classified as stylommatophoran pulmonates. They have radiated into about 30,000 species, have highly successfully colonized habitats from polar regions to the tropics, and some are well-known invasive species or pests across the world^{1–5}. Stylommatophoran pulmonates are among the few representatives of mollusks that have colonized the terrestrial environment. The changes in the physical and chemical properties of the environment are immense for animals moving from aquatic to terrestrial environments, and these changes could affect all possible life processes, from respiration and excretion to methods of movement, the functioning of sense organs, and reproduction⁶. Overcoming drought, for example, is one of the biggest challenges in water-land transition⁷. Compared to land snails, the lack of a protective shell in land slugs seems to have further increased the difficulty in coping with external stimuli, predators, sun exposure, and drought. Land slugs have evolved certain innovations, such as defense by chemical compounds or behavior, to counteract these challenges^{1,8}. However, the lack of shell also gives advantages such as reduced weight and lower energy costs, reduced dependence on calcium uptake, better mobility, and ability to occupy small spaces. Recently, comparative genomics methods have provided key perspectives for revealing the process of water-land transition and illuminated adaptive mechanisms^{9–11}. With the rapid development of genome sequencing, several land snail genomes have been published (Supplementary data 1), however, the genomic resources for land slugs are still lacking.

In recent years, the notorious “Spanish” slug, *Arion vulgaris* Moquin-Tandon, 1855, has attracted widespread attention due to its invasiveness and negative impact on the economy, ecology, health, and social system¹². As a major defoliator of plants, *A. vulgaris* causes serious damage in orchard cultivation, gardens, and agriculture resulting in financial losses^{13–15}. *Arion vulgaris* also transmits plant pathogens, contaminates silage, and might

¹SNSB-Bavarian State Collection of Zoology, Münchhausenstr. 21, 81247 Munich, Germany. ²Department Biology II, Ludwig-Maximilians-Universität, Planegg-Martinsried, 82152 Munich, Germany. ³Department of Molecular Biology and Genetics, Faculty of Science, Sivas Cumhuriyet University, Sivas, Turkey. ⁴Evolutionary Biology and Ecology, Université Libre de Bruxelles, 1050 Brussels, Belgium. ⁵INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, 35650 Le Rheu, France. ⁶Univ. Rennes, CNRS, Inria, IRISA-UMR 6074, 35000 Rennes, France. ⁷GeoBio-Center LMU, 80333 Munich, Germany. email: Chen.Z@snsb.de

cause health problems in animals^{16,17}. It also outcompetes native slug species and reduces biodiversity¹⁸. Delivering Alien Invasive Species Inventoried for Europe (DAISIE) has listed *A. vulgaris* as one of the 100 worst alien species in Europe¹⁹, and it is the only land gastropod in the list. Although recent studies disputed the origin and invasiveness of *A. vulgaris* based on the genetic diversity patterns of mitochondrial and nuclear loci^{20–22}, its outstanding adaptability and mass occurrences are undeniable.

Here, we assembled and annotated the first land slug genome—*A. vulgaris*. By comparing *A. vulgaris* with two stylommatophoran land snails, and stylommatophoran species with other aquatic or marine gastropods, the well-annotated genome provides a broader perspective to decipher the water-land transition process of stylommatophoran species. The *A. vulgaris* genome also provides insights into how shell-less *A. vulgaris* adapted to terrestrial environments and the underlying molecular mechanisms (e.g., whole-genome duplication (WGD), small scale gene duplication, transposable elements explosion). Moreover, the high-quality genome provides an important reference for future research on *A. vulgaris* population genetics and mollusk evolutionary trajectories, e.g., the loss and evolution of mollusk shells²³.

Results

Arion vulgaris genome assembly and annotation. The genome size of *A. vulgaris* (Fig. 1a) is estimated to be around 1.45 Gb from *k*-mer analysis with short reads (Supplementary Fig. S1; Supplementary Table S1). We sequenced 75 Gb (52x) of long reads (mean length 19.39 kb, N50 length 25.80 kb) using Oxford Nanopore sequencing technology to produce a draft genome assembly. The draft assembly was polished using a combination of 57 Gb (40x) Illumina short reads and 138 Gb (95x) 10X Genomics linked reads. Next, the polished assembly was scaffolded using linked reads and then improved into a chromosome-level assembly with 135 Gb (93x) Hi-C data (Supplementary Table S1). Finally, we obtained an assembly with a total length of 1.54 Gb, a contig N50 of 8.6 Mb, and a scaffold N50 of 63.3 Mb, and with 93.8% of the sequences anchored onto 26 scaffolds (Supplementary Fig. S1–2; Supplementary Table S2). The number of chromosome-scale scaffolds is consistent with the species' determined chromosome number based on karyotype studies²⁴. We assessed the quality of the genome assembly in three aspects: (1) more than 95.99% of the Illumina short reads could be mapped to the assembly; (2) a total of 886 (90.59%) conserved genes in BUSCO's metazoan (odb9) benchmark set²⁵ were present and complete in the genome (Supplementary Table S2); (3) the *k*-mer distribution showed a relatively collapsed assembly including mostly single copies of the homozygous content and a partial representation of the heterozygous content, as is expected in a haploid assembly²⁶ (Supplementary Fig. S3). These results all suggested a high-quality genomic resource of this initial genome assembly of *A. vulgaris*, which is comparable to other mollusk genomes, especially in a high level of heterozygosity and repeats content (Supplementary Fig. S4; Supplementary data 1). Gene annotation combining the evidence from transcripts, homologous proteins, and ab initio prediction revealed 32,518 predicted genes with an average length of 15,429 bp (Supplementary Table S4–5). The length distribution of transcripts, coding sequences, exons and introns, and the distribution of exon numbers per gene were comparable to that of other gastropods (Supplementary Fig. S5). Among the predicted protein-coding genes, 97.6% could be annotated through at least one of the following protein-related databases: the EggNOG²⁷ database (51.64%), the Swiss-Prot²⁸ protein database (97.57%), the Translated European Molecular Biology Laboratory (TrEMBL)²⁸ database (96.65%), the protein families (Pfam)²⁹ database (81.55%), and the Kyoto Encyclopedia of Genes and Genomes (KEGG)³⁰ database (29.65%) (Supplementary Table S6).

Phylogenetic relationships within gastropod lineages. The relationship of early gastropods has been controversial for a long time as different datasets and methodology show different topologies^{31–33}. By means of comparing whole genomic data, a total of 223 single-copy orthologous genes (158,094 amino acid sites) were identified from 14 gastropod species that cover five main gastropod subclasses and 2 bivalve species (Supplementary Table S3). Both concatenated and coalescent-based methods produced an identical strongly supported topology (bootstrap value = 100, posterior probabilities = 1), except for the position of the Patellogastropoda and the Vetigastropoda + Neomphalina clades (Fig. 2a; Supplementary Fig. S6a). Our results show Patellogastropoda as sister to all other gastropods, and monophyletic Vetigastropoda + Neomphalina as sister to the clade Apogastropoda (Heterobranchia + Caenogastropoda) with relatively higher support compared with the other two topologies: Patellogastropoda as sister to Vetigastropoda + Neomphalina, and Patellogastropoda as sister to Heterobranchia + Caenogastropoda (Fig. 2a; Supplementary Fig. S6). The results thus favor the hypothesis of a clade Orthogastropoda (the united clade of Heterobranchia, Caenogastropoda, Vetigastropoda, and Neritimorpha), which is congruent with morphology-based and recently reported mitogenomic phylogenies^{32,34–36}; but see Chen and Schrödl³⁷.

Molecular dating suggests that *A. vulgaris* diverged from the most recent common ancestor with the land snails *Lissachatina (Achatina) fulica* and *Li. immaculata* about 126 million years ago (Mya, 95% confidence interval: 92–159 Mya) (Fig. 2a). The estimated divergence time is close to a previous estimate (132 Mya) based on mitochondrial genomes³⁸. Stylommatophora split from Hygrophila around 199 Mya (95% confidence interval: 159–228 Mya), and Panpulmonata split from Sacoglossa around 235 Mya (95% confidence interval: 191–260 Mya) (Fig. 2a).

Analysis of gene family evolution provides insights into *A. vulgaris* terrestrial adaptation. Recently, the changes of gene families have been recognized as a primary driver of phenotypic diversity and adaptive evolution⁹. Hence, we investigate the genetic basis of species adaptive evolution by defining the relationship of gene families. Based on pairwise sequence similarities, we identified 26,693 putative orthologous gene families composed of 378,381 genes among *A. vulgaris*, other gastropods, and outgroup species, of which 1610 gene clusters were shared by all gastropod species, representing ancestral gastropod gene families (Fig. 2b;

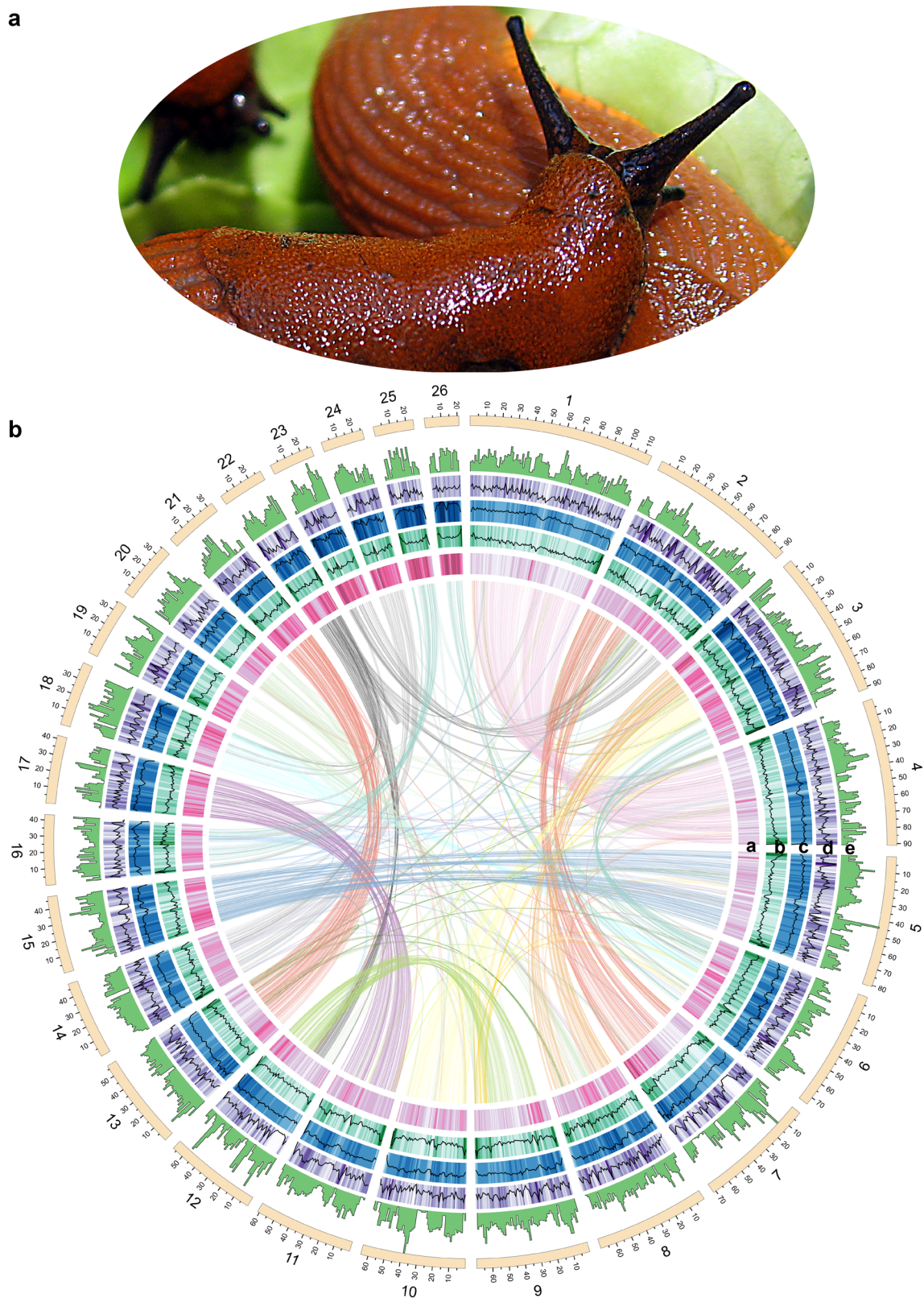
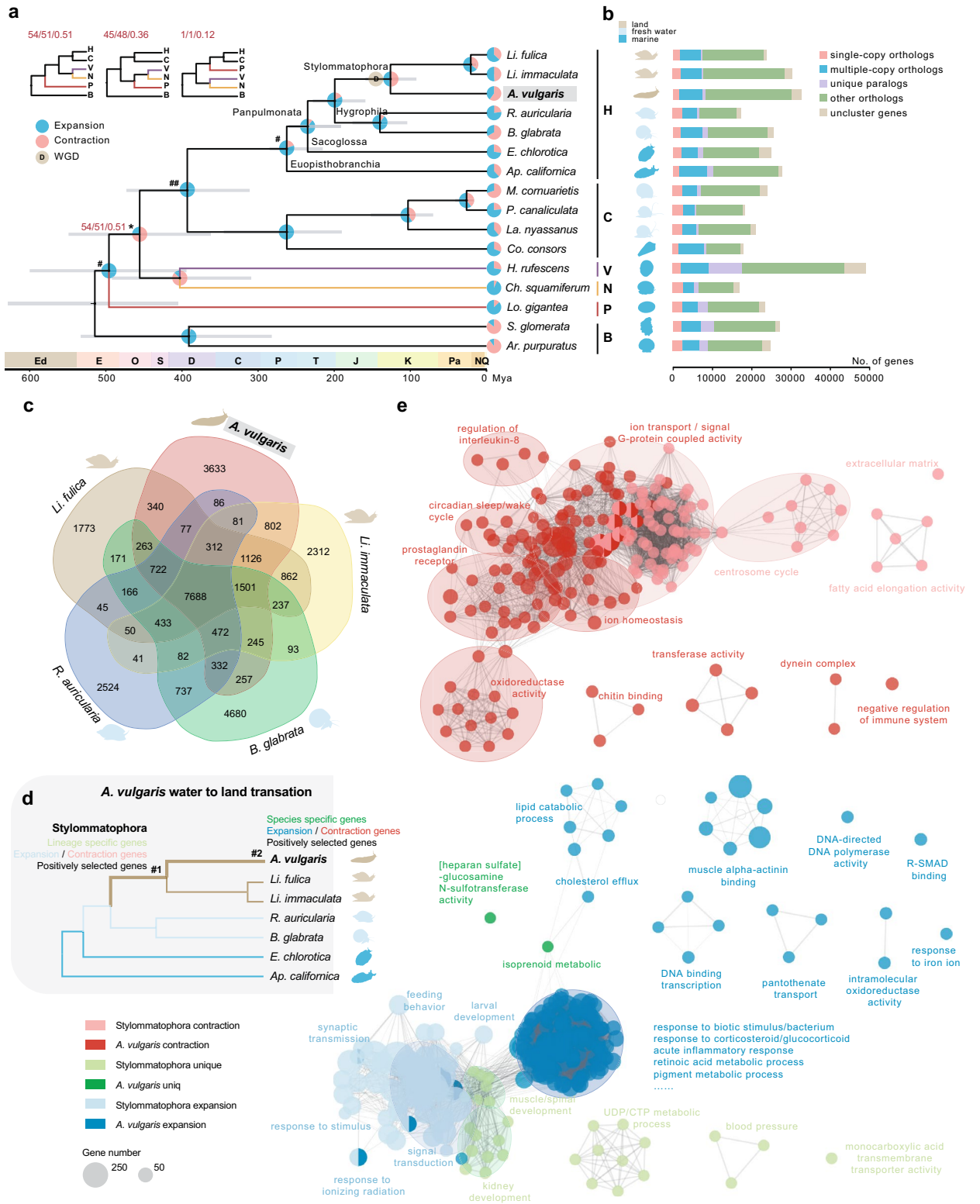


Figure 1. Genome features of *Arion vulgaris*. (a) Adult *A. vulgaris*. (b) General characteristics of the *A. vulgaris* genome. Tracks from inside to outside correspond to (a) GC content, (b) LTRs density, (c) TEs density, (d) genes density, and (e) heterozygosity in sliding windows of 1 Mb across each of the 26 pseudo-chromosomes. Inner lines connect syntenic genes due to ancestral whole-genome duplication events.



◀ **Figure 2.** Phylogeny within gastropod lineages and gene family evolution. (a) Dated Maximum-likelihood phylogenetic tree among gastropod species using 223 single-copy orthologous genes. Grey lines indicate the 95% confidence intervals for the time of divergence between different clades. “#” indicates that nodes are constrained with fossil calibration and “##” indicate node is constrained with secondary calibration data. All nodes have bootstrap support values of 100 and posterior probabilities of 1 in all analyses, except the node with an asterisk. Three alternative topologies are shown on top left. Red numbers are bootstrap support percentages and posterior probabilities (from left to right, inferred by RaxML with GTR+L, IQ-TREE and ASTRAL respectively). H, Heterobranchia; C, Caenogastropoda; V, Vetigastropoda; N, Neomphalina; P, Patellogastropoda; B, Bivalve. The pie diagram on each branch of the tree represents the proportion of gene families undergoing expansion (blue) or contraction (red) events. (b) The distribution of single-copy, multiple-copy, unique, and other orthologs in the 16 mollusks. (c) Venn diagram represents the number of shared and unique gene families among five Panpulmonata species. (d) A simplified diagram showing the evolution of *A. vulgaris* water-land transition. (e) Gene Ontology (GO) enrichment map summarizing major biological networks of *Arion vulgaris* and Stylommatophora specific, expanded and contraction genes. Each node represents one GO term with adjusting P-value < 0.05. Node sizes indicate the number of genes within the corresponding GO term. The thickness of the edges represents the number of genes shared by two terms. Striking groups were manually circled and labeled.

Supplementary Table S7). A total of 10,311 orthologous gene families were shared by all Heterobranchia species and 7688 orthologous gene families were shared by five Panpulmonata species (Fig. 2b, c; Supplementary Table S7).

To explore the genetic basis of terrestrial adaptability shared by stylommatophoran species, we considered the properties of the 1126 gene families exclusively shared by three stylommatophoran species (Fig. 2c, d). GO enrichment analyses of these lineage-specific genes demonstrated that they were mainly assigned to kidney development, CTP/UDP metabolic processes, regulation of blood pressure, muscle growth, and spinal development (Fig. 2e; Supplementary Fig. S7; Supplementary Table S8). Molluscan kidneys are involved in the secretion of waste and the resorption of metabolites from the urinary fluid³⁹. The enrichment of a series of genes related to kidney and ureteric bud development suggests the improvement of the efficiency of maintaining water balance and nutrients re-absorption in stylommatophoran species (Supplementary Table S8). In addition, the enriched biological process related to blood pressure regulation might be responsible for overcoming the gravity problem during landing⁸. Moreover, the enriched functions of muscle growth and spinal development might also improve the movement and flexibility in terrestrial life (Fig. 2e; Supplementary Fig. S7; Supplementary Table S8). There were 2140 and 1958 gene families expanded and contracted in the Stylommatophora lineage respectively (Fig. 2a). The expanded genes were functionally enriched in response to stimulus, response to radiation, signaling, larval development, and regulation of feeding/eating behavior (Fig. 2e; Supplementary Fig. S8; Supplementary Table S9). Meanwhile, genes related to transmembrane transport, fatty acid elongation, and centrosome cycle were contracted in both *A. vulgaris*, *Li. fulica* and *Li. immaculata* (Fig. 2e; Supplementary Fig. S9; Supplementary Table S10). A total of 251 genes are likely positively selected in Stylommatophora, and their function mainly refers to the regulation of myelination (Supplementary Fig. S10; Supplementary Table S11).

Considering the specific adaptations of shell-less *A. vulgaris* (Fig. 2d), we identified a total of 2763 genes unique to *A. vulgaris*, of which 2629 (95.2%) have known InterPro domains (Supplementary Table S7). We found *A. vulgaris* specific genes were significantly enriched in functional categories related to isoprenoid metabolic process and organelle cell components (Fig. 2e; Supplementary Fig. S11; Supplementary Table S12). In comparison with two *Lissachatina* land snail species, *A. vulgaris* expanded genes exhibited significant enrichment in various aspects, including immune system, response to biotic/radiation stress, excretion, etc., which are very likely beneficial for its land adaptation (Fig. 2e; Supplementary Fig. S12; Supplementary Table S13). Specifically, genes related to response to corticosteroids and glucocorticoid pathways are highly increased. Corticosteroids are involved in a wide range of physiologic systems such as stress response, immune response, and regulation of inflammation⁴⁰, glucocorticoids act primarily on carbohydrate and protein metabolism, and have anti-inflammatory effects^{41,42}. Moreover, processes in acute inflammatory response and regeneration are also enriched. All of these might highly improve the ability of *A. vulgaris* to recover from damage. We found *A. vulgaris* expanded genes were also enriched in response to molecules of bacterial origin and response to lipopolysaccharide, which might improve its ability in response to biotic stress. The enrichment of pigment metabolism processes might advance the ability of shell-less *A. vulgaris* to reduce solar radiation damage. Furthermore, genes related to excretion, uronic acid metabolism, and larval development are expanded in *A. vulgaris*. Surprisingly, we also found an enrichment of genes related to pesticides, which might be the result of interaction with human agricultural activity. Similar to genes contracted in Stylommatophora, a high proportion of contracted genes were functionally related to transmembrane transport processes. In addition, contracted genes that regulate circadian rhythm and oxidase activity are also enriched (Fig. 2e; Supplementary Fig. S13; Supplementary Table S14). Strikingly, we found that genes involved in the positive regulation of interleukin-8 production were enriched in *A. vulgaris* contracted genes, and genes related to Interleukin-3,4,9,10,12,21,23,27,35 were likely positively selected in *A. vulgaris* (Supplementary Fig. S14; Supplementary Table S15). This adaptive immune response again might highly increase the ability of *A. vulgaris* in response to stress and stimuli.

Whole-genome duplication events shared by Stylommatophora species. Whole-genome duplication (WGD) events are proposed to be a key evolutionary event driving phenotypic complexity, functional novelty, and ecological adaptation⁴³. An earlier study suspected a WGD event somewhere at the base of Stylommatophora by comparison of chromosome numbers among closely related mollusks⁴⁴, and a recent genomic study of *Li. immaculata* and *Li. fulica* proved the WGD event using genomic analysis and deduced the WGD

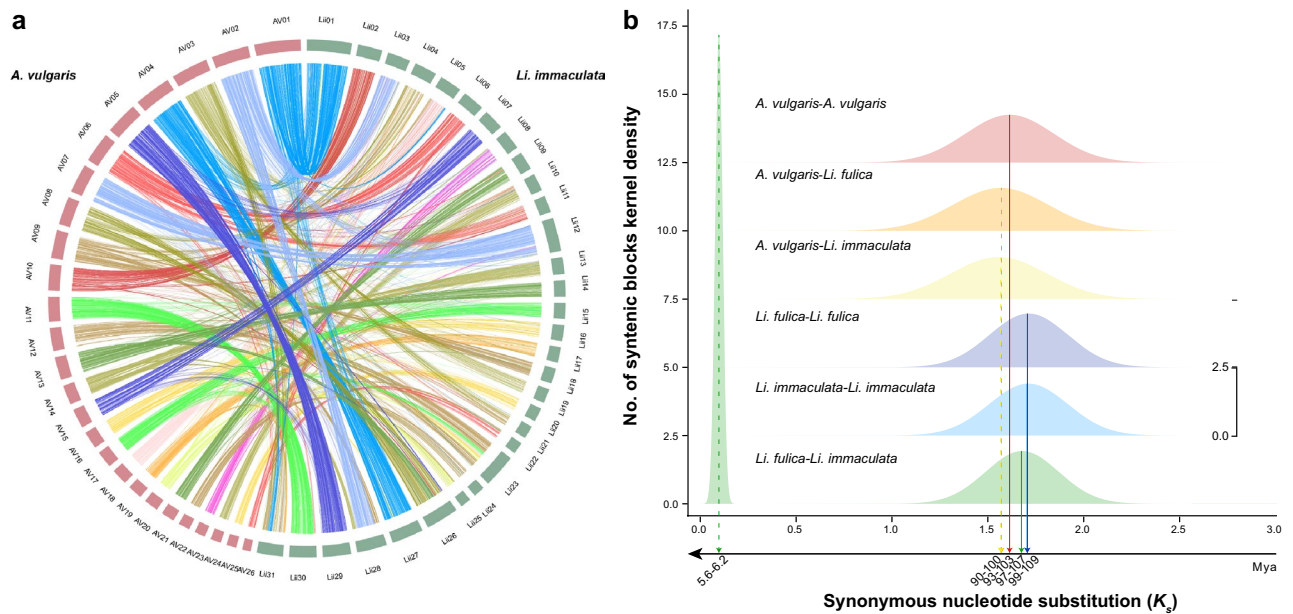


Figure 3. Whole-genome duplication shared by *Arion vulgaris*, *Lissachatina (Achatina) fulica*, and *Li. immaculata*. **(a)** An approximately one to one corresponding relationship in the comparison of *A. vulgaris* and *Li. immaculata* genomes. **(b)** Frequency distributions are shown of values of synonymous substitutions (K_s) for homologous gene pairs in comparisons of *A. vulgaris*, *Li. fulica* and *Li. immaculata*.

event occurred around 70 Mya¹¹. However, 70 Mya is much later than the divergence time that we estimated between *A. vulgaris* and *Lissachatina* (126 Mya, Fig. 2a). Therefore, we raised two questions: (a) whether *A. vulgaris* also experienced a WGD event, and (b) if it has happened, whether it happened independently after divergence from *Lissachatina* or it was shared by their common ancestor.

Our results of chromosome macrosynteny show that most of the chromosomes found a corresponding one in the *A. vulgaris* genome (Fig. 1b). In addition, we detected an approximately one-to-one corresponding relationship in the comparison of *A. vulgaris* ($n = 26$) and *Li. immaculata* ($n = 31$) chromosomes (Fig. 3a) and a one-to-two corresponding relationship in the comparison of *A. vulgaris* and *Aplysia californica* ($n = 17$) chromosomes (Supplementary Fig. S15). In both *Lissachatina* and *A. vulgaris*, the distribution of synonymous substitutions (K_s) shows a clear peak, which represents WGD events. There is also an overall slower synonymous substitution rate in *A. vulgaris* (max K_s : 1.61) than in *Li. fulica* (max K_s : 1.71) and *Li. immaculata* (max K_s : 1.71) (Fig. 3b; Supplementary Fig. S16). Based on our results and the previous karyotype research, we conclude that a WGD event did occur in the ancestry of *A. vulgaris*.

To figure out when the WGD event happened, we further compared the synteny gene pairs between *A. vulgaris* and two *Lissachatina* snails. First, the results showed the best BLASTP hits of homologous gene pairs are from interspecies comparisons instead of intraspecies comparisons (Supplementary Figs. S17a–19a), which implies that the WGD event seems to have occurred before the divergence of *A. vulgaris* and *Lissachatina*. Moreover, the distribution of K_s of *A. vulgaris*-*Li. immaculata* gene pairs and *A. vulgaris*-*Li. fulica* gene pairs show only one peak each, respectively. The K_s values corresponding to the peak are smaller between species (*A. vulgaris*-*Li. fulica*: 1.56; *A. vulgaris*-*Li. immaculata*: 1.57) than within *A. vulgaris* (*A. vulgaris*-*A. vulgaris*: 1.61) (Fig. 3b). This result could be explained by the species differentiation event occurring shortly after the WGD event. Such a short time is reflected in our results as the peak of species differentiation coinciding with the peak of the WGD event in the K_s distribution of *A. vulgaris*-*Li. fulica* and *A. vulgaris*-*Li. immaculata*, and the overall distribution has moved towards small K_s (Fig. 3b). Assuming that the mutation rate of Mollusca is 1.645×10^{-9} per site per year⁴⁵, we estimated the WGD event happened at approximately 93–109 Mya, and the species differentiation of *A. vulgaris* and *Lissachatina* occurred a very short time after the WGD, with molecular dating estimates for this at approximately 90–103 Mya.

After WGD, the two sets of chromosomes evolved differently with one set of chromosomes being more structurally stable and conserved compared to the other (Supplementary Figs. S17–19b), and this imbalance might provide a rich genomic resource for rapid evolution and adaptation⁴⁶. Since the differentiation of Arionoidea and Achatinoidea is almost at the base of Stylommatophora differentiation⁴⁷, we further speculate that all Stylommatophora species shared the common WGD event. The newly generated chromosome set provided abundant evolutionary resources in functional novelty and ecological adaptation, which may have led to the successful territorialization and diversity of Stylommatophora species.

Evolution of gene duplication and adaptability. Gene duplication is another important evolutionary mechanism to provide new genetical material and opportunities to acquire new gene functions for an organism⁴⁸. We found that Heterobranchia species have an abundance of duplicate genes. In our analysis, between 55% (*Ely-*

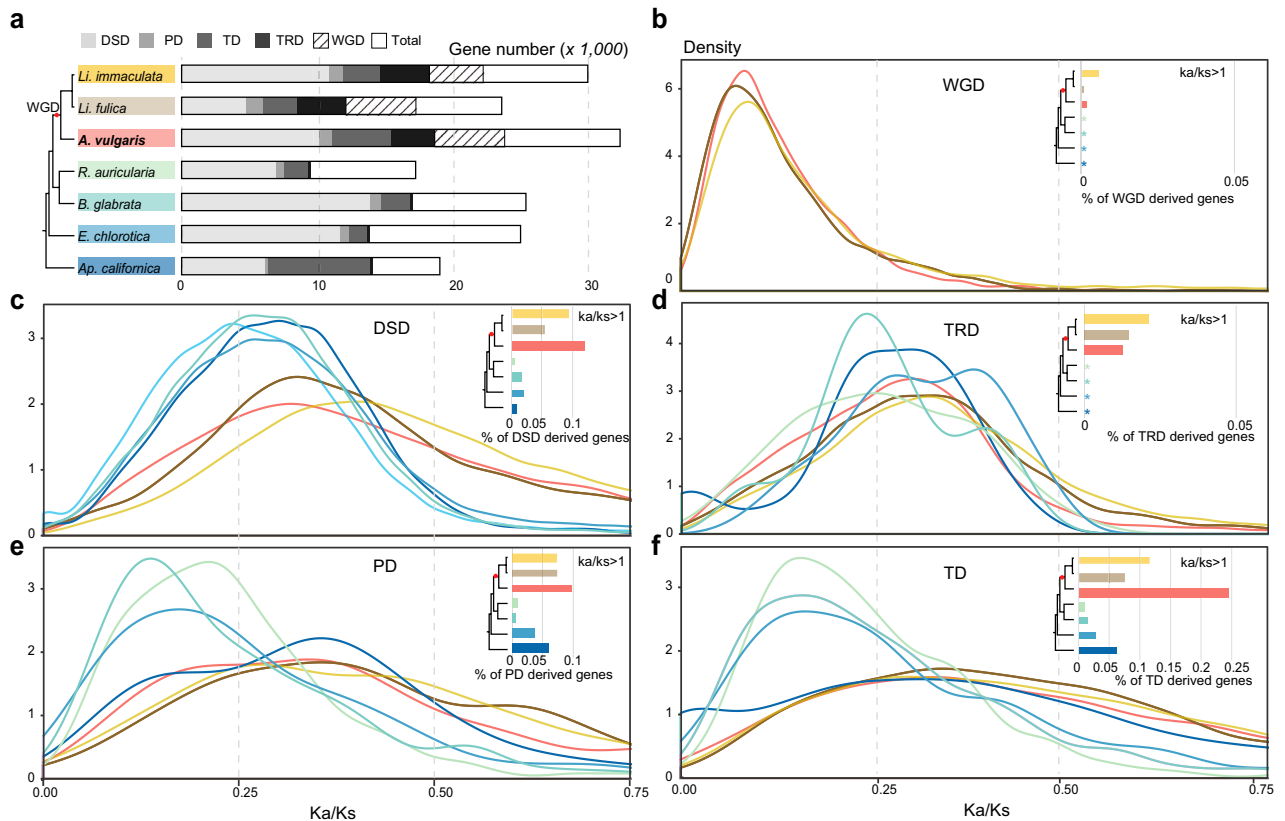


Figure 4. Gene duplication and evolution in Heterobranchia species. **(a)** The number of gene pairs derived from different modes of duplication. **(b–f)** The Ka/Ks ratio distributions of gene pairs derived from different modes of duplication in different species. Bar graphs show the proportion of positively selected genes ($Ka/Ks > 1$) that were duplicated by different mechanisms. Colors represent different species. WGD, whole-genome duplication; TD, tandem duplication; PD, proximal duplication; TRD, transposed duplication; DSD, dispersed duplication; Total, total annotated genes.

sia chlorotica) and 75% (*Ap. californica*, *A. vulgaris*) of genes were identified as paralogous (Fig. 4a). Three Stylommatophora species (*A. vulgaris*, *Li. fulica*, and *Li. immaculata*) have an average of 16.8% more duplicate genes than other species. Among them, WGD events contribute 13% (*Li. immaculata*) to 22% (*Li. fulica*) to existing duplicates, and WGD-derived gene pairs are the most conserved among all types of duplicated genes (Fig. 4a, b; Supplementary Table S16). Another type of duplicate gene that has increased significantly in both Stylommatophora species is transposed duplication (TRD) gene, which is on average 48 times more frequent than in other species (Fig. 4a; Supplementary Table S16). Dispersed duplication (DSD) accounts for a high proportion (mean 56% of all duplicated genes, $SD = 24\%$), while proximal duplication (PD) generates a small proportion (mean 5% of all duplicated genes, $SD = 1.8\%$) of gene copies in all Heterobranchia species. Strikingly, tandem duplication (TD) gene pairs account for the highest proportion in *Ap. californica*, which is about 1.7–5.3 times that of other species (Fig. 4a; Supplementary Table S16).

The evolutionary pattern of duplicated genes is similar within three Stylommatophora species. The overall age of duplicated genes is young and under a weak purifying selection ($Ka/Ks < 1$) in both duplicated modes compared to other species (Fig. 4b–f; Supplementary Fig. S20). For *Ap. californica*, the TD- and PD-derived gene pairs have similar distribution of Ka/Ks (mode: TD-0.30; PD-0.35) between Stylommatophoran species (mean of the modes: TD-0.32; PD-0.36) when compared to other more closely related species (mean of the modes: TD-0.15; PD-0.17) suggesting that tandem and proximal duplicates happened recently and experienced relatively relaxed purifying selection (Fig. 4e, f). However, DSD- and TRD-derived gene pairs are more conserved in *Ap. californica*, which is more similar to *Radix auricularia*, *Biomphalaria glabrata*, and *E. chlorotica* (Fig. 4c, d; Supplementary Fig. S20).

We further explored the roles of positive selection ($Ka/Ks > 1$) in the evolution of duplicated genes in seven Heterobranchia species. As expected, *A. vulgaris*, *Li. fulica* and *Li. immaculata* experienced stronger positive selection than other species, reflected by the high percentages of gene pairs showing $Ka/Ks > 1$ in all kinds of duplicated gene pairs (Fig. 4c, d; Supplementary Table S17). Among all duplicate genes, TD-, PD-, DSD-derived gene pairs have experienced stronger positive selection compared with genes generated by other duplication mechanisms (Fig. 4b–f; Supplementary Table S17). In *A. vulgaris*, 24% TD-derived genes were likely positively selected, which is 2–3 times that of *Li. immaculata* and *Li. fulica*, and 4–29 times that of other species (Fig. 4c, e, f; Supplementary Table S17). Interestingly, we found that the functional enrichment of genes caused by TD in *A. vulgaris* concerns response to external stress, pigment catabolism, and acute inflammatory process, which

echoes the previous enrichment results of *A. vulgaris* unique and expanded genes and is related to its unique adaptation (Supplementary Fig. S21, 22). On the other hand, only 0.3% WGD—which only exists in *A. vulgaris*, *Li. fulica*, and *Li. immaculata*, and 1.6% TRD—which are highly expanded in *A. vulgaris*, *Li. fulica*, and *Li. immaculata*, Fig. 4a) derived gene pairs were likely positively selected in *A. vulgaris* (Fig. 4b, d; Supplementary Table S17). The TRD-derived gene pairs which were functionally enriched mostly refer to cell components (Supplementary Fig. S23), and WGD-derived genes were prone to be enriched in basic biological functions such as signal transduction, ion transport, muscle development (Supplementary Fig. S24).

Massive expansion of transposable elements in *A. vulgaris* genome. Repeat content analysis showed that the repeat sequences occupy approximately 75.09% (1.15 Gb) of the *A. vulgaris* assembly (Supplementary Table S18), which is the highest value among all studied gastropod species²³. We also found that species in Heterobranchia have a higher repeat content than other gastropod groups (i.e., Caenogastropoda, Vetigastropoda, Neomphalina, and Patellogastropoda) (Supplementary Fig. S25). In all types of repetitive sequences, transposable elements (TEs) account for 61.08% of the *A. vulgaris* assembly, and among them, long interspersed elements (LINEs), DNA transposons (DNAs), and short interspersed elements (SINEs) account for 36.39%, 5.44%, 1.78% of the assembly, respectively.

A high proportion of unclassified TEs (17.76%) was also detected in the *A. vulgaris* genome (Fig. 5a; Supplementary Table S18). Overall, the composition of TEs of *A. vulgaris* is similar to *Li. fulica* and *Li. immaculata*, in which LINEs are dominant, whereas in other Heterobranchia species DNA transposons are most abundant (except *B. glabrata*, see below). Most of the LINEs in *A. vulgaris* showed a low divergence rate, indicating a recent explosion of LINEs in the *A. vulgaris* genome (peak % divergence to consensus = 3). However, *Li. fulica* and *Li. immaculata* LINEs were not recent invaders since they exhibit a large divergence from the consensus (the distributions peak at 31% divergence for *Li. fulica* and 33% for *Li. immaculata*) (Fig. 5a). Two freshwater snails (*R. auricularia*, *B. glabrata*) and *Ap. californica* also showed recent expansion of LINEs, which even resulted in LINEs that replaced DNAs and became the dominant TE type in the *B. glabrata* genome (Fig. 5a). We found that although the total TE number of *A. vulgaris* is 1.35–6.09 times greater than in the other species considered, the insertion of TEs was very conservative. Specifically, genes with TEs distributed in putatively functional regions, i.e., 2 kb upstream, 1 kb downstream, or intron, exon regions in *A. vulgaris* were 1.21–1.87 times that of all other species. However, the number of TEs inserted into exons in *A. vulgaris* only accounts for 51% and 66% of that of *Li. fulica* and *B. glabrata*, respectively (Fig. 5b; Supplementary Table S19). Among all species, TEs were mainly inserted into introns in different degrees of divergence from consensus (Fig. 5b; Supplementary Fig. S26). The insertion of *A. vulgaris* TEs in intron regions greatly increased compared to other species (1.29–6.74 times), especially young TEs with a low divergence rate (%divergence to consensus < 16, Supplementary Fig. S26). The insertion into upstream and downstream is also increased, by 1.81–3.72 and 2.06–5.39 times that of other species, respectively. Previous reports have shown that TEs are powerful facilitators of rapid adaptation to novel environments^{49–51}. The recent expansion of LINEs in *A. vulgaris* may also have played an important role in promoting potential plasticity and stress resistance correlated with its invasiveness and competitiveness.

Recent studies showed that TEs have driven massive changes in genome size^{52–54}. In our results, although we found TE coverage is slightly positively correlated with genome size, the correlation is not significant (Supplementary Fig. S27a). In further analyses, we determined that these positive contributions all come from the LINEs (Fig. 5c, Supplementary Fig. S27b–d), but are still not significantly related to genome size. However, *A. vulgaris* and *Lissachatina* have larger genome sizes compared to other species, thus we assume that the changes in the Heterobranchia species genome size might be the result of the expansion of LINEs and the WGD event.

Population dynamics of *A. vulgaris*. We observed an average genome-wide heterozygosity rate of 1.55 per hundred base pairs in *A. vulgaris*, which is about three times of the invasive land snail, *Li. fulica* (0.47 per hundred base pair)⁵⁵, but comparable to freshwater snails *Pomacea canaliculata* (Caenogastropoda, 1.41%) and *P. maculata* (1.22%) which are also notable invaders⁵⁶. We further compared the population dynamic history of *A. vulgaris* with relatively closely related invasive species *Li. fulica* and *B. glabrata*. We found that *A. vulgaris* and *B. glabrata* populations exhibited similar demographic histories, with a high N_e (4×10^5) ~ 1.2 Mya and both increased between 1.2 and 0.8 Mya (Fig. 6). The *A. vulgaris* population continuously declined after the Pre-Pastonian glaciation and dramatically decreased ~ 40,000 years ago, which is consistent with the sharp temperature drop. *Li. fulica* population shows a relatively small N_e (3×10^5) ~ 1.2 Mya and a continuously prolonged decline until ~ 10,000 years ago, and then the population increased significantly to almost the initial level (Fig. 6). The very recent expansion of *Li. fulica* from a relatively small effective population size can also explain that the heterozygosity of *Li. fulica* is much smaller than that of *A. vulgaris*. On the other hand, the relatively long-term large effective population size of *A. vulgaris* may cause the complexity in its population structure, thereby increasing the difficulty of research on population expansion/invasion studies^{22,57,58}.

Discussion

Whole-genome duplication (WGD) is a common phenomenon in plants and has been shown in invertebrate species^{59,60}. It plays an important role in providing evolutionary novelties and promoting speciation^{43,61}. Based on chromosome-level genomic analysis of two land snails, Liu et al.¹¹ first reported the WGD on the Sigmurethra-Orthurethra branch within Stylommatophora at ~ 70 Mya. However, our results indicate that the WGD is most likely an event shared by all Stylommatophora species, which we have dated back to 90–103 Mya (Fig. 3). The inconsistency in timing inference may be caused by the identification of paralogous gene pairs derived by the WGD event. In the study of Liu et al., MScanX⁶² was used with default parameters to identify the collinearity blocks in *Li. immaculata* and *Li. fulica* and the K_s distribution was calculated using the gene pairs in the

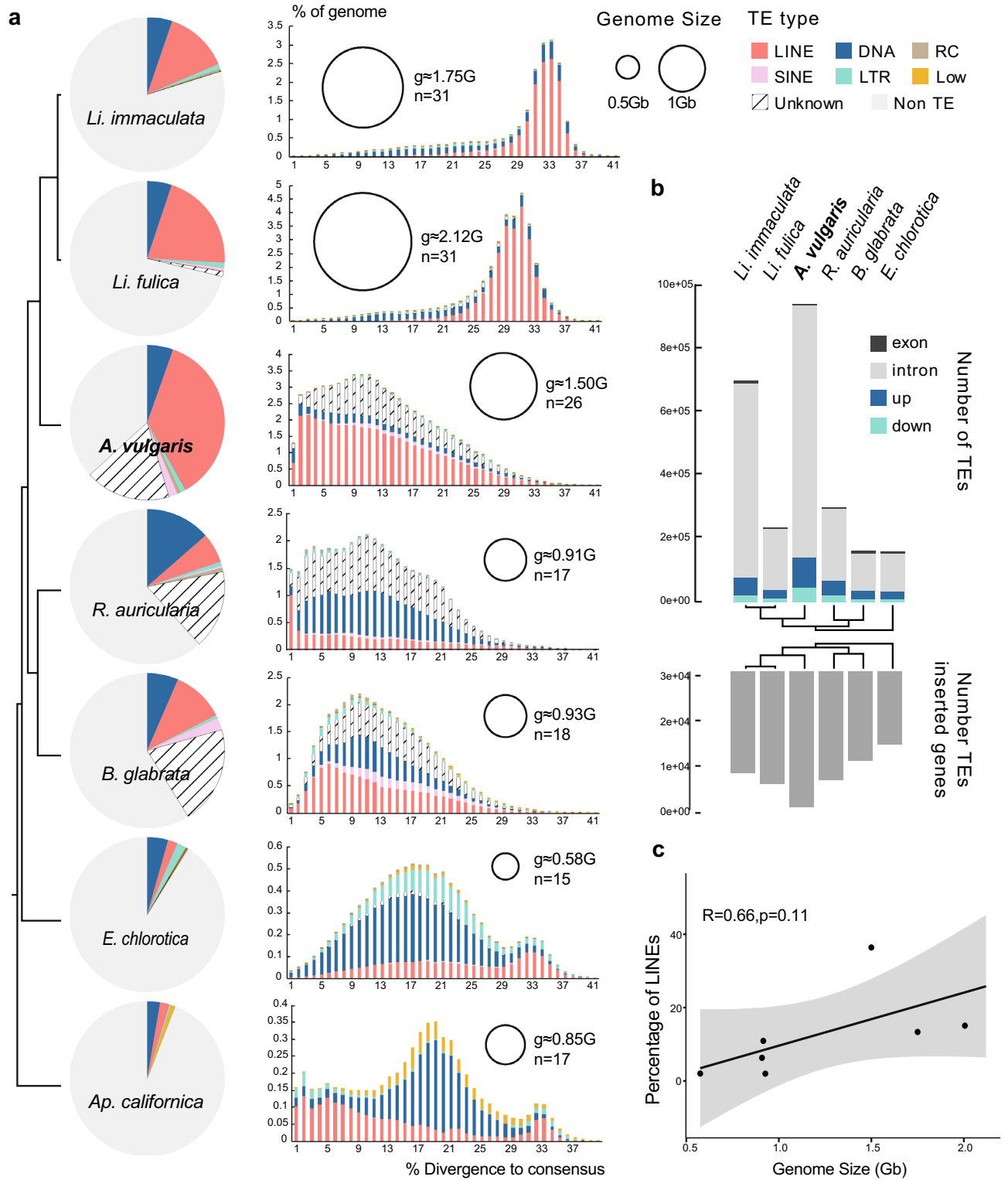


Figure 5. Massive expansion of transposable elements (TEs) in *Arion vulgaris* genome. **(a)** Left: TE composition by class among Heterobranchia species. Right: TE accumulation history corresponding to each species. The size of the circle indicates the estimated genome size of each species. n represents the number of haploid chromosomes; g represents the estimated genome size. **(b)** Bar graph shows the number of TEs in different genic regions and the number of TEs-inserted genes in each species. **(c)** Relationship between LINES coverage and genome size in Heterobranchia species. Numbers on linear regressions correspond to adjusted r^2 coefficients (Pearson's test).

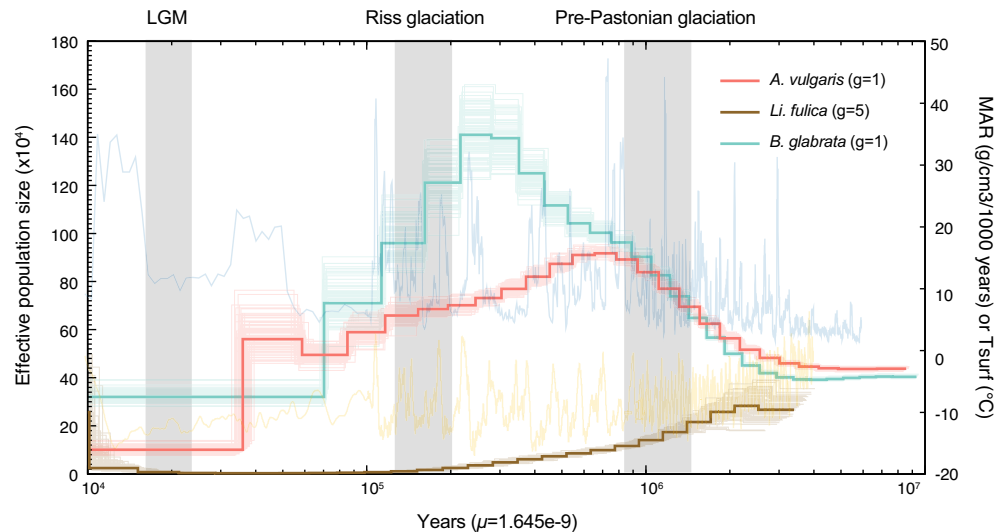


Figure 6. Demographic histories of *Arion vulgaris*, *Lissachatina (Achatina) fulica* and *Biomphalaria glabrata*. Solid bold lines indicate inferred ancestral effective population sizes for three species, while background blue and yellow lines represent the mass accumulation rate (MAR) and the atmospheric surface air temperature (Tsurf) relative to the present, respectively.

collinearity blocks. In our initial analysis, we used the same method as Liu et al., described. We did observe *Ks* peaks shared by *A. vulgaris* and *Li. fulica* which represents the WGD event, however, the *Ks* distribution of *Li. immaculata* has a relatively large deviation (Supplementary Fig. S28). From our results of the syntenic dot plots, *Li. immaculata* exhibits lower synteny in comparisons to *A. vulgaris* than the *Li. fulica*-*A. vulgaris* comparison (Supplementary Figs. S17, 18), implying that *Li. immaculata* has experienced more genome reconfiguration and chromosome rearrangement, and this may increase errors and difficulty in the identification of collinearity gene pairs within *Li. immaculata*. We addressed this problem by implementing WGD⁶³, a new tool which can identify collinearity more accurately and comprehensively. In the *Ks* distribution obtained by WGD, the three species have relatively consistent *Ks* peaks (Fig. 3b). Therefore, we suppose that the estimation based on *Ks* distribution derived from WGD can more accurately represent the older than expected time of the WGD event.

In our results, after WGD, the extra chromosome copy shows a release of selective pressure with large structural variations and increased synonymous mutation rate (Supplementary Figs. S17–19), which might serve as an abundant resource for mutations and novo functions, and may have facilitated the stylommatophoran transition from water to land. For example, expansion genes derived by WGD duplications are enriched in nerve and muscle development, which might enhance the locomotion and movement ability in terrestrial environments (Supplementary Fig. S24). Moreover, we also detected genes related to kidney development, response to stimulation, radiation, larval development, and dietary habits that were expanded in both *A. vulgaris* and shelled land pulmonates (Fig. 2d). These genes might have contributed to the stylommatophoran ancestor's ability to overcome challenges such as gravitational pressure and water loss brought by the terrestrial environment.

The split of *A. vulgaris* and *Lissachatina* land snail lineages happened in a very short time after WGD (Fig. 3b). However, the slug *A. vulgaris* has evolved its unique adaptability in further improving water re-absorption and resistance to external stimuli (Fig. 2d). For example, a series of interleukin genes were positively selected in *A. vulgaris* genome, which might enhance the immune response; genes related to acute inflammatory processes were expanded, which might improve the innate defense; the expansion of genes related to regeneration might help to quickly recover from body/organ damage and increase the survival rate, and the expansion of genes and pathways in pigment biosynthesis might protect *A. vulgaris* from solar radiation.

In addition to WGD, small scale gene duplication also plays an important role in providing new genetic material for mutation, drift, and selection⁶¹. We found Heterobranchia species have an abundance of duplicated genes. Among all types of duplicated genes, the largest category is dispersed duplication (DSD) genes (Fig. 4). The proportion of tandem duplication (TD) genes has greatly increased in *A. vulgaris* as compared to other Heterobranchia species, and ~24% of them were positively selected ($Ka/Ks > 1$). Enrichment analysis showed the functions of TD derived genes largely overlapped with *A. vulgaris* expansion gene functions, e.g., response to external stress, pigment catabolism, acute inflammation, which thus implies that tandem duplication of genes might be one of the forces driving evolution, adaptation, and potential invasiveness of *A. vulgaris*.

Previous studies have shown that transposable element (TE) insertions play a critical role in rapid phenotypic variation and might help invasive species to successfully adapt to a novel environment⁴⁹. In our results, the recent massive expansion of TEs (more precisely, LINES) in *A. vulgaris* might act as potent insertional mutagens, greatly enhancing the adaptive success, invasiveness, and the ability to outcompete other land slugs.

All in all, our genomic analysis reveals the powerful potential of *A. vulgaris* for adaptation and evolution, which may explain why *A. vulgaris* is considered as an invasive species in central Europe. However, there is ongoing controversy about its native range and invasiveness. According to the record of first discovery in many

European countries, it was believed that the slug originated on the Iberian Peninsula and expanded its range into central and eastern Europe over the last five decades¹². However, the very similar external appearance with other closely related native large arionids as well as (potential) hybrid species between *A. vulgaris*, *A. ater*, and *A. rufus*^{64–66}, might have caused the misidentification of *A. vulgaris*, obscured the specimen records, and made it difficult to trace its origin and monitoring the spread only by morphological identification^{66,67}. Recent studies based on the genetic diversity patterns of mitochondrial and nuclear loci suggested that *A. vulgaris* is native in central Europe rather than alien/invasive while probably invasive in other parts of Europe^{20–22}. Our *A. vulgaris* individual was collected in Munich, Southern Germany and has a relatively rich genetic diversity, which implies a large effective population size. This result seems to support the point of view that *A. vulgaris* is more likely native rather than invasive at least in south Germany (Fig. 6). However, more robust conclusions still require extensive sampling and more population data. Our high-quality *A. vulgaris* genome will promote future population studies from the use of single/multiple molecular markers to the use of whole genome-wide polymorphism and will help us to understand its origin, expansion, and potential invasiveness more comprehensively.

Materials and methods

Sample collection and sequencing. An adult *A. vulgaris* was collected in the garden of the Zoologische Staatssammlung München, Germany. Genomic DNA was extracted from the foot muscle tissue with MagAttract HMW DNA Kit and CTAB method⁶⁹. Quality was checked using agarose gel electrophoresis. Four different sequencing technologies were used to obtain the genome sequence (Supplementary Table S1). First, one Illumina paired-end sequencing library was generated following the manufacturer's standard protocol (Illumina) with an insert size of 350 bp. Also, high molecular weight DNA was separated and loaded onto the 10X Genomics Chromium microfluidics controller for barcoding and generated two 10X Genomics linked-read libraries with an insert size of 350 bp. Those reads not only provided the long-range positional information to assemble contigs into scaffolds but were also used for the genome survey analysis and final base-level genome sequence correction⁷⁰. One Hi-C library digested with MboI and with an insert size of 350 bp was constructed for providing long-range information on the grouping and linear organization of sequences along entire chromosomes to assemble the scaffolds into chromosome-level scaffolds⁷¹. The Illumina paired-end sequencing library, 10X Genomics linked-read libraries, and Hi-C library were sequenced on an Illumina HiSeqX Ten platform (Illumina, San Diego, CA, USA) with 150 bp paired-end reads. The raw reads generated by Illumina HiSeqX Ten platform were all filtered with the following criteria: reads with adapters, reads with N bases more than 5%, and reads with more than 65% of low-quality bases (\leq seven) using Fastp v0.20.0⁷². Meanwhile, Nanopore libraries were prepared using SQK-LSK109 kit and sequenced in the platform Nanopore PromethION (Oxford Nanopore Technologies). We performed a base calling of the raw Nanopore data with Guppy v2.2.3⁷³.

Total RNA was extracted from the 'head' part of the sample which includes tentacles, mantle, inner head and anterior visceral organs, and foot and sequenced on an Illumina NovaSeq platform with paired-end 150 bp.

Genome feature estimation and assembly. The genome size and heterozygosity were estimated by GenomeScope v1.0.0⁷⁴ using the quality-controlled paired-end Illumina sequence data and linked reads. We combined reads generated using different sequencing platforms to generate a high-quality de novo genome assembly (Supplementary Table S2). Specifically, long reads, generated with the Nanopore PromethION platform, were assembled into contigs using the wtdbg2 v2.2 assembler⁷⁵. The contigs were subsequently polished by ntEdit v1.3.1⁷⁶ using Illumina short reads and linked reads. The resulting contigs were then connected into scaffolds by 10X Genomics linked-read data using Scaff10X v4.2⁷⁷. Hi-C reads were mapped to the draft assembly and processed using the hicstuff v2.2.2 pipeline⁷⁸ with the parameters --aligner bowtie2 --enzyme MboI --iterative --matfmt graal --quality-min 30 --size 0. We ran instaGRAAL v0.1.2⁷⁹ on the resulting matrix and the draft assembly with parameters --level 5 --cycles 100 --coverage-std 1 --neighborhood 5 and the module instagraal-polish for refinement. After building the interaction map of the final scaffolds with hicstuff, we noticed an intra-chromosomal translocation on chromosome 9 which could have been due to a misassembly. In the subsequent analysis, we mapped all reads to the assembly 'chromosome 9' and identified two breakpoints (at site 21,000,000 and 27,600,000 respectively) based on the read' depth and gene distribution. We corrected the orders manually and reconnected sequences with 10 N's at the new junction sites.

Gene prediction. Protein-coding genes were predicted using the following approaches: ab initio prediction, homology-based prediction, and transcriptome-based prediction. For ab initio prediction, RNA-seq reads were first aligned to the *A. vulgaris* genome sequence using STAR v2.7.2b⁸⁰, then the read alignment information was merged and used for Braker2 v2.1.5⁸¹ gene prediction pipeline. For homology-based prediction, we selected six gastropods from closely to distantly related to *A. vulgaris*, namely *Li. fulica*, *B. glabrata*, *Ap. californica*, *E. chlorotica*, *P. canaliculata*, *Haliotis rufescens* (Supplementary Table S3). The protein sequences of the six species were downloaded from NCBI and aligned against the assembled genome with MMseqs v11.e1a1c⁸². These results were then combined into gene models separately with GeMoMa v1.3.1⁸³ using mapped RNA-seq data for splice site identification. The resulting gene annotation sets were further filtered using the GeMoMa module GAF with default parameters. For the transcriptome-based prediction, RNA-seq data had been assembled using both de novo and genome-guided approaches with Trinity vr20140413p1⁸⁴, and the gene predictions were carried out with PASA v2.0.2⁸⁵. All gene annotations were combined with EVM v1.1.1⁸⁶ (Supplementary Table S4). Partial genes and genes with a coding length of less than 150 bp were removed from further analysis.

The predicted genes were functionally annotated by aligning them to the eggNOG²⁷, SWISS-PROT²⁸, TrEMBL²⁸, KEGG³⁰, and InterPro²⁹ databases using BLAST v2.2.31⁸⁷ with a maximal e-value of $1e-5$ and by

aligning to the Pfam database using HMMer v3.0⁸⁸. Gene Ontology (GO) terms (Gene Ontology, RRID:SCR002811) were assigned to the genes using the BLAST2GO v2.5 pipeline⁸⁹.

Gene family cluster and terrestrial adaptation analysis. To resolve the early phylogeny of gastropods, we selected the species according to the following rules: (1) coverage of as many subclasses as possible; (2) the lineage diversity within each subclass should be covered; (3) in case of closely related species, those with high-quality genomes or better gene annotations were preferred. As a result, fourteen Gastropoda species were selected, including six Heterobranchia: *Li. fulica*, *Li. immaculata*, *B. glabrata*, *R. auricularia*, *Ap. californica*, *E. chlorotica*; four Caenogastropoda: *P. canaliculata*, *Marisa cornuarietis*, *Lanistes nyassanus*, *Conus consors*; one Vetigastropoda: *H. rufescens*; one Neomphalina: *Chrysomallon squamiferum*, and one Patellogastropoda: *Lottia gigantea*. Two bivalve species: *Argopecten purpuratus* and *Saccostrea glomerata* were selected as outgroups (Supplementary Table S3). Protein sequences were extracted from each species and an all-against-all comparison was performed using BLASTP v2.9.0⁹⁰ with an e-value cut-off of $1e-5$. OrthoFinder v2.4.0⁹¹ was used to cluster gene families.

Based on the clustered gene families, we explored the terrestrial adaptation of *A. vulgaris* from two aspects. One is the genetic basis of adaptability shared by stylommatophoran species relative to other aquatic or marine Heterobranchia species, another is the specific adaptations of shell-less *A. vulgaris* compared to two land snails, *Li. fulica* and *Li. immaculata*. For both cases, we tested lineage/species specific genes, expansion/contraction genes, and positively selected genes (PSGs) and performed Gene Ontology (GO) enrichment analysis.

We applied the CAFÉ v4.2.1⁹² program to examine gene family expansion and contraction across entire genomes with default parameters. To identify PSGs, OrthoFinder v2.4.0⁹¹ was used to cluster gene families from five Heterobranchia species: *Li. fulica*, *Li. immaculata*, *B. glabrata*, *R. auricularia*, *Ap. californica* (Supplementary Table S3). Single-copy orthologous genes were extracted based on the results of clustered gene families. MAFFT v7.455⁹³ was used for multiple sequence alignments and converted to codon sequences by PAL2NAL v14⁹⁴. Poorly aligned positions were removed with Gblocks v0.91b⁹⁵ with parameters “-b2 = 85% alignment length -b3 = 6 -b4 = 10 -b5 = h -t = c”. The PSGs were identified by comparing the null model (fix_omega = 1) to the alternative model (fix_omega = 0) using codeml branch-site model in the PAML package⁹⁶. The foreground branch was set to (1) the node of the most common ancestor of *A. vulgaris*, *A. vulgaris* and *Li. immaculata*, to identify putative PSGs shared by stylommatophoran species, and (2) *A. vulgaris*, for the detection of potential PSGs of *A. vulgaris*. Chi-square tests were performed for each pair and genes with a 5% significance level were selected as putative PSGs⁹⁶. Cytoscape v3.8.2⁹⁷ was used for visualizing molecular interaction networks and biological pathways.

Phylogenetic analysis. Gene families with only one copy from each of 16 species were selected as single-copy genes and were concatenated and aligned by MUSCLE v3.8.1551⁹⁸ with default parameters. The maximum likelihood (ML) trees were inferred using both RAXML v8.2.8⁹⁹ with the GTR+ Γ model and IQ-TREE v1.6.9¹⁰⁰, which automatically selected the best-fit substitution model using ModelFinder¹⁰¹. For coalescent-based analysis, gene trees were first estimated using RAXML v8.2.8⁹⁹ with 100 replicates from each single copy gene. The best tree was then selected as input to ASTRAL v5.6.1¹⁰² to infer the species tree with default parameters. Gene trees were visualized using DensiTree v2.01¹⁰³.

Divergence time was computed using the MCMCTREE program implemented in the PAML v4.8⁹⁶ package. For calibration, we used the soft bounds of Euopisthobranchia—Panpulmonata (divergence time between 190 and 270 MY)³³, the fossil of Sublitoidea (418 MY) constraints on the node of Heterobranchia and Caenogastropoda¹⁰⁴, and the fossil of *Fordilla troyensis* (530 MY) for the root^{105,106}.

Identification of whole-genome duplication event. For macrosynteny analysis, LASTZ v1061¹⁰⁷ was used to perform whole-genome alignments between chromosome-level assemblies of *A. vulgaris*, *Li. immaculata*¹¹ and *Ap. californica*¹⁰⁸. The alignments were visualized by Circos v0.69-6¹⁰⁹. For synteny analysis of homologous gene pairs, the protein sequences of *A. vulgaris*, *Li. fulica* and *Li. immaculata* were first searched against themselves and also between species using BLASTP v2.9.0⁹⁰, then subjected to WGDI v0.5.1⁶³ to determine syntenic blocks, estimate K_s values for each block and calculate K_s distributions of gene pairs in collinearity blocks. Curves were fitted using the Gaussian approximation function in the WGDI package.

Identifying gene duplications. The different modes of gene duplication were identified using the Dup-Gen_finder v1.07 pipeline¹¹⁰ using *P. canaliculata* as a reference⁵⁶. Gene pairs were further filtered to remove overlaps between different duplicate modes. For each duplicated gene pair, we aligned their protein sequences using MAFFT v7.455⁹³ and converted the protein alignment into a codon alignment using PAL2NAL v14⁹⁴. Then, the resulting codon alignment was formatted into an AXT format and K_a (number of substitutions per nonsynonymous site) and K_s (number of substitutions per synonymous site) values were calculated by KaKs_Calculator v2.0¹¹¹.

Repeat prediction and expansions of transposable elements. TRF v4.09¹¹² was used for tandem repeats identification with default parameters. Transposable elements (TEs) were annotated using a combination of ab initio and homology-based approaches. First, repeat elements were identified de novo using RepeatModeler v2.0.1¹¹³. The database predicted by RepeatModeler, the RepBase¹¹⁴ (RepBase-20170127) and the Dfam¹¹⁵ (Dfam_Consensus-20170127) libraries were then merged together and used as a custom library for RepeatMasker v4.0.7¹¹³ to identify repeats comprehensively. The repeat divergence rate was measured by the percentage of substitutions in the corresponding regions between annotated repeats and consensus sequences in the RepBase database. For species with incomplete TE annotations (e.g., *Li. fulica*, *R. auricularia*), we predicted their TEs

using the same approaches as just described. We regarded genes with TEs inserted in introns, exons or with 2-kb upstream or 1 kb downstream of the terminal exons as likely to be affected by these insertions and compared the number of genes affected by TEs in different insertion regions.

Genome heterozygosity and reconstruction of effective population size (N_e). Heterozygosity was estimated by the following steps. First, the clean Illumina reads and linked reads were merged and mapped onto the *A. vulgaris* assembly by BWA-MEM v0.7.17-r1188¹¹⁶ with default parameters. The sequence alignment/map (SAM) file format was processed using SAMtools v1.9¹¹⁷, and Picard v2.23.3¹¹⁸ was used to mark duplicates. Finally, single nucleotide polymorphisms (SNPs) calling was implemented in GATK v4.1.6.0¹¹⁹ using default parameters, and several filtering steps were performed to reduce false positives, including (1) remove SNPs with more than two alleles; (2) remove SNPs with a quality score less than 30; (3) remove SNPs at or within 5 bp from any InDels; (4) remove sites with extremely low (less than one-third average depth) or extremely high (more than three-fold average depth) coverage.

We inferred the demographic history by applying the Pairwise Sequentially Markovian Coalescence model using PSMC v0.6.5-r67¹²⁰ with the following parameters: -N25 -t15 -r5 -p '4 + 25 × 2 + 4 + 6'. This method reconstructs the history of changes in population size over time using the distribution of the most recent common ancestor (tMRCA) between two alleles in an individual. The generation time of *A. vulgaris* and *B. glabrata* was assumed to be 1 year^{121,122} and *Li. fulica* was assumed to be 5 years¹²³.

Ethics declarations. No specific permits were required for the described field studies, no specific permissions were required for these locations/activities, and the field studies did not involve endangered or protected species.

Data availability

The *A. vulgaris* genome project of this study was deposited at the National Center for Biotechnology Information (NCBI) under BioProject number PRJNA680311. Genomic and transcriptome sequence reads was deposited in the SRA database with BioSample: SAMN16874494. The assembled genome had been deposited at GenBank with accession number: GCA_020796225.1. In addition, the genome annotation files had been submitted at the Figshare: <https://doi.org/10.6084/m9.figshare.15022212.v1>; <https://doi.org/10.6084/m9.figshare.15022203.v1>.

Received: 19 February 2022; Accepted: 5 August 2022

Published online: 20 August 2022

References

- Barker, G. M. In *The Biology of Terrestrial Molluscs* (ed. Braker, G.M.) 20 (CABI Publishing, England, 2001).
- Kano, Y., Brenzinger, B., Nützel, A., Wilson, N. G. & Schrödl, M. Ringiculid bubble snails recovered as the sister group to sea slugs (Nudipleura). *Sci. Rep.* **6**, 30908 (2016).
- Klussmann-Kolb, A., Dinapoli, A., Kuhn, K., Streit, B. & Albrecht, C. From sea to land and beyond—new insights into the evolution of euthyneuran Gastropoda (Mollusca). *BMC Evol. Biol.* **8**, 57 (2008).
- Romero, P. E., Weigand, A. M. & Pfenninger, M. Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life. *BMC Evol. Biol.* **16**, 164 (2016).
- Romero, P. E., Pfenninger, M., Kano, Y. & Klussmann-Kolb, A. Molecular phylogeny of the Ellobiidae (Gastropoda: Panpulmonata) supports independent terrestrial invasions. *Mol. Phylogenet. Evol.* **97**, 43–54 (2016).
- Baur, B. Parental care in terrestrial gastropods. *Experientia* **50**, 5–14 (1994).
- Takei, Y. From aquatic to terrestrial life: evolution of the mechanisms for water acquisition. *Zool. Sci.* **32**, 1–7 (2015).
- Mordan, P. & Wade, C. In *Phylogeny and Evolution of the Mollusca* (eds. Ponder, W. & Lindberg, D.R.) 409–426 (University of California Press, Cambridge, 2008).
- Ohta, T. Gene conversion and evolution of gene families: an overview. *Genes* **1**, 349–356 (2010).
- Wang, K. *et al.* African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* **184**, 1362–1376 (2021).
- Liu, C. *et al.* Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. *Mol. Ecol. Resour.* **21**, 478–494 (2020).
- Zajac, K. S., Gaweł, M., Filipiak, A. & Kramarz, P. *Arion vulgaris* Moquin-Tandon, 1855—the aetiology of an invasive species. *Folia Malacol.* **25**, 81–93 (2017).
- Frank, T. Slug damage and number of slugs (Gastropoda: Pulmonata) in winter wheat in fields with sown wildflower strips. *J. Molluscan Stud.* **64**, 319–328 (1998).
- Gren, I. M., Isacs, L. & Carlsson, M. Costs of alien invasive species in Sweden. *Ambio* **38**, 135–140 (2009).
- Kozłowski, J. The significance of alien and invasive slug species for plant communities in Agroecosystems. *J. Plant Prot. Res.* **52**, 67–76 (2012).
- Gismervik, K. *et al.* Invading slugs (*Arion vulgaris*) can be vectors for *Listeria monocytogenes*. *J. Appl. Microbiol.* **118**, 809–816 (2015).
- Slotsbo, S. *et al.* Cold tolerance and freeze-induced glucose accumulation in three terrestrial slugs. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **161**, 443–449 (2012).
- Frank, T. Influence of slug herbivory on the vegetation development in an experimental wildflower strip. *Basic. Appl. Ecol.* **4**, 139–147 (2003).
- Rabitsch, W. DAISIE—Delivering Alien Invasive Species Inventoried for Europe. <http://www.europe-aliens.org> (2006).
- Pfenninger, M., Weigand, A., Balint, M. & Klussmann-Kolb, A. Misperceived invasion: The Lusitanian slug (*Arion lusitanicus* auct. non-Mabille or *Arion vulgaris* Moquin-Tandon 1855) is native to Central Europe. *Evol. Appl.* **7**, 702–713 (2014).
- Zajac, K. S. *et al.* A comprehensive phylogeographic study of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Pulmonata: Arionidae) in Europe. *Org. Divers. Evol.* **20**, 37–50 (2019).
- Zemanova, M. A., Knop, E. & Heckel, G. Phylogeographic past and invasive presence of *Arion* pest slugs in Europe. *Mol. Ecol.* **25**, 5747–5764 (2016).
- Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L. F. C. & Froufe, E. Molluscan genomics: The road so far and the way forward. *Hydrobiologia* **847**, 1705–1726 (2019).
- Beeson, G. E. Chromosome numbers of slugs. *Nature* **186**, 257–258 (1960).

25. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
26. Mapleson, D., Garcia-Accinelli, G., Kettleborough, G., Wright, J. & Clavijo, B. J. KAT: A K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**, 574–576 (2017).
27. Huerta-Cepas, J. *et al.* eggNOG 50: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
28. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
29. Hunter, S. *et al.* InterPro: The integrative protein signature database. *Nucleic Acids Res.* **37**, D211–215 (2009).
30. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
31. Cunha, T. J. & Giribet, G. A congruent topology for deep gastropod relationships. *Proc. R. Soc. B* **286**, 20182776 (2019).
32. Uribe, J. E., Irisarri, I., Templado, J. & Zardoya, R. New patellogastropod mitogenomes help counteracting long-branch attraction in the deep phylogeny of gastropod mollusks. *Mol. Phylogenet. Evol.* **133**, 12–23 (2019).
33. Zapata, F. *et al.* Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda. *Proc. R. Soc. B* **281**, 20141739 (2014).
34. Haszprunar, G. On the origin and evolution of major gastropod groups, with special reference to the Streptoneura. *J. Molluscan Stud.* **54**, 367–441 (1988).
35. Ponder, W. F. & Lindberg, D. R. Gastropod phylogeny—challenges for the 90s. In *Origin and evolutionary radiation of the mollusca*. (ed. Taylor, J.D.) 135–154 (Oxford University Press, Oxford, 1996).
36. Ponder, W. F. & Lindberg, D. R. Towards a phylogeny of gastropod molluscs: An analysis using morphological characters. *Zool. J. Linn. Soc.* **119**, 83–265 (1997).
37. Chen, Z. & Schrödl, M. How many single-copy orthologous genes from whole genomes reveal deep gastropod relationships?. *PeerJ* **10**, e13285 (2022).
38. Doğan, O., Schrödl, M. & Chen, Z. The complete mitogenome of *Arion vulgaris* Moquin-Tandon, 1855 (Gastropoda: Stylommatophora): Mitochondrial genome architecture, evolution and phylogenetic considerations within Stylommatophora. *PeerJ* **8**, e8603 (2020).
39. Klobucar, G., Lajtner, J. & Erben, R. Increase in number and size of kidney concretions as a result of PCP exposure in the freshwater snail *Planorbarius corneus* (Gastropoda, Pulmonata). *Dis. Aquat. Organ* **44**, 149–154 (2001).
40. Barnes, P. J. How corticosteroids control inflammation: Quintiles prize lecture 2005. *Br. J. Pharmacol.* **148**, 245–254 (2006).
41. Cain, D. W. & Cidlowski, J. A. Immune regulation by glucocorticoids. *Nat. Rev. Immunol.* **17**, 233–247 (2017).
42. Munck, A., Guyre, P. M. & Holbrook, N. J. Physiological functions of glucocorticoids in stress and their relation to pharmacological actions. *Endocr. Rev.* **5**, 25–44 (1984).
43. Van de Peer, Y., Mizrachi, E. & Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
44. Hallinan, N. M. & Lindberg, D. R. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. *Genome Biol. Evol.* **3**, 1150–1163 (2011).
45. Allio, R., Donega, S., Galtier, N. & Nabholz, B. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* **34**, 2762–2772 (2017).
46. Gillard, G. B. *et al.* Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. *Genome Biol.* **22**, 103 (2021).
47. Wade, C. M., Mordan, P. & Naggs, F. Evolutionary relationships among the Pulmonate land snails and slugs (Pulmonata, Stylommatophora). *Biol. J. Linn. Soc.* **87**, 593–610 (2006).
48. Lallemand, T., Leduc, M., Landes, C., Rizzon, C. & Lerat, E. An overview of duplicated gene detection methods: Why the duplication mechanism has to be accounted for in their choice. *Genes* **11**, 5 (2020).
49. Niu, X. M. *et al.* Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc. Natl. Acad. Sci. USA* **116**, 6908–6913 (2019).
50. Schrader, L. *et al.* Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. Commun.* **5**, 5495 (2014).
51. Schrader, L. & Schmitz, J. The impact of transposable elements in adaptive evolution. *Mol. Ecol.* **28**, 1537–1549 (2019).
52. Canapa, A., Barucca, M., Biscotti, M. A., Forconi, M. & Olmo, E. Transposons, genome size, and evolutionary insights in animals. *Cytogenet. Genome Res.* **147**, 217–239 (2015).
53. Naville, M. *et al.* Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr. Biol.* **29**, 1161–1168 (2019).
54. Zhou, W., Liang, G., Molloy, P. L. & Jones, P. A. DNA methylation enables transposable element-driven genome expansion. *Proc. Natl. Acad. Sci. USA* **117**, 19359–19366 (2020).
55. Guo, Y. *et al.* A chromosomal-level genome assembly for the giant African snail *Achatina fulica*. *Gigascience* **8**, giz124 (2019).
56. Sun, J. *et al.* Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Mol. Biol. Evol.* **36**, 1507–1520 (2019).
57. Zemanova, M. A. *et al.* Slimy invasion: Climatic niche and current and future biogeography of *Arion* slug invaders. *Divers. Distrib.* **24**, 1627–1640 (2018).
58. Zemanova, M. A., Knop, E. & Heckel, G. Introgressive replacement of natives by invading *Arion* pest slugs. *Sci. Rep.* **7**, 14908 (2017).
59. Alix, K., Gerard, P. R., Schwarzacher, T. & Heslop-Harrison, J. S. P. Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. *Ann. Bot.* **120**, 183–194 (2017).
60. Sacerdot, C., Louis, A., Bon, C., Berthelot, C. & Roest Crollius, H. Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol.* **19**, 166 (2018).
61. Crow, K. D. & Wagner, G. P. What is the role of genome duplication in the evolution of complexity and diversity?. *Mol. Biol. Evol.* **23**, 887–892 (2006).
62. Wang, Y. *et al.* MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
63. Sun, P. *et al.* WGDI: A user-friendly toolkit for evolutionary analyses of whole2 genome duplications and ancestral karyotypes. <https://www.biorxiv.org/content/https://doi.org/10.1101/2021.04.29.441969v1> (2021).
64. Dreijers, E., Reise, H. & Hutchinson, J. M. C. Mating of the slugs *Arion iusitanicus* auct. non Mabile and *A. rufus* (L.): Different genitalia and mating behaviours are incomplete barriers to interspecific sperm exchange. *J. Molluscan Stud.* **79**, 51–63 (2013).
65. Hatteland, B. A. *et al.* Introgression and differentiation of the invasive slug *Arion vulgaris* from native *A. ater*. *Malacologia* **58**, 303–321 (2015).
66. Roth, S., Hatteland, B. A. & Solhoy, T. Some notes on reproductive biology and mating behaviour of *Arion vulgaris* Moquin-Tandon 1855 in Norway including a mating experiment with a hybrid of *Arion rufus* (Linnaeus 1758) x *ater* (Linnaeus 1758). *J. Conchol.* **41**, 249–258 (2012).
67. Barr, N. B. *et al.* Application of a DNA barcode using the 16S rRNA gene to diagnose pest *Arion* species in the USA. *J. Molluscan Stud.* **75**, 187–191 (2009).

68. Quinteiro, J., Rodríguez-Castro, J., Castillejo, J., Iglesias-Piñero, J. & Rey-Méndez, M. Phylogeny of slug species of the genus *Arion*: Evidence of monophyly of Iberian endemics and of the existence of relict species in Pyrenean refuges. *J. Zool. Syst.* **43**, 139–148 (2005).
69. Arseneau, J. R., Steeves, R. & Laflamme, M. Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. *Mol. Ecol. Resour.* **17**, 686–693 (2017).
70. Mostovoy, Y. *et al.* A hybrid approach for de novo human genome sequence assembly and phasing. *Nat. Methods* **13**, 587–590 (2016).
71. Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
72. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
73. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
74. Vurture, G. W. *et al.* GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
75. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
76. Warren, R. L. *et al.* ntEdit: scalable genome sequence polishing. *Bioinformatics* **35**, 4430–4432 (2019).
77. Ning, Z. H. In *Scaff10X v4.2: Pipeline for Scaffolding and Breaking a Genome Assembly Using 10x Genomics Linked-Reads*. <https://github.com/wtsi-hpag/Scaff10X> (2020).
78. Matthey-Doret, C. *et al.* hicstuff: Simple library/pipeline to generate and handle Hi-C data. Zenodo. 10.5281/zenodo.4066363 (2020).
79. Baudry, L. *et al.* instaGRAAL: Chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffold. *Genome Biol.* **21**, 148 (2020).
80. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
81. Brúna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **3**, 108 (2021).
82. Hauser, M., Steinegger, M. & Soding, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323–1330 (2016).
83. Keilwagen, J., Hartung, F. & Grau, J. GeMoMa: Homology-Based gene prediction utilizing intron position conservation and RNA-seq data. *Methods Mol. Biol.* **1962**, 161–177 (2019).
84. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
85. Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
86. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
87. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
88. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).
89. Conesa, A. *et al.* Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
90. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
91. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
92. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
93. Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490–2492 (2018).
94. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–612 (2006).
95. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
96. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **13**, 555–556 (1997).
97. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
98. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
99. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
100. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
101. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
102. Zhang, C., Scornavacca, C., Molloy, E. K. & Mirarab, S. ASTRAL-Pro: Quartet-based species-tree inference despite paralogy. *Mol. Biol. Evol.* **37**, 3292–3307 (2020).
103. Bouckaert, R. R. & Heled, J. DensiTree 2: Seeing trees through the forest. <https://www.biorxiv.org/content/https://doi.org/10.1101/012401v1> (2014).
104. Fryda, J., Nützel, A. & Wagner, P. J. In *Phylogeny and Evolution of the Mollusca*. (eds. Ponder W. & Lindberg D. R.) 238–270 (University of California Press, Berkeley, 2008).
105. Nützel, A., Erwin, D. H. & Mapes, R. H. Identify and phylogeny of the Late Paleozoic Subulitoidea (Gastropoda). *J. Paleont.* **74**, 575–598 (2000).
106. Pojeta, J. & Runnegar, B. *Fordilla troyensis* and the early history of pelecypod mollusks: Early cambrian fossils from New York State provide important clues to the evolution of the class. *Am. Sci.* **62**, 706–711 (1974).
107. Harris, R. S. In *Improved pairwise alignment of genomic DNA*, Pennsylvania State University (2007).
108. Di Palma, F. *et al.* In *The Draft Genome of Aplysia californica*, <https://www.ncbi.nlm.nih.gov/nucore/AASC00000000.3> (2014).
109. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
110. Qiao, X. *et al.* Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* **20**, 38 (2019).
111. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinform.* **8**, 77–80 (2010).
112. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
113. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **25**, 41011–141014 (2009).

114. Bao, W., Kojima, K. K. & Kohany, O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
115. Wheeler, T. J. *et al.* Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* **41**, D70–82 (2013).
116. Li, H. In *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM*. <https://www.arxiv-vanity.com/papers/1303.3997> (2013).
117. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
118. *Picard toolkit*. In *Broad Institute, GitHub Repository*. <http://broadinstitute.github.io/picard> (2019).
119. do-Valle, I. F. *et al.* Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data. *BMC Bioinform.* **17**, 341 (2016).
120. Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol. Ecol.* **25**, 1058–1072 (2016).
121. Davies, S. M. *Arion flagellus* Collinge and *A. lusitanicus* Mabilie in the British Isles: A morphological, biological and taxonomic investigation. *J. Conchol.* **32**, 339–354 (1987).
122. Pimentel, D. Life history of *Australorbis glabratus*, the intermediate snail host of *Schistosoma mansoni* in Puerto rico. *Ecology* **38**, 576–580 (1957).
123. Pawson, P. A. & Chase, R. The life-cycle and reproductive of *Achatina fulica* (bowdich) in laboratory culture. *J. Molluscan Stud.* **50**, 85–91 (1984).

Acknowledgements

We thank Gert Wörheide and Michael Eitel (Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität München, LMU) for organizing the IGNITE Comparative Genomics of Non-Model Invertebrates program; Denis Tagu (INRAE), Fabrice Legeai (INRAE, INRIA), Claire Lemaitre (INRIA), Grace McCormack and Kenneth Sandoval (National University of Ireland) for the guidance and advice on genome assembly during secondments; Ramon Rivera (LMU) for software installation and maintenance; Juan Moles (ZSM) for polishing an earlier version of the article; Isabella Stöger (ZSM) for laboratory work and other ZSM mollusk teammates as well as colleagues from IGNITE for valuable feedback and comments. The bioinformatics analyses were performed at Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources) and GenOuest IRISA-INRIA, France. This study has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764840.

Author contributions

Z.C. and M.S. conceived the study. Z.C., Ö.D., N.G., A.G. performed genome assembly. Z.C. performed genome annotation and evolution analysis. Z.C. and M.S. wrote the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-18099-7>.

Correspondence and requests for materials should be addressed to Z.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022