Whole Genome Duplication and Gene Evolution in the Hyperdiverse Venomous Gastropods

Sarah Farhat [1],*[,1] Maria Vittoria Modica [2] and Nicolas Puillandre[1]

[1]Institut Systématique Evolution Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France

[2]Department of Biology and Evolution of Marine Organisms (BEOM), Stazione Zoologica Anton Dohrn, Roma, Italy

***Corresponding author:** E-mail: sarah.farhat@mnhn.fr.

Associate editor: Jun Gojobori

Abstract

The diversity of venomous organisms and the toxins they produce have been increasingly investigated, but taxonomic bias remains important. Neogastropods, a group of marine predators representing almost 22% of the known gastropod diversity, evolved a wide range of feeding strategies, including the production of toxins to subdue their preys. However, whether the diversity of these compounds is at the origin of the hyperdiversification of the group and how genome evolution may correlate with both the compounds and species diversities remain understudied. Among the available gastropods genomes, only eight, with uneven quality assemblies, belong to neogastropods. Here, we generated chromosome-level assemblies of two species belonging to the Tonnoidea and Muricoidea superfamilies (*Monoplex corrugatus* and *Stramonita haemastoma*). The two obtained high-quality genomes had 3 and 2.2 Gb, respectively, and 92–89% of the total assembly conformed 35 pseudochromosomes in each species. Through the analysis of syntenic blocks, Hox gene cluster duplication, and synonymous substitutions distribution pattern, we inferred the occurrence of a whole genome duplication event in both genomes. As these species are known to release venom, toxins were annotated in both genomes, but few of them were found in homologous chromosomes. A comparison of the expression of ohnolog genes (using transcriptomes from osphradium and salivary glands in *S. haemastoma*), where both copies were differentially expressed, showed that most of them had similar expression profiles. The high quality of these genomes makes them valuable reference in their respective taxa, facilitating the identification of genome-level processes at the origin of their evolutionary success.

Key words: **neogastropods, genomes, whole-genome duplication, venom, neofunctionalization.**

Introduction

Venomous organisms display a high level of diversification, both at the taxonomic and the functional levels (Casewell et al. 2013; Fry et al. 2015; Zancolli and Casewell 2020). The production of venom has been reported across at least 101 independent lineages in animals (Schendel et al. 2019), among which some well-known examples are snakes, spiders, and scorpions, with multiple instances of convergent evolution (King et al. 2008; Barua and Mikheyev 2019; Forde et al. 2022). Venom is a mix of toxins (mostly proteins and peptides) and other molecules, produced and stored in specialized anatomical structures of an organism and injected in another organism to disrupt its normal physiological processes (Fry et al. 2009). Venomous animals thus developed complex systems adapted to their target organisms, comprising not only toxic molecules but also the self-resistance mechanism needed to deal with their production, storage, and delivery, as well as specialized morphological structures including secretory glands and delivery weapons (Schendel et al. 2019). Venoms can be used for multiple purposes, including predation, defense, or intraspecific competition (Casewell et al. 2013). The specific functions of the toxins that compose the venoms are extremely diversified, and in most cases, toxins act synergistically to incapacitate the target organism by inducing paralysis through targeting neuromuscular receptors, causing blood loss by injecting anticoagulant, or causing intense pain by triggering pain receptors, among other mechanisms (Bohlen et al. 2011; Zancolli and Casewell 2020). Given the adaptive value of venom, understanding the origin and evolution of such diversified molecules and the organisms that produce them have become crucial in evolutionary biology.

Neogastropoda is a group of hyperdiverse predatory marine mollusks. More than 15,000 species were described so far, representing 22% of the gastropod diversity (Lemarcis et al. 2022). Beside their high phylogenetic and taxonomic diversity, they are also characterized by a remarkable disparity in feeding strategies which include—but are not restricted to—envenomation, asphyxia, and vampirism (Taylor et al. 1980; Modica and Holford 2010; Olivera et al. 2017). Indeed, several neogastropod species were found to produce molecular compounds to subdue their preys (Modica et al. 2018; Turner et al. 2018), with

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

Article

Conus species undeniably representing the most studied lineage of neogastropods, and one of the gold standards in venom research. Cone snails represent around 1,000 species (WoRMS Editorial Board 2022), are widely distributed, and have different hunting targets (Zhao and Antunes 2022). Thus, more and more toxins ("conotoxins") were found (~200 toxins produced per species) in *Conus* species (Balaji et al. 2000; Buczek et al. 2005; Akondi et al. 2014), some of which entered clinical trials that, in one case, led to the market: the synthetic analog of a conotoxin, Ziconotide, is currently commercialized as an anesthetic (Layer and McIntosh 2006; Fedosov et al. 2012; Safavi-Hemami et al. 2019; Reynaud et al. 2020; Zhao and Antunes 2022).

Conotoxins are small peptides (usually shorter than 40aa) and possess a conserved gene structure composed of an N-terminal signal sequence, a pro-region, and a mature peptide (Kaas et al. 2010; Puillandre et al. 2012; Fedosov et al. 2021). The majority of them are rich in cysteine residues, arranged in specific patterns ('cysteine frameworks'), resulting in disulfide bounds that determine the precise 3D conformation and high stability of the toxins themselves (Jin et al. 2019). Conotoxins were shown to have diverged, at least partly, through the onset of gene variants, especially due to insertions and deletions (Yang and Zhou 2020). Toxins closely related to conotoxins or sharing the same structural conformation were also found in other Conoidea (e.g., in Turridae, with longer mature peptide regions [Watkins et al. 2006], in Terebridae [Gorson et al. 2015], and in Drilliidae [Lu et al. 2020]). Beside conotoxins, a large variety of toxins were described in the Neogastropoda, acting as pore forming (e.g., colutoxins [Gerdol et al. 2018], hemolytic [e.g., echotoxins; Shiomi et al. 2002]) or anesthetic toxins like ShKT (Gerdol et al. 2019), to cite only few of their putative functions. Toxins were found in species belonging to multiple neogastropod families, including Babyloniidae, Muricidae, Buccinidae, Colubrariidae (Turner et al. 2018), Mitridae (Reynaud et al. 2020), and Costellaridae (Kuznetsova et al. 2022) but were reported in some of the closely related Caenogastropoda as well, namely the Tonnoidea families, Cassidae, Ranellidae, and Charoniidae (Hwang et al. 2021; Jia et al. 2022). However, the diversity of these toxins, the genomic mechanisms that may have triggered their evolution, and their potential adaptive and evolutionary implications remain understudied in non-Conoidea neogastropods. In particular, for a comprehensive evaluation of the molecular mechanisms of venom toxins evolution, a genomic perspective is indeed required. However, only a few comparative genomic studies are available to date on venomous organisms (Drukewitz and von Reumont 2019; Zancolli and Casewell 2020; Almeida et al. 2021; Barua and Mikheyev 2021; Koludarov et al. 2022), suggesting that duplication and neofunctionalization events constitute crucial steps in the onset of toxins from ancestral genes (Wong and Belov 2012; Giorgianni et al. 2020). If gene duplication can lead to evolutionary novelties even when restricted to a single gene (Magadum et al. 2013), large-scale duplication events indeed can dramatically increase the adaptative potential of species (Ren et al. 2018).

Polyploidy or whole genome duplication (WGD) is a large-scale duplication event that generates additional copies of an entire genome within a species. This phenomenon is widespread in plant genomes (del Pozo and Ramirez-Parra 2015) but has also been recognized through multiple other eukaryotic lineages (Sankoff 2001). WGD is commonly followed by massive gene loss or duplication and other genomic rearrangements (Sémon and Wolfe 2007), which would imply the emergence of evolutionary novelties in such species. A hypothesis was suggested in a prior study that WGD may have taken place in Neogastropoda and some closely related superfamilies of Caenogastropoda. This was determined through a predictive model based on karyological data, which employed a likelihood method that modeled the background rate of chromosome number evolution as a birth-death process (Hallinan and Lindberg 2011). A recent publication confirmed WGD in *Conus ventricosus* genome, where 35 pseudochromosomes with a total genome size of 3.5 Gb were described, evidenced by the presence of syntenic blocks with *Pomacea canaliculata*, an Ampullarioidea species, where one pseudochromosome of *P. canaliculata* represented two to four pseudochromosomes of *C. ventricosus*, a duplicated cluster of Hox genes, and a Ks (synonymous substitutions) distribution peak (Pardos-Blas et al. 2021).

Given the high potential of WGD to foster diversification (Moriyama and Koshiba-Takeuchi 2018), it is crucial to understand whether this particular event is restricted to *Conus* or if it occurred in other lineages across the Neogastropoda, possibly related to the hyperdiversity of the entire group. To answer this question, high-quality genomes, assembled at the chromosomal level, are required.

As sequencing cost is decreasing and sequencing quality is improving, it became easier to obtain chromosome-level assemblies in nonmodel organisms with the use of long-read technologies (PacBio or Nanopore) combined with Hi-C or Omni-C links (Ghurye et al. 2019). Thus, gastropod genomes are increasingly being sequenced, summing up to ~45 genomes, eight of which being of neogastropods (Stockwell et al. 2010; Hu et al. 2011; Barghi et al. 2016; Pardos-Blas et al. 2021; Peng et al. 2021); however, their taxonomic coverage is mostly restricted to Conoidea species, and their quality is uneven. The only available chromosome-level genome to date is from *C. ventricosus* (Pardos-Blas et al. 2021), even though the BUSCOs values for the annotation were low. Outside Neogastropoda, four Ampullarioidea species have good-quality genome assemblies (Sun et al. 2019), among which *P. canaliculata* is the only one with chromosome-level assembly, which makes it a good candidate for comparative analysis (Liu et al. 2018).

Here, we generated chromosome-level assemblies of two additional gastropod species: *Stramonita haemastoma* (Linnaeus, 1767) from the neogastropod superfamily

Muricoidea and *Monoplex corrugatus* (Lamarck, 1816), belonging to the Tonnoidea superfamily, closely related to the Neogastropoda (Lemarcis et al. 2022) and also including species able to produce venom (Bose et al. 2017). Both species are known to be venomous marine predators and are good candidates to unveil the molecular diversity of venomous gastropods, as they are taxonomically distant both from each other and from *Conus* (Lemarcis et al. 2022). We first investigated WGD in both genomes by analyzing synteny, hox gene clusters, and Ks distribution to confirm the taxonomic span of such event in Neogastropoda but also in Tonnoidea. Then, we looked for potential toxin-related genes in both genomes using toxin databases as Conoserver (Kaas et al. 2012). We also performed gene expression analysis of *S. haemastoma* genes in order to obtain insights about their evolutionary patterns.

Results

Chromosome-Level Genomes

Respectively, 359 (120 ×) and 195 Gb (89 ×) long reads (PacBio) were generated to assemble a 3 and 2.2 Gb draft genomes of *M. corrugatus* and *S. haemastoma*, falling within the size range of other neogastropod genomes (table 1) and significantly higher than other Caenogastropoda, Patellogastropoda, Neomphalina, and Vetigastropoda genomes available to date (supplementary table S1, Supplementary Material online). We used the Dovetail Omni-C method to achieve chromosome-level assemblies for both species, resulting in 35 pseudochromosomes each. No previous report was available in the literature on the karyotype of Tonnoidea species, including *M. corrugatus*, whereas only 36 karyotypes have been described for neogastropod species, none of them being *S. haemastoma* (supplementary table S2, Supplementary Material online). Among the species belonging to the Muricidae family (which includes *S. haemastoma*), most have 35 chromosomes (12 species over a total of 21, see supplementary table S2, Supplementary Material online). Actually, most of the neogastropods possess between 30 and 43 chromosomes, with the exception of *Strigatella litterata*, *Nucella lapillus*, and *Myurella nebulosa* ($n = 19$, 14–16, and 17, respectively). The single chromosome-level neogastropod genome (of the 7 available; see supplementary table S1, Supplementary Material online), from *C. ventricosus* (Pardos-Blas et al. 2021), also presented 35 assembled pseudochromosomes, congruently with our results (supplementary table S2, Supplementary Material online).

We retrieved an N50 value of 87 Mb for *M. corrugatus* and 58 Mb for *S. haemastoma* (table 1) indicating a high contiguity of the assemblies when compared for example with *C. tribblei* with a N50 value of 2.7 Kb. In total, 91% and 89% metazoan BUSCO genes, respectively, were found in these assemblies, indicating a valuable completeness of the genomes, even higher than that obtained for *C. ventricosus* (84.7%). When looking at single genes only (supplementary table S1, Supplementary Material online), found in the assemblies using BUSCO, the values are similar to *C. ventricosus* (82%, 84%, and 85% in *C. ventricosus*, *S. haemastoma*, and *M. corrugatus*, respectively), whereas duplicated genes number doubles in *M. corrugatus* (7%) and *S. haemastoma* (5%) compared with *C. ventricosus* (3%). This might indicate a higher purge of genes after the WGD *in C. ventricosus* than in *M. corrugatus* and *S. haemastoma*. After masking repeated elements, representing 61% and 40% of each genome, respectively, we predicted 36,924 and 34,865 protein-coding genes (table 1 and supplementary table S1, Supplementary Material online for more details). The global statistics of the annotations were similar to *C. ventricosus*, including the number of predicted genes which is, however, slightly higher than what has been reported in the other caenogastropods (e.g., in *C. ventricosus* or *Batillaria attramentaria*) and significantly higher than in *P. canaliculata* (2 and 1.8 times higher for *M. corrugatus* and *S. haemastoma*, respectively). The average number of exons per coding sequence (CDS), number of introns per CDS, and CDS length were lower in *M. corrugatus* and *S. haemastoma* compared with *P. canaliculata* and other Ampullarioidea (table 1). We also observed a higher proportion of CDS in our genomes, with CDS BUSCO scores of 84% and 79%, respectively, compared with the

Table 1. Assembly and Annotation Statistics of *Monoplex corrugatus*, *Stramonita haemastoma* (from This Study) Compared with *Conus ventricosus*, *Pomacea canaliculata*, and *Achatina fulica*.

Species	Assembly Length (Gb)	N50 Mb (L50)	BUSCO Score on Assembly	GC Content (%)	Number of Protein-Coding Genes	Exons/ cds	Introns/ CDS	CDS Length	Exon Length	Intron Length	BUSCO Score on Annotation
Monoplex corrugatus	3.075	86.8 (15)	C: 91.4% (S: 84.6%, D: 6.8%)	41	36,924	4.1	3.1	1,128	281	1,621	C: 83.6% (S: 83.3%, D: 0.3%)
Stramonita haemastoma	2.236	58.4 (16)	C: 89.1% (S: 83.8%, D: 5.3%)	42	34,865	4.8	3.8	1,141	243	1,229	C: 78.7% (S: 77.1%, D: 1.6%)
Conus ventricosus	3.592	93.5 (16)	C: 84.7% (S: 81.8%, D: 2.9%)	43	31,591	4.6	3.6	1,126	246	1,603	C: 32.2% (S: 31.6%, D: 0.6%)
Pomacea canaliculata	0.447	32.6 (6)	C: 97.3% (S: 96.5%, D: 0.8%)	39	18,265	9.3	8.3	1,671	263	1,542	C: 89.1% (S: 88.4%, D: 0.7%)
Achatina fulica	1.856	59.6 (13)	C: 93.4% (S: 87.7%, D: 5.7%)	39	23,726	7.1	6.1	1,559	271	3,793	C: 89.8% (S: 82.9%, D: 6.9%)

NOTE.—N50 values are in Mega base, BUSCO score column reports complete BUSCO sequences (C) number, that is the sum of single BUSCO sequences (S) and duplicated BUSCO sequences (D).

reported score of 32% for *C. ventricosus*, which could be attributed in part to sequencing errors for the *C. ventricosus* genome, as the sequencing depth was of 54X only (table 1; Pardos-Blas et al. 2021).

Duplication Events in Neogastropoda and Tonnoidea

We detected paralogous genes in *M. corrugatus*, *S. haemastoma*, *C. ventricosus*, and *P. canaliculata* using the best reciprocal hit (BRH) method, as well as orthologs shared between species pairs (fig. 1). We plotted the distribution of the identity percent for each comparison of the orthologs between the four species and found that orthologs between *M. corrugatus* and *S. haemastoma* had higher median and average identity percent than orthologs between *C. ventricosus* and the two other species *M. corrugatus* and *S. haemastoma* (supplementary fig. S1, Supplementary Material online).

To further detect duplication events, we searched for conserved synteny among the chromosome-level assemblies and compared them with *P. canaliculata* (fig. 2). MCScanX detected collinear orthologs (called segmental) in each pairwise comparison (see numbers on fig. 1). Among these species, almost all pseudochromosomes of each species had homologous regions detected in the two other species; for example, all pseudochromosomes numbered 1 (largest pseudochromosomes of *M. corrugatus*, *S. haemastoma* and *C. ventricosus*) matched each other's, with some translocation events in pseudochromosomes 2 or 3 (fig. 2). More interestingly, each of *P. canaliculata* pseudochromosomes displayed synteny with more than one pseudochromosome in *C. ventricosus*, *S. haemastoma*, and *M. corrugatus* but the chromosome 14. For example, syntenic blocks of *P. canaliculata* pseudochromosome 7 were retrieved in pseudochromosomes 1 and 2 in *M. corrugatus*, 1 and 3 in *S. haemastoma*, and 1 and 2 in *C. ventricosus*; pseudochromosome 3 had syntenic blocks occurring in pseudochromosomes 6, 10, 19, and 24 in *M. corrugatus*, 6, 11, 21, and 22 in *S. haemastoma* and 7, 8, 20, and 22 in *C. ventricosus*. In summary, ten pseudochromosomes of *P. canaliculata* had homologous regions in exactly two pseudochromosomes of *C. ventricosus*, *S. haemastoma*, and *M. corrugatus*, and three pseudochromosomes had homologous regions in exactly four pseudochromosomes of *C. ventricosus*, *S. haemastoma*, and *M. corrugatus*. Genes found in chromosome 14 of *P. canaliculata* (990 annotated genes) had 136, 140, and 122 orthologs in *M. corrugatus*, *S. haemastoma*, and *C. ventricosus*, respectively. Although syntenic blocks were not conserved in this chromosome, most of the orthologs were found in the pseudochromosomes 29 and 34 of *M. corrugatus*, in the pseudochromosomes 24 and 31 of *S. haemastoma*, and in the pseudochromosomes 32 and 34 of *C. ventricosus*, all pseudochromosomes which had none or very few syntenic blocks (fig. 2). Although a strong synteny was found between each of these species' pseudochromosomes (fig. 2), very few syntenic blocks were detected among the homologous pseudochromosomes within a species (supplementary fig. S2, Supplementary Material online). This is in contrast to *Achatina fulica* (Liu et al. 2021), where synteny is strongly conserved between homologous pseudochromosomes within the species.

To further validate the presence of a putative WGD as suggested by our results, showing a higher number of pseudochromosomes in Neogastropoda and in closely related Caenogastropoda, and in agreement with reported chromosome counts retrieved from the previously published in the relevant literature (supplementary table S2, Supplementary Material online), we employed an independent approach based on the analysis of the number of Ks (synonymous substitutions) between paralogs within a genome. This approach is used to estimate the timing of gene duplication events by comparing the number of synonymous substitutions between paralogs (synonymous divergence) to estimate the time since the gene duplication event occurred (Van de Peer et al. 2009). Synonymous substitutions, which are changes in the DNA sequence that do not result in a change in the amino acid sequence of the protein, are used because they are less likely to be affected by natural selection than nonsynonymous substitutions, which do change the amino acid sequence of the protein. Here, the distribution of Ks between all paralogous genes in each of the neogastropods and *M. corrugatus* (fig. 3*B*) showed a first high peak corresponding to recent events of duplication as very few substitutions were detected. As very few genes were conserved between two putative homologous pseudochromosomes (ohnologs, see fig. 3*A*)

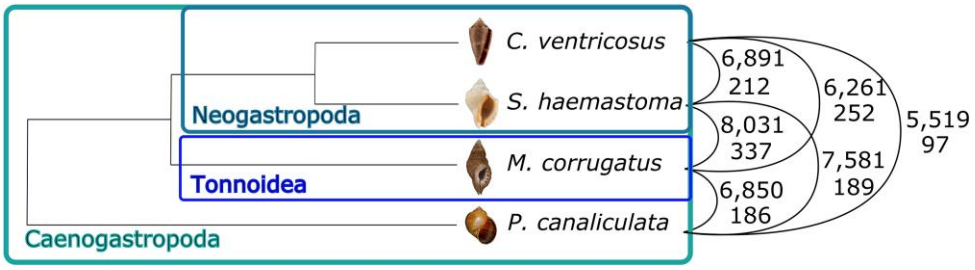

Fig. 1. Schematic representation of the phylogenetic relationships between the species compared in this study. Simplified tree of the four Caenogastropoda species studied in this work. Numbers for each pair correspond to the number of orthologs detected between the two species (top) and the number of syntenic regions detected between the two genomes, all chromosomes considered (bottom). The three first pictures credits: MNHN CC BY 4.0. The picture of *P. canaliculata* was taken from Ng et al. 2017.

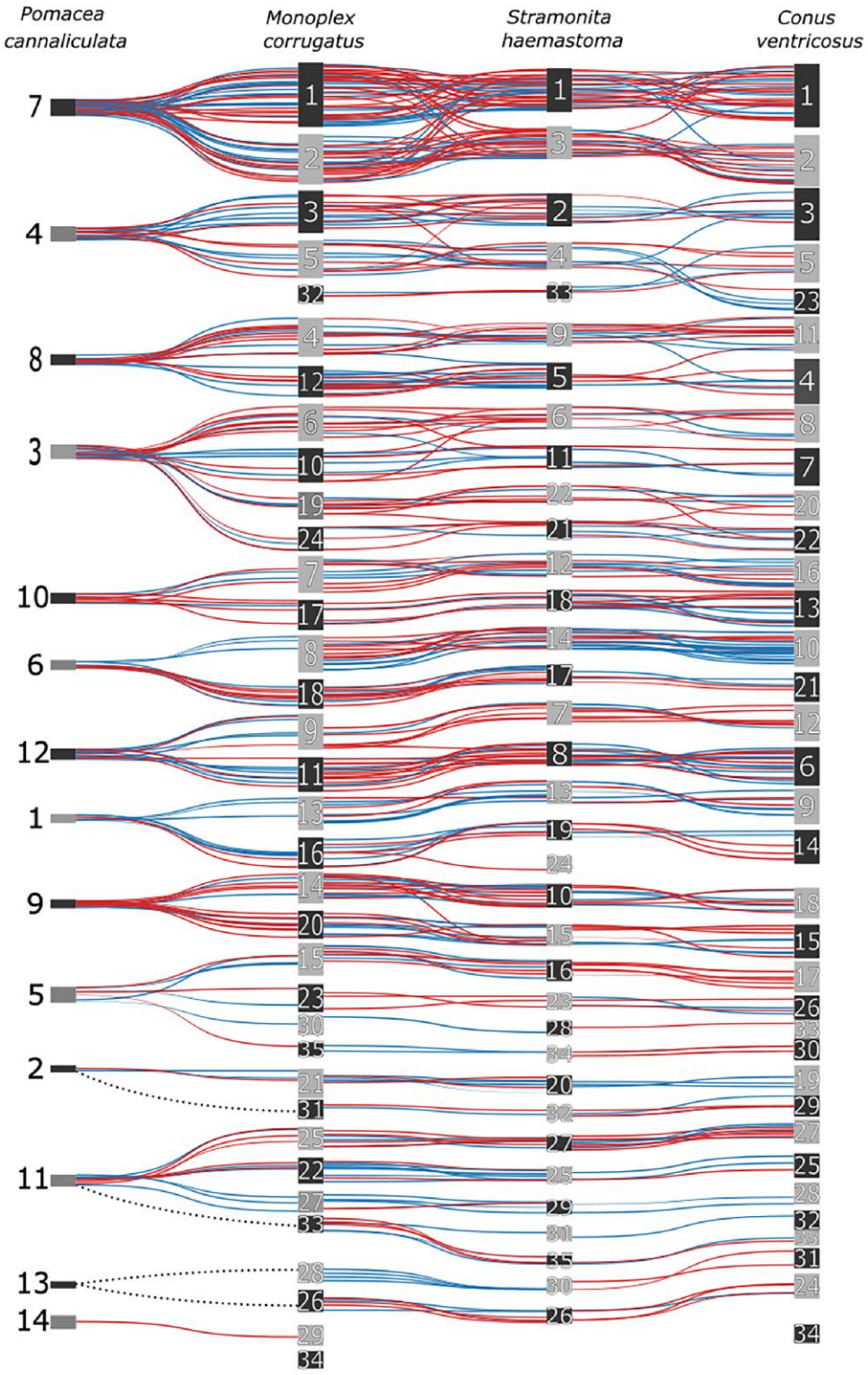

Fig. 2. Syntenic blocks in Caenogastropoda. Representation of the pseudochromosomes (black and gray rectangles with their pseudochromosome numbers) of the four Caenogastropoda generated using Synvisio (Bandi and Gutwin 2020). Each line represents a syntenic block in reverse or in forward orientation. Only links between side-by-side species are represented by full lines, whereas the dotted lines represent links between not juxtaposed species (no synteny detected between the two juxtaposed pseudochromosomes).

in *M. corrugatus* and *S. haemastoma*, we plotted the distribution of Ks between orthologs in a pairwise comparison (fig. 3*B*) to identify in which order the events occurred. Three peaks were detected, where the first peak appears for paralogous genes in each of the three species, the second one for the paralog and ortholog genes within and between the three species, and the third one for the orthologs between *P. canaliculata* and each of the three species.

Homeobox Genes

Homeobox genes are involved in the early stages of embryonic development (Gehring et al. 1994). For our target species, no RNAseq data of development stages from previous studies were available yet, and thus the automatic gene annotation failed to detect them. Therefore, we manually annotated them, not only in the two genomes generated in this study but also in other genomes (*P. canaliculata*,

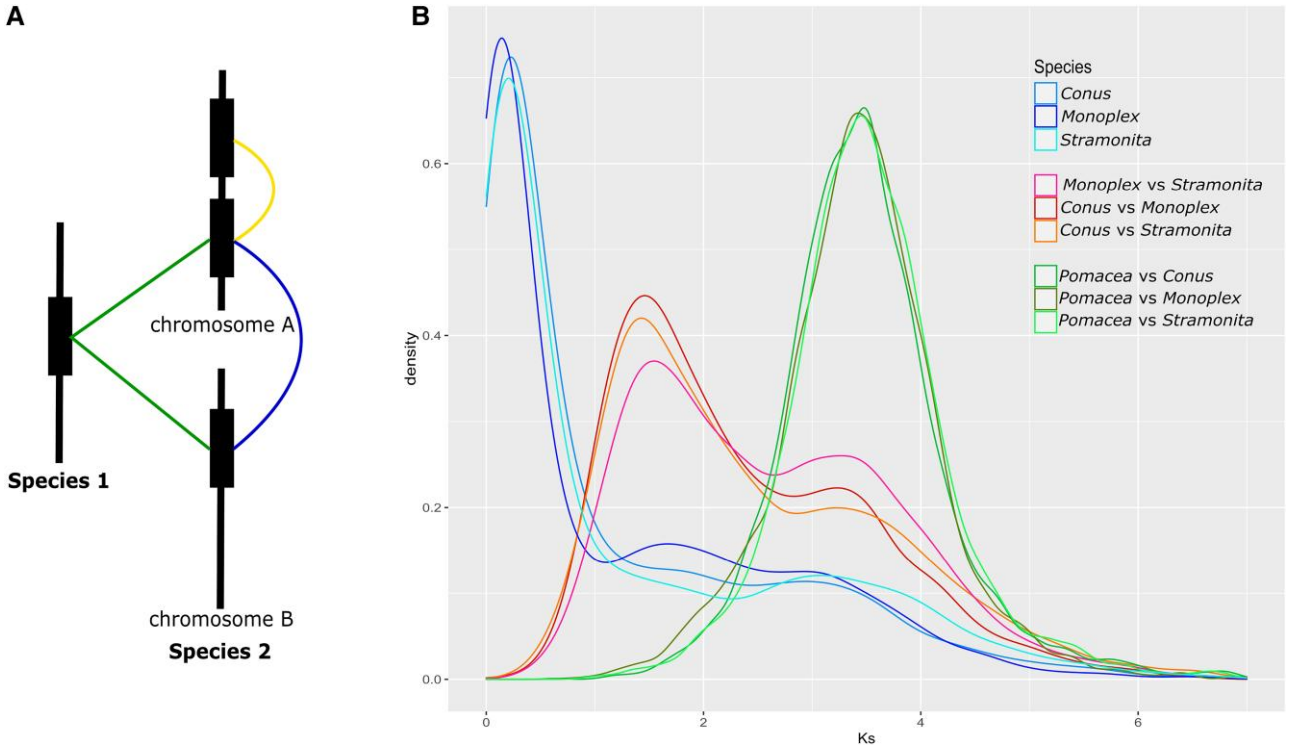

Fig. 3. (*A*) Orthologs, paralogs, and ohnologs. The rectangle on a line represents a gene on a chromosome. Lines from species 1 to species 2 identify orthologs; the upper curved line defines paralogs (in tandem on the same chromosome); the lower curved line identify ohnologs (in two homologous chromosomes within a single species). (*B*) Ks density plot. Distribution of synonymous divergence between pairs of paralogs (first three lines, *Conus*, *Monoplex* and *Stramonita*) and orthologs (for all remaining lines). The first peak represents the newest events of duplication of each species (*Conus ventricosus*, *Monoplex corrugatus*, and *Stramonita haemastoma*); the middle peak corresponds to the whole genome duplication event; and the third peak represents the point of divergence between *P. canaliculata* and the three other species.

Chrysomallon squamiferum, *Haliotis rubra*, *Lottia gigantea*, and *Pecten maximus*) for which homeobox genes were not annotated yet. These new annotations showed that the early diverged gastropods *Ch. squamiferum*, *H. rubra*, *L. gigantea*, and the bivalve *P. maximus* had exactly one cluster of homeobox genes with the same order on specific superscaffolds/pseudochromosomes. Anterior group genes (Ant, Lox) are placed on the same superscaffold/pseudochromosome and are followed by posterior group genes, whereas parahox genes (Xlox, Gsx, and Cdx genes) are found on a separate superscaffold/pseudochromosome (fig. 4). Neither *A. fulica* nor *C. ventricosus* homeobox genes were reannotated here (since already described in Liu et al. 2021 and Pardos-Blas et al. 2021), but both show a duplication of homeobox components on two different pseudochromosomes. Although *A. fulica* shows a duplication of a part of the anterior group genes (*Lox5*, *Antp*, *Lox4*, and *Lox2*), Hox1 and 5 in addition to Lox2 and Lox4 were also found as duplicated in *C. ventricosus*. Beside this duplication, *C. ventricosus* genes are arranged in a similar order as in *P. canaliculata* and opposed to *A. fulica*: posterior group genes are in between two types of anterior group genes in *C. ventricosus*, whereas anterior group genes are clustered together and followed by the posterior group genes in *A. fulica*. In *M. corrugatus*, most genes were found duplicated (we could not detect Hox4 and Post1 in Scaffold_30) and arranged in a different order when comparing each hox cluster in this species genome. Fewer genes were found duplicated in hox cluster in *S. haemastoma* (we could not find Lox5, Antp, Lox4, and Lox2 in Scaffold_28), but we found three tandem duplication events (for Hox4 in scaffold_23 and Hox1 and Post2 in Scaffold_28). Interestingly, in addition to the two copies of Hox5, we also detected a homeobox gene closely related to Hox5 in both *M. corrugatus* and *S. haemastoma* (see fig. 4A, "Hox5-like" in *S. haemastoma* and *M. corrugatus*). Our annotated homeobox genes were further verified with a phylogenetic analysis, in which orthologous genes of a given group are clustered together (fig. 4*B*).

Putative Toxins Found within Syntenic Blocks

To target toxin-coding genes potentially resulting from a post-WGD sub or neofunctionalization, we focused on the genes found in syntenic blocks with *P. canaliculata*. From the 166 and 130 putative toxin genes annotated using the in-house toxin database (see Materials and Methods and supplementary table S3, Supplementary Material online) in *M. corrugatus* and *S. haemastoma*, respectively, 16 and 5 genes were found in syntenic blocks in *M. corrugatus* and *S. haemastoma*, respectively (table 2). From the 16 *M. corrugatus* genes, two were discarded as they could not be reliably identified as toxins or toxin-related genes. Then, for each of the 14 remaining genes, we

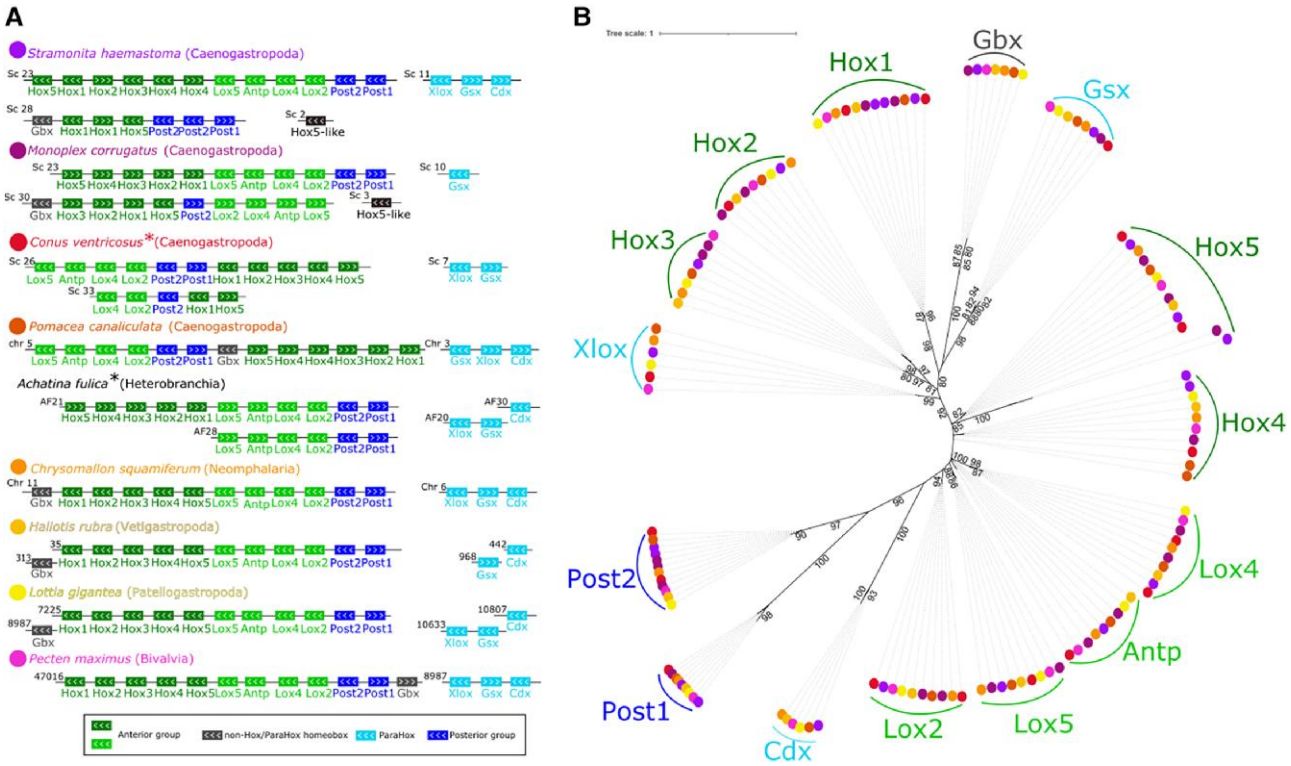

Fig. 4. Hox genes clusters. (*A*) Representation of the gene order of Hox genes; colored boxes with the arrow indicating the gene direction on the pseudochromosome (Chr) or superscaffold (Sc) represented by a line, as detected during the annotation. Star marks species for which data were retrieved from literature (Liu et al. 2020; Pardos-Blas et al. 2021). (*B*) Tree representing the phylogenetic relationships between all the sequences of the analysis; color dots correspond to the species as listed on the (*A*) part of the figure.

searched for their putative orthologs in *S. haemastoma* (if not already found in the toxin annotation results) and in *P. canaliculata*, as well as for the possible paralogs in the homologous pseudochromosome of *M. corrugatus* (as defined by the *P. canaliculata* synteny, fig. 2; see Materials and Methods). Finally, 12 groups of genes related to toxins were analyzed in more details (table 2).

In most cases, we found at least one copy in each species but for the group 5, where the *P. canaliculata* orthologous gene was not detected (table 2). The groups 5, 7, and 9 had exactly one copy in each genome: we could not detect a second copy in the other pseudochromosomes. Genes from group 5 are probable Kunitz domain toxins; group 7 are probable enzyme responsible for posttranscriptional modification (PTM enzyme) known to be related to toxins (Soares and Oliveira 2009), whereas group 9 are probable neurotransmitters. The groups 1, 2, 4, 6, and 12 had all exactly two copies in *M. corrugatus* and *S. haemastoma* genomes, in homologous pseudochromosomes. All genes from group 1 had the best matches with allatotropin precursor genes (Gruber and Muttenthaler 2012) except for *P. canaliculata* gene, which had the best match with itself as it was already been annotated in NR as "hypothetical protein C0Q70_00675." Genes from group 2 had best matches with portions of three genes, which are functionally annotated differently, but all had an astacin domain. This domain is common in venom metalloprotease toxins from different organisms (Trevisan-Silva et al. 2010; Ponce et al. 2016; Gerdol et al. 2019). All genes from group 4 had matches with conkunitzin-M16 that encode for Kv channels-blocking conotoxins (Finol-Urdaneta et al. 2020). However, the sequence from *P. canaliculata* was annotated as "hypothetical protein C0Q70_05870". Genes from group 6 had all their best matches with a FMRFamide precursor from *Charonia tritonis*, a neurotransmitter whose function is unknown but is highly conserved through molluscs (Klein et al. 2021). Finally, genes from group 12 were annotated as neuropeptide as they had all their best matches with sequences annotated as precursor of achatin, a neuromodulator described in *A. fulica* (Liu and Takeuchi 1993), except for *P. canaliculata* in which the function is unknown ("uncharacterized protein LOC112557935"). For the group 10, we found two gene copies in *M. corrugatus* on the respective homologous pseudochromosomes, whereas only one copy was found in *S. haemastoma*. Group 10 genes had hits with precursors of ConoCAP, a cardioactive neuropeptide from cone snails (Möller et al. 2010), where the copies in the homologous pseudochromosomes in *M. corrugatus* are similar, but the only copy found in *S. haemastoma* did not conserve an open reading frame in the genome and had no hits in NR database. The group 11 included duplicated genes in the pseudochromosome 3 of *M. corrugatus*, whereas no other copies were detected in the other pseudochromosomes, similarly to *S. haemastoma* having two copies in the same pseudochromosome 2. They were all annotated

Table 2. Toxins and Related Genes Detected in Syntenic Blocks in *Monoplex corrugatus, Stramonita haemastoma*, and *Pomacea canaliculata*.

Group	Nb Genes Per Species	Description of the Best Match against NR	Putative Function
1	2 *M. corrugatus*	Allatotropin-1 precursor (*Charonia tritonis*)	Conotoxin
	2 *S. haemastoma*		
	1 *P. canaliculata*	Hypothetical protein (*P. canaliculata*)	
2	2 *M. corrugatus*	Hypothetical protein BaRGS_000210 (*Batillaria attramentaria*)	Enzyme toxin
	2 *S. haemastoma*		
	1 *P. canaliculata*	Zinc metalloproteinase nas-13-like (*P. canaliculata*)	
3	8 *M. corrugatus*	Chitinase-like lectin, partial (*Crepidula fornicata*)	Chitinase: enzyme/toxin
	3 *S. haemastoma*	Probable chitinase 10 (*P. canaliculata*)	
	3 *P. canaliculata*	Probable chitinase 10 (*P. canaliculata*)	
4	2 *M. corrugatus*	Conotoxin precursor conkunitzin (*Conus ebraeus*)	Toxin
	2 *S. haemastoma*	Conotoxin superfamily conkunitzin-8 (*C. magus*)	
	1 *P. canaliculata*	Hypothetical protein (*P. canaliculata*)	
5	1 *M. corrugatus*	Kunitz-type peptide 1 (*Bitis atropos*)	Toxin
	1 *S. haemastoma*	Carboxypeptidase inhibitor SmCI-like isoform X2 (*Limulus polyphemus*)	
6	2 *M. corrugatus*	FMRFamide precursor (*Ch. tritonis*)	Neurotrasmitter/toxin
	2 *S. haemastoma*		
	1 *P. canaliculata*		
7	1 *M. corrugatus*	Lysosomal alpha-mannosidase-like (*P. canaliculata*)	Enzyme for PTM: N-linked glycosylation, often present in venom
	1 *S. haemastoma*		
	1 *P. canaliculata*		
8	4 *M. corrugatus*	Protein disulfide isomerase (*C. vexillum*)	Enzyme involved in toxin folding, often present in venom
	3 *S. haemastoma*		
	2 *P. canaliculata*	Protein disulfide-isomerase-like isoform X1 (*P. canaliculata*)	
9	1 *M. corrugatus*	conopressin/conophysin (*Haliclona caerulea*)	Neurotransmitter/toxin
	1 *S. haemastoma*		
	1 *P. canaliculata*		
10	2 *M. corrugatus*	CCAP-1 precursor (*Ch. tritonis*)	Toxin
	1 *S. haemastoma*		
	1 *P. canaliculata*	ConoCAP-like (*P. canaliculata*)	
11	3 *M. corrugatus*	Probable serine carboxypeptidase CPVL (*P. canaliculata*)	Enzyme toxin
	2 *S. haemastoma*		
	1 *P. canaliculata*		
12	2 *M. corrugatus*	Achatin-2 precursor (*Ch. tritonis*)	Neuropeptide, maybe a toxin?
	2 *S. haemastoma*		
	1 *P. canaliculata*	Uncharacterized protein (*P. canaliculata*)	

NOTE.—See supplementary table S3, Supplementary Material online for more details.

as probable serine carboxypeptidase CPVL, known as main enzyme acting in blood digestion (Motobu et al. 2007). Finally, groups 3 and 8 had more than two copies in each of the three species where group 3 is a member of the chitinase gene family and group 8 of the thioredoxin gene family. As an illustration, we choose to highlight group 4 sequences, closely related to conotoxin, and group 1 by representing the syntenic blocks where it was found (fig. 5*A*) and the multiple alignment of their sequences with conotoxin sequences from *C. betulinus* and *C. magus* for group 4 sequences (fig. 5*B*).

Differential Gene Expression in the Neogastropoda

To assess the evolutionary origin of Neogastropoda genes, we performed differential gene expression on *S. haemastoma* transcriptomes. We sequenced three replicates for two tissues, the osphradium and the salivary glands, retrieving a total of 6,412 differentially expressed genes (DEGs) between the two tissues, 2,439 of which were up-regulated in the salivary gland. We also realigned the DEGs on the genome of *S. haemastoma* to detect potential ohnolog genes not previously found in our annotation. For most of the genes (66% and 61%) that were up- and down-regulated, we did not detect a paralog in the whole genome (table 3). For others, we found duplicates in the homologous chromosome (ohnologs). Although the majority of ohnologs had no differential expression, most of the ones found to be differentially expressed had similar expression pattern, that is both copies were either up- or downregulated. Few of the ohnolog copies had opposite expression profiles.

Among the toxins found in the genomes and described in table 2, only three pairs of ohnologs were found differentially expressed (supplementary table S3, Supplementary Material online). The two copies of the genes of *S. haemastoma* in groups 1 and 6 were both downregulated, and the two copies of genes from group 8 were both upregulated in the salivary glands.

Discussion

First Assembled Genomes of Tonnoidea and Muricoidea

The present work is a significant addition to the genome data available to date for Neogastropoda and

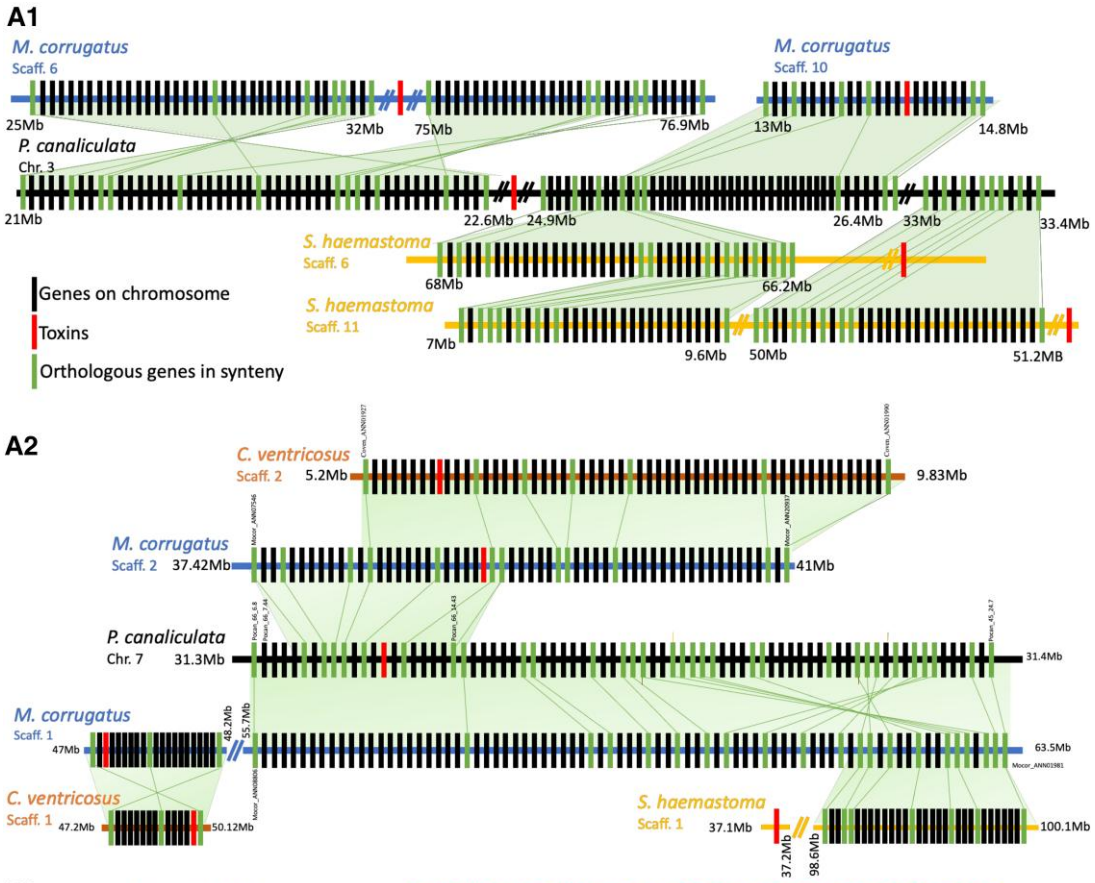

Fig. 5. Synteny block and multiple alignment of toxins similar to conotoxins. (*A*) Representation of syntenic blocks (light shadow) on each superscaffold/pseudochromosome and each species with putative toxins from group 4 (*A1*) and 1 (*A2*). Each horizontal line of a superscaffold/pseudochromosome represents one species. Red blocks are the detected toxins, black blocks are the genes found in the syntenic region, and green blocks are ortholog genes found in synteny (indicated by green lines). A toxin from group 1 found in *Monoplex corrugatus* (scaffold 2) was found in synteny with *Pomacea canaliculata* in a region that also shares a syntenic block with scaffold 1 of *M. corrugatus* close to the region of the paralog toxin. (*B*) Multiple alignment of group 4 (*B1*) and group 1 (*B2*) sequences (supplementary table S3, Supplementary Material online) manually annotated and found in a syntenic block with the addition of two conotoxins from *Conus betulinus* and *C. magus*, generated using MUSCLE v.3.8 (Edgar 2004).

Table 3. Numbers of Differentially Expressed Genes and Their Copies in the Genome.

	Upregulated Genes	Downregulated Genes
Total number of DEG	**2,439**	**3,973**
Number of copies found per DEG	**1,611**	**2,410**
Total copies found as ohnologs	**267**	**376**
Expression of the copies: up | down | not DEG	**62 (1) | 26 | 179**	**19 | 102 (2) | 255**

NOTE.—Up and downregulated in salivary gland compared with osphradium tissues. Numbers between parenthesis in table line "Expression of the copies" are the predicted toxin-coding genes.

Caenogastropoda. These previously included only eight neogastropod species, six of which belonging to Conoidea and two to Buccinoidea superfamily, and six caenogastropod species, four of which from a same family, the Ampullariidae (supplementary table S1, Supplementary Material online). Here, we added two genomes, one from the Muricidae family (Neogastropoda) and one for the Tonnoidea superfamily (Littorinimorpha), closely related to neogastropods (Lemarcis et al. 2022), thus broadening the taxonomic span of the available genomes. With the further addition of the recently sequenced genome of *C. ventricosus* (Pardos-Blas et al. 2021), we can now rely on reference genomes that effectively represent the diversity of predatory gastropods and remarkably expand the available molecular databases. Furthermore, these genomes were assembled at a chromosomal level with high BUSCO scores and thus constitute valuable assets to reconstruct evolutionary processes and patterns at the genomic level.

In addition to the high genome assembly quality, gene annotations provide a reliable nearly complete dataset of genes, whose completeness is comparable with available data on the Ampullarioidea superfamily (see annotation BUSCO score, table 1 and supplementary table S1, Supplementary Material online). General features of these two new genomes do not significantly differ from previously published neogastropods genomes, even though more genes were found, likely as a consequence of the higher quality of the sequencing (fewer base errors). Although gene numbers were 1.8–2 times higher in our two genomes than in Ampullarioidea genomes, almost twice less exons were found per CDS, with a length 1.5 times shorter, indicating a tendency to favor smaller genes in Caenogastropoda, a feature potentially associated to an increased transcription and expression frequency (Lopes et al. 2021). The overall sizes of our newly sequenced genomes were five to six times larger than Ampullarioidea genomes; this observation, coupled with the fact that more than twice pseudochromosomes were detected, strongly implies the occurrence of a WGD event. Together with the proliferation of repeated elements observed in the newly sequenced genomes (from 11% in *P. canaliculata* to 60% and 40% in *M. corrugatus* and *S. haemastoma*), these processes may have considerably increased genetic variation and diversity of gene expression regulation patterns in these species.

WGD in Neogastropoda and Tonnoidea

WGD is an important evolutionary event where all chromosomes are doubled in a genome. It provides increased opportunities for rearrangement, deletion, and insertion events that could contribute to evolvability, the capacity of species to adapt to different environments and diversify (Phuong et al. 2019), ultimately leading to the emergence of adaptive evolutionary novelties (Flagel and Wendel 2009; Van de Peer et al. 2009). Indeed, WGDs have been detected in many species-rich lineages, including, for example teleost fishes (Glasauer and Neuhauss 2014) and orchids (Moriyama and Koshiba-Takeuchi 2018), to name only few of them. In caenogastropods, karyotype analyses have been carried out on several individual species, providing data for a subsequent study that predicted a WGD event in Neogastropoda, Tonnoidea, Cypraeidae, and Capulidae based on genome size and chromosome counts (Hallinan and Lindberg 2011). WGD has been subsequently confirmed in *C. ventricosus* (Pardos-Blas et al. 2021), based on the detection of synteny and colinear blocks, the distribution of Ks peak, and the duplication of Hox gene clusters. Other WGD events were described in gastropods within the Heterobranchia subclass, as for example in *A. immaculata* and *A. fulica* (Liu et al. 2021), each possessing 31 chromosomes. In this work, we detected synteny between the different species (fig. 2), and the comparison with *P. canaliculata* showed a doubling (or sometimes quadrupling) of the pseudochromosome number in each of our target species. Having more than two copies of a chromosome raises the question of multiple WGD, but it can also be explained by duplication events limited to some chromosomes only. To clarify this issue, genome comparison between more species within neogastropods and in closely related caenogastropods like, for example Tonnoidea, Calyptraeidae, or Cypraeidae families (Lemarcis et al. 2022), would be needed. In addition, the exact timing of the WGD event remains to be determined. Our data suggest that the WGD event took place at least prior to the divergence of the Tonnoidea + Neogastropoda lineage (Lemarcis et al 2022) from the rest of the Caenogastropoda. However, according to the hypothesis proposed by Hallinan and Lindberg (2011), the WGD event may have also affected Cypraeidae and Capulidae, implying that it could have occurred earlier in the evolutionary history of Caenogastropoda, if confirmed in other lineages potentially closely related to neogastropods such as Calyptraeoidea and Velutinoidea. Although sequencing the genomes for the missing lineages would provide valuable insights going well beyond the scope of this work, such an endeavor would be highly costly and time consuming, given the high species diversity of the group; an estimation of their chromosome number and/or genome size could be a more affordable approach to determine the origin and nature of this WGD (Hallinan and Lindberg 2011).

The occurrence of syntenic blocks between the analyzed species (*M. corrugatus* and *S. haemastoma*) and *P. canaliculata*, the Ks distribution, and the duplication of Hox genes found within the species we analyzed strongly suggest WGD, with the same evidences found in *Achatina* species and *C. ventricosus* (Liu et al. 2020; Pardos-Blas et al. 2021). Although the first higher peak of the Ks distribution would correspond to the newly duplicated genes within each species, the second smaller peak would correspond to a WGD in neogastropods and tonnoideans, whereas the third peak would correspond to divergence between neogastropods, Tonnoidea, and the other caenogastropods. Hox clusters were found in both *M. corrugatus* and *S. haemastoma*, and both showed a duplication in the pseudochromosomes found to be homologs with those of *P. canaliculata*. In those clusters, we can identify some differences, possibly due to rearrangements, deletions, and/or insertions. All these evidences confirm the WGD event previously suggested (Bose et al. 2017; Pardos-Blas et al. 2021) in neogastropods and in the Tonnoidea, for which data concerning karyotypes or genome sequencing were completely lacking until the present study. The presence of only few colinear blocks within *M. corrugatus* and *S. haemastoma* (supplementary fig. S2, Supplementary Material online), in opposition of what was found in *Achatina*, strongly suggests a major role for other evolutionary events after the WGD. Indeed, genomic doubling would imply failure of cell division; thus, interchromosomal rearrangements are suggested to be the most frequent events after a WGD (Otto and Whitton 2000). Additionally, a WGD implies the retention of two copies of a gene or a repertoire of genes (Sémon and Wolfe 2007) which could lead to neofunctionalization and subfunctionalization events, as it has been reported in plants (He and Zhang 2005).

When a gene is duplicated, one of the copies can acquire a new function (neofunctionalization), retain a subset of the original gene's functions (subfunctionalization), or degenerate to the point of losing its function (nonfunctionalization) (Force et al. 1999; Kuzmin et al. 2022). Gene expression analysis (Pasquier et al. 2017), structural analysis (Voordeckers et al. 2012), or experimental manipulation (Hittinger and Carroll 2007) can be applied to determine whether gene duplicates have undergone subfunctionalization, neofunctionalization, or nonfunctionalization. Since it is still a challenge to maintain neogastropods alive in the laboratory and available data are too limited to reliably predict protein structure in such nonmodel organisms, we focused on gene expression in two available tissues from *S. haemastoma* to gain insights on the fate of duplicated genes, the salivary glands (which are often involved in venom secretion, (Ponte and Modica 2017), and the osphradium, a sensory organ located in the molluscan mantle (Lindberg and Sigwart 2015). However, as synteny was rarely conserved between the homologous chromosomes within a given species, in most cases we were not able to detect the second copy of a gene, suggesting that the copy diverged enough to be detectable or has been depleted and lost its function, which is the most common event after WGD (Pasquier et al. 2017). Nonetheless, when the two copies were detected, most of the ohnologs that were found differentially expressed in the salivary gland compared with the osphradium had similar expression patterns. Whether these similar expression profiles correspond to sub- or neofunctionalization would require expression data from species that diverged before the WGD event, as described in snakes, whose venom evolution occurred through gene duplication and subfunctionalization of genes encoding salivary proteins (Hargreaves et al. 2014). Furthermore, our gene expression analyses were limited to two tissues, whereas expression profiles in a wide range of tissues and/or different developmental stages would be needed to fully understand how expression patterns of ohnologs are correlated. Future studies would ideally include structural characterization of proteins produced by each gene copy and functional assays to test for differences in their biological activity. Such a multidisciplinary approach would provide a more comprehensive understanding of the roles and functions of the duplicated genes in the organism.

Toxin Evolution

Toxins are peptides or proteins, produced or stored by an organism, which interfere with physiological or biochemical processes in another organism (Olivera et al. 2015; Arbuckle 2018). They were found to be produced by multiple species across the eukaryote kingdom (Casewell et al. 2013; Fry et al. 2015) and were known to serve multiple functions. In this work, we were attentive to find putative toxin candidates in both *M. corrugatus* and *S. haemastoma*. Multiple hits of toxins or related genes were found in both genomes. Among these genes, at least one group of genes (group 1) was found closely related to a conotoxin (fig. 5*B*), in both *M. corrugatus* and *S. haemastoma*; its occurrence was also reported in another neogastropod species belonging to the genus *Vexillum* (Costellariidae; Kuznetsova et al. 2022). This finding might indeed suggest that the appearance of conotoxin-related peptide could be a shared evolutionary novelty of predatory gastropods, predating the recruitment of such toxins before the origin of Conoidea, where they might indeed have gone through multiple duplication-neofunctionalization rounds to attain the observed diversity. More studies are needed also in other neogastropods (and related predatory marine snails) to confirm that the capacity to produce toxins is not a key innovation limited to cone snails or conoideans, as traditionally seen, but can be extended to the whole neogastropods (and close relatives) making them one of the most, if not the most, diverse group of venomous animals on Earth.

As we focused only on genes found in syntenic blocks, it should be noted that our findings constitute a limited and preliminary assessment of toxin repertoire in *M. corrugatus* and *S. haemastoma*. Moreover, our analysis was based on

sequence similarity of toxin databases against the genomes and gene expression in *S. haemastoma*, but gene expression studies and proteomic analyses are needed in order to confirm their expression in secretory organs used in prey capture. Our DEG analysis actually suggests that many genes are up or down-regulated in the salivary glands that are involved in toxin production (Ponte and Modica 2017). Furthermore, the functions of these genes related to conotoxins, toxins, and related genes found in *M. corrugatus* and *S. haemastoma* are yet to be confirmed by testing their potential functions in vitro or in vivo. Nonetheless, identifying such sequences in the genome helps in understanding their evolutionary origin and history. Interestingly, the toxin genes on which we focused all had sequence similarities with *P. canaliculata*, the apple snail. This freshwater snail is extremely polyphagous, feeding mostly on vegetal material (Estebenet and Martin 2002) but no evidences indicate a possible toxin production. Thus, it is likely that these genes encode for proteins having other functions in *P. canaliculata*, and their duplication during the WGD followed by neofunctionalization events led to the apparition of new functions, including toxins. Duplication events, followed by sub and/or neofunctionalization, has been proposed as the main mechanism driving the evolution of venom genes in different clades, including spiders (Haney et al. 2016), snakes (Hargreaves et al. 2014; Giorgianni et al. 2020), and cone snails (Duda and Palumbi 1999; Chang and Duda 2012), especially in presence of positive selective forces (Kordiš and Gubenšek 2000). Growing evidence support the hypothesis of a relationship between WGD events and the emergence of evolutionary novelties, despite other factors, such as local duplications (Zhang et al. 1998) and co-option of single genes (Martinson et al. 2017), might be involved in the onset of adaptive traits. It has been argued that the ability of WGD to convey selective advantages and drive the appearance of evolutionary novelties would only emerge in presence of drastic environmental changes, remodeling ecological communities and increasing the availability of new niches (Moriyama and Koshiba-Takeuchi 2018). Indeed, in amphibian genomes, a potential correlation between WGD, adaptive radiation, and the well-known mass extinction event at the K-Pg boundary has been reported (Mable et al. 2011). In fishes and plants genomes, WGD that could have predated adaptive radiation could have been triggered by subsequent environmental changes, as proposed in the "WGD radiation lag-time model" (Eric Schranz et al. 2012). Interestingly, fossil data strongly suggest that the emergence of Neogastropoda and carnivorous caenogastropods, and their rapid radiation, could be placed in the context of the major reorganization of the benthic marine communities known as the Mesozoic marine revolution (Vermeij 1977; Bouchet et al. 2002; Harper 2003), where the evolutionary potential offered by the WGD could have been adaptive to fully exploit novel niches. Expanding genomic characterization to a broader array of taxa within the caenogastropods could help shedding light into the role of WGD and major genomic rearrangements in the origin of hyperdiverse groups.

Materials and Methods

DNA and RNA Samples

One female (MNHN-IM-2019-21838) and one male (MNHN-IM-2019-21839) specimen of *M. corrugatus* were collected in Corsica, France, whereas one female (MNHN-IM-2019-21840) and two male (MNHN-M-2019-21841 and MNHN-IM-2019-21842) specimens of *S. haemastoma* were collected in Capo Linaro, Italy. The foot (muscular tissue only) and the columellar muscle associated with the cephalopodium (pooled samples 1020.1 and 1020.5) of the *M. corrugatus* MNHN-IM-2019-21838 were used for DNA sequencing, whereas the heart (samples 1020.2 and 1020.15), kidney (samples 1020.3 and 1020.16), and ctenidium (samples 1020.4 and 1020.17) of each of the two specimens, respectively, were kept for RNA sequencing. For *S. haemastoma*, the heart, the capsule gland, the gland of Leiblein, the kidney, and ctenidium samples (pooled samples 0620.59, 0620.61, 0620.65, 0620.70, and 0620.72) from the specimen MNHN-IM-2019-21840 were used for DNA sequencing, and the osphradium (samples 0620.38, 0620.51 and 0620.71) and salivary gland (samples 0620.43, 0620.54, and 0620.68) of each of the three specimens, respectively, were kept for RNA sequencing. Samples were sent to Dovetail (Scotts Valley, California, USA) for library preparation, sequencing, assembly, and annotation.

PacBio Library Preparation, Sequencing, and Assembly

Both *M. corrugatus* and *S. haemastoma* DNA samples were generated following extraction protocol performed at Dovetail (Grohme et al. 2018) and were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20 kb) for PacBio Sequel was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) using the manufacturer-recommended protocol. The library was bound to polymerase using the Sequel II Binding Kit 2.0 (PacBio) and loaded onto PacBio Sequel II. Sequencing was performed on three PacBio Sequel II 8 M SMRT cells generating 359 and 195 Gb of data (*M. corrugatus* and *S. haemastoma*, respectively). The reads were used as an input to Wtdbg2 v2.5 (Ruan and Li 2020) with a genome size of 3.0gb, a minimum read length of 20 Kb, and a minimum alignment length of 8,192. Additionally, purge_dups v1.2.3 (Guan et al. 2020) was used to remove haplotigs and contig overlaps in the resulted assembly.

Omni-C Library Preparation, Sequencing, and Scaffolding

For each Dovetail Omni-C library, chromatin was fixed in formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DNAse I, chromatin ends

were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After proximity ligation, crosslinks were reversed and the DNA purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeqX platform to produce an approximately 30 × sequence coverage. The input de novo assemblies and Dovetail OmniC libraries reads were used as input data for HiRise with MQ > 50 as parameter, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). Dovetail OmniC library sequences were aligned to the draft input assemblies using BWA (Li and Durbin 2009). The separations of Dovetail OmniC read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and to make joins above a threshold.

RNA Sequencing

Total RNA extraction was done using the QIAGEN RNeasy Plus Kit following manufacturer protocols. Total RNA was quantified using Qubit RNA Assay and TapeStation 4200. Prior to library prep, a DNase treatment was performed, followed by AMPure bead cleanup and QIAGEN FastSelect HMR rRNA depletion. Library preparation was done with the NEBNext Ultra II RNA Library Prep Kit following manufacturer protocols. Then, these libraries were run on the NovaSeq6000 platform in 2 × 150 bp configuration.

Gene Annotation Pipeline

Repeat families found in the genome assemblies of *M. corrugatus* and *S. haemastoma* were identified de novo and classified using the software package RepeatModeler, v.2.0.1 (Flynn et al. 2020). The custom repeat libraries obtained from RepeatModeler, v.2.0.1 (Flynn et al. 2020), were used to discover, identify, and mask the repeats in the assembly files using RepeatMasker, v.4.1.0 (Smit et al. 2013). CDSs from published annotations (*Aplysia californica*, *Biomphalaria glabrata*, *Crassostrea virginica*, *L. gigantea*, and *Octopus bimaculoides)* were used to train the initial ab initio model for both species using the AUGUSTUS software, v.2.5.5 (Stanke et al. 2006). Six rounds of prediction optimization were done with the software package provided by AUGUSTUS, v.2.5.5 (Stanke et al. 2006). The same CDSs were also used to train a separate ab initio model for each genome using SNAP, v.2006-07-28 (Korf 2004). RNAseq reads were mapped onto the genome using the STAR aligner software, v.2.7 (Dobin et al. 2013), and intron hints generated with the bam2hints tool within the AUGUSTUS software, v.2.5.5. MAKER, v3.01.03 (Cantarel et al. 2008), SNAP and AUGUSTUS (with intron-exon boundary hints provided from RNA-Seq) were then used to predict genes in the repeat-masked reference genome. To help guide the prediction process, Swiss-Prot peptide sequences from the UniProt database were downloaded and used in conjunction with the published protein sequences listed above to generate peptide evidence in the Maker pipeline. Only genes that were predicted by both SNAP and AUGUSTUS were retained in the final gene sets.

Duplications, Syntenic Regions, and Ks Distribution

Paralogs and orthologs were detected using the BRH method. Briefly, using BLASTp v2.13.0 (Camacho et al. 2009), all proteomes of *M. corrugatus*, *S. haemastoma*, *C. ventricosus*, and *P. canaliculata* were aligned to each other's with a maximal evalue of 1e−10 and at least 50% of one of the two sequences length aligned. Duplications within a species genome (*M. corrugatus*, *S. haemastoma*, *C. ventricosus*, *P. canaliculata*) and syntenic regions between two different species (pairwise comparison) were assessed using MCScanX v.2 (Wang et al. 2012). First, alignments between proteins from the same species and proteins from two different species were computed using BLASTp v2.13.0 (Camacho et al. 2009) with the parameters "number of alignments" equal to 5 and with a "maximal e-value" of 1e−10. The resulted BLAST output and GFF files were given to MCScanX v2 (Wang et al. 2012) scripts (McscanX, duplicate_gene_classifier, detect_collinear_tandem_arrays and dot_plotter) with default parameters. To visualize syntenic regions, online synvisio program (Bandi and Gutwin 2020) was used to generate the figures using the collinearity file resulted from MCScanX v.2. We defined homologous pseudochromosomes within *C. ventricosus*, *M. corrugatus*, and *S. haemastoma* resulting from the syntenic regions shared with *P. canaliculata*. All orthologs defined by the BRH methods and found in homologous pseudochromosomes had their protein sequences aligned between them using MAFFT v.7.471 with default parameters (Katoh and Frith 2012). Then, the protein multiple alignments were associated to the corresponding DNA sequence into a codon alignment using pal2nal with -nogap option. The number of synonymous substitutions per synonymous sites (Ks) for each BRH were calculated using KaKs_Calculator, v2.0 (Wang et al. 2010). The Ks distribution was plotted using ggplot package from R v3.6.3 (R Core Team 2022) with density function.

Homeobox Gene Annotations

Hox genes were specifically annotated in *M. corrugatus*, *S. haemastoma*, and in four species representatives of different gastropod subclasses, having their genomes sequenced and available: *P. canaliculata* (Caenogastropods; Architaenioglossa), *Ch. squamiferum* (Neomphalina), *H. rubra* (Vetigastropoda), and *L. gigantea* (Patellogastropoda). In addition, we annotated Homeobox genes in a Bivalvia species, *P. maximus*. First of all, we created a database of homeobox genes from different species including *Cr. gigas*

(Zhang et al. 2012), *L. goshimai* (Huan et al. 2020), *Acanthochitona rubrolineata* (Huan et al. 2020), *C. ventricosus* (Pardos-Blas et al. 2021), *Elysia marginata* (Maeda et al. 2021), and *P. canaliculata* (full NCBI annotation). Then, the database was aligned against all predicted proteomes considered in this study. From the five best matches, we assigned the predominant function (minimum of three same functions over five). This constituted a second database of homeobox genes, and was aligned again on the same proteomes. Proteins not found in the first round and having more than 60% of identity and aligned on more than 60% of the sequence length were kept. Then, all the proteins detected were mapped on the seven genomes of interest using exonerate v2.2 (Slater and Birney 2005) and all matches were given to Gmove v2 (Dubarry et al. 2016) for automatic annotation. Each predicted homeobox gene was further annotated manually. Briefly, we verified the presence of homeobox domains using CDDsearch (online version, (Marchler-Bauer et al. 2015) and manually annotated the genes regarding the mapping of proteins from the reference database using IGV v2.5.3 (Thorvaldsdottir et al. 2013). The phylogeny of homeobox domains predicted was performed on their multiple alignment using the MAFFT webserver tool with default parameters (Katoh and Frith 2012) and IQ-tree v1.6.12 (Minh et al. 2020).

Putative Toxins Identification

All sequences from ConoServer (Kaas et al. 2012) and the sequences annotated with the key words "toxin" or "venom" from uniprotKB were mapped on both genomes using BLASTx v2.13.0 (Camacho et al. 2009) with a maximal e-value of 0.001. Next, Exonerate v.2.2 (Slater and Birney 2005) was launched to better align the exons boundaries of the hits identified using BLAST v2.13.0. We then used Gmove (Dubarry et al. 2016) in order to detect possible complete ORF in both genomes. We discarded reverse transcriptases from the analysis after running BLASTp v2.13.0 against NR. Only the putative toxin-coding genes detected in *M. corrugatus* and *S. haemastoma*, respectively, and found within syntenic blocks between each species and *P. canaliculata* were manually annotated using the combination of the online Introproscan domain search (Jones et al. 2014), ConoPrec (Kaas et al. 2012), and the online BLASTp v2.13.0 against NR database (Sayers et al. 2022). From the detected *M. corrugatus*' putative toxin genes, we further mapped them on the genomes of *S. haemastoma*, *P. canaliculata*, and itself to span all possible ortholog and paralogous genes using BLASTx v2.13.0 with an e-value of 1E−5 followed by exonerate with default parameters.

Differential Gene Expression in *S. haemastoma*

The analysis of differential gene expression was conducted on the available RNAseq data from *S. haemastoma* as three replicates were sequenced for each tissue (osphradium and salivary gland). First quality control was performed on each replicates raw reads using fastqc v.0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and quality filtered and trimmed using fastp v.0.23.1with default parameters (Chen et al. 2018). Then, cleaned reads were mapped onto the genome sequences of *S. haemastoma* using STAR v.2.7.10b (Dobin et al. 2013) with the following parameters: –alignIntronMax 1,000 –alignMatesGapMax 10,000, and the addition of the gene annotation file. Each bam file was indexed using samtools v.1.15.1 (Danecek et al. 2021). Read counts were assessed using featureCounts from subread package v.2.0.1 (Liao et al. 2014). Differential gene expression analysis was performed using the R software v.4.2.2 (R Core Team 2022), Bioconductor (Gentleman et al. 2004) packages including DESeq2 v.1.38.1 (Anders and Huber 2010; Love et al. 2014), and the SARTools package v.1.8.1 (Varet et al. 2016). The parameters used for this analysis were the following: fitType parametric, cooksCutoff TRUE, independentFiltering TRUE, alpha 0.05, pAdjustMethod BH, typeTrans VST, and locfunc median. All genes detected as differentially expressed were then realigned on the genome using exonerate v.2.4.0 (Slater and Birney 2005) to check for potential undetected copies in the genomes and compared with our available annotation.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The present work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 865101) to N.P. Samples of *M. corrugatus* were collected during the expedition CORSICABENTHOS 2, conducted by Muséum National d'Histoire Naturelle in partnership with Parc Naturel Marin du Cap Corse et de l'Agriate and Université de Corse Pasquale Paoli. It is the marine component of the "Our Planet Reviewed" expedition in Corsica, made possible by fundings from Agence Française pour la Biodiversité and Collectivité Territoriale de Corse. Samples of *S. haemastoma* were collected at Capo Linaro (RM), Italy, and kindly provided by Prof. Marco Oliverio, Dept. of Biology and Biotechnologies Charles Darwin, Sapienza University of Roma, Italy. Sampling operated under the regulations then in force in the countries in question and satisfy the conditions set by the Nagoya Protocol for access to genetic resources. All biocomputing were done on the MNHN cluster: Plateforme de Calcul Intensif et Algorithmique PCIA, Muséum national d'histoire naturelle, Centre national de la recherche scientifique, UAR 2700 2AD, CP 26, 57 rue Cuvier, F-75231 Paris Cedex 05, France. We thank the associate editor and the reviewers for their careful reading and thoughtful comments.

Data Availability

Monoplex corrugatus and *Stramonita haemastoma* genome and transcriptome raw sequences and assembly data are available on SRA under the BioProject PRJNA912994 and PRJNA913004, respectively. Gene annotations and toxins for both *M. corrugatus* and *S. haemastoma* as well as Hox genes annotation and predicted proteins of all seven species are available at https://doi.org/10.48579/PRO/CEWICN.

References

Akondi KB, Muttenthaler M, Dutertre S, Kaas Q, Craik DJ, Lewis RJ, Alewood PF. 2014. Discovery, synthesis, and structure–activity relationships of conotoxins. *Chem Rev*. **114**:5815–5847.

Almeida DD, Viala VL, Nachtigall PG, Broe M, Gibbs HL, Serrano SMT, Moura-da-Silva AM, Ho PL, Nishiyama-Jr MY, Junqueira-de-Azevedo ILM. 2021. Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the *Bothrops jararaca* genome. *Proc Natl Acad Sci*. **118**:e2015159118.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol*. **11**:R106.

Arbuckle K. 2018. Phylogenetic comparative methods can provide important insights into the evolution of toxic weaponry. *Toxins (Basel)*. **10**:518.

Balaji RA, Ohtake A, Sato K, Gopalakrishnakone P, Kini RM, Seow KT, Bay B-H. 2000. λ-Conotoxins, a new family of conotoxins with unique disulfide pattern and protein folding. *J Biol Chem*. **275**:39516–39522.

Bandi V, Gutwin C. 2020. Interactive Exploration of Genomic Conservation 10.

Barghi N, Concepcion GP, Olivera BM, Lluisma AO. 2016. Structural features of conopeptide genes inferred from partial sequences of the *Conus tribblei* genome. *Mol Genet Genomics*. **291**:411–422.

Barua A, Mikheyev AS. 2019. Many options, few solutions: over 60 my snakes converged on a few optimal venom formulations. *Mol Biol Evol*. **36**:1964–1974.

Barua A, Mikheyev AS. 2021. An ancient, conserved gene regulatory network led to the rise of oral venom systems. *Proc Natl Acad Sci*. **118**:e2021311118.

Bohlen CJ, Chesler AT, Sharif-Naeini R, Medzihradszky KF, Zhou S, King D, Sánchez EE, Burlingame AL, Basbaum AI, Julius D. 2011. A heteromeric Texas coral snake toxin targets acid-sensing ion channels to produce pain. *Nature* **479**:410–414.

Bose U, Wang T, Zhao M, Motti CA, Hall MR, Cummins SF. 2017. Multiomics analysis of the giant triton snail salivary gland, a crown-of-thorns starfish predator. *Sci Rep*. **7**:6000.

Bouchet P, Lozouet P, Maestrati P, Heros V. 2002. Assessing the magnitude of species richness in tropical marine environments: exceptionally high numbers of molluscs at a New Caledonia site. *Biol J Linn Soc*. **75**:421–436.

Buczek O, Bulaj G, Olivera BM. 2005. Conotoxins and the posttranslational modification of secreted gene products. *Cell Mol Life Sci*. **62**:3067–3079.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. **10**:421.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. **18**:188–196.

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol*. **28**:219–229.

Chang D, Duda TF. 2012. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. *Mol Biol Evol*. **29**:2019–2029.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**:giab008.

del Pozo JC, Ramirez-Parra E. 2015. Whole genome duplications in plants: an overview from *Arabidopsis*. *J Exp Bot*. **66**:6991–7003.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-Seq aligner. *Bioinformatics* **29**:15–21.

Drukewitz SH, von Reumont BM. 2019. The significance of comparative genomics in modern evolutionary venomics. *Front Ecol Evol*. **7**:163.

Dubarry M, Noel B, Rukwavu T, Farhat S, Silva CD, Seeleuthner Y, Lebeurrier M, Aury J-M. 2016. Gmove a tool for eukaryotic gene predictions using various evidences. *[version 1; not peer reviewed]. F1000Research, 5(ISCB Comm J):681 (poster)*. Available from: https://f1000research.com/posters/5-681

Duda TF, Palumbi SR. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod Conus. *Proc Natl Acad Sci*. **96**:6820–6823.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **32**:1792–1797.

Eric Schranz M, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr Opin Plant Biol*. **15**:147–153.

Estebenet AL, Martin PR. 2002. *Pomacea canaliculata* (Gastropoda: Ampullariidae): life-history traits and their plasticity. *Biocell* **26**:83–89.

Fedosov AE, Moshkovskii SA, Kuznetsova KG, Olivera BM. 2012. Conotoxins: from the biodiversity of gastropods to new drugs. *Biochem Mosc Suppl Ser B Biomed Chem*. **6**:107–122.

Fedosov A, Zaharias P, Puillandre N. 2021. A phylogeny-aware approach reveals unexpected venom components in divergent lineages of cone snails. *Proc R Soc B Biol Sci*. **288**:20211017.

Finol-Urdaneta RK, Belovanovic A, Micic-Vicovac M, Kinsella GK, McArthur JR, Al-Sabi A. 2020. Marine toxins targeting kv1 channels: pharmacological tools and therapeutic scaffolds. *Mar Drugs*. **18**:173.

Flagel LE, Wendel JF. 2009. Gene duplication and evolutionary novelty in plants. *New Phytol*. **183**:557–564.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *PNAS* **117**:9451–9457.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**:1531–1545.

Forde A, Jacobsen A, Dugon MM, Healy K. 2022. Scorpion species with smaller body sizes and narrower chelae have the highest venom potency. *Toxins (Basel)*. **14**:219.

Fry BG, Koludarov I, Jackson TNW, Holford M, Terrat Y, Casewell NR, Undheim EAB, Vetter I, Ali SA, Low DHW, *et al.* 2015. CHAPTER 1. Seeing the woods for the trees: understanding venom evolution as a guide for biodiscovery. In: King GF, editor. *Drug discovery*. Cambridge: Royal Society of Chemistry. p. 1–36.

Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, *et al.* 2009. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genomics Hum Genet*. **10**:483–511.

Gehring WJ, Affolter M, Bürglin T. 1994. Homeodomain proteins. *Annu Rev Biochem*. **63**:487–526.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, *et al.* 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. **5**:R80.

Gerdol M, Cervelli M, Mariottini P, Oliverio M, Dutertre S, Modica M. 2019. A recurrent motif: diversity and evolution of ShKT domain containing proteins in the vampire snail *Cumia reticulata*. *Toxins (Basel)*. **11**:106.

Gerdol M, Cervelli M, Oliverio M, Modica MV. 2018. Piercing fishes: porin expansion and adaptation to hematophagy in the vampire snail *Cumia reticulata*. *Mol Biol Evol*. **35**:2654–2668.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol*. **15**:e1007273.

Giorgianni MW, Dowell NL, Griffin S, Kassner VA, Selegue JE, Carroll SB. 2020. The origin and diversification of a novel protein family in venomous snakes. *Proc Natl Acad Sci*. **117**:10911–10920.

Glasauer SMK, Neuhauss SCF. 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics*. **289**:1045–1060.

Gorson J, Ramrattan G, Verdes A, Wright EM, Kantor Y, Rajaram Srinivasan R, Musunuri R, Packer D, Albano G, Qiu W-G, *et al*. 2015. Molecular diversity and gene evolution of the venom arsenal of terebridae predatory marine snails. *Genome Biol Evol*. **7**:1761–1778.

Grohme MA, Vila-Farré M, Rink JC. 2018. Small- and large-scale high molecular weight genomic DNA extraction from planarians. In: Rink JC, editor. *Planarian regeneration: methods and protocols. Methods in molecular biology*. New York (NY): Springer. p. 267–275.

Gruber CW, Muttenthaler M. 2012. Discovery of defense- and neuropeptides in social ants by genome-mining. *PLoS One*. **7**:e32559.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**:2896–2898.

Hallinan NM, Lindberg DR. 2011. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. *Genome Biol Evol*. **3**:1150–1163.

Haney RA, Clarke TH, Gadgil R, Fitzpatrick R, Hayashi CY, Ayoub NA, Garb JE. 2016. Effects of gene duplication, positive selection, and shifts in gene expression on the evolution of the venom gland transcriptome in widow spiders. *Genome Biol Evol*. **8**:228–242.

Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. 2014. Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. *Genome Biol Evol*. **6**:2088–2095.

Harper EM. 2003. The mesozoic marine revolution. In: Kelley PH, Kowalewski M, Hansen TA, editors. *Predator—prey interactions in the fossil record. Topics in geobiology*. Boston (MA): Springer US. p. 433–455.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**:1157–1164.

Hittinger CT, Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**:677–681.

Hu H, Bandyopadhyay PK, Olivera BM, Yandell M. 2011. Characterization of the conus bullatus genome and its venom-duct transcriptome. *BMC Genomics*. **12**:60.

Huan P, Wang Q, Tan S, Liu B. 2020. Dorsoventral decoupling of Hox gene expression underpins the diversification of molluscs. *Proc Natl Acad Sci*. **117**:503–512.

Hwang HJ, Patnaik BB, Chung JM, Sang MK, Park JE, Kang SW, Park SY, Jo YH, Park HS, Baliarsingh S, *et al*. 2021. De novo transcriptome sequencing of triton shell *Charonia lampas sauliae*: identification of genes related to neurotoxins and discovery of genetic markers. *Mar Genomics*. **59**:100862.

Jia H, Zhang G, Zhang C, Zhang H, Yao G, He M, Liu W. 2023. Identification and expression of the conotoxin homologous genes in the giant triton snail (*Charonia tritonis*). *J Ocean Univ China* **22**:213–220.

Jin A-H, Muttenthaler M, Dutertre S, Himaya SWA, Kaas Q, Craik DJ, Lewis RJ, Alewood PF. 2019. Conotoxins: chemistry and biology. *Chem Rev*. **119**:11510-11549.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, *et al*. 2014. Interproscan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240.

Kaas Q, Westermann J-C, Craik DJ. 2010. Conopeptide characterization and classifications: an analysis using ConoServer. *Toxicon* **55**:1491–1509.

Kaas Q, Yu R, Jin A-H, Dutertre S, Craik DJ. 2012. Conoserver: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res*. **40**:D325–D330.

Katoh K, Frith MC. 2012. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28**:3144–3146.

King GF, Gentz MC, Escoubas P, Nicholson GM. 2008. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon* **52**:264–276.

Klein A, Motti C, Hillberg A, Ventura T, Thomas-Hall P, Armstrong T, Barker T, Whatmore P, Cummins S. 2021. Development and interrogation of a transcriptomic resource for the giant triton snail (*Charonia tritonis*). *Mar Biotechnol*. **23**:501–515.

Koludarov I, Velasque M, Timm T, Greve C, Ben Hamadou A, Gupta D K, Lochnit G, Heinzinger M, Vilcinskas A, Gloag R, *et al*. 2022. A common venomous ancestor? Prevalent bee venom genes evolved before the aculeate stinger while few major toxins are bee-specific. *bioRxiv* 2022.01.21.477203.

Kordiš D, Gubenšek F. 2000. Adaptive evolution of animal toxin multigene families. *Gene* **261**:43–52.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics*. **5**:59

Kuzmin E, Taylor JS, Boone C. 2022. Retention of duplicated genes in evolution. *Trends Genet*. **38**:59–72.

Kuznetsova KG, Zvonareva SS, Ziganshin R, Mekhova ES, Dgebuadze P, Yen DTH, Nguyen THT, Moshkovskii SA, Fedosov AE. 2022. Vexitoxins: conotoxin-like venom peptides from predatory gastropods of the genus Vexillum. *Proc R Soc B Biol Sci*. **289**:20221152.

Layer R, McIntosh J. 2006. Conotoxins: therapeutic potential and application. *Mar Drugs*. **4**:119–142.

Lemarcis T, Fedosov AE, Kantor YI, Abdelkrim J, Zaharias P, Puillandre N. 2022. Neogastropod (Mollusca, Gastropoda) phylogeny: a step forward with mitogenomes. *Zool Scr*. **51**:550–561.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.

Liao Y, Smyth GK, Shi W. 2014. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma Oxf Engl*. **30**:923–930.

Lindberg DR, Sigwart JD. 2015. What is the molluscan osphradium? A reconsideration of homology. *Zool Anz—J Comp Zool*. **256**:14–21.

Liu C, Ren Y, Li Z, Hu Q, Yin L, Qiao X, Zhang Y, Xing L, Xi Y, Jiang F, *et al*. 2020. Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic-terrestrial transition. bioRxiv 2020.02.02.930693 .

Liu C, Ren Y, Li Z, Hu Q, Yin L, Wang H, Qiao X, Zhang Y, Xing L, Xi Y, *et al*. 2021. Giant African snail genomes provide insights into molluscan whole-genome duplication and aquatic–terrestrial transition. *Mol Ecol Resour*. **21**:478–494.

Liu GJ, Takeuchi H. 1993. Modulation of neuropeptide effects by achatin-I, an achatina endogenous tetrapeptide. *Eur J Pharmacol*. **240**:139–145.

Liu C, Zhang Y, Ren Y, Wang H, Li S, Jiang F, Yin L, Qiao X, Zhang G, Qian W, *et al*. 2018. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. *GigaScience* **7**:giy101.

Lopes I, Altab G, Raina P, de Magalhães JP. 2021. Gene size matters: an analysis of gene length in the human genome. *Front Genet*. **12**:559998.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol*. **15**:550.

Lu A, Watkins M, Li Q, Robinson SD, Concepcion GP, Yandell M, Weng Z, Olivera BM, Safavi-Hemami H, Fedosov AE. 2020.

Transcriptomic profiling reveals extraordinary diversity of venom peptides in unexplored predatory gastropods of the genus Clavus. *Genome Biol Evol*. **12**:684–700.
Mable BK, Alexandrou MA, Taylor MI. 2011. Genome duplication in amphibians and fish: an extended synthesis. *J Zool*. **284**:151–182.
Maeda T, Takahashi S, Yoshida T, Shimamura S, Takaki Y, Nagai Y, Toyoda A, Suzuki Y, Arimoto A, Ishii H, *et al.* 2021. Chloroplast acquisition without the gene transfer in kleptoplastic sea slugs, *Plakobranchus ocellatus*. *eLife* **10**:e60176.
Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. 2013. Gene duplication as a major force in evolution. *J Genet*. **92**:155–161.
Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, *et al.* 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. **43**:D222–D226.
Martinson EO, Kelkar YD, Chang C-H, Werren JH. 2017. The evolution of venom by co-option of single-copy genes. *Curr Biol*. **27**: 2007–2013.e8.
Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. **37**:1530–1534.
Modica MV, Holford M. 2010. The neogastropoda: evolutionary innovations of predatory marine snails with remarkable pharmacological potential. In: Pontarotti P, editor. *Evolutionary biology—concepts, molecular and morphological evolution*. Berlin (Heidelberg): Springer Berlin Heidelberg. p. 249–270.
Modica MV, Reinoso Sánchez J, Pasquadibisceglie A, Oliverio M, Mariottini P, Cervelli M. 2018. Anti-haemostatic compounds from the vampire snail *Cumia reticulata*: molecular cloning and in-silico structure-function analysis. *Comput Biol Chem*. **75**: 168–177.
Möller C, Melaun C, Castillo C, Díaz ME, Renzelman CM, Estrada O, Kuch U, Lokey S, Marí F. 2010. Functional hypervariability and gene diversity of cardioactive neuropeptides*. *J Biol Chem*. **285**: 40673–40680.
Moriyama Y, Koshiba-Takeuchi K. 2018. Significance of whole-genome duplications on the emergence of evolutionary novelties. *Brief Funct Genomics*. **17**:329–338.
Motobu M, Tsuji N, Miyoshi T, Huang X, Islam MK, Alim MA, Fujisaki K. 2007. Molecular characterization of a blood-induced serine carboxypeptidase from the ixodid tick *Haemaphysalis longicornis*. *FEBS J*. **274**:3299–3312.
Ng TH, Tan SK, Yeo D. 2017. *South American apple snails, Pomacea spp. (Ampullariidae), in Singapore*. in book 221–239.
Olivera BM, Fedosov A, Imperial JS, Kantor Y. 2017. Physiology of envenomation by conoidean gastropods. In: Saleuddin S, Mukai S, editors. *Physiology of molluscs*. Vol. 1. 1st ed. New Jersey: Apple Academic Press, Inc., p. 153–188.
Olivera BM, Seger J, Horvath MP, Fedosov AE. 2015. Prey-capture strategies of fish-hunting cone snails: behavior, neurobiology and evolution. *Brain Behav Evol*. **86**:58–74.
Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet*. **34**:401–437.
Pardos-Blas JR, Irisarri I, Abalde S, Afonso CML, Tenorio MJ, Zardoya R. 2021. The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *GigaScience* **10**:giab037.
Pasquier J, Braasch I, Batzel P, Cabau C, Montfort J, Nguyen T, Jouanno E, Berthelot C, Klopp C, Journot L, *et al.* 2017. Evolution of gene expression after whole-genome duplication: new insights from the spotted gar genome. *J Exp Zoolog B Mol Dev Evol*. **328**:709–721.
Peng C, Huang Y, Bian C, Li J, Liu J, Zhang K, You X, Lin Z, He Y, Chen J, *et al.* 2021. The first Conus genome assembly reveals a primary genetic central dogma of conopeptides in *C. betulinus*. *Cell Discov*. **7**:11.
Phuong MA, Alfaro ME, Mahardika GN, Marwoto RM, Prabowo RE, von Rintelen T, Vogt PWH, Hendricks JR, Puillandre N. 2019. Lack of signal for the impact of conotoxin gene diversity on speciation rates in cone snails. *Syst Biol*. **68**:781–796.
Ponce D, Brinkman DL, Potriquet J, Mulvenna J. 2016. Tentacle transcriptome and venom proteome of the pacific sea nettle, *Chrysaora fuscescens* (cnidaria: scyphozoa). *Toxins (Basel)*. **8**:102.
Ponte G, Modica MV. 2017. Salivary glands in predatory mollusks: evolutionary considerations. *Front Physiol*. **8**:580.
Puillandre N, Koua D, Favreau P, Olivera BM, Stöcklin R. 2012. Molecular phylogeny, classification and evolution of conopeptides. *J Mol Evol*. **74**:297–309.
Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, *et al.* 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res*. **26**:342–350.
R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/
Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol Plant*. **11**: 414–428.
Reynaud S, Ciolek J, Degueldre M, Saez NJ, Sequeira AF, Duhoo Y, Brás JLA, Meudal H, Cabo Díez M, Fernández Pedrosa V, *et al.* 2020. A venomics approach coupled to high-throughput toxin production strategies identifies the first venom-derived melanocortin receptor agonists. *J Med Chem*. **63**:8250–8264.
Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nat Methods*. **17**:155–158.
Safavi-Hemami H, Brogan SE, Olivera BM. 2019. Pain therapeutics from cone snail venoms: from Ziconotide to novel non-opioid pathways. *J Proteomics*. **190**:12–20.
Sankoff D. 2001. Gene and genome duplication. *Curr Opin Genet Dev*. **11**:681–684.
Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, *et al.* 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. **50**:D20–D26.
Schendel V, Rash LD, Jenner RA, Undheim EAB. 2019. The diversity of venom: the importance of behavior and venom system morphology in understanding its ecology and evolution. *Toxins (Basel)*. **11**:666.
Sémon M, Wolfe KH. 2007. Consequences of genome duplication. *Curr Opin Genet Dev*. **17**:505–512.
Shiomi K, Kawashima Y, Mizukami M, Nagashima Y. 2002. Properties of proteinaceous toxins in the salivary gland of the marine gastropod (Monoplex echo). *Toxicon* **40**:563–571.
Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. **6**:31.
Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org
Soares SG, Oliveira LL. 2009. Venom-sweet-venom: N-linked glycosylation in snake venom toxins. *Protein Pept Lett*. **16**:913–919.
Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. **34**:W435–W439.
Stockwell T, Baden-Tillson H, Favreau philippe, Mebs D, Ducancel F, Stöcklin R. 2010. Sequencing the genome of Conus consors: preliminary results. https://hal.science/hal-00738636v1/file/AJ_Ebook-RT18-2010-signets.pdf#page=11
Sun J, Mu H, Ip JCH, Li R, Xu T, Accorsi A, Sánchez Alvarado A, Ross E, Lan Y, Sun Y, *et al.* 2019. Signatures of divergence, invasiveness, and terrestrialization revealed by four apple snail genomes. *Mol Biol Evol*. **36**:1507–1520.
Taylor JD, Morris NJ, Taylor CN. 1980. Food specialization and the evolution of predatory prosobranch gastropods. *Palaeontology* **23**:375–409.
Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. **14**:178–192.

Trevisan-Silva D, Gremski LH, Chaim OM, da Silveira RB, Meissner GO, Mangili OC, Barbaro KC, Gremski W, Veiga SS, Senff-Ribeiro A. 2010. Astacin-like metalloproteases are a gene family of toxins present in the venom of different species of the brown spider (genus Loxosceles). *Biochimie* **92**:21–32.

Turner A, Craik D, Kaas Q, Schroeder C. 2018. Bioactive compounds isolated from neglected predatory marine gastropods. *Mar Drugs*. **16**:118.

Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nat Rev Genet*. **10**:725–732.

Varet H, Brillet-Guéguen L, Coppée J-Y, Dillies M-A. 2016. SARTools: a DESeq2- and EdgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. *PLoS One*. **11**:e0157022.

Vermeij GJ. 1977. The mesozoic marine revolution: evidence from snails, predators and grazers. *Paleobiology* **3**:245–258.

Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ. 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. *PLoS Biol*. **10**: e1001446.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, *et al.* 2012. MCScanx: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. **40**:e49.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. Kaks_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**:77–80.

Watkins M, Hillyard DR, Olivera BM. 2006. Genes expressed in a turrid venom duct: divergence and similarity to conotoxins. *J Mol Evol*. **62**:247–256.

Wong ESW, Belov K. 2012. Venom evolution through gene duplications. *Gene* **496**:1–7.

WoRMS Editorial Board. 2022. World Register of Marine Species. Available from https://www.marinespecies.org at VLIZ. Accessed 2022. https://doi.org/10.14284/170

Yang M, Zhou M. 2020. Insertions and deletions play an important role in the diversity of conotoxins. *Protein J*. **39**:190–195.

Zancolli G, Casewell NR. 2020. Venom systems as models for studying the origin and regulation of evolutionary novelties. *Mol Biol Evol*. **37**:2777–2790.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, *et al.* 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**:49–54.

Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci*. **95**:3708–3713.

Zhao Y, Antunes A. 2022. Biomedical potential of the neglected molluscivorous and vermivorous conus species. *Mar Drugs*. **20**:105.