Evolution of venom production in marine predatory snails

Giulia Zancolli[1,2*], Maria Vittoria Modica[3], Nicolas Puillandre[4], Yuri Kantor[5], Agneesh Barua[1,2], Giulia Campli[1,2], Marc Robinson-Rechavi[1,2]

[1]Department of Ecology and Evolution, University of Lausanne, 1015, Lausanne, Switzerland

[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland

[3]Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, 1 Via Gregorio Allegri, 00198, Roma, Italy

[4]Institut Systématique Evolution Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier, CP51, 75005, Paris, France

[5]Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, 33 Leninsky prospect, 119034, Moscow, Russian Federation

*Corresponding author: giulia.zancolli@gmail.com

Keywords

Neogastropoda, transcriptome evolution, ancestral reconstruction, toxin, venom, UPR, ER stress, secretion

Abstract

Venom is a widespread secretion in nature, extensively studied for its toxin components and application potential. Yet, the evolution of venom production remains poorly understood. To address this question, we conducted a comparative transcriptomics analysis of the oesophagus-associated glands in marine predatory gastropods, among which the cone snail venom gland represents a pinnacle of specialisation. We found that the functional divergence and specialisation of the venom gland was achieved through a redistribution of its ancestral digestive functions to other organs, specifically the oesophagus. This entailed concerted expression changes and accelerated transcriptome evolution across the entire digestive system. The increase in venom gland secretory capacity was achieved through the modulation of an ancient secretory machinery, particularly genes involved in endoplasmic reticulum stress and unfolded protein response. On the other hand, the emergence of novel genes, involving transposable elements, contributed to the gland regulatory network. Our analysis provides new insights into the genetic basis of functional divergence and highlights the remarkable plasticity of the gastropod digestive system.

Introduction

Across diverse branches of the animal kingdom, organisms have independently evolved the ability to produce and deliver venom, a complex cocktail of bioactive toxin molecules. In many venomous animals, these toxins are synthetised in specialised exocrine organs known as venom glands[1,2]. While significant research has explored the molecular evolution of toxin genes and venom composition, the genetic basis underlying the evolution of venom-producing organs and their specialised function remains largely unexplored[3]. Recent studies on snakes have begun to elucidate the regulatory network co-opted by toxin genes to achieve high levels of expression in their venom glands[4–6]. These findings indicate that trans-regulatory factors from the extracellular signal-regulated kinase (ERK) and unfolded protein response (UPR) pathways are specifically involved in coordinating venom toxin expression. Moreover, the upregulation of UPR and endoplasmic reticulum stress pathways in the venom glands of several distinct venomous taxa suggests that similar molecular solutions may have evolved independently across different lineages[4,7,8]. However, a comprehensive understanding of how these glands have become specialised and optimised for the efficient mass production of toxins is still lacking.

Studying homologous structures with divergent functions provides valuable insights into the genetic basis of functional innovation, such as the evolution of limbs[9] or of feathers[10] in vertebrates. At a smaller evolutionary scale, a notable example of this is found in the glands associated with the mid-oesophagus of marine predatory snails within the subclass Caenogastropoda (Fig. 1). In some caenogastropods, such as those in the family Naticidae[11], the oesophageal gland is a dilated section of the oesophageal wall that is believed to secrete digestive enzymes and mucus[12]. In contrast, other caenogastropods in

the order Neogastropoda have evolved this glandular tissue into a distinct organ known as the gland of Leiblein, which connects to the oesophagus via a duct[12,13]. Ultrastructural observations in species from the families Muricidae and a Nassariidae suggest that the gland of Leiblein plays a role in food processing, particularly nutrient absorption and storage, as indicated by the presence of numerous lysosomes[14]. In the superfamily Conoidea, which comprises 18 families[15] including the well-known cone snails (family Conidae)[16], the oesophageal gland has undergone further modification into a long, convoluted duct that secretes a mixture of hundreds of primarily neurotoxic peptides[13,15,17]. The venom gland in these snails is attached to a large muscular bulb that contracts to push venom through the duct and into the buccal cavity, where the radula is modified to varying degrees to inject venom into prey or predators[13,18,19]. Interestingly, some neogastropods, such as those in the family Mitridae[20] and Terebridae[21], have entirely lost the mid-oesophageal gland.

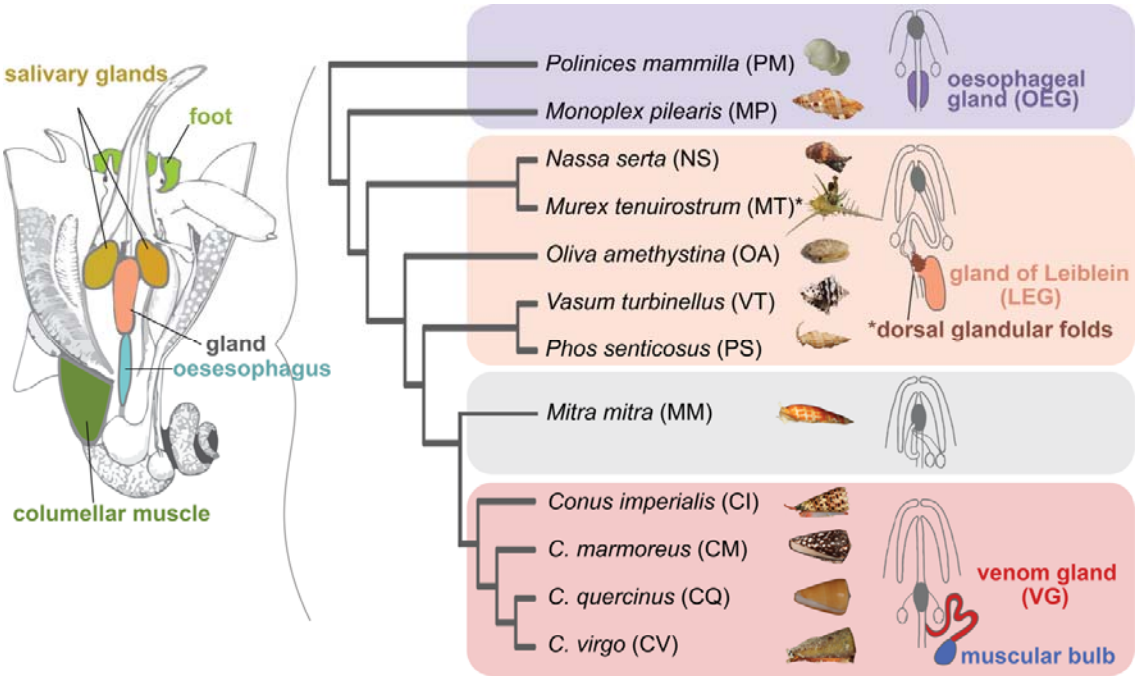


Figure 1: Species and tissues investigated. Overview of the anatomy of a marine snail with the color-coded sampled tissues and the phylogeny of the species used in this study based on (Fedosov et al. 2024). The abbreviations used for the tissues and species used in other figures are shown. The anatomy drawing is modified after [13], while the schematic of the foregut apparatus of caenogastropods with their mid-oesophageal glands was modified after [76].

The evolution of venom production from digestive-related functions represents a pinnacle of specialisation for the mid-oesophageal glands. First, while the secretion targets of the oesophageal gland and the gland of Leiblein are internal, within the gland itself (e.g., lysosomes)[14], the targets of conoidean venom are external, acting on receptors like ion channels in other organisms[17]. Second, because venom must be injected into prey or predators, or sometimes released into the surrounding water[22], the venom gland likely evolved mechanisms to detect external stimuli and convert them into signals that trigger venom release. Finally, this novel chemical weapon provided a significant adaptive

advantage, enabling cone snails to diversify their diet to include fast-moving prey, such as fish, and to defend themselves against powerful predators[23]. The processes by which an organ originally dedicated to digestive functions transformed into a specialised toxin-producing factory remain unknown.

Here, we analyse gene expression data from the mid-oesophageal glands and other tissues of 12 marine caenogastropod species to investigate the genetic basis of venom gland evolution. Specifically, we aim to address the following questions: i) Do oesophageal glands share similar global gene expression profiles? ii) Considering the divergent function of the venom gland, does its transcriptome evolve more rapidly than that of the other mid-oesophageal glands? iii) Did novel genes contribute the evolution of venom production?

To answer these questions, we first performed species-level analyses to characterize sets of overexpressed genes and to better understand the extent of functional specialization in the organs investigated. We then conducted between-species comparisons to investigate the gene expression dynamics that led to the evolution of venom production. Our findings reveal that the venom gland exhibit extremely high expression levels, unmatched by any other tissue analysed. The specialisation of the mid-oesophageal gland for toxin secretion was facilitated by the redistribution of ancestral digestive functions to other organs in the digestive system. This shift was accompanied by high evolutionary rates not only in the venom gland itself but throughout the entire digestive system, reflecting concerted evolutionary changes across these organs. While the remarkable secretory capability of the venom gland was achieved through expression changes of a conserved secretory machinery, the regulatory network was achieved by the evolution of novel genes.

Results

Summary of species and tissues investigated

We sampled ten species of Neogastropoda and two outgroup species within the same subclass Caenogastropoda (Fig. 1). The two outgroup species, the Cymatidae *Monoplex pilearis* (Linnaeus, 1758) and the Naticidae *Polinices mammilla* (Linnaeus, 1758), possess a simple oesophageal gland (OEG) attached to the oesophagus, representing the ancestral state. Among the Neogastropoda, two species from the Muricidae family, *Nassa serta* (Bruguière, 1789) and *Murex tenuirostrum* (Lamarck, 1822), have a gland of Leiblein (LEG), with *Murex tenuirostrum* also displaying dorsal glandular folds on the mid-oesophagus, often referred to as glande framboisée (Lobo Cunha 2019). Additionally, we sampled the gland of Leiblein in the Olividae *Oliva amethystina* (Rőding, 1798), the Vasidae *Vasum turbinellus* (Linnaeus, 1758), and the Nassaridae *Phos senticosus* (Linnaeus, 1758). The venomous species, possessing a venom gland (VG), belonged to the family Conidae, including *Conus imperialis* (Linnaeus, 1758), *C. marmoreus* (Linnaeus, 1758), *C. virgo* (Linnaeus, 1758), and *C. quercinus* (Lightfoot, 1786). We also collected tissue samples from the Mitridae *Mitra mitra* (Linnaeus, 1758), which lacks a mid-oesophageal gland[20]. Besides the glands, we took samples of either the foot or the columellar muscle, oesophagus, salivary glands, dorsal glandular folds in *M. tenuirostrum*, and muscular venom bulb in cone snails.

Sequencing and de novo assembly statistics

A total of 150 libraries were sequenced, yielding an average of 45 million reads per library across 12 species, with an average of three biological replicates per species (Supplementary Data 1). We generated *de novo* assemblies by pooling all libraries within each species and processing them through our quality-filtering pipeline (see Methods). After filtering, the assemblies contained between 26,215 and 59,431 annotated, non-redundant protein-coding genes, with an average of 40,192 sequences per assembly (Supplementary Data 2). The completeness of these assemblies ranged from 82% to 93%, with an average of 89%.

Prior to data analysis, we evaluated the quality of read alignment. Samples with low mapping rates or those that did not cluster with other samples of the same tissue type in a Principal Component Analysis (PCA) plot were excluded (see Methods). After this quality filtering step, we retained a total of 140 samples (Supplementary Fig. 1).

Within-species analysis

Characterisation of tissue-specific gene sets

To better understand the functional specialisation of the mid-oesophageal glands, we identified sets of genes that were overexpressed in each organ. Genes were classified as tissue-specific if their expression in a given tissue was at least twice that of the second most highly expressed tissue. On average, 14% (range: 10-19%) of genes were tissue-specific, with 43% (range: 38-51%) of these genes exclusively expressed in one tissue. Tissue specificity was consistent across tissues and species, although we observed a significantly higher number of tissue-specific genes in the venom glands compared to the glands of Leiblein and oesophageal glands (Fig. 2a). Notably, the smallest set observed was the oesophagus of *M. tenurostrum* (N = 318). This species though also possesses dorsal glandular folds on the oesophagus (Fig. 1). The tissue-specificity of these gene sets was further validated by differential expression analysis using a likelihood ratio test across all tissues (Supplementary Note 1).

Gene Ontology (GO) enrichment analysis of the tissue-specific gene sets revealed marked differences between the mid-oesophageal glands (Fig. 2b), as well as lineage-specific patterns (Supplementary Note 2, Supplementary Fig. 2-10). In OEG-specific gene sets, we observed enrichment for terms related to intracellular trafficking and communication (e.g., ‘AP-5 adaptor complex’) and transmembrane transport. Additionally, terms related to detoxification, homeostasis, and response to external stimuli (e.g., ‘response to inorganic substance’, ‘superoxide dismutase activity’) were enriched (Supplementary Fig. 2-4). In contrast, gene sets specific to the gland of Leiblein across multiple species were enriched for terms related to intracellular digestion, particularly protein digestion and absorption, with several lysosome-related terms (e.g., ‘endopeptidase activity’, ‘WASH complex’), aligning with previous ultrastructural studies in Muricidae[14]. Terms related to iron homeostasis and metabolism were also enriched. However, *O. amethystina* showed some unique enrichments, including extracellular rather than intracellular compartment terms, reflecting possible lineage-specific adaptations linked to diet (Supplementary Fig. 5-7). In venom glands, GO terms associated with toxin and neurotoxic activity (e.g., ‘toxin activity’, ‘host cell postsynaptic membrane’) were over-represented in all cone snail species, as expected (Fig. 2b). Additionally, we found terms related to protein synthesis (e.g., ‘signal peptide processing’, ‘Golgi organization’) and to post-translational modifications (e.g., ‘hydroxylation’,

‘disulfide oxidoreductase activity’). Notably, terms related to endoplasmic reticulum (ER) stress and UPR were also enriched (e.g., ‘response to unfolded proteins’, ‘response to ER stress’) (Supplementary Fig. 8-10).

In summary, while the non-venomous mid-oesophageal glands are primarily involved in digestion, homeostasis, absorption, and storage, the homologous venom gland has specialised into a factory for toxin synthesis and secretion. Despite these functional differences, all three gland types shared common enriched terms typically associated with the vertebrate liver, although not homologous, such as those related to iron homeostasis and metabolism (e.g., ‘iron binding’) and methylation (‘methionine adenosyltransferase activity’, ‘betaine-homocysteine S-methyltransferase activity’). Additionally, we found terms related to hormone response across all three gland types, including ‘thyroid hormone generation’ in OEG- and LEG-specific gene sets, and ‘response to thyroglobulin triiodothyronine’ in VG-specific sets.

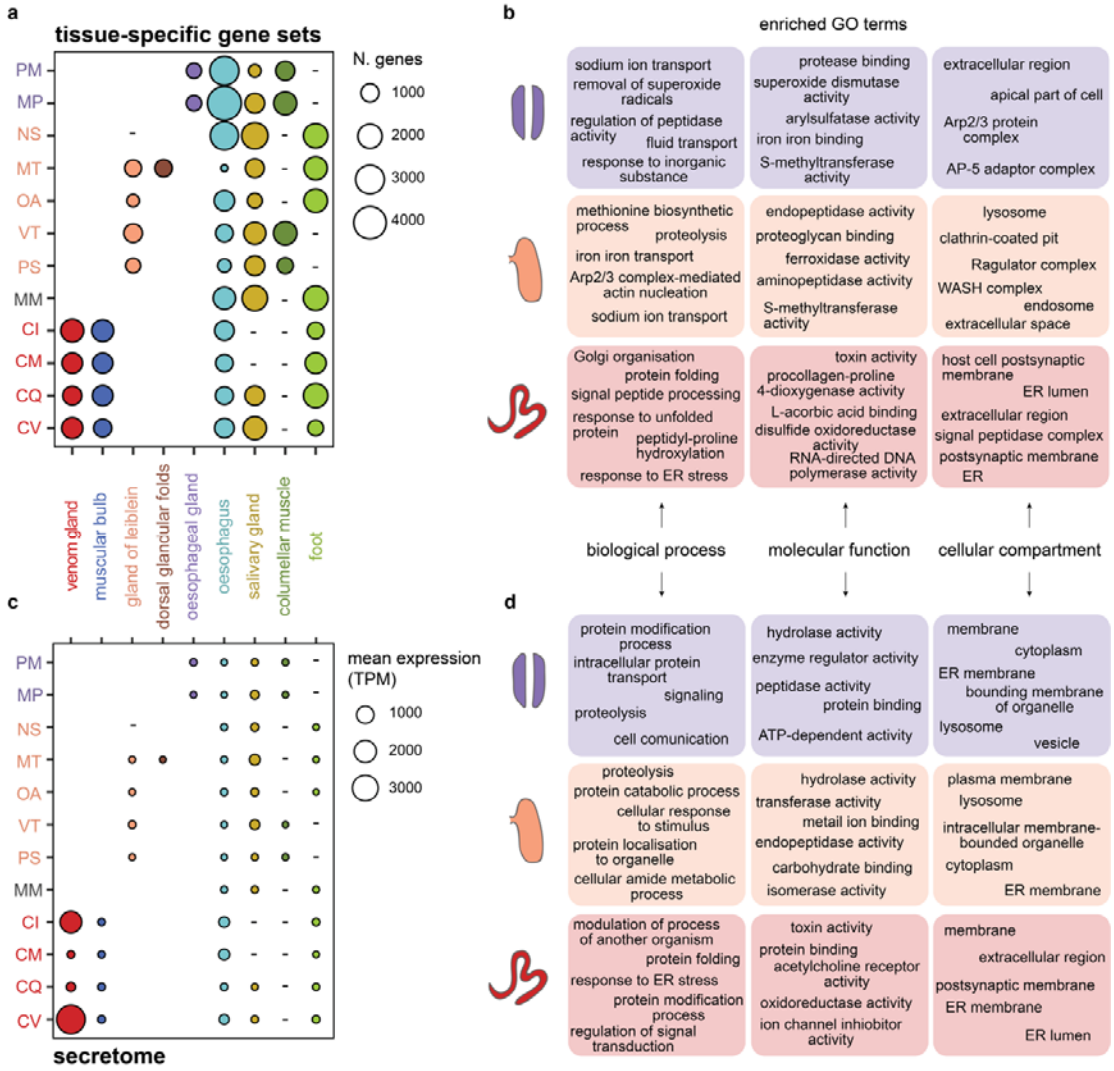


Figure 2: Overview of tissue-specific gene sets and gland secretomes. a) Number of genes within each tissue-specific set across all organs and species. The venom glands have a significant

higher number of tissue-specific genes (mean N = 1,424) compared to the gland of Leiblein (mean N = 771; t = 3.6, df = 6, p = 0.005) and oesophageal gland (mean N = 699; t = 5.6, df =3, p = 0.005). Missing tissue samples are marked with a "-". b) Gene Ontology (GO) enrichment results of the OEG-, LEG-, and VG-specific gene sets. c) Mean expression levels as Transcript per Million (TPM) of genes possessing a signal peptide, therefore comprising the secretomes, expressed in the organs. Missing tissue samples are marked with a "-". d) GO enrichment results of the oesophageal gland, gland of Leiblein, and venom gland secretomes. Species abbreviations as in Fig. 1.

Characterisation of the glands' secretomes

Given the fundamental role of secretion in the evolution of the venom gland, we analysed the 'secretome' of the mid-oesophageal glands by assessing the diversity and expression levels of genes predicted to have a signal peptide with SignalP[24].

The number of expressed genes coding for secreted proteins was relatively consistent across tissue types, although it was higher in the venom gland compared to the gland of Leiblein and the oesophageal gland, although the latter was not statistically significant (Supplementary Fig. 11a). When focussing on tissue-specific genes, the venom gland secretome was also significantly more diverse than the other glands (Supplementary Fig. 11b). Interestingly, the salivary glands did not show particularly high secretome diversity despite being exocrine glands (Supplementary Fig. 11).

In terms of expression, the differences between tissues were more pronounced (Fig. 2c). The venom gland displayed markedly higher expression levels than any other tissue, with the topmost expressed conotoxin in *C. virgo* (707,515 TPM). Expression levels varied among cone snail species (Fig. 2c), likely reflecting differences in the venom replenishment circle at the time of sampling. In non-venomous species, the highest expression levels were observed in the salivary glands, with a serine protease reaching 24,619 TPM. Interestingly, the oesophagus in non-venomous species exhibited generally lower expression levels (mean = 21 TPM) compared to venomous species (mean = 260; t = -3, df = 3 TPM, p = 0.02), suggesting higher secretory activity in the latter.

As anticipated, the secretomes of the oesophageal gland and gland of Leiblein were enriched in hydrolases and peptidases, enzymes essential for digestive processes (Fig. 2d). Additionally, *O. amethystina* showed enrichment in 'toxin activity' (see next section), while *M. pilearis* was enriched in terms related to communication and transport (e.g., 'metal ion transport'). Although the salivary glands also expressed hydrolases and peptidases, these enzymes were extracellular, whereas those in the gland of Leiblein were primarily intracellular, consistent with enrichment in cellular compartments like 'lysosome' and 'organelle lumen". In contrast, the venom gland secretome was dominated by toxins released in the extracellular space and of genes involved in the ER function (Fig. 2d).

Identification and characterisation of conotoxins

Cone snail venom is composed primarily of small, disulfide-rich peptides known as conotoxins. To identify conotoxins in our dataset, we used the ConoPrec tool from the Conoserver website[25]. The number of putative conotoxin transcripts predicted in the venomous species' assemblies was consistent with previous findings for cone snail venom gland *de novo* transcriptomes[26–29], with 150-250 toxins predicted per species. Notably, conotoxin-like genes were also identified in other species, with numbers ranging from 33 in

M. mitra to 88 in *O. amethystina*. However, when restricting the analysis to sequences with both a signal peptide and a predicted conotoxin domain, the numbers were significantly reduced, ranging from 16 to 108 in venomous species, and 7 to 28 in non-venomous species.

As expected, conotoxins in venomous species were predominantly expressed in the venom gland (Supplementary Fig. 12). A few were also highly and specifically expressed in the salivary glands, consistent with previous reports in other cone snail species[30]. Interestingly, putative conotoxins were also expressed in other tissues; however, these sequences lacked the predicted conotoxin domain and instead contained domains such as Kunitz and Kazal-type protease inhibitor. In other organisms, such as spiders, the secretion of protease inhibitors is thought to neutralise the digestive enzymes like trypsin in the body of ingested prey[31]. Therefore, the putative conotoxins expressed outside of the two exocrine glands likely represent physiological proteins with domains resembling those of animal toxins. In non-venomous species, putative conotoxins with both, a signal peptide and a conotoxin predicted domain, were expressed at much lower levels and across multiple tissues, although a preference for salivary glands was observed (Supplementary Fig. 13-14).

It is important to note that ConoPrec is design to predict only conoidean toxins. Consequently, toxins like echotoxin, a haemolytic compound found in *Monoplex* and expressed in the salivary glands[32], were not predicted, leading to an underestimation of the toxic potential of gastropod salivary glands[33]. Our focus on conotoxins stems from their critical role as the primary weapon of cone snails, whose massive gene expansion and diversification likely drove the evolution of a specialised organ for their production - the venom gland.

Between-species analysis

Orthogroup assignment

We assigned 330,491 genes (68% of the total) to 41,720 orthogroups (OGs), with 2,588 OGs shared across all species and 197 OGs present as single-copy genes. To compare expression patterns across tissues and species, we created a multi-species expression matrix using the 2,588 shared OGs common to all species. Since many OGs contained multiple genes per species, we selected a representative gene for each OG using two methods: 1) randomly selecting a single transcript's TPM value, and 2) calculating the mean TPM across all the transcripts within an OG. Both approaches produced similar outcomes, therefore the results presented here are based on the first method. Results from the second approach are provided in the Supplementary Information.

Transcriptome similarity and shared tissue specificity between organs and species

To determine whether homologous glands share similar global gene expression profiles or if their functional specialisations align them more closely with non-homologous organs, we analysed whole transcriptome similarity patterns by means of correlation matrix and PCA. Overall, samples primarily clustered by tissue type both based on (Fig. 3a, Supplementary Fig. 15-16). The oesophageal gland and gland of Leiblein grouped together, while the venom gland clustered with salivary glands rather than with its homologous counterparts. Interesting patterns were observed when comparing tissue specificity across species. We found that the oesophageal glands on average shared more tissue-specific OGs with the

glands of Leiblein (mean N = 61) and the oesophagus of the glandless *M. mitra* (mean N. = 56) than with each other (mean N. = 41). This suggests substantial variation between the two OEG-species, as corroborated by the GO enrichment analysis (Supplementary Fig. 2-4). The glands of Leiblein, in contrast, shared more tissue-specific OGs among themselves (mean N. 103), and with the oesophagus of *M. mitra* (mean N. = 100), similar to the oesophageal gland. In contrast, venom glands shared the highest number of tissue-specific OGs among themselves (mean N = 111) and with the salivary glands of *M. mitra* (mean N = 71), while very few with the oesophageal gland (mean N = 11) and the gland of Leiblein (mean N = 16).

These results confirm a strong similarity between the oesophageal gland and the gland of Leiblein. The significant number of tissue-specific OGs in common between these glands and the oesophagus of the glandless species suggests that the latter may have re-acquired gland functions following the loss of a separate gland. In contrast, the venom gland showed greater similarity to the salivary glands, particularly those of the glandless species, indicating functional convergence.

For subsequent analyses, we defined OGs as tissue-specific if they were upregulated in at least two species, resulting in 41 OEG-specific OGs, 343 LEG-specific OGs, and 405 VG-specific OGs.

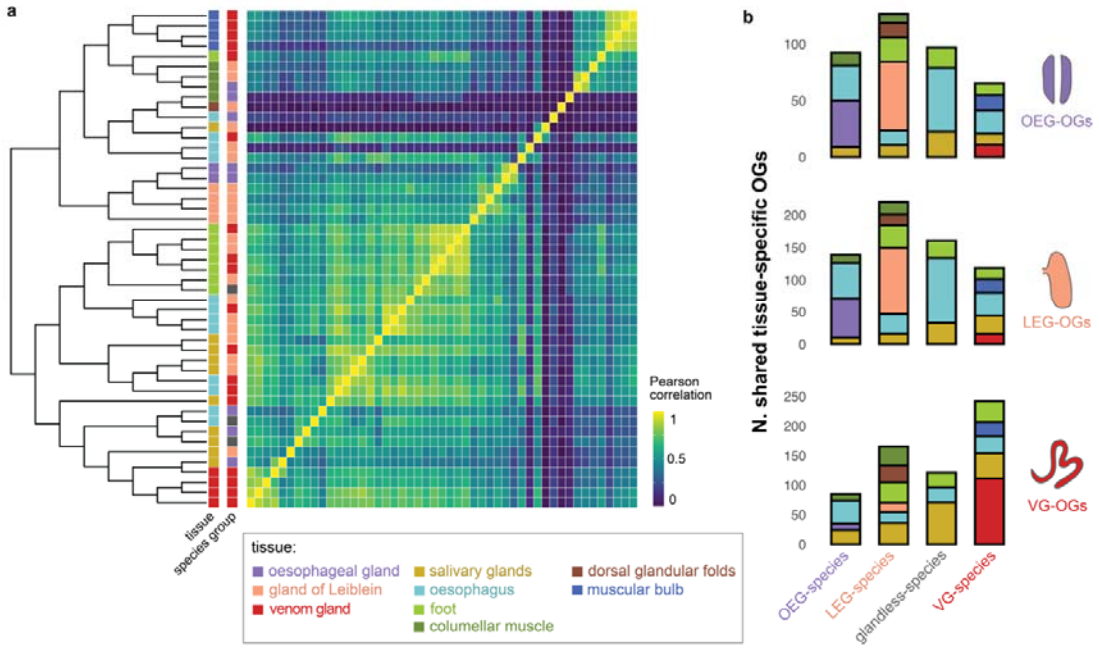


Figure 3. Transcriptome similarity and shared tissue specificity. A) Heatmap of Pearson correlation coefficients between tissues and species. The expression tree was made using neighbour-joining based on the correlation matrix. The samples are colored based on the tissue and the gland type that the species possess. B) Average number of OEG-, LEG-, and VG-specific OGs shared with other tissue-specific OG sets, where species have been grouped by their gland type (OEG-, LEG-, VG- species and glandless).

Rates of gene expression evolution

Next, we investigated the dynamics of gene expression evolution across the phylogeneny to determine whether shifts in organ function were accompanied by significant changes in gene expression. Given the marked functional divergence of the venom gland compared to its homologous glands, we hypothesised that the venom gland transcriptome evolves faster compared to its homologous glands. We tested this hypothesis using CAGEE[34], which employs a bounded Brownian motion model to estimate the most likely value of the evolutionary rate parameter (σ^2) consistent with an ultrametric species tree and observed expression values at the tip of the tree. We ran CAGEE for the mid-oesophageal glands, salivary glands, and oesophagus, as these tissues were sampled across most species representing all three gland types.

We evaluated four different evolutionary models (Supplementary Fig. 17). The first model estimated a single rate σ^2 across all species. The second model estimated two distinct rates, one for the venomous clade and one for all other species. In the third model, species were grouped by gland type and the rates were estimated for each group separately. The final model also calculated three rates but assigned them randomly across the phylogeny. Overall, the third model, which estimated different rates for the OEG-, LEG-, and VG-species, had the best fit (Table 1, Supplementary Table 1). Notably, the model with the poorest fit was the first one, indicating that some degree of variation in evolutionary rates, even if random, is more consistent with the data than assuming a uniform rate across all lineages. The venom gland had the highest σ^2 value, supporting our hypothesis of accelerated evolution in the venom gland compared to the other homologous glands. A similar trend was observed in the other organs, with higher rates in the venomous species compared to the non-venomous ones. However, when comparing across organs, the venom gland did not have the highest σ^2 value, suggesting concerted evolution across the digestive system.

Table 1. Evolutionary rates for the tested models.

tissue	model	N. of evolutionary rates (σ2)	likelihood (-ln L)	OEG-species σ2	LEG-species σ2	glandless-species σ2	VG-species σ2
gland	random	3	17109.7				
	1	1	17649.4	1.02	1.02	-	1.02
	2	2	16525.1	0.59	0.59	-	1.79
	3	**3**	**16018.7**	**0.23**	**0.72**	**-**	***1.80***
salivary glands	random	3	17480				
	1	1	17953.2	0.93	0.93	0.93	0.93
	2	2	15434.8	0.51	0.51	0.51	2.93
	3	**3**	**15314.9**	**0.35**	**0.58**	**0.35**	***2.86***
oesophagus	random	3	22018.2				
	1	1	22728.1	1.15	1.15	1.15	1.15
	2	2	21316.9	0.70	0.70	0.70	2.11
	3	**3**	**21089.5**	**0.39**	**0.81**	**0.39**	***2.10***

For each organ is reported: the number of evolutionary rates estimated (σ^2), the likelihood, and the σ^2 estimates for each group.

Expression changes of tissue-specific orthogroups along the phylogeny

Given the better fit of the third evolutionary model, we used this one for ancestral state reconstruction to assess the number and direction of expression changes at each node of the phylogenetic tree (Fig. 4b). We observed significant changes at node 18, which leads to the glandless *M. mitra* and the venomous clade (Fig. 4a, Supplementary Fig. 18a). At this node, many OEG- and LEG-specific OGs showed a marked decrease in expression in the ancestral gland, coupled with an increase in expression in the oesophagus and salivary glands. Conversely, VG-specific OGs underwent upregulation in the gland as well as across other tissues (Fig. 4a).

Among the OEG- and LEG-specific OGs, the most significant downregulation in the gland at node 18 was a galectin, which simultaneously showed the highest upregulation in the oesophagus (Fig. 4d), alongside with a member of the ependymin family. The top three VG-specific OGs that were upregulated in the gland at node 18 included an integral membrane protein of the DAD family, a disulfide-isomerase, and a selenoprotein F (Fig. 4d), all involved in the ER.

We then utilised the ancestral state reconstruction from CAGEE to calculate tissue specificity at each node, akin to our approach for extant species. The goal was to understand whether gland-specific OGs were tissue-specific in the ancestral lineages. Our findings reveal that some OEG- and LEG-specific OGs were gland-specific in ancestral lineages until node 18, where they shifted mainly to the oesophagus (Fig. 4c, Supplementary Fig. 18c). The galectin and ependymin mentioned earlier were among the genes that transitioned to oesophagus specificity, alongside other genes involved in gluconeogenesis. In contrast, VG-specific OGs were primarily salivary gland-specific in early nodes but shifted to VG-specific at node 18 (Fig. 4c). These shifts suggest that the organs of the digestive system had to adapt to the new function, or loss of, the mid-oesophageal gland.

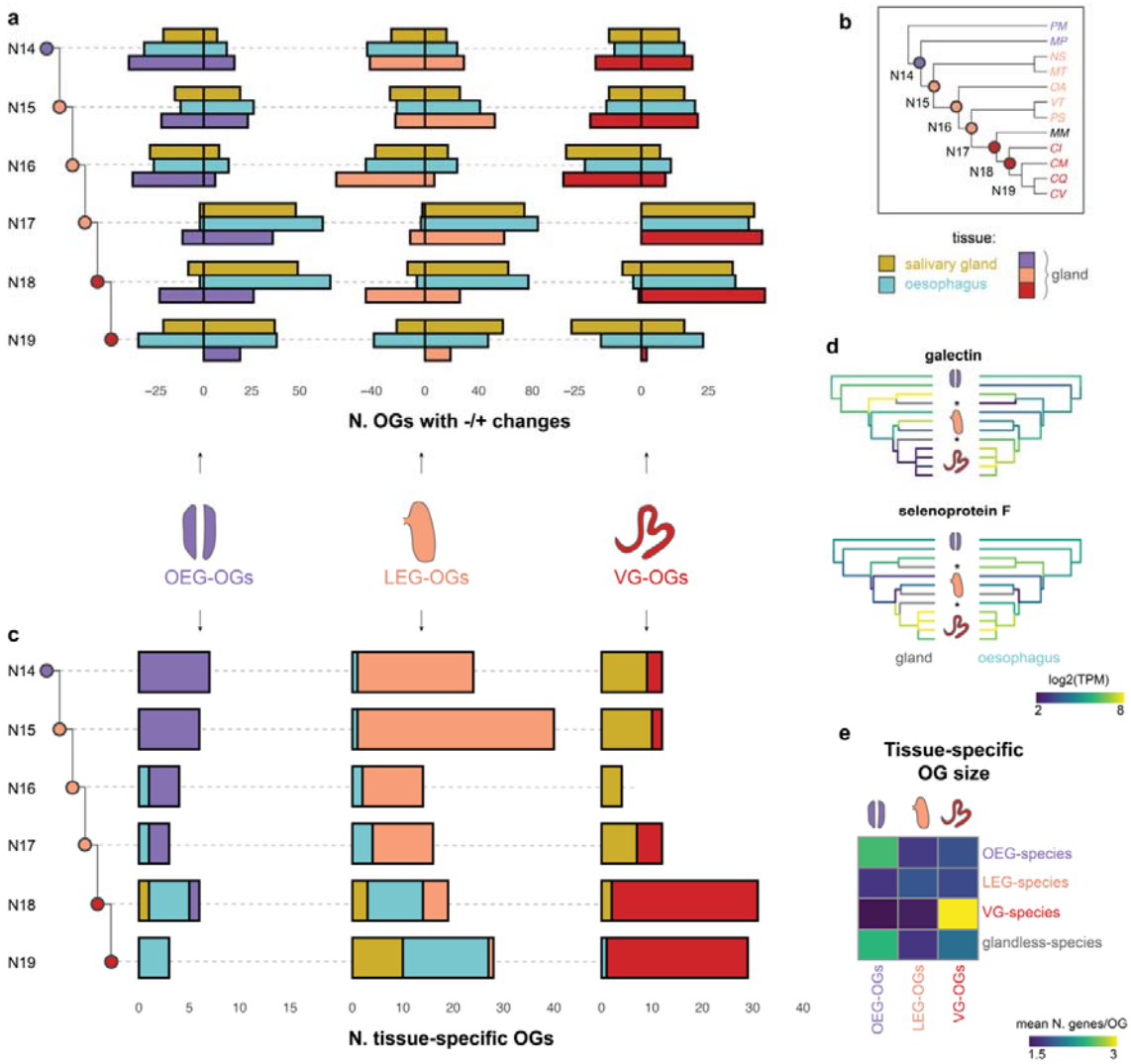


Figure 4. Gene expression dynamics across the phylogeny. a) Number of OEG-, LEG, and VG-specific OGs decreasing and increasing their expression levels in the ancestral salivary glands, oesophagus, or mid-oesophageal gland at each internal node of the gastropod phylogeny. The expression changes were calculated based on gene expression reconstruction at each node of the phylogeny. b) Species phylogeny with the number of the internal nodes. c) Number of OEG-, LEG-, and VG-specific OGs which were tissue-specific in the ancestral mid-oesophageal gland, salivary glands, and oesophagus at each node of the phylogeny. d) Ancestral reconstruction of gene expression of a galectin (OG2631) and selenoprotein F (OG465) in the mid-oesophageal gland and oesophagus. The missing tissues are marked with a *. e) Heatmap representing the mean number of genes within OEG-, LEG, and VG-specific OGs. VG-specific OGs are have significantly more genes in the VG-species (N = 3.2;) compared to the OEG-species (N = 1.8; t = 7.4m df = 1216.3, p = 1.19e-13), LEG-species (N = 1.9; t = 8.4, df = 1986.8, p = 2.2e-16), and the glandless species (N = 2; t = 5.6, df = 566.8, p = 1.408e-08).

Evolution of novel genes for venom production

The final question we sought to address was whether cone snails evolved new genes specifically for the novel function of venom production. Such genes could either appear *de*

novo within the venomous clade or originate from duplication events. In the first scenario, we would expect some VG-specific OGs to be found exclusively in the venomous clade. In the second scenario, the VG-specific OGs would be larger in venomous species compared to non-venomous ones.

To test the first scenario, we examined VG-specific OGs found exclusively in the venomous clade. These represent a small proportion of the gene set (16%), and included conotoxins, uncharacterised proteins, or genes with encoding collagen alpha-5 with a von Willebrand factor type A domain. It is important to note that our transcriptome assemblies were derived from a limited number of organs, therefore some of these OGs may be expressed in tissues not investigated in the non-venomous species.

To test the second scenario, we compared the average size (i.e., number of genes per OG) of gland-specific OGs between venomous and non-venomous species. As shown in Fig. 4e, VG-specific OGs were significantly larger in the venomous species (mean N = 3.2) compared to all the others. Overall, 155 VG-specific OGs (38%) underwent expansion in the venomous clade, with many of these OGs annotated as uncharacterised proteins. Among those with known annotations, we found a predominance of genes related to regulation, such Zinc finger transcriptional regulators (e.g., ZNF862, Yy1), and transposition, including retrovirus-related Pol polyproteins from retrotransposable elements, transposon TX1, and NFX1-type zinc finger-containing proteins.

The discovery of genes related to transposable elements (TEs) warranted further investigation since a relationship between toxins and TEs has been documented[35–37]. We therefore quantified the number of gland-specific genes annotated as TEs across organs and species. The venom gland contained significantly more tissue-specific TEs (mean N = 20.2) than the gland of Leiblein (mean N = 6; t = 3.8, df = 5.9, p = 0.004) and the oesophageal gland (mean N = 4.5; t = 5.1, df = 3.9, p = 0.003), but not compared to other tissues (Supplementary Fig. 19a). Similarly, the expression levels were higher in the venom glands, although not statistically significant (Supplementary Fig. 19a). The most highly expressed families in venom glands were Pol and LINE-1 retrotransposable element ORF2 (Supplementary Fig. 19b).

Discussion

Animal venom glands are highly specialised exocrine organs focused solely on toxin production[1]. Our functional enrichment analyses of gland-specific genes indeed revealed that venom glands are uniquely characterised by highly upregulated protein secretion pathways, consistent with other lineages[4,7,8]. In contrast, genes specifically expressed in the gland of Leiblein, and, to a lesser extent, in the oesophageal gland, are associated with intracellular digestion, particularly of proteins, as well as uptake and storage of lipids and carbohydrates[14]. Additionally, the oesophageal gland appears to play a role in homeostasis and transport of fluid and nutrients, along with secretion of digestive enzymes[11].

The phylogenetic distribution of glands, and the function of their transcriptomes, suggest a scenario of functional and morphological evolution in neogastropods. First, the role of the mid-oesophageal glands in absorption and intracellular digestion requires nutrients to be either pre-digested by salivary gland digestive enzymes or easily derived from soft tissues.

Indeed, species that drill holes and consume soft tissues, such as members of the Muricidae family, possess a well-developed gland of Leiblein[14], and the salivary glands open in the buccal cavity at the tip of the proboscis, in close proximity with the prey (Kantor 1996). Second, as snail diversified their diet and adopted macrophagous feeding habits, there was a trend towards the reduction and eventual loss of the gland. For instance, Buccinoidea species ingest prey whole or in large chunks[38] and possess a simple gland[14]. In species where the gland no longer serves as a site of absorption, it either disappear entirely, as in the Mitridae, or specialised into something else, as seens in conoideans, including cone snails. These changes were facilitated by the coordinated evolution of other digestive organs which adapted by shifting their functions.

Our ancestral state reconstruction reveals extensive changes in gene expression across the phylogeny, reflecting the dynamic nature of this system and the adaptation of these snails to diverse ecological niches and feeding strategies. Notably, the venomous clade experienced more substantial expression changes and higher evolutionary rates than non-venomous species, consistent with the drastic functional divergence of the venom gland from its homologous counterparts. However, the venom gland itself did not exhibit the highest evolutionary rates, suggesting that other organs co-evolved to support its new function. In the ancestor of the venomous species and the glandless *M. mitra* (node 18), several genes became downregulated in the gland while upregulated in the oesophagus (Fig. 4a, c). Among these were genes involved in gluconeogenesis, such as galectin and PEPCK, indicating that the ancestral gland's role in producing energy from non-carbohydrate substrates was reassigned to other digestive tissues. This concerted adaptation is particularly evident in the glandless species, where many OEG- and LEG-specific genes are now oesophagus-specific (Fig. 3b), especially those related to lysosomal functions. Mitridae snails feed on soft-bodied Sipuncula worms which they may ingest whole[39,40], or by pumping the worm's viscera[41], including coelomic fluids and eggs[42], into the buccal cavity. In such cases, nutrients may be pre-digested by salivary enzymes, allowing the oesophagus to take over functions previously performed by the gland of Leiblein.

During the evolution of conoideans, the gland, while losing its original digestive functions, increased its secretory capacity, a change driven by the modulation of genes involved in pre-existing secretory machinery. Our analysis identified key upregulated genes in the ancestral venom gland (node 18) that encode proteins active in the ER, such as DAD1, essential for N-glycosylation and protein translocation[43], disulfide isomerases, and a selenoprotein F which may be involved in ER protein folding quality control[44]. Additionally, genes linked to the UPR and ER stress response are enriched in cone snail venom glands, consistent with observations across distinct venomous lineages[4,8]. Studies in snakes have led to a model where venom production activates the UPR, creating a positive feedback loop that enhances venom production through up-regulation and binding of UPR transcription factors targeting toxin genes[4–6]. Future comparative studies, leveraging the increasing availability of genomes from venomous species[45], including conoideans[46–49], combined with functional genomics (e.g., ATAC- and ChIP-seq) could further advance our understanding of the involvement of this stress-related pathway in venom production.

In earlier nodes of the tree, OEG- and LEG-specific genes were expressed in the ancestral gland, while VG-specific genes were found in salivary glands, a pattern that we observed also in modern species (Fig. 3, Fig. 4c). The convergence between venom and salivary

glands is logical, as both are exocrine organs secreting products into the extracellular environment. Notably, some neogastropods secrete toxins in their salivary glands[33], further emphasising their similarity. Interestingly, the salivary glands of the glandless species *M. mitra* share tissue-specific genes with the venom glands. Unlike other snails, *Mitra*'s salivary ducts open at the tip of the epiproboscis, an extendible muscular rod situated within the proboscis[20], likely facilitating the delivery of secretions to prey[41]. We observed overexpression of cysteine-rich secreted proteins, peptidases, and serine proteases in *Mitra*'s salivary glands, indicating a role in tissue degradation and potential toxin activity. Despite this functional similarity, *Mitra*'s salivary glands did not evolve into specialised venom glands. One possible explanation lies in their diet – Mitridae snails have adapted to feed on Sipuncula worms, a niche food source that is scarcely exploited by other taxa and consisted of relatively inactive prey. This reduced competition likely diminished the selective pressure to evolve a specialised venom apparatus, unlike cone snails, which faced greater competition and diversified to prey on fast-moving organisms that required paralysing toxins for successful capture.

While the venom gland's secretory role evolved through changes in the expression of ancient genes, its secretory products, i.e., toxins, evolved through the emergence of novel genes. Conotoxin diversity is likely driven by duplications followed by diversification through insertions and deletions mainly in the pro and mature regions[50], although the original physiological protein from which conotoxins derived remain unknown[36,37]. We identified transcripts with conotoxin domains in non-venomous species, consistent with other studies[32,49,51–53], suggesting that conotoxin-related peptides may represent a shared evolutionary novelty among predatory gastropods. However, these related putative toxins do not experience the same explosive diversification as conotoxins, nor are they produced at high levels in a dedicated organ, such as the venom gland.

Besides conotoxins, we observed that many VG-specific gene families that underwent expansion in the venomous clade are annotated as transcriptional regulators, aligning with previous observations that venom gland regulation is lineage-specific[8], and with retrotransposition. A relationship between toxin gene diversity and transposable elements (TEs) has been previously documented in cone snails, where high density of repetitive elements was found in conotoxin gene flanking regions[37,47]. Similarly, in snakes, TEs were found implicated in non-homologous recombination of toxin genes[35]. Our study not only confirms the expansion of TEs in cone snails, but also reveals their tissue-specificity to the venom gland. Further genomic analysis coupled with advanced sequencing technologies will be essential to elucidate the role of TEs in venom production and the mechanisms underlying toxin gene regulation.

Methods

Sample collection and sequencing

Specimens from Neogastropoda and two outgroup species were collected during the expedition Koumac 2.3 of the Natural History Museum Paris in 2018 and 2019 in Koumac, New Caledonia under permit N°609011- 55 /2019/DEPART/JJC. Individuals were collected by searches during scuba dives and kept for a maximum of three days in aquaria with fresh sea water before dissection. The following tissues were dissected and preserved in RNA*later*

(Invitrogen): salivary glands, foot, columellar muscle, oesophagus, oesophageal gland, gland of Leiblein, dorsal glandular folds (only in *M. tenuirostrum*), venom gland, muscular bulb (only in cone snails) (Supplementary Data 1).

Tissue samples were homogenised with Trizol (Invitrogen) and total RNA purified with the PureLink™ RNA Mini kit (Invitrogen) with an additional DNase I treatment following manufacturer's protocol. cDNA libraries were constructed using the NEBnext Ultra II Directional RNA Library Preparation Kit with polyA selection and dUTP method (New England BioLabs) and sequenced on an Illumina NovaSeq 2x150bp. Raw paired-end reads were checked with FastQC 0.11.9[54], quality-filtered and trimmed with FastP 0.20.1[55], and the outputs combined with multiQC 1.10.1[56].

De novo assembly and annotation

All reads from all specimens of each species were pooled to generate *de novo* transcriptome assemblies using rnaSpades 3.15.2[57]. To reduce assembly size and redundancy, and remove spurious transcripts, we adopted a series of filtering steps. First, we translated the transcripts to amino acid sequences using Borf[58] and kept only those with a complete open read frame. Second, we compared the translated sequences with BlastP[59] against a suite of databases including UniprotKB/Swiss-Prot[60], a set of 11 Gastropoda genomes (Supplementary Table 2), Conoserver[25] and Tox-Prot[61]. Protein domains were identified with PSIblast[59] against the Pfam[62] and Cdd[63] databases. Only hits with evalue < 1e-05 were retained. Additionally, we run signalP[24] to annotate signal peptides which are relevant in the identification of toxin transcripts. Putative conotoxin were predicted with ConoPrec[25]. All sequences with a UniprotKB/Swiss-Prot hit to a non-Metazoa organism were removed. We then trimmed the retained transcripts to their coding region, reduced redundancy by clustering sequences with 98% or more identical nucleotide sequences with Cd-hit[64], and kept only transcripts with more than one read count in at least one library. Finally, we evaluated assembly completeness using Omark[65].

For each species we performed GO annotation by combining the annotation from Pannzer2[66] and DeepGOPlus 1.0.2[67]. Based on the score values distribution, we used 0.3 as a score threshold for both methods.

Within-species analysis

Expression levels

We mapped all libraries to the respective species assembly with Kallisto 0.48.0[68] with 100 bootstrap and quantified gene expression using the package Sleuth 0.30.1[69] in RStudio 2023.6.2.561[70] using R 4.2.2[71]. We employed a suite of quality control steps to identify and remove outliers. First, we analysed read count distribution with *vioplot* 0.4.0 and removed samples with particularly low distributions. Then, we utilise dimensionality reduction techniques, including principal component analysis (*dudi_PCA*) and multidimensional scaling (*plotMDS*) to verify that samples were clustering by tissue type. If a sample had unusual clustering, i.e., it was far from the others of the same tissue, we flagged it as problematic and excluded. Normalised estimated counts and TPM abundances were then re-calculated after outlier libraries' removal. As in some species samples still tended to cluster by individual specimen rather than tissue type, we corrected the expression values to account for the specimen effect using an empirical Bayes method implemented via the *ComBat_seq* function in the package sva 3.46.0[72] which specifically targets RNA-Seq data. We build a full model as ~ specimen + tissue_type, and a reduced model as ~ specimen.

Tissue-specific gene sets

For each species, we identified tissue-specific genes based on their fold change (FC) calculated as the ratio between the TPM value of the first most-highly expressed tissue and the TPM value of the second-most highly expressed tissue. A gene with TPM ≥ 2 and FC ≥ 2 was classified as specific of the top tissue. We validated our FC method by confirming that the tissue-specific genes identified were also significantly differentially expressed when analysed using the likelihood ratio test in Sleuth. Differences in the number of tissue-specific genes between glands were tested by means of one-tailed t-tests. Functional enrichment of tissue-specific gene sets was performed with TopGO 2.50.0[73] using the *elim* algorithm and Fisher test. The foreground was the list of tissue-specific genes while the background included all the genes expressed in that species.

Between-species analysis

Orthogroup expression matrix

Protein sequences were assigned to orthogroups with the OrthoDB standalone pipeline OrthoLoger 3.0.2[74] using default parameters. Since most orthogroups (OGs) included more than one gene per species (i.e., one-to-many or many-to-many orthologs), we randomly selected one representative sequence for each OG in each species and used the TPM value estimated for that gene as the orthogroup expression value. This method was shown to be robust[8]. Alternatively, we calculated the mean TPM values across all genes belonging to the same OG. All samples were then merged into a multi-species multi-tissue matrix and the expression levels corrected for the species effect using *ComBat_seq*. All the downstream analyses reported in the main text are based on the random matrix, while the results from the mean-based matrix are reported in Supplementary Information.

To have an overview of global transcriptome similarity across tissues and species, we calculated pairwise distances as 1- Spearman correlation and used it to reconstruct a gene expression tree using the neighbour-joining method. Additionally, we calculated the proportion of shared OGs among tissue-specific genes. For downstream analyses, OGs were classified tissue-specific if at least one gene within that OG was specific in a tissue, and if it was found specific in at least two species.

Gene expression evolution analysis

We analysed changes in gene expression with the program CAGEE 1.1.1[34] which uses Brownian motion to model gene expression across a phylogenetic tree. The tree was derived from the phylogeny of Fedosov et al.[75] by excluding the families not encompassed in this study and retaining only the branches most closely related to the species that we examined. The rates of expression changes (σ^2) were calculated for the mid-oesophageal glands, salivary glands, and oesophagus separately. We fit a series of nested models in which σ^2 varies across branches of the species tree to test for different hypotheses. In the first model we assigned a single rate for the entire tree. In the second model, the venomous lineage had a distinct rate from all the other branches. In the third model, we assigned the same rate for species with the same gland type. In the last model, we assigned branches randomly to three group rates. We then used the best fit model to performed ancestral state reconstruction. The reconstructed transcriptomes at the inner nodes were used to calculate ancestral tissue specificity for each OG as we did for extant species.

Evolution of novel genes

To test whether new genes evolved along with the venom gland, we examined the number of OGs that were found exclusively in the venomous species, as well as whether venom gland-specific OGs had, on average, a higher number of genes in venomous species compared to the other species (i.e., mean OG size in VG-species > mean OG size of all the other species groups).

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie Grant Agreement 845674 to G.Z. We are grateful to Philippe Buchet and the team of the expedition Koumac 2.3 for providing the logistic support and expertise, David Massemin and Patrick for their skills in finding the species needed. We thank Alexander Fedesov for providing the phylogenetic tree, Consolée Aletti for support with the RNA extraction, the personnel of the Genomic Technologies Facility and Genewiz for the RNA-seq experiment, and members of the Robinson Rechavi's lab for bioinformatics support and scientific discussions.

Author contributions

G.Z., M.V.M., N.P., Y.K., and M.R.-R. conceived the study design; Y.K., and M.V.M perform the dissections; G.Z. generated the RNA-seq data, perfomed data analysis, and wrote the manuscript; A.B. performed the GO annotation and G.C. the orthogroup assignment. All the authors contributed to the interpretation and discussion of the results, and to the final version of the manuscript.

Competing interests

The authors declaire no competing interests.

Data availability

The RNA-seq data generated in this study have been deposited in the NCBI SRA archive with the accession number PRJNA1158673. Additional data generated in this study and the R scripts used to analyse the data are available at zenodo.org. Additional data are provided as Supplementary Information files.

References

1. Schendel, V., Rash, L. D., Jenner, R. A. & Undheim, E. A. B. The diversity of venom: The importance of behavior and venom system morphology in understanding its ecology and evolution. *Toxins* **11**, 666 (2019).
2. Casewell, N. R., Wüster, W., Vonk, F. J., Harrison, R. A. & Fry, B. G. Complex cocktails: the evolutionary novelty of venoms. *Trends in Ecology & Evolution* **28**, 219–229 (2013).

3. Zancolli, G. & Casewell, N. R. Venom systems as models for studying the origin and regulation of evolutionary novelties. *Molecular Biology and Evolution* **37**, 2777–2790 (2020).

4. Perry, B. W., Schield, D. R., Westfall, A. K., Mackessy, S. P. & Castoe, T. A. Physiological demands and signaling associated with snake venom production and storage illustrated by transcriptional analyses of venom glands. *Scientific Reports* **10**, 18083 (2020).

5. Perry, B. W. *et al.* Snake venom gene expression is coordinated by novel regulatory architecture and the integration of multiple co-opted vertebrate pathways. *Genome Res.* **32**, 1–16 (2022).

6. Westfall, A. K. *et al.* Single-cell heterogeneity in snake venom expression is hardwired by co-option of regulators from progressively activated pathways. *Genome Biology and Evolution* **15**, evad109 (2023).

7. Barua, A. & Mikheyev, A. S. An ancient, conserved gene regulatory network led to the rise of oral venom systems. *Proceedings of the National Academy of Sciences* **118**, e2021311118 (2021).

8. Zancolli, G., Reijnders, M., Waterhouse, R. M. & Robinson-Rechavi, M. Convergent evolution of venom gland transcriptomes across Metazoa. *PNAS* **119**, e2111392119 (2022).

9. Zuniga, A. Next generation limb development and evolution: old questions, new perspectives. *Development* **142**, 3810–3820 (2015).

10. Benton, M. J., Dhouailly, D., Jiang, B. & McNamara, M. The early origin of feathers. *Trends in Ecology & Evolution* **34**, 856–869 (2019).

11. Reid, R. & Friesen, J. The digestive system of the moon snail *Polinices lewisii* (Gould, 1847) with emphasis on the role of the oesophageal gland. *Veliger* **23**, 25–34 (1980).

12. Lobo-da-Cunha, A. Structure and function of the digestive system in molluscs. *Cell Tissue Res* **377**, 475–503 (2019).

13. Modica, M. V. & Holford, M. The Neogastropoda: Evolutionary innovations of predatory marine snails with remarkable pharmacological potential. in *Evolutionary Biology – Concepts, Molecular and Morphological Evolution: 13th Meeting 2009* (ed. Pontarotti, P.) 249–270 (Springer, Berlin, Heidelberg, 2010). doi:10.1007/978-3-642-12340-5_15.

14. Andrews, E. B. & Thorogood, K. E. An ultrastructural study of the gland of Leiblein of muricid and nassariid neogastropods in relation to function, with a discussion on its homologies in other caenogastropods. *Journal of Molluscan Studies* **71**, 269–300 (2005).

15. Abdelkrim, J. *et al.* Exon-capture-based phylogeny and diversification of the venomous gastropods (Neogastropoda, Conoidea). *Molecular Biology and Evolution* **35**, 2355–2374 (2018).

16. Nguyen, L. T. T., Craik, D. J. & Kaas, Q. Bibliometric review of the literature on cone snail peptide toxins from 2000 to 2022. *Marine Drugs* **21**, 154 (2023).

17. Morales Duque, H., Campos Dias, S. & Franco, O. L. Structural and functional analyses of cone snail toxins. *Marine Drugs* **17**, 370 (2019).

18. Safavi-Hemami, H., Young, N. D., Williamson, N. A. & Purcell, A. W. Proteomic interrogation of venom delivery in marine cone snails: novel insights into the role of the venom bulb. *J Proteome Res* **9**, 5610–5619 (2010).

19. Salisbury, S. M., Martin, G. G., Kier, W. M. & Schulz, J. R. Venom kinematics during prey capture in *Conus*: the biomechanics of a rapid injection system. *Journal of Experimental Biology* **213**, 673–682 (2010).

20. Harasewych, M. G. Anatomy and biology of *Mitra cornea* Lamarck, 1811 (Mollusca, Caenogastropoda, Mitridae) from the Azores. *Açoreana* **6**, 121–135 (2009).

21. Holford, M. *et al.* Evolution of the Toxoglossa venom apparatus as inferred by molecular phylogeny of the Terebridae. *Molecular Biology and Evolution* **26**, 15–25 (2009).

22. Safavi-Hemami, H. *et al.* Specialized insulin is used for chemical warfare by fish-hunting cone snails. *PNAS* **112**, 1743–1748 (2015).

23. Olivera, B. M. *et al.* Adaptive radiation of venomous marine snail lineages and the accelerated evolution of venom peptide genes. *Annals of the New York Academy of Sciences* **1267**, 61–70 (2012).

24. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* **40**, 1023–1025 (2022).

25. Kaas, Q., Yu, R., Jin, A.-H., Dutertre, S. & Craik, D. J. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Research* **40**, D325–D330 (2012).

26. Abalde, S., Tenorio, M. J., Afonso, C. M. L. & Zardoya, R. Conotoxin diversity in *Chelyconus ermineus* (Born, 1778) and the convergent origin of piscivory in the Atlantic and Indo-Pacific cones. *Genome Biology and Evolution* **10**, 2643–2662 (2018).

27. Gao, B. *et al.* High throughput identification of novel conotoxins from the vermivorous oak cone snail (*Conus quercinus*) by transcriptome sequencing. *International Journal of Molecular Sciences* **19**, 3901 (2018).

28. Dutt, M. *et al.* Venomics reveals venom complexity of the piscivorous cone snail, *Conus tulipa*. *Marine Drugs* **17**, 71 (2019).

29. Prashanth, J. R., Dutertre, S., Rai, S. K. & Lewis, R. J. Venomics reveals a non-compartmentalised venom gland in the early diverged vermivorous *Conus distans*. *Toxins* **14**, 226 (2022).

30. Fedosov, A., Tucci, C. F., Kantor, Y., Farhat, S. & Puillandre, N. Collaborative expression: Transcriptomics of *Conus virgo* suggests contribution of multiple secretory glands to venom production. *J Mol Evol* **91**, 837–853 (2023).

31. Neto, O. B. S. *et al.* Spiders' digestive system as a source of trypsin inhibitors: functional activity of a member of atracotoxin structural family. *Sci Rep* **13**, 2389 (2023).

32. Shiomi, K., Kawashima, Y., Mizukami, M. & Nagashima, Y. Properties of proteinaceous toxins in the salivary gland of the marine gastropod (*Monoplex echo*). *Toxicon* **40**, 563–571 (2002).

33. Ponte, G. & Modica, M. V. Salivary glands in predatory mollusks: Evolutionary considerations. *Frontiers in Physiology* **8**, (2017).

34. Bertram, J. *et al.* CAGEE: Computational analysis of gene expression evolution. *Molecular Biology and Evolution* **40**, msad106 (2023).

35. Dowell, N. L. *et al.* The deep origin and recent loss of venom toxin genes in rattlesnakes. *Current Biology* **26**, 2434–2445 (2016).

36. Phuong, M. A. & Mahardika, G. N. Targeted sequencing of venom genes from cone snail genomes improves understanding of conotoxin molecular evolution. *Mol Biol Evol* **35**, 1210–1224 (2018).

37. Zheng, J.-W. *et al.* Systematic dissection of genomic features determining the vast diversity of conotoxins. *BMC Genomics* **24**, 598 (2023).

38. Kantor, Y. I. Comparative anatomy of the stomach of Buccinoidea (Neogastropoda). *J Molluscan Stud* **69**, 203–220 (2003).

39. Taylor, J. D. The diet of coral-reef Mitridae (Gastropoda) from Guam; with a review of other species of the family. *Journal of Natural History* **23**, 261–278 (1989).

40. Taylor, J. D. Dietary and anatomical specialization of mitrid gastropods (Mitridae) at Rottnest Island, Western Australia. in *Proceedings of the fifth international marine biological workshop: the marine flora and fauna of Rottnest Island, Western Australia* (eds. Wells, F. E., Walker, D. I., Kirkman, H. & Lethdridge, R.) 583–599 (Western Australian Museum, Perth, 1993).

41. West, T. L. Feeding behavior and functional morphology of the epiproboscis of *Mitra idae* (Mollusca: Gastropoda; Mitridae). *Bulletin of Marine Science* **46**, 761–779 (1990).

42. West, T. L. Functional morphology of the proboscis of *Mitra catalinae* Dall 1920 (Mollusca: Gastropoda: Mitridae), and the evolution of the mitrid epiproboscis. *Bulletin of Marine Science* **48**, 702–718 (1991).

43. Zhang, Y., Cui, C. & Lai, Z.-C. The defender against apoptotic cell death 1 gene is required for tissue growth and efficient N-glycosylation in *Drosophila melanogaster*. *Developmental Biology* **420**, 186–195 (2016).

44. Takeda, Y. *et al.* Both isoforms of human UDP-glucose:glycoprotein glucosyltransferase are enzymatically active. *Glycobiology* **24**, 344–350 (2014).

45. von Reumont, B. M. *et al.* Modern venomics—Current insights, novel methods, and future perspectives in biological and applied animal venom research. *GigaScience* **11**, giac048 (2022).

46. Herráez-Pérez, A., Pardos-Blas, J. R., Afonso, C. M. L., Tenorio, M. J. & Zardoya, R. Chromosome-level genome of the venomous snail *Kalloconus canariensis*: a valuable model for venomics and comparative genomics. *GigaScience* **12**, giad075 (2023).

47. Pardos-Blas, J. R. *et al.* The genome of the venomous snail *Lautoconus ventricosus* sheds light on the origin of conotoxin diversity. *GigaScience* **10**, (2021).

48. Liu, Z. *et al.* Chromosome-level genome assembly of the deep-sea snail *Phymorhynchus buccinoides* provides insights into the adaptation to the cold seep habitat. *BMC Genomics* **24**, 679 (2023).

49. Farhat, S., Modica, M. V. & Puillandre, N. Whole genome duplication and gene evolution in the hyperdiverse venomous gastropods. *Molecular Biology and Evolution* **40**, msad171 (2023).

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

51. Lu, A. *et al.* Transcriptomic profiling reveals extraordinary diversity of venom peptides in unexplored predatory gastropods of the genus *Clavus*. *Genome Biology and Evolution* **12**, 684–700 (2020).

52. Gerdol, M., Cervelli, M., Oliverio, M. & Modica, M. V. Piercing fishes: Porin expansion and adaptation to hematophagy in the vampire snail *Cumia reticulata*. *Molecular Biology and Evolution* **35**, 2654–2668 (2018).

53. Turner, A. H., Craik, D. J., Kaas, Q. & Schroeder, C. I. Bioactive compounds isolated from neglected predatory marine gastropods. *Marine Drugs* **16**, 118 (2018).

54. Andrews, S. FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).

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

56. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

57. Bushmanova, E., Antipov, D., Lapidus, A. & Prjibelski, A. D. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience* **8**, giz100 (2019).

58. Signal, B. & Kahlke, T. Borf: Improved ORF prediction in de-novo assembled transcriptome annotation. 2021.04.12.439551 Preprint at https://doi.org/10.1101/2021.04.12.439551 (2021).

59. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).

60. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **51**, D523–D531 (2023).

61. Jungo, F., Bougueleret, L., Xenarios, I. & Poux, S. The UniProtKB/Swiss-Prot Tox-Prot program: A central hub of integrated venom protein data. *Toxicon* **60**, 551–557 (2012).

62. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412–D419 (2021).

63. Wang, J. *et al.* The conserved domain database in 2023. *Nucleic Acids Res* **51**, D384–D388 (2022).

64. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

65. Nevers, Y. *et al.* Quality assessment of gene repertoire annotations with OMArk. *Nat Biotechnol* 1–10 (2024) doi:10.1038/s41587-024-02147-w.

66. Törönen, P., Medlar, A. & Holm, L. PANNZER2: a rapid functional annotation web server. *Nucleic Acids Res* **46**, W84–W88 (2018).

67. Kulmanov, M. & Hoehndorf, R. DeepGOPlus: improved protein function prediction from sequence. *Bioinformatics* **36**, 422–429 (2020).

68. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).

69. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods* **14**, 687–690 (2017).

70. Posit team. RStudio: Integrated Development Environment for R. Posit Software, PBC (2023).

71. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. (2019) doi:URL http://www.R-project.org/.

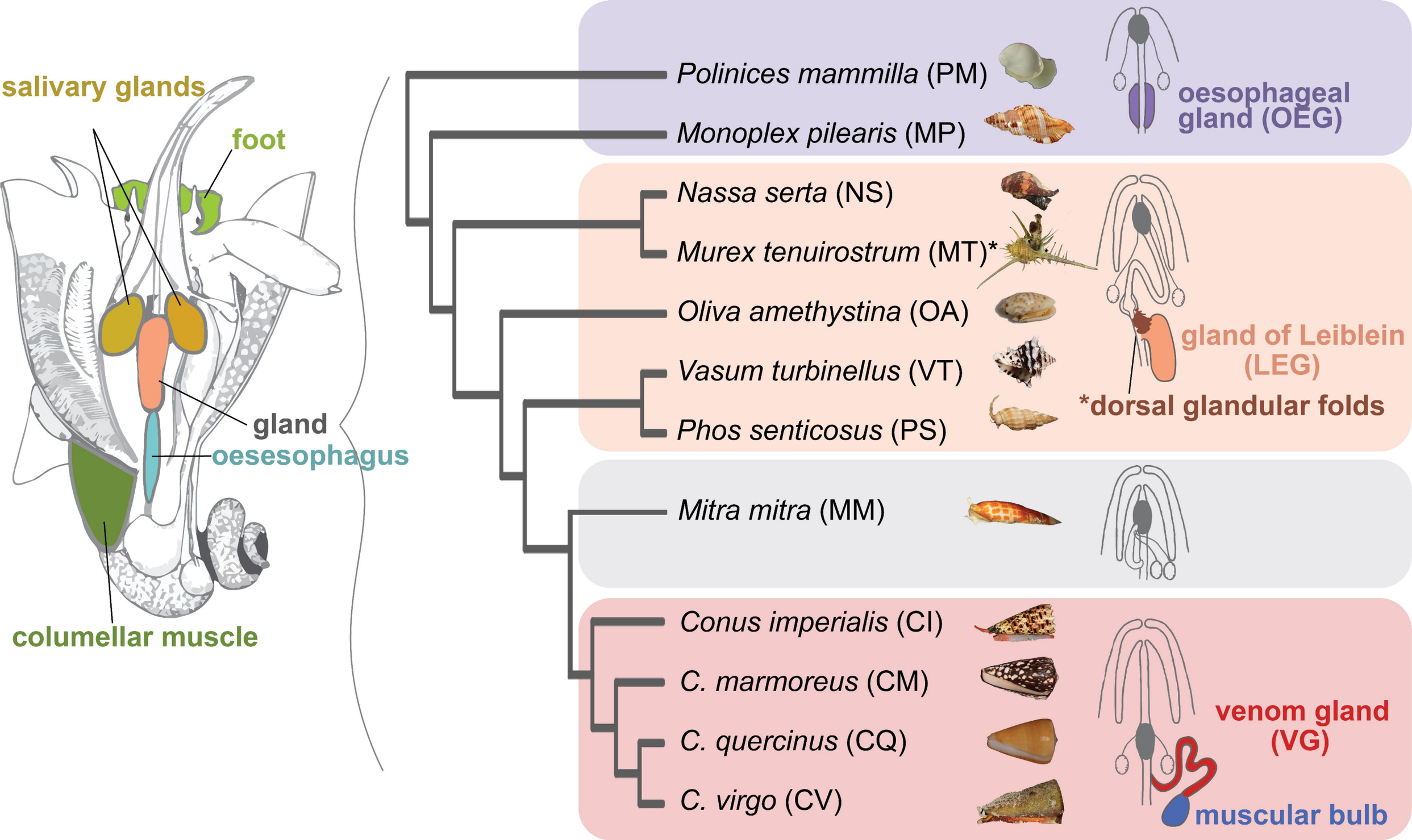
72. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

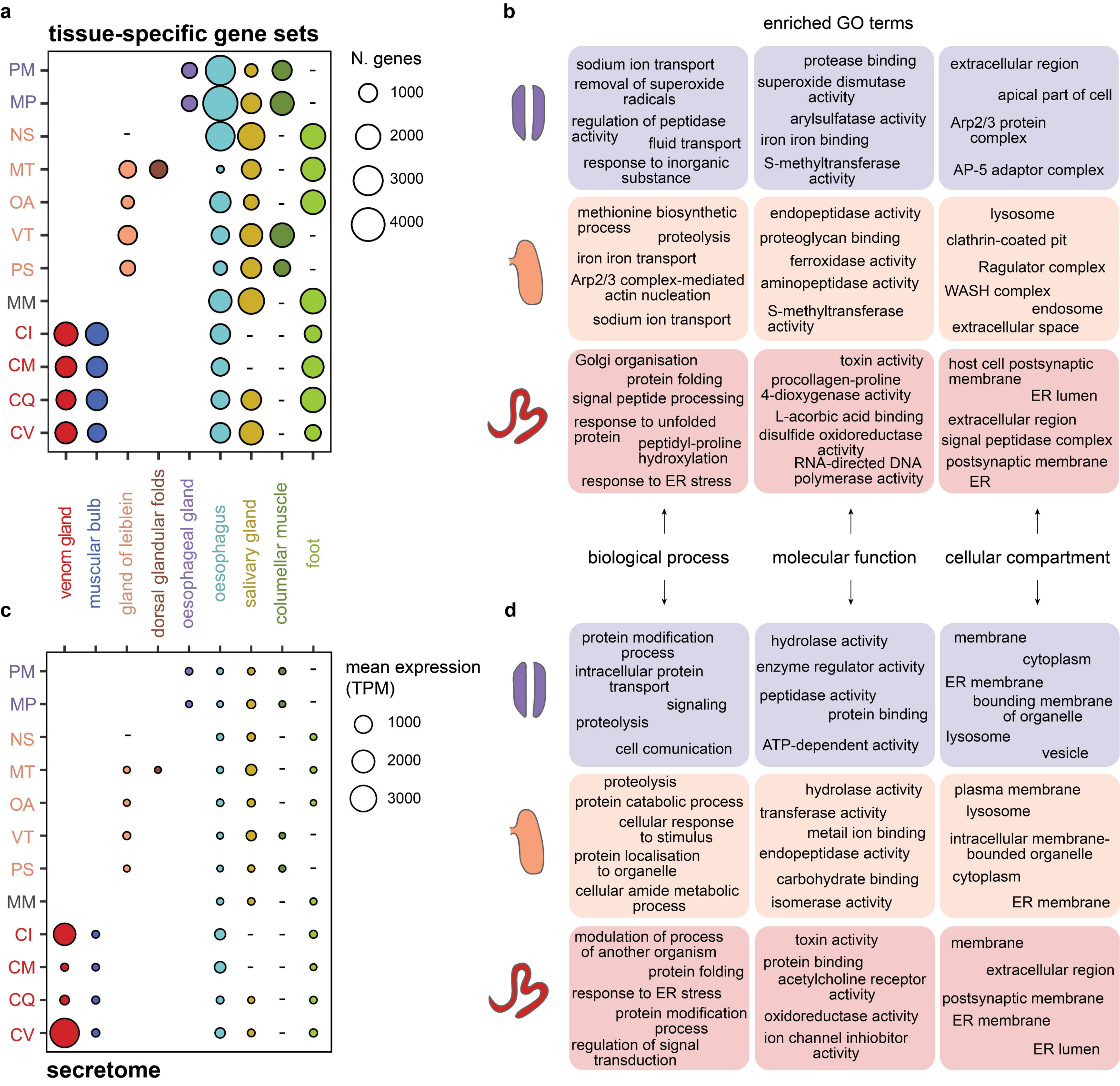
73. Alexa, A. & Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology. R package version 2.38.1. (2019).

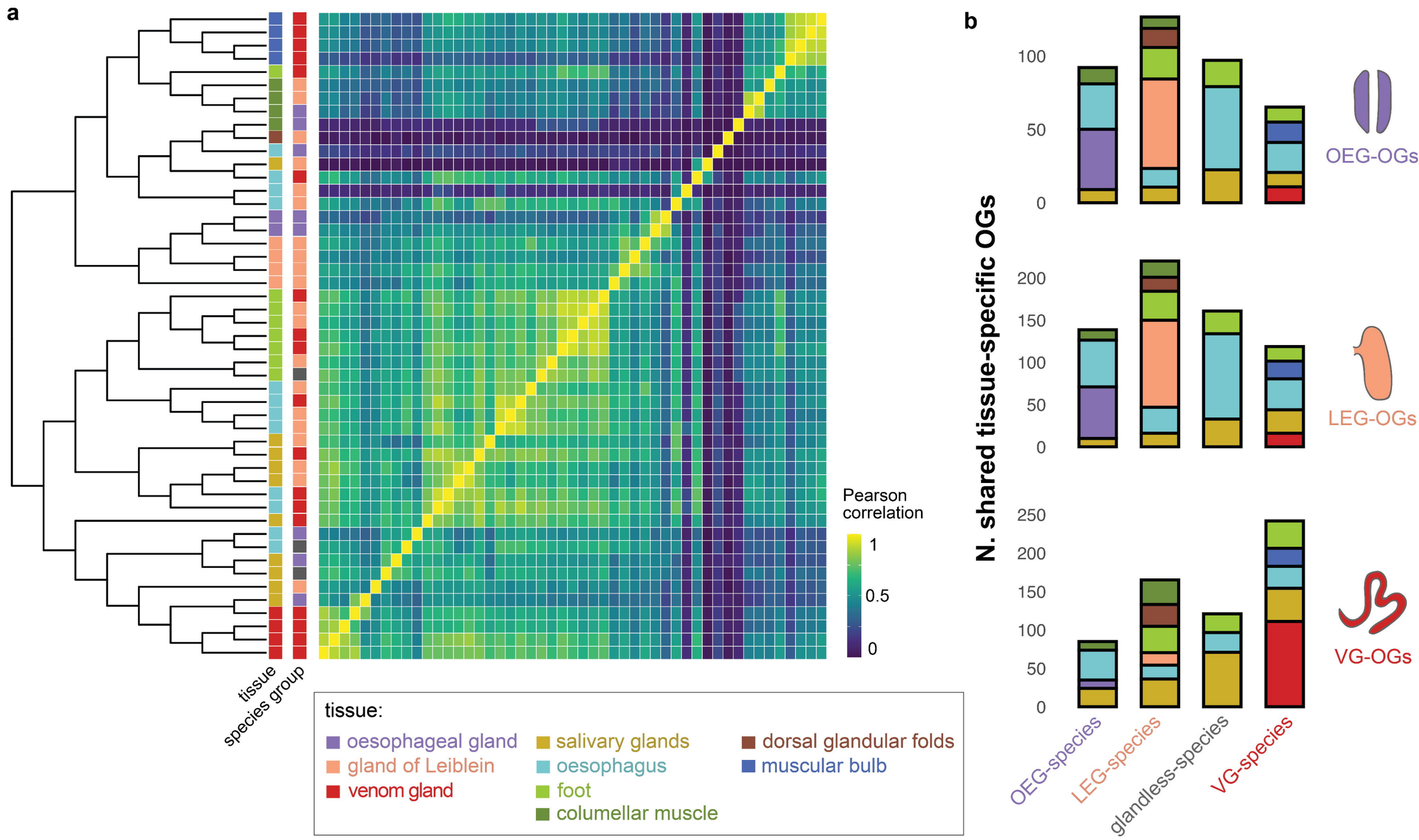
74. Kuznetsov, D. *et al.* OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research* **51**, D445–D451 (2023).

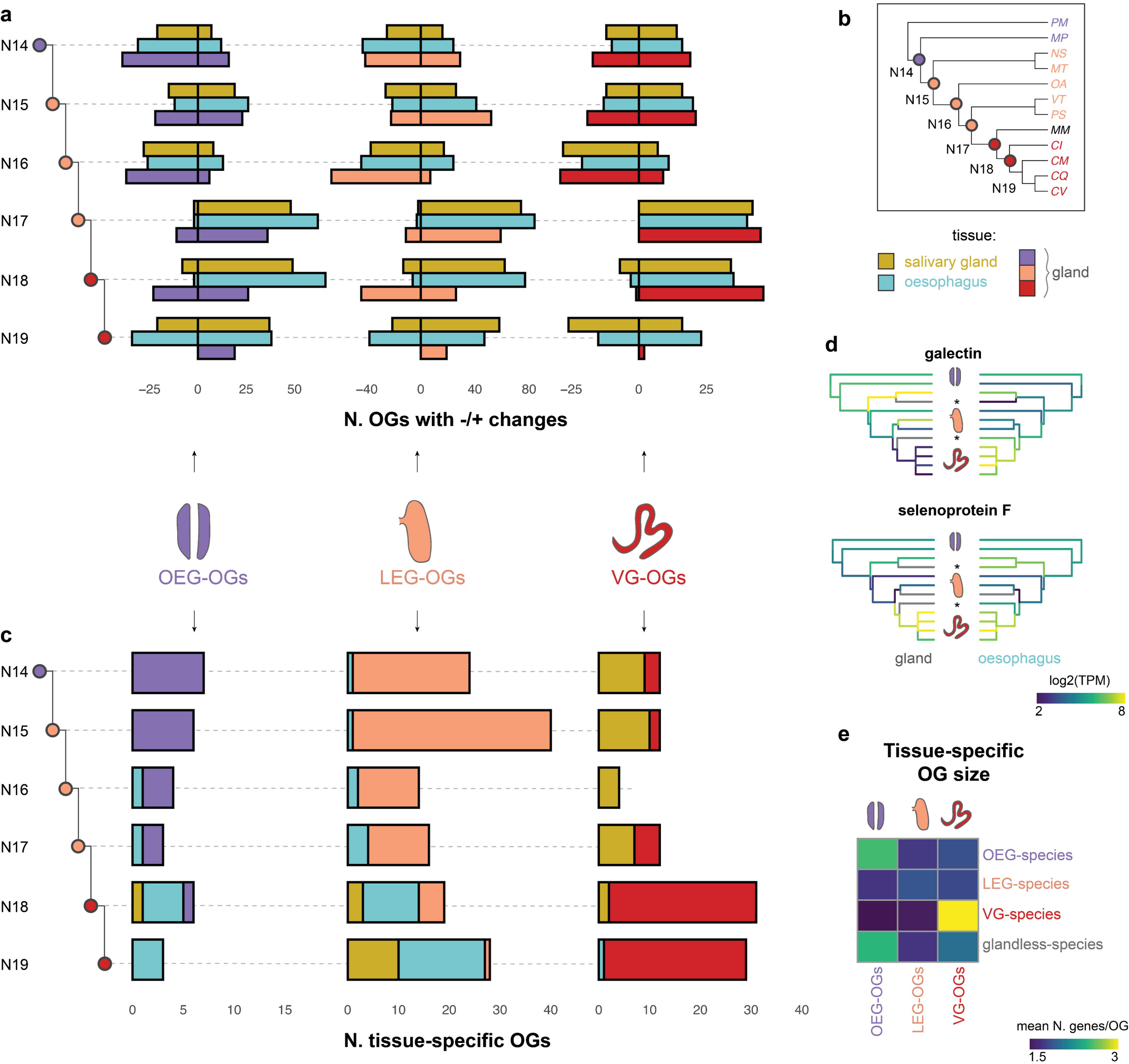
75. Fedosov, A. E. *et al.* Phylogenomics of Neogastropoda: The backbone hidden in the bush. *Systematic Biology* syae010 (2024) doi:10.1093/sysbio/syae010.

76. Page, L. R. Metamorphic remodeling of a planktotrophic larva to produce the predatory feeding system of a cone snail (Mollusca, Neogastropoda). *The Biological Bulletin* **221**, 176–188 (2011).


salivary glands
foot
gland
oesesophagus
columellar muscle
Polinices mammilla (PM)
Monoplex pilearis (MP)
oesophageal gland (OEG)
Nassa serta (NS)
Murex tenuirostrum (MT)*
Oliva amethystina (OA)
Vasum turbinellus (VT)
Phos senticosus (PS)
gland of Leiblein (LEG)
*dorsal glandular folds
Mitra mitra (MM)
Conus imperialis (CI)
C. marmoreus (CM)
C. quercinus (CQ)
C. virgo (CV)
venom gland (VG)
muscular bulb


a
tissue-specific gene sets
N. genes
1000
2000
3000
4000
PM
MP
NS
MT
OA
VT
PS
MM
CI
CM
CQ
CV
venom gland
muscular bulb
gland of leiblein
dorsal glandular folds
oesophageal gland
oesophagus
salivary gland
columellar muscle
foot
b
enriched GO terms
sodium ion transport
removal of superoxide radicals
regulation of peptidase activity
fluid transport
response to inorganic substance
protease binding
superoxide dismutase activity
arylsulfatase activity
iron iron binding
S-methyltransferase activity
extracellular region
apical part of cell
Arp2/3 protein complex
AP-5 adaptor complex
methionine biosynthetic process
proteolysis
iron iron transport
Arp2/3 complex-mediated actin nucleation
sodium ion transport
endopeptidase activity
proteoglycan binding
ferroxidase activity
aminopeptidase activity
S-methyltransferase activity
lysosome
clathrin-coated pit
Ragulator complex
WASH complex
endosome
extracellular space
Golgi organisation
protein folding
signal peptide processing
response to unfolded protein
peptidyl-proline hydroxylation
response to ER stress
toxin activity
procollagen-proline 4-dioxygenase activity
L-acorbic acid binding
disulfide oxidoreductase activity
RNA-directed DNA polymerase activity
host cell postsynaptic membrane
ER lumen
extracellular region
signal peptidase complex
postsynaptic membrane
ER
biological process
molecular function
cellular compartment
c
mean expression (TPM)
1000
2000
3000
secretome
d
protein modification process
intracellular protein transport
signaling
proteolysis
cell comunication
hydrolase activity
enzyme regulator activity
peptidase activity
protein binding
ATP-dependent activity
membrane
cytoplasm
ER membrane
bounding membrane of organelle
lysosome
vesicle
proteolysis
protein catabolic process
cellular response to stimulus
protein localisation to organelle
cellular amide metabolic process
hydrolase activity
transferase activity
metail ion binding
endopeptidase activity
carbohydrate binding
isomerase activity
plasma membrane
lysosome
intracellular membrane-bounded organelle
cytoplasm
ER membrane
modulation of process of another organism
protein folding
response to ER stress
protein modification process
regulation of signal transduction
toxin activity
protein binding
acetylcholine receptor activity
oxidoreductase activity
ion channel inhiobitor activity
membrane
extracellular region
postsynaptic membrane
ER membrane
ER lumen


a
tissue
species group
Pearson correlation
1
0.5
0
tissue:
oesophageal gland
salivary glands
dorsal glandular folds
gland of Leiblein
oesophagus
muscular bulb
venom gland
foot
columellar muscle
b
N. shared tissue-specific OGs
100
50
0
200
150
100
50
0
250
200
150
100
50
0
OEG-OGs
LEG-OGs
VG-OGs
OEG-species
LEG-species
glandless-species
VG-species


a
N14
N15
N16
N17
N18
N19
–25
0
25
50
–40
0
40
80
–25
0
25
N. OGs with -/+ changes
OEG-OGs
LEG-OGs
VG-OGs
b
PM
MP
NS
MT
OA
VT
PS
MM
CI
CM
CQ
CV
N14
N15
N16
N17
N18
N19
tissue:
salivary gland
oesophagus
gland
c
N14
N15
N16
N17
N18
N19
0
5
10
15
0
10
20
30
40
0
10
20
30
40
N. tissue-specific OGs
d
galectin
selenoprotein F
gland
oesophagus
log2(TPM)
2
8
e
Tissue-specific
OG size
OEG-species
LEG-species
VG-species
glandless-species
OEG-OGs
LEG-OGs
VG-OGs
mean N. genes/OG
1.5
3