REGULAR ARTICLE (to be submitted)

Comparative transcriptomics of the venoms of continental and insular radiations of West African cone snails

Samuel Abalde[1] | Manuel J. Tenorio[2] | Carlos M. L. Afonso[3] | Rafael Zardoya[1]

[1]Museo Nacional de Ciencias Naturales (MNCN-CSIC), José Gutiérrez Abascal 2, 28006, Madrid, Spain;
[2]Departamento CMIM y Q. Inorgánica-INBIO, Facultad de Ciencias, Universidad de Cadiz; 11510 Puerto Real, Cádiz, Spain;
[3]Centre of Marine Sciences (CCMAR), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

Correspondence

Rafael Zardoya, Museo Nacional de Ciencias Naturales-CSIC, Madrid, Spain
Email: rafaz@mncn.csic.es

Abstract

The transcriptomes of the venom glands of 12 closely related species of vermivorous cones endemic to West Africa from genera *Africonus* and *Lautoconus* were sequenced. These cones belong, respectively, to insular and continental radiations, for which robust phylogenies were available, allowing comparative evolutionary studies. The total number of conotoxin precursors, hormones and associated venom proteins per species varied between 95 and 210, and larger repertoires could indicate broader diets. We were able to perform parsimony ancestral reconstructions and find shared peptides at the individual, species and genus levels, as well as instances of convergent evolution. Individuals of the same species shared half to one third of the total conotoxin precursors. Due to the high variability of these secreted peptides, the number of common sequences was drastically reduced in the pairwise comparisons between closely related species and virtually almost no sequence was shared at the genus level. The two genera showed distinct catalogues of conotoxins precursors in terms of type of superfamilies, abundance of members per superfamily, and relative expression levels. Yet, a common set of six superfamilies (T, O1, O2, M, Cerm_03, and conkunitzin) was found to be expanded in all studied cone species. We detected significant overexpression of B1 superfamily in *Africonu*s species with respect to *Lautoconus* species, and of A superfamily in the piscivorous *Chelyconus ermineus* with respect to the vermivorous species.

KEYWORDS
conotoxin precursors, transcriptomes, *Africonus*, *Lautoconus*

1 | INTRODUCTION

Cone snails (Gastropoda: Conidae) are key predators in marine ecosystems that actively hunt on worms, snails, and fish (Kohn, 1959). Cones present a sophisticated venom system: a radular sac produces hollow radular teeth, which are loaded with venom generated in a convoluted venom duct (Tucker and Tenorio, 2009). The venom of cone snails is a cocktail constituted by hundreds of peptides named conotoxins, together with hormones and other proteins that participate in the synthesis or enhance the activity of the venom (Olivera, 2006; Barghi et al., 2015; Safavi-Hemami et al., 2015). Once secreted and inoculated into the prey, conotoxins can interact with different targets such as ionic channels and neurotransmitter receptors, triggering different physiological responses: from sedation to tetanic paralysis by muscle hyperactivity (Olivera et al., 1990; Lopez-

Vera et al., 2007; Robinson and Norton, 2014; Olivera et al., 2015; Ahorukomeye et al., 2019). Conotoxin precursors typically present a three domain structure, consisting of signal, pro-peptide, and mature (which constitutes, after cleavage of the other domains, the functional toxin) regions (Kaas et al., 2010). Sometimes a post-peptide region is also found. The signal region is highly conserved, and it has been used to classify these peptides into different toxin "superfamilies" (Robinson and Norton, 2014).

The composition of the venom secreted by cone snails is highly variable among species, specimens, and even within the same individual depending on its physiological status (Prator et al., 2014; Chang and Duda, 2016; Peng et al., 2016; Li et al., 2017; Abalde et al., 2018). This striking variability has been proposed to be generated through different mechanisms, including gene duplication (Duda and Palumbi, 1999; Espiritu et al., 2001), accelerated substitution rates (Conticello et al., 2001), recombination (Espiritu et al., 2001), differential expression (Duda and Palumbi, 2004), and/or post-translational modifications (Bergeron et al., 2013; Dutertre et al., 2013).

Thus far, cone snail venomics has been driven preferentially by the pharmacological potential of conotoxins, and the different studies were mostly limited to the purification of mature peptides and the identification of their function, thus lacking the wider phylogenetic perspective that is already being applied in the study of other venomous animals (Binford, 2001; Gibbs et al., 2013; Lomonte et al., 2014). Comparing venom cocktails in different cone species within a phylogenetic framework should provide insights on venom evolution, including how the rich diversity of conotoxins was generated and is maintained (Chang and Duda, 2012; Dutertre et al., 2014), to what extent the distinct repertoires are adapted to different diet specializations (Remigio and Duda, 2008; Chang and Duda, 2016; Phuong et al., 2016), which are the functional constraints and levels of convergence imposed by this coevolutionary arms race system (Conticello et al., 2001; Abalde et al., 2018), and which is the ultimate influence (if any) of conotoxin diversity in the extraordinary rates of species diversification of the group (Phuong et al., 2019).

Several recent studies have started exploring the evolutionary processes underlying the adaptive nature of venom composition in cones (Aman et al., 2014; Phuong et al., 2016; Abalde et al., 2018; Jin et al., 2019). As in other venomous animals (Pahari et al., 2007; Pekar et al., 2018), dietary breadth has been proposed to be a main factor triggering venom evolution in cones (Remigio and Duda, 2008; Elliger et al., 2011; Phuong et al., 2016). Since cone snail hunting performance relies on the specificity of their venom, shifts in diet can trigger changes in venom composition (Duda, 2008; Duda et al., 2009; Chang and Duda, 2016) and in general, species with more generalized diets tend to have more complex venoms (Phuong et al., 2016). Moreover, instances of functional convergence have been shown in the cocktails of Atlantic versus Indo-Pacific piscivorous cones (Abalde et al., 2018). In addition, another level of complexity comes from the capacity of cone snails to modulate the composition of their venom depending on its final use, whether to subdue preys or defend themselves against predators (Dutertre et al., 2014; Prashanth et al., 2016; Prashanth et al., 2017; Jin et al., 2019). Despite the extraordinary variability of the venom cocktails, and thus the great potential for ecological adaptation and species diversification, a recent study found no significant correlation between conotoxin gene diversity and speciation rates (Phuong et al., 2019), suggesting that other traits hampering gene flow may have been more critical in promoting the astonishing species diversity of cones (Cunha et al., 2005).

All the above-mentioned studies explored general venom evolutionary trends at the family (Conidae) level, comparing distantly related lineages. Here, we propose to analyze venom evolution within two radiations of closely related cone species inhabiting West Africa. This region is a hotspot of cone

diversity, including approximately 10% of all described species thus far (Tucker and Tenorio, 2013). This species diversity was generated through independent radiation events, leading to high rates of endemicity (Pin and Leung Tack, 1995; Cunha et al., 2005; Duda and Rolán, 2005). In particular, we focused on two species-rich lineages; one comprising cones endemic to the Cabo Verde archipelago and the other including cones endemic to Senegal (plus one closely related species inhabiting Canary Islands). Recently, robust phylogenies based on mitogenomes were reconstructed for both clades (ascribed to the genera *Africonus* and *Lautoconus*, respectively), providing the necessary framework for evolutionary studies (Abalde et al., 2017a; Abalde et al., 2017b). The ancestor of the genus *Africonus* arrived at the archipelago of Cabo Verde about 23 mya and diversified about nine mya into four main clades and at least 40 endemic species (Abalde et al., 2017a). The lineage of *Lautoconus* endemic to Senegal and Canary Islands diversified about six mya into three main clades and at least 15 endemic species (Abalde et al., 2017b). All of the cones in both clades are vermivorous. However, while cones endemic to Cabo Verde show no apparent differences in radular tooth morphology, the three clades of *Lautoconus* have each distinct radular teeth (Abalde et al., 2017b). All cones have non-planktotrophic larvae with restricted dispersal capacities. Therefore, it has been proposed that diversification of West African cones was in allopatry and mainly triggered by eustatic sea level changes during the Miocene-Pliocene (Cunha et al., 2005; Abalde et al., 2017a). Since *Africonus* species are normally restricted to single islands, the difference in number of species between the archipelago and the continent would be explained in terms of more chances to restrict gene flow in the former. However, the genus *Lautoconus* has only one species in Canary Islands, *Lautoconus guanche*, contradicting the pattern found in Cabo Verde. In this case, differences in the mean temperature of the water (in the limits of tolerance for cones in Canary Islands) and the proximity of Canary Islands to the continent (Fuerteventura Island is 120 km away from the coast of Morocco) could explain the lack of diversification (Cunha et al., 2014). No study, to our knowledge, has analysed the composition of the venoms of the vermivorous cones endemic to Cabo Verde, Canary Islands, and Senegal, nor their potential contribution to the observed enhanced rates of speciation in these areas.

Here, we sequenced the transcriptomes of the venom glands from ten individuals representing nine species of *Africonus* and four specimens of *Lautoconus* representing two species from Senegal and one from the Canary Islands. By sequencing these transcriptomes, we aimed to: 1) describe the venom composition for all these species in terms of presence and diversity of the conotoxin precursor superfamilies as well as relative abundance of the transcripts as proxy of expression levels; 2) assess the levels of divergence in venom composition at different hierarchical levels (within species, between species from the same clade within *Africonus*, between species from different clades within a genus, and between genera) and discern between shared derived peptides and cases of functional convergence; 3) determine whether there could be instances of differential expression between the two genera as footprint of adaptation; 4) compare venom compositions of these vermivorous species to that of the Atlantic piscivorous *Chelyconus ermineus* (Abalde et al., 2018) to further understand the connections between venom evolution and diet specialization; and 5) to determine whether there is any influence of venom variability in enhancing rates of diversification.

2 | Materials and Methods

2.1. Taxon sampling

A total of 14 specimens of cone snails were included in this study, four belonging to *Lautoconus* and ten to *Africonus*. Taxon selection was aimed at maximizing lineage representation and based on reconstructed phylogenies of the two genera (Abalde et al.,

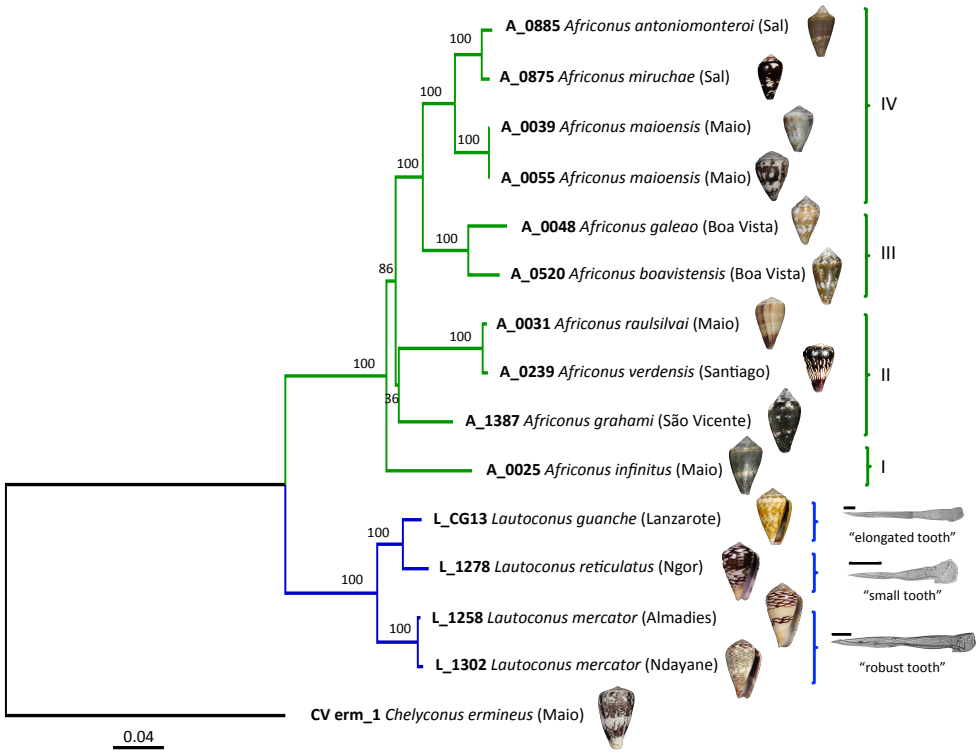

Figure 1. Phylogenetic relationships among the species included in this study (*Africonus* in green; *Lautoconus* in blue), including the sampling location. The four clades of the genus *Africonus* are identified in roman numbers (Abalde et al., 2017a) whereas the three clades of Lautoconus are identified by their typical radular tooth (Abalde et al., 2017b).

2017a; Abalde et al., 2017b). Thus, the three main clades in the genus *Lautoconus* and the four main clades recovered in *Africonus* were represented. For the genus *Lautoconus*, we studied one specimen of *Lautoconus guanche* (L_CG13) and *Lautoconus reticulatus* (L_1278) and two of *Lautoconus mercator* (L_1258 and L_1302; representing two different shell phenotype formerly classified as distinct species). In the case of *Africonus*, we included one specimen of *Africonus infinitus* (A_0025), *Africonus raulsilvai* (A_0031), *Africonus galeao* (A_0048), *Africonus verdensis* (A_0239), *Africonus boavistiensis* (A_0520), *Africonus miruchae* (A_0875), *Africonus antoniomonteroi* (A_0885) and *Africonus grahami* (A_1387), and two of *Africonus maioensis* (A_0055 and A_0039; representing two different shell phenotype formerly classified as distinct species). Sampling localities and voucher numbers for shells are shown in Table 1. Phylogenetic relationships including sequence divergences (branch lengths) are depicted in Fig. 1. All the specimens were adults and were dissected in a resting stage to remove the venom duct, which was preserved in RNALater (Invitrogen, Life technologies) at 4ºC during the sampling and -20ºC for the long term.

2.2. RNA extraction and sequencing

RNA extraction and sequencing were essentially performed as in (Abalde et al., 2018). First, each venom duct was incubated in a 2 ml eppendorf with 500 µl of TRIzol LS Reagent (Invitrogen, Life Technologies) and grinded with ceramic beads in a Praecellys Evolution tissue homogeneizer. Then, the solution was mixed with 100 µl of chloroform and centrifuged at 12000xg for 15 min at 4ºC. The supernatant was collected, mixed with 250 µL of isopropanol and left for precipitation overnight at -80ºC. Total RNA

Table 1. Specimens of *Africonus* and *Lautoconus* here analysed and main statistics of Illumina sequencing and assembly.

ID	Species	Country	Locality/Island	Voucher MNCN	SRA_accesion	Sequencing_date	Number_reads	%_clean_reads	Number_contigs	Number_blast_hits	Number_proteins	Number_conotoxins
A_0025	Africonus infinitus	Cabo Verde	Ponta do Pau Seco, Maio	15.05/78650	to be provided	13/3/14	35854397	98.52	76339	1095	206	175
A_0031	Africonus raulsilvai	Cabo Verde	Praia da Soca, Maio	15.05/78656	to be provided	28/10/13	56718528	100	99699	1249	223	189
A_0039	Africonus maioensis	Cabo Verde	Praia Santana, Maio	15.05/78664	to be provided	28/10/13	52523501	100	77336	783	169	140
A_0048	Africonus galeao	Cabo Verde	Navío Quebrado, Maio	15.05/78673	to be provided	21/12/16	28109709	100	50811	803	177	154
A_0055	Africonus maioensis	Cabo Verde	Navío Quebrado, Maio	15.05/78680	to be provided	28/10/13	44748977	100	102227	850	173	143
A_0239	Africonus verdensis	Cabo Verde	Tarrafal, Santiago	15.05/78864	to be provided	28/10/13	40237424	100	77906	1266	239	205
A_0520	Africonus boavistensis	Cabo Verde	Ervatao, Boa Vista	15.05/80413	to be provided	21/12/16	26715260	100	39935	797	199	175
A_0875	Africonus miruchae	Cabo Verde	Terrinha Fina, Sal	15.05/79784	to be provided	21/12/16	24097307	100	62006	615	131	108
A_0885	Africonus antoniomonteroi	Cabo Verde	Pedra Lume, Sal	15.05/79794	to be provided	21/12/16	26026957	100	84001	731	157	122
A_1387	Africonus grahami	Cabo Verde	Calhau, São Vicente	15.05/78549	to be provided	21/12/16	22718525	100	51601	850	180	159
L_1258	Lautoconus mercator	Senegal	Almadies	15.05/78419	to be provided	8/3/16	28883175	100	69501	631	143	126
L_1278	Lautoconus reticulatus	Senegal	Ngor	15.05/78439	to be provided	21/12/16	24263358	100	52496	479	109	89
L_1302	Lautoconus mercator	Senegal	Ndayane	15.05/78463	to be provided	8/3/16	28392465	100	78580	783	176	157
L_CG13	Lautoconus guanche	Spain	Playa del Cable, Lanzarote	–	to be provided	8/3/16	29973740	100	87516	815	175	150

was purified using the Direct-Zol RNA miniprep kit (Zymo Research, Irvine) following manufacturer's instructions.

Dual-indexed cDNA libraries were constructed for each sample using the TruSeq RNA library Prep kit v2 (Illumina, San Diego) at Sistemas Genómicos (Valencia, Spain) following manufacturer's instructions. The quality of the libraries was checked with the TapeStation 4200, High Sensitivity Assay, and the quantity determined by real-time PCR in LightCycler 480 (Roche). The pool of libraries (including samples of other cone snail species) was split into several runs of paired-end sequencing (2x100bp) in an Illumina HiSeq2500 (two flowcells per run) following the standard procedures at Sistemas Genómicos (Valencia, Spain).

2.3. RNA assembly and conotoxin identification

The raw reads corresponding to the different individuals were sorted using the corresponding library indices, which were removed using Cutadapt v.1.3 (Martin, 2011). Raw read quality was checked using FastQC v.0.10.1 (Andrews, 2010), and the assembly was performed using Trinity v.2.6.6 (Grabherr et al., 2011) with default settings (minimum contig length = 200bp, sequence identity threshold = 0.95) and the --trimmomatic option active with default parameters. The raw reads of all transcriptomes are available at the SRA database (Table 1).

All conotoxin precursors, hormones, and associated proteins publicly available in different databases (GenBank release 222 (Benson et al., 2013), Uniprot release 2017_09 (UniProt, 2015) and ConoServer release 30/10/2017 (Kaas et al., 2012) were downloaded in October 30th, 2017 and concatenated into a single fasta file. Duplicated sequences were removed, and the resulting file was formatted as a Blast database using Blast+ (Camacho et al., 2009) to create the custom reference database.

All putative conotoxin precursor, hormones, and associated protein sequences in the assembled transcriptomes were identified using BLASTX over the reference database (e-value: 1e-5). The selected sequences were manually inspected and compared against the most similar sequences in the reference database, and translated into the appropriate open reading frame (ORF). All the sequences considered as false positives or assembly artifacts (showing internal stop codons and chimeras), those that were duplicated or highly truncated (missing >55% of the estimated length of the reference protein), and those showing low coverage values were discarded. We implemented an extra curation step consisting on TBLASTX searches over the nr database in GenBank to discard wrong ORF assignations.

The remaining sequences constituted our working list of conotoxin precursors, hormones, and associated proteins (see Suppl. Mat. File 1 and Suppl. Mat. Table 1). The three domain structure and cysteine frameworks of conotoxin precursor alignments were inferred using Conoprec (Kaas et al., 2010). The different proteins were assigned to a given superfamily by comparison with best-hit results using BLASTP searches against GenBank, and in the case of the conotoxin precursors, taking into consideration the percentage of identity in the signal region using a general threshold

Table 2. New signal sequences of conotoxin precursors here described, including the cysteine framework and main Blast resul

				Best-hit known superfamily		
Unassigned superfamily	Signal	Cysteine framework	Also found in:	Superfamily	% coverage	% Identity
1	MNCLQPLLVLLLISTITA	XIII	betulinus	M	65	28.81
2	MSGTMIVLLAVLLLVDLSTS	VI/VII	*betulinus*	O3	82	37.29
3	MPGSRVALLAFLLLLSLVTNLQG	VI/VII	*betulinus, leopardus*	O3	25	62.5
4	MTMDMKMTFSGFVLVVLVTTVVG	VIII	*betulinus, praecellens, andremenezi*	—	—	—
5	MMTLRHVLLFTLLLLPLATIR	XXII	*betulinus*	A	68	31.37
6	MLSVFTVVWVLTTAMMMTDVTFQ	C7C6C8C3C1C5C23C8C18C1C10C	*praecellens, andremenezi*	I2	45	24.59
7	MWSGKDQAAFLALVLMVVGAHSTTA	IX	*praecellens, andremenezi*	—	—	—

of 70% (Robinson and Norton, 2014). We further checked the correct identification of all conotoxin precursor superfamilies by aligning all the signal regions and building with neighbor-joining a guide tree (Supp. Mat. Fig. 1) based on uncorrected *p* distances on ClustalW (Thompson et al., 1994). Within each superfamily, sequences were assigned to different groups of paralogy based on the sequence divergence at the pro-peptide region, different cysteine frameworks in the mature peptide, and the presence of clades in the reconstructed guide tree (Supp. Mat. Fig. 1). Those sequences that did not match any previously reported conotoxin precursor superfamily were considered unassigned superfamilies and described here. The nucleotide sequences of all venom proteins here identified are available at Genbank under accession numbers XXXX-XXXX.

2.*4. Comparative analyses of venom composition*

The final list of conotoxin precursors for each species was pairwise compared and common sequences were detected using the ClustalW algorithm as implemented in Geneious® 8.0.3. All sequences that were common to two or more species were mapped onto the phylogeny using parsimony ancestral character reconstruction as implemented in Mesquite v 3.6 (Maddison and Maddison, 2018).

In order to infer venom composition similarities between species and genera, we run in R (R Core Team, 2013), six principal component analyses (PCAs) comparing at the paralog group and superfamily levels: 1) the presence or absence of conotoxin precursor superfamilies; 2) the relative abundance of each superfamily in terms of number of different sequences; and 3) the relative expression level of each superfamily measured as transcript per million (TPMs) estimates (see below).

2.5. Expression analyses

Relative expression levels for each individual were calculated by mapping the raw reads to the nucleotide sequence of each contoxin precursor using Bowtie 2 (Langmead and Salzberg, 2012), and the values transformed to TPM estimates using RSEM (Li and Dewey, 2011) as implemented in Trinity v.2.6.6 (Grabherr et al., 2011). In order to identify those conotoxin superfamilies that could be differentially expressed between *Lautoconus* and *Africonus*, we run the EBSeq software (Leng et al., 2013), that estimates the posterior probability of being differentially expressed (PPDE), using all the specimens of each genera as biological replicates. We considered as differentially expressed all those conotoxin precursor superfamilies with a PPDE > 0.95 and with a fold change above 32 (calculated as $\log_2$ RealFC $\geq$ 5). The same type of analysis was performed to identify those superfamilies significantly expressed in the comparison between vermivory, (using the 14 specimens of West Africa as replicates) and piscivory (using the three individuals of *C. ermineus* from (Abalde et al., 2018). In both comparisons, we run an ANOVA test in R (R Core Team, 2013) over those superfamilies identified as differentially expressed to take into consideration variance among replicates and further confirm the statistical significance of the results.

3 | RESULTS

3.1. Sequencing and assembly

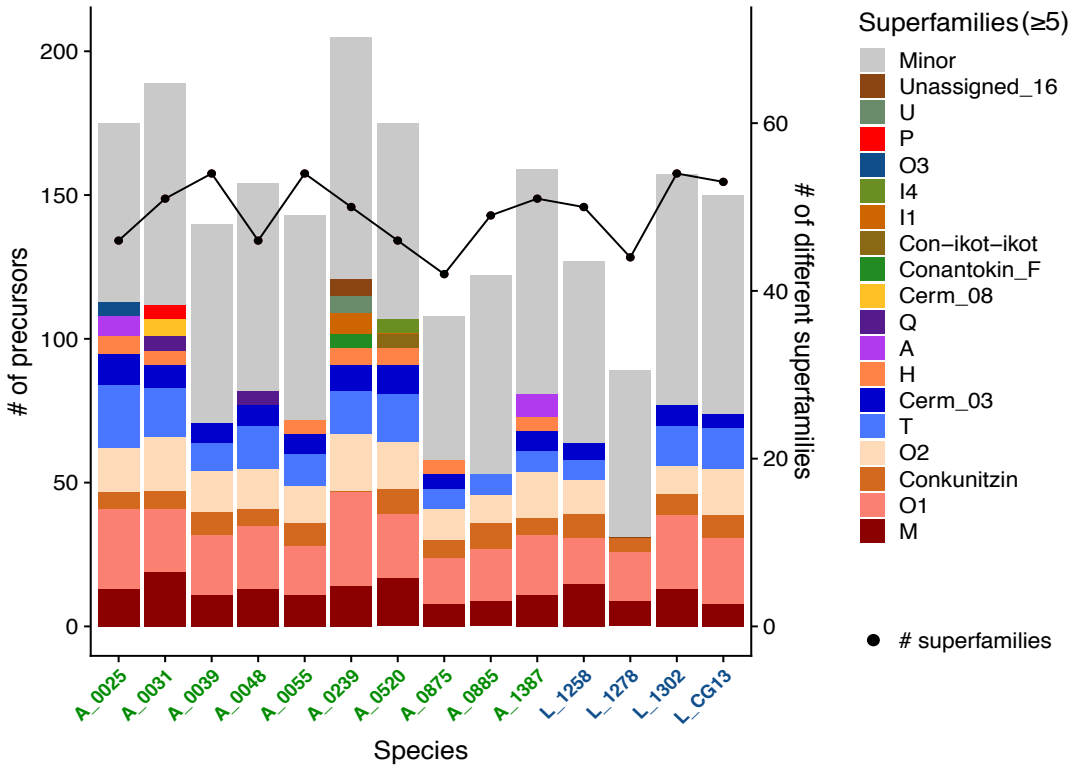

Figure 2. Venom composition of each of the 14 specimens studied. The bars represent the total number of conotoxin precursors and hormones. The proportion of those superfamilies with more than five members is shown in colours. The bar line represents the number of different superfamilies identified in the venom. The species codes in green belong to the genus *Africonus*, and those in blue to *Lautoconus*.

The RNAs of 14 samples corresponding to 12 species of the genera *Lautoconus* and *Africonus* were sequenced and their transcriptomes assembled. The main statistics associated to the sequencing and assembly procedures are summarized in Table 1. The number of raw reads sequenced per sample varied between 22.7 and 56.7 millions with a mean of 33.5 millions, and most reads were kept after cleaning (Table 1). The number of contigs generated after the assembly varied between 39,935 and 102,227 with a mean of 72,139. The BLASTX searches retrieved between 479 and 1269 putative conotoxin precursors, hormones and other venom proteins per species, and after all curation steps, we kept between 109 and 239 of them with a mean of 175,5 (Table 1). The number of conotoxin precursors varied between 89 and 205 with a mean of 149.4.

3.2. Venom cataloguing of West African cones

A total of 2,575 transcripts were identified in the venom gland transcriptomes of the 14 cone specimens: 2,070 were conotoxin precursors, 71 were hormones, and 434 were designed as other venom proteins. The species that presented the highest diversity of contoxin precursors were *A. verdensis* (205) and *A, raulsilvai* (189) whereas the least diversity was found in *L. reticulatus* (89; Table 1 and Fig. 2). All conotoxin precursors were classified into 60 distinct superfamilies and 142 groups of paralogy (hereafter "families") taking into consideration sequence divergences in the signal rand pro-peptide regions and the clades recovered in the reconstructed guide tree. Most of the conotoxin precursor superfamilies were recovered as monophyletic in this guide tree, with the exception of the M superfamily (two lineages), Cerm_03 (three), Q (three), and N (two; Supp. Mat. Fig. 1). The superfamilies that presented more diversity of members

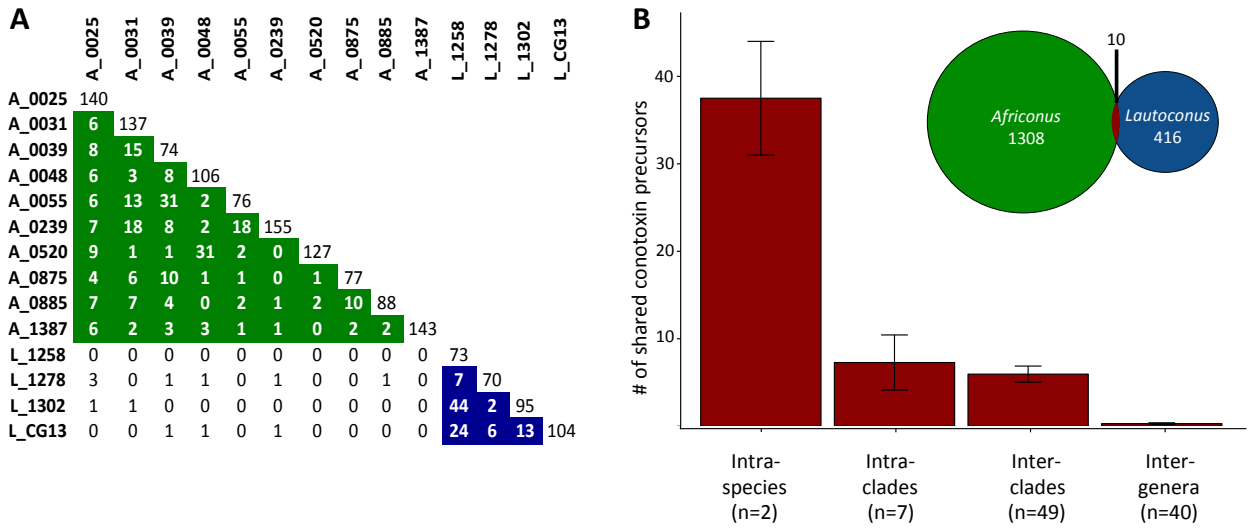

Figure 3. Differences in venom composition along the phylogeny. A) pairwise comparisons of the number of shared identical conotoxin precursors. The number of conotoxins exclusive for an individual is represented in the diagonal. B) Average number of identical conotoxins shared by individuals, intra- and interclade species, and genera. The number of comparisons is shown below each bar. A Venn diagram representing the number of unique and common conotoxin precursors for both genera is shown as inset.

were conkunitzin (13), O2 (11), and O1 (9). A total of 70 peptides could not be assigned to any known superfamily, and were grouped into seven new unassigned superfamilies (their signal sequences, cysteine frameworks and other features are reported in Table 2). Several precursors previously reported as valid conotoxin precursor superfamilies such as R, W, Z (Lavergne et al., 2013) and Cerm_17 (Abalde et al., 2018), among others, were found to be fragments of other proteins once the right ORFs were identified using TBLASTX.

The species that presented the highest diversity of conotoxin precursor superfamilies were *A. maioensis* (A_0055, 54; A_0039, 54) and *L. mercator* (54) whereas the least complex venom was found in *A. miruchae* (42). The superfamilies with highest number of members were O1, O2, T, M, Conkunitzin, and Cerm_03 (Fig. 2). They were found in all the species with the following exceptions: *A. verdensis* lacked the Conkunitzin; *A. antoniomonteroi* had no Cerm_03; and *L. reticulatus* lacked the O2, T and Cerm_03 (Fig. 2). All other conotoxin precursor superfamilies in the species of the genus *Lautoconus* had less than five members. However, the species in the genus *Africonus* (except *A. antoniomonteroi* and the specimen A_0039 of *A. maioensis*) showed another 12 superfamilies with five or more distinct conotoxin precursors. These expanded superfamilies were differently distributed: four were exclusive to *A. verdensis* (Conantokin F, I1, U, and Unassigned_16); two to *A. raulsilvai* (Cerm_08 and P); two to *A. boavistiensis* (Con-ikot-ikot and I4); and one to *A. infinitus* (O3); the remaining (H, A, and Q) were found in more than one species (Fig. 2). The 14 species presented up to 71 hormone sequences that were classified into six superfamilies (none of them present five or more members): Conopressin, Conorfamide, Insulins 1-5, Prohormone-4, Thyrostimulin hormone alpha, and Thyrostimulin hormone beta 5 (Suppl. Mat. File 1 and Suppl. Mat. Table 1).

Finally, we identified up to 434 transcripts that were assigned to 28 protein families of various function. Among them, the Protein Disulfide Isomerase (81 sequences), Ferritin (50) and Conodipine (41) were the most diverse. Interestingly, we found in the venom, several members of a conserved superfamily of cysteine-rich secretory, antigen 5, and pathogenesis-related 1 proteins (CAP) that are often secreted and have a

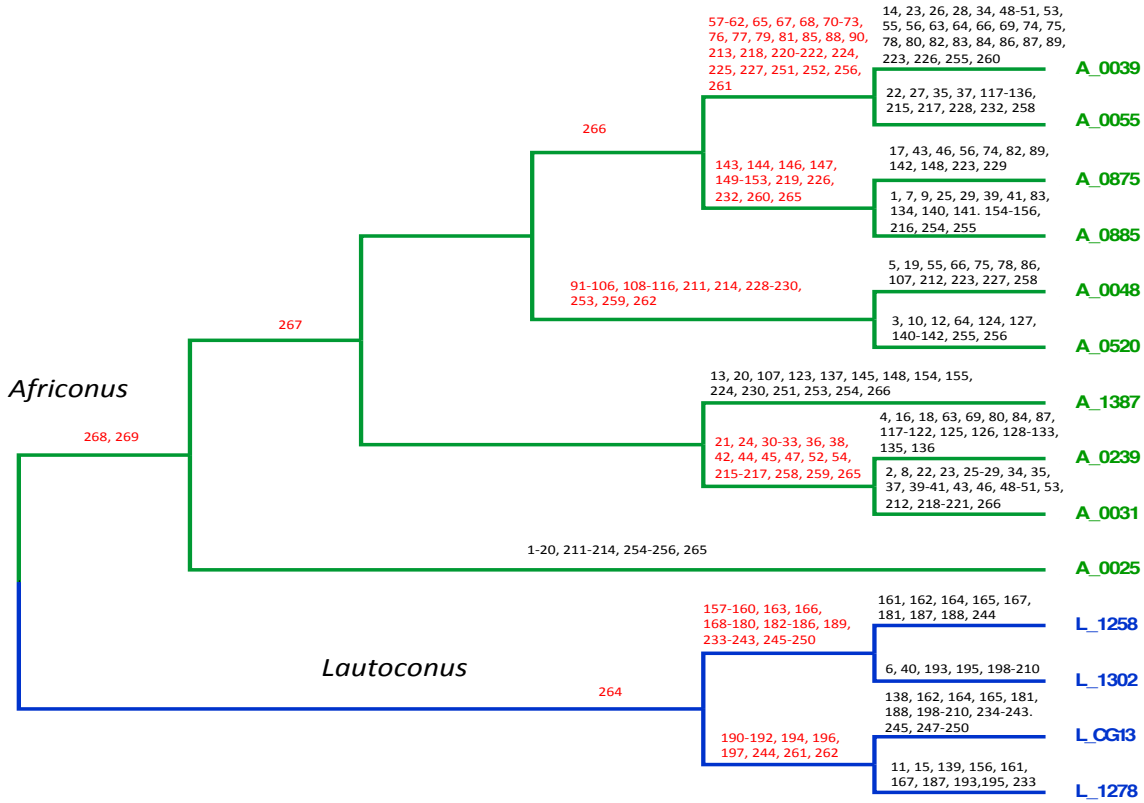

Figure 4. Reconstruction under parsimony of the conotoxin precursors present in the different common ancestors in the phylogeny of cones from West Africa. The numbers in the nodes correspond to conotoxin precursors shared by at least two taxa and shown in Suppl. Mat. Table 2.

protease activity with an extracellular endocrine or paracrine function in a wide range of organisms including venomous ones such as ants, wasps and snakes (Bateman et al., 2004). These proteins were previously reported in the molluscivorous *Cylinder textile* and the vermivorous *Conus marmoreus* and may be important for venom function ((Milne et al., 2003; Hansson et al., 2006); Suppl. Mat. Fig. 2).

3.3. Variations in venom composition according to phylogenetic divergence

The venom composition of the 14 specimens was pairwise compared at different hierarchical levels (Fig. 3). For two species (*A. maioensis* and *L. mercator*), we could compare venom composition between individuals. The two specimens of *A. maioensis* shared 31 conotoxin precursors, which represent 37-39% of the total sequences. The two specimens of *L. mercator* had 44 conotoxin precursors in common (46-60%; Fig. 3). The number of shared sequences between pairs of species from the same clade within *Africonus* varied between 1 to 31, with a mean of 8.7 (8% of the mean total sequences). Four pairwise comparisons at this level rendered common sequences more than average: *A. boavistensis* versus *A. galeao* (clade III; 31 shared sequences), *A. verdensis* versus *A. raulsilvai* (Clade II; 18), *A. maioensis* A_0039 versus *A. miruchae* (Clade IV; 10), and *A. antoniomonteroi* versus *A. miruchae* (Clade IV; 10). The number of shared sequences between pairs of species from different clades within the same genus varied between 0 and 18 with a mean of 4.61 (4.1% of the mean total sequences). Four pairwise comparisons at this level rendered common sequences more than average: *A. verdensis* versus *A. maioensis* A_0055 (18 shared sequences), the two specimens of *A. maioensis* versus *A. raulsilvai* (13 and 15), and the two specimens of *L. mercator* versus *L. guanche* (13-24; Fig. 3). The species *A. grahami* was the one sharing fewer sequences with other species (1-6; Fig. 3). The number

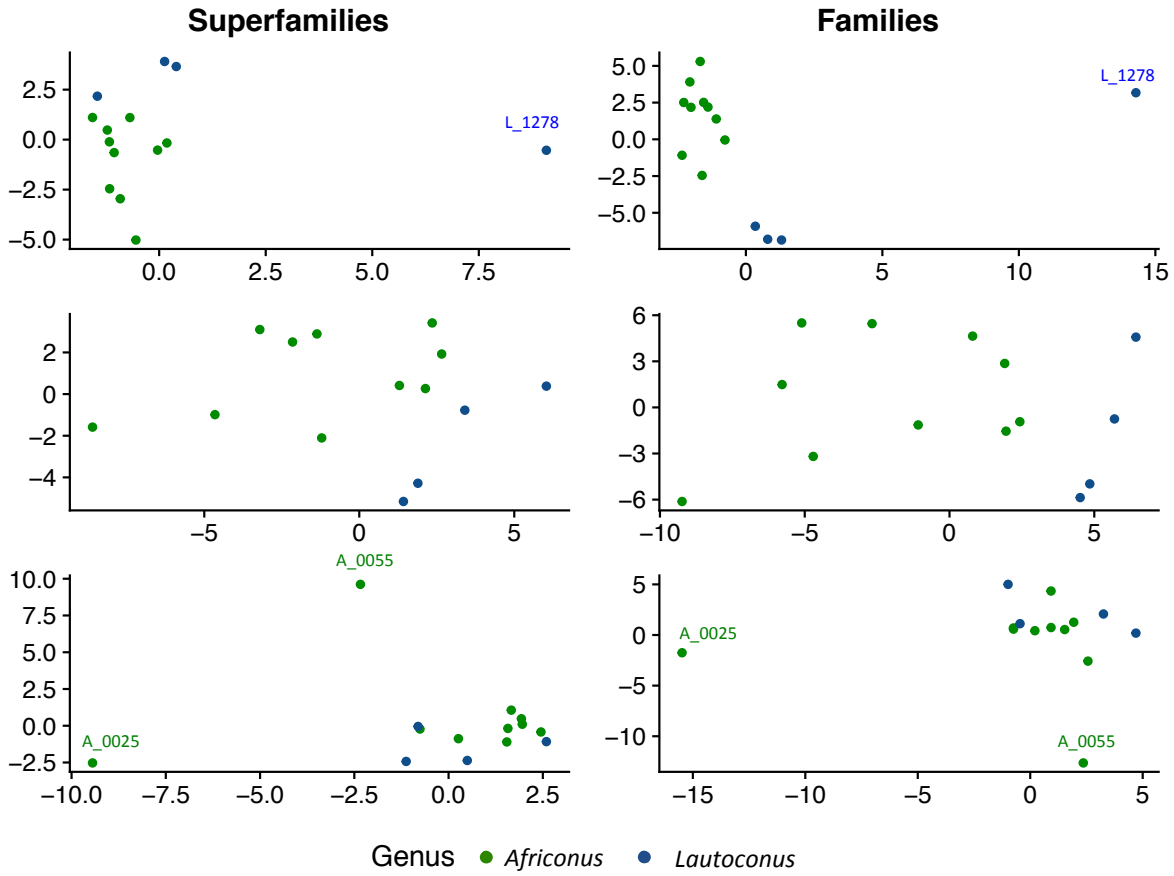

Figure 5. Principal Component Analysis comparing the venom composition among species and genera. The venom composition was defined as presence/ absence of conotoxin precursors (first row), number of members (second row), and relative expression levels (third row). The three comparisons were made at the superfamily and family (groups of paralogy) levels.

of shared sequences between genera was nine (0.7-2% of the mean total sequences). The total diversity of conotoxin precursors in *Africonus* triplicated that of *Lautoconus* (1367 versus 426; Fig. 3).

An ancestral reconstruction analysis under parsimony was performed to determine contoxin precursors in most recent common ancestors in the phylogeny and detect instances of convergence (Fig. 4). According to the reconstruction, the ancestor of the species belonging to clade IV of *Africonus* had the Cerm_10 superfamily. The ancestor of Clade III had members of B2, C, I1, J, M, O1, O2, P, T, U, V, Y, Conkunitzin, Con-ikot-ikot, Rimp_01, Rimp_03, Rmil_01, Tpra_06, Pamg_02, Cerm_02, Cerm_03 and Cerm_011 superfamilies (Fig. 4). The ancestor of *A. raulsilvai* and *A. verdensis* in Clade II had members of Conantokin F, B2, F, H, M, O1, O2, O3, T, Conkunitzin, Cerm_03, Cerm_10, and Unassigned_07 superfamilies (Fig. 4). The ancestor of clades II-IV had a member of the O2 superfamily. The ancestor of all *Africonus* species had a member of O2 and Pmag_02 superfamilies. The ancestor of *L. guanche* and *L. reticulatus* had members of A2, M, O1, O3, Conkunitzin, Pmag_02, Rimp_01, Rimp_04, and Rmil_02 superfamilies (Fig. 4). The ancestor of all *Lautoconus* species had a member of the O2 superfamily. Finally, *A. raulsilvai* (clade II; Maio) and *A. maioensis* A_0039 (clade IV; Maio) shared aeveral conotoxin precursors (members of B1, I3, M, O1, Con-ikot-ikot, Crerm_08, Cerm_011, and Unaasigned_04 superfamilies) despite distantly related in the phylogeny.

Pairwise comparisons were also performed taking superfamily or family classification into consideration and subjected to PCAs. Three different metrics were

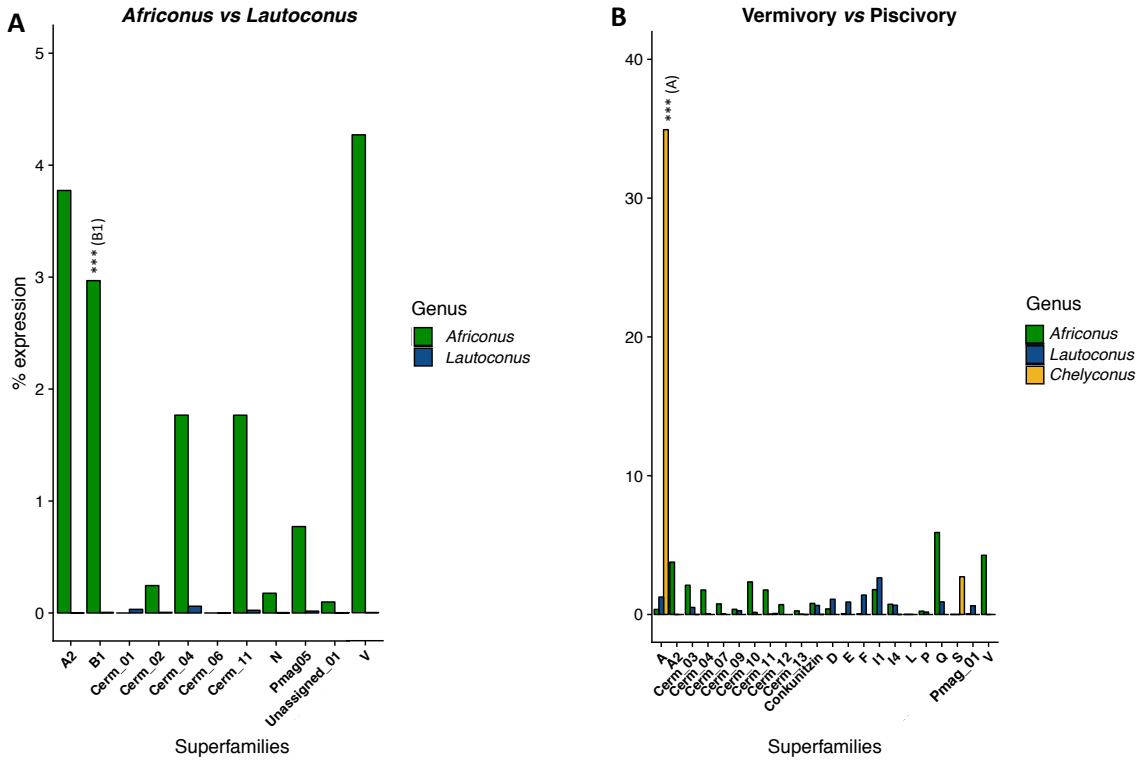

Figure 6. A) Differential Expression (measured in TPMs) of superfamilies between A) *Africonus* and *Lautoconus* and B) the 14 specimens here sequenced as biological replicates for vermivory and the three individuals of *Chelyconus ermineus* from (Abalde et al., 2018) representing piscivory, Bar plot depicting the superfamilies differentially expressed are shown. The genus *Africonus*, *Lautoconus*, and *Chelyconus* are depicted in green, blue, and yellow, respectively. The three asteriks represent a significant *p*-value = 0 following the ANOVA analyses.

analyzed at both levels: presence/absence, number of members, and the relative expression levels (calculated as TPMs) (Fig. 5). According to the PCAs, species from each genera cluster together when the presence/ absence of conotoxin superfamilies and families is tested, although *L. reticulatus* is clearly an outlier; Fig. 5). The PCA of the relative abundance (number of members) of each superfamily/ family revealed no overlapping between genera (Fig. 5). Finally, relative expression levels are similar in both genera, although *A. infinitus* and *A. maioensis* A_0055 are outliers (Fig. 5).

3.4. Differential expression of conotoxins and hormones in the two genera and in vermivores versus piscivores

We estimated whether any superfamily was differentially expressed between the two genera using the species of each genus as biological replicates and the sums of the expression of each superfamily as variables. We detected 11 superfamilies differentially expressed between genera (Fig. 6): A2, B1, Cerm_01, Cerm_02, Cerm_04, Cerm_06, Cerm_11, N, Pmag_05, Unassigned_01, and V. All these superfamilies were overexpressed in *Africonus*, except Cerm_01 and Cerm_06, which were in *Lautoconus*. After an ANOVA test, only the overexpression of B1 superfamily in *Africonu*s remained significant (p-value = 0).

The same tests were performed to detect differential expression between the 14 individuals of vermivorous species here studied and three individuals of the Atlantic piscivorous species *C. ermineus* (Fig. 6). Using the same threshold as above, we found 22 superfamilies differentially expressed. Among them, only the A and S superfamilies were overexpressed in *Chelyconus ermineus*. The ANOVA test only identified the expression levels of the A superfamily as significantly different between diets (*p*-value = 0).

4 | DISCUSSION

At present, high-throughput sequencing techniques are producing massive amounts of raw read sequence data, which are assembled and automatically annotated using a variety of bioinformatic pipelines. However, correct annotation of genes relies entirely on the quality of the reference database (Salzberg, 2019). In particular, in the case of conotoxins, and given their intrinsic variability, several authors have warned about the need of inspecting carefully automated annotation results, since assembly artifacts could over-estimate final conotoxin diversity (Phuong et al., 2016; Abalde et al., 2018) and, even more dangerous, they could propagate once incorporated into updated reference databases (Li et al., 2017; Abalde et al., 2018; Salzberg, 2019). Here, we carefully inspected all the ORFs initially rendered by the BLASTX searches through manual comparison to previously published sequences obtained from cone venoms or from venom gland transcriptomes. At the assembly level, after mapping the reads back to the transcripts, we found in many instances, regions of the assembled transcript (particularly at the tails) mapped only by very few reads (sometimes even only one), which could represent errors of the sequencing process and led to frame shifts that generated spurious variability (Phuong et al., 2016). At the annotation level, we implemented a TBLASTX step that found instances of wrong ORF assignation. For example, the ORF of the R superfamily (Lavergne et al., 2013) when translated in a different frame corresponds to the proteasome subunit alpha, and the conotoxin precursor "Bt-23" from (Peng et al., 2016) corresponds to a 60S ribosomal protein (other examples are shown in Suppl. Mat. File 1)..

Remarkably, we found in different species several sequences that were identical, except for the presence of one or several copies in tandem of a duplicated fragment within the pro-peptide region. It is the case, for example, of the B1 superfamily (Conantokin) sequences A_0239_306, A_0039_179 and A_0055_201 (these and other examples are found in Suppl. Mat. file 1). The mapping of the reads over the corresponding regions of the transcript confirmed high coverage and, in some cases, the reads covered the entire duplicated region, thus discarding an assembly artefact (not shown). The signal, mature and post regions of the precursors are encoded each by a single exon in the genome, whereas the pro-peptide region could be encoded by up to six different exons (Phuong and Mahardika, 2018). Hence, one potential explanation for the observed duplications could be the inclusion more than once of the same exon during the formation of the mRNA, although confirming this hypothesis would require experimental validation.

The number of raw reads sequenced per sample varied between 23-57 millions. Generated sequence data fit well within the range recommended by previous studies testing the optimal sequencing depth for *de novo* assembly, which concluded that 20-30 million reads would report most of the expressed genes in a given tissue while minimizing the number of artifacts (Francis et al., 2013). The number of raw reads generated per species did not correlate neither with the number of assembled contigs nor with the number of conotoxin precursors, hormone, and other venom proteins. For example, the assembly of the *L. mercator* (L_1302) and *A. maioensis* (A_0039) transcriptomes rendered similar number of contigs (77-78,000) but started from 28 and 52 million raw reads, respectively. Similarly, *A. miruchae* and *L. reticulatus* started from a similar sequencing depth (24 million reads) but presented a 17.5% difference in the number of conotoxin precursors (108 and 89, respectively).

The number of conotoxin precursors present on the venom varied from 89 in *L. reticulatus* to 205 in *A. verdensis*. These numbers are in good agreement with those reported for other species of cones (Hu et al., 2011; Barghi et al., 2015; Peng et al., 2016; Li et al., 2017; Abalde et al., 2018; Jin et al., 2019). It has ben proposed that larger sets of conotoxins are associated to broader diets (Elliger et al., 2011; Phuong et al., 2016;

Phuong and Mahardika, 2018). Hence, the richer diversity of conotoxins of *A. verdensis* could reflect a broader diet, which hitherto is known to be based on worms but largely unstudied. In this regard, ecological studies on *Miliariconus miliaris* showed that the individuals of this species inhabiting the remote Eastern Island presented a considerably broader diet of worms, which could have evolved through ecological release in the absence of congeners (Duda and Lee, 2009). This could also be the case for *A. verdensis*, which is the only species inhabiting the island of Santiago in Cabo Verde. However, this plausible explanation does not seem to fully apply to *A. raulsilvai* and *A. infinitus*, with 189 and 175 different conotoxin precursors, respectively. In these cases, while the observed conotoxin precursor diversity could still reflect a broader diet, it could not have evolved by ecological release, as both species are endemic to Maio, one of the islands of Cabo Verde where more cone species cohabit (Abalde et al., 2017a).

Conotoxins are well known for their accelerated rates of evolution, which in turn generate high sequence divergences even between individuals of the same species (Peng et al., 2016; Abalde et al., 2018; Jin et al., 2019), and are the basis of the reported general lack of common peptides between cone species, and the extended notion that virtually each species has produces a unique venom cocktail (e.g., (Gao et al., 2017). However, thus far, these conclusions were based mostly on comparisons between distantly related species (for an exception see (Li et al., 2017), whereas the present study brings the opportunity of comparing two clades (genera) of up to 12 closely related species sharing relatively recent common ancestors. Individuals of the same species (although representing distinct shell phenotypes previously described as species) showed only half (*L. mercator*) to one-third (*A. maioensis*) common conotoxin precursor sequences. These proportions are similar to those reported for intraspecific comparisons in *D. betulinus* (Peng et al., 2016), *Rhombiconus imperialis* (Jin et al., 2019) or *C. ermineus* (Abalde et al., 2018). The proportion of shared sequences decreased substantially for the pairwise comparisons between species (but within the range of 2-9% reported for sister species of the genus *Turriconus*; (Li et al., 2017), although there were still enough common sequences to infer which conotoxin precursors were likely present in most recent common ancestors along the phylogeny. In particular, O2 superfamily was found to be the only one already present in the ancestors of *Africonus* and *Lautoconus*. Therefore, our results support that a phylogenetic signal exists in conotoxins above the species level, but it is quickly eroded as lineages diverge and virtually almost no sequence is shared between closely-related genera comparisons (Phuong et al., 2016). On the other hand, there are some instances in which identical conotoxin precursor sequences are found even in species from distantly related genera (see Suppl. Mat. File 1), indicating that those sequences either are subjected to strong balancing selection or reflect cases of convergent evolution. The rather erratic distribution of these sequences in the phylogeny of cones favors the latter hypothesis. Moreover, the presence of common conotoxin precursors in *A. raulsilvai* and *A. maioensis* A_0039, which are distantly related in the phylogeny of cones of Cabo Verde (clades II and IV within *Africonus*; (Abalde et al., 2017a) but inhabit two close bays (Soca and Santana, respectively; Table 1) in the northwest of Maio Island, suggests functional convergence to similar diets (Remigio and Duda, 2008).

The analyses of the composition of the venoms of West African cones in terms of number and type of superfamilies showed that almost all species had a core set of six superfamilies, which are characterized by having five or more members: M, O1, O2, T, Conkunitzin, and Cerm_03. The wider presence of the former four superfamilies in any cone and always showing similar levels in diversity of members (e.g. (Terrat et al., 2012; Lavergne et al., 2013; Peng et al., 2016; Phuong et al., 2016; Li et al., 2017; Robinson

et al., 2017; Abalde et al., 2018) may suggest that the ancestor of living cones already had this core set and that having members of these superfamilies is essential for triggering the minimum physiological responses leading to the capture of a prey, regardless of whether it is a worm, a snail or a fish.

In addition, in West African cones, we found up to 13 different groups of paralogy or families of conkunitzins present in most cases both in *Africonus* and *Lautoconus*. The conkunitzins block voltage-gated potassium channels and were first described in *Pionoconus striatus* (Bayrhuber et al., 2005). They have been also identified in other piscivorous cones such as *C. ermineus* (Abalde et al., 2018) and *G. geographus* (Dutertre et al., 2014) as well as in the vermivore *Dendroconus betulinus* (Peng et al., 2016). Conkunitzins belong to a larger superfamily of kunitz-type fold peptides, which are ubiquitous serine protease inhibitors found in different animals (Ranasinghe and McManus, 2013). We also found a great diversity of paralog groups belonging to Cerm_03 superfamily. This superfamily was recently described in *C. ermineus* (Abalde et al., 2018), and has a typical precursor structure with the three domains and a mature peptide with a cysteine framework type XIV. It has been also found in several vermivore species including *D. betulinus*, *Turriconus praecellens*, and *Elisaconus litteratus* but without the variety of paralogs identified here and in *C. ermineus*. The function of the mature peptide remains to be determined.

While *Lautoconus* species showed an expanded number of members only for the above-mentioned six superfamilies, most *Africonus* species presented in addition the expansion of H superfamily. Not much is known about the function of the conotoxins (with a cysteine frame work VI/VII) of this superfamily, which was first identified in the vermivorous *C. marmoreus* (Dutertre et al., 2013) and constitutes a large proportion of conotoxin mRNA transcripts in the venom gland of the molluscivorous *Cylinder victoriae* (Robinson et al., 2014). Interestingly, *A. verdensis*, the only cone from Santiago Island, is the species with more number of conotoxin precursor superfamilies showing expansion of their members, and these expanded superfamilies do not coincide with those of *A. raulsilvai* from Maio Island, which is its sister species (together with *Africonus gonsaloi*), and the second species with highest diversity of expanded families. The larger set of expanded superfamilies in *Africonus* versus *Lautoconus* could be explained by the species radiation in the archipelago and the potential opportunities for local diet specializations, thus increasing the variability in the compositions of the venoms of these insular species (Abalde et al., 2017a). This would be in agreement with the lack of diversification of *L. guanche* in Canary Islands, which is reflected in a lack of diversity of expanded superfamilies.

PCA has been used in several other animal groups to summarize the information related to venom composition (Gibbs et al., 2013; Lomonte et al., 2014) but not in cones to the best of our knowledge. All PCAs recovered generally non-overlapping patterns for *Africonus* and *Lautoconus*, indicating that species more closely related tend to have the same conotoxin precursor superfamilies, in similar proportions and expressed at similar levels (contradicting the results of (Duda and Remigio, 2008), who reported that differences in expression of conotoxins of closely related Indo-Pacific vermivorous species could not be explained by phylogeny but by functional convergence). The species *L. reticulatus* was as an outlier in the PCA of presence/absence of superfamilies, which could be because it was the one with the least number of conotoxin precursors. Two outliers were found in the PCA of relative expression levels. In the case of *A. infinitus*, the differences in expression could be partly related to the fact that this species from Maio Island was the only representative of its lineage. In the case of the individual A_0055 of *A. maioensis* the interpretation is not that straightforward since the other individual of this species clusters with most species, and they come from neighboring populations.

Although the exact worm species eaten by the different species of *Lautoconus* and *Africonus* are unknown, at least the three clades described within genus *Lautoconus* correlate with different morphologies of the radular teeth suggesting subtle diet specializations (Abalde et al., 2017b). Moreover, the two genera have a dissimilar geographic distribution, with *Africonus* confined to the archipelago of Cabo Verde, and *Lautoconus* present in both the continental coast and the Canary Islands, opening the possibility of different processes of local adaptation. We tested whether the two genera showed differential expression of their venom components, which could be correlated with diet adaptations. Up to 11 superfamilies were differentially expressed, most of them overexpressed in *Africonus*. Most of these superfamilies were only recently described through comparative transcriptomics and only three have been known for longer time. The B1 superfamily (Conantokin) was originally described in the piscivorous *G. geographus* and reported to provoke a "sleeping" phenotype in vertebrates (Olivera et al., 1985). Although this superfamily has been identified in other vermivorous species, its function has not been characterized (Lu et al., 2014; Robinson and Norton, 2014). The N superfamily has been described in the molluscivorous *C. marmoreus*, although its function is unknown (Dutertre et al., 2013; Robinson and Norton, 2014). Finally, the V superfamily was first identified in the venom of the vermivorous *Virgiconus virgo* (Peng et al., 2008) and later in other species. To date, there is no information regarding the structure or function of this superfamily (Robinson and Norton, 2014). Nonetheless, results on differential expression should be interpreted with caution as in some instances the captured signal could rely on the high expression level of just one biological replicate, and may not be considering the variance among replicates. To correct such potential bias, we run an ANOVA test and found that only the B1 superfamily presented significantly different expression between genera.

Similarly, we tested the presence of differential expression between piscivorous (three individuals of *C. ermineus*; (Abalde et al., 2018) and vermivorous cones in West Africa. We found two superfamilies differentially overexpressed in *C. ermineus* (A and S) and 20 in the vermivorous cones. The importance of having different members of the A superfamily for hunting fish has been highlighted previously for several cone species (Lopez-Vera et al., 2007; Olivera et al., 2015; Hoggard et al., 2017; Abalde et al., 2018). The S superfamily was first identified in *G. geographus* and found to inhibit neurotransmitter receptors (England et al., 1998). Later, it was reported as minor component of different cone species, not all necessarily hunting on fish. After the ANOVA test, only the expression of the A superfamily was identified as significantly different between diets.

Acknowledgements

We are grateful to Cabo Verde biology students Paulo Vasconcelos and Stiven Delgado for their valuable help during sampling in Santa Luzia (Cabo Verde). We thank Dr Rui Freitas from the Universidade de Cabo Verde (UniCV) for his continuous support of our research in Cabo Verde. We are indebted to Dr Iderlindo Silva dos Santos and Dra. Sonia Monteiro de Pina Araujo from the Direcçao Nacional do Medio Ambiente of the Ministério do Ambiente, Habitação e Ordenamento do Território (MAHOT) of the Republic of Cabo Verde for their help with collecting permits (Autorizações 07/2013, 26/2013, 01/2104, 04/2015, and 03/2016). We thank Jesús Marco and Aida Palacio, who provided access to the supercomputer Altamira at the Institute of Physics of Cantabria (IFCA-CSIC), member of the Spanish Supercomputing Network, for performing assembling and phylogenetic analyses.

Funding

This work was funded by the Spanish Ministry of Economy, Industry and Competitiveness (CGL2013-45211-C2-2-P

and CGL2016-75255-C2-1-P [AEI/FEDER, UE]) to R.Z.; BES-2014-069575 to S.A.

Authors' contributions

R.Z. conceived the study; M.J.T., C.M.L.A., and R.Z. obtained the samples; S.A. generated, assembled and annotated the trancriptome sequences; S.A. and R.Z. analyzed the sequence data. All authors participated in the writing of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Data availability

Sequence data are available at NCBI (SRA and GenBank). Supplementary Material available at:

https://docs.wixstatic.com/ugd/42cff7_b64fdd27335a488a9a97ce9040f7b6a9.pdf
https://docs.wixstatic.com/ugd/42cff7_b790267e56194c11b3136a9dfdcd94e1.xlsx?dn=Suppl.%20Mat.%20Table%201.xlsx
https://docs.wixstatic.com/ugd/42cff7_3c420c46f2c2489fb540ff115fbc1d9e.xlsx?dn=Suppl.%20Mat.%20Table%202.xlsx
https://docs.wixstatic.com/ugd/42cff7_f2729f1eba304ab8b5023ed5c27d56fc.pdf

5 | REFERENCES

Abalde, S., Tenorio, M.J., Afonso, C.M.L., Uribe, J.E., Echeverry, A.M., Zardoya, R., 2017a. Phylogenetic relationships of cone snails endemic to Cabo Verde based on mitochondrial genomes. BMC Evolutionary Biology 17, 231.

Abalde, S., Tenorio, M.J., Afonso, C.M.L., Zardoya, R., 2017b. Mitogenomic phylogeny of cone snails endemic to Senegal. Molecular Phylogenetics and Evolution 112, 79-87.

Abalde, S., Tenorio, M.J., Afonso, C.M.L., Zardoya, R., 2018. 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.

Ahorukomeye, P., Disotuar, M.M., Gajewiak, J., Karanth, S., Watkins, M., Robinson, S.D., Florez Salcedo, P., Smith, N.A., Smith, B.J., Schlegel, A., Forbes, B.E., Olivera, B., Hung-Chieh Chou, D., Safavi-Hemami, H., 2019. Fish-hunting cone snail venoms are a rich source of minimized ligands of the vertebrate insulin receptor. Elife 8, e41574.

Aman, J.W., Imperial, J.S., Ueberheide, B., Zhang, M.-M., Aguilar, M., Taylor, D., Watkins, M., Yoshikami, D., Showers-Corneli, P., Safavi-Hemami, H., Biggs, J., Teichert, R.W., Olivera, B.M., 2014. Insights into the origins of fish hunting in venomous cone snails from studies of Conus tessulatus. Proceedings of the National Academy of Sciences USA 112, 5087-5092.

Andrews, S., 2010. FastQC: a quality control for high throughput sequence data. available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; last accessed September 05, 2018.

Barghi, N., Concepcion, G.P., Olivera, B.M., Lluisma, A.O., 2015. High conopeptide diversity in *Conus tribblei* revealed through analysis of venom duct transcriptome using two High-Throughput sequencing platforms. Marine Biotechnology 17, 81-98.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., Eddy, S.R., 2004. The Pfam protein families database. Nucleic Acids Research 32, D138-141.

Bayrhuber, M., Vijayan, V., Ferber, M., Graf, R., Korukottu, J., Imperial, J., Garrett, J.E., Olivera, B.M., Terlau, H., Zweckstetter, M., Becker, S., 2005. Conkunitzin-S1 is the first member of a new Kunitz-type neurotoxin family. Structural and functional characterization. Journal of Biological Chemistry 280, 23766-23770.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2013. GenBank. Nucleic Acids Research 41, D36-42.

Bergeron, Z.L., Chun, J.B., Baker, M.R., Sandall, D.W., Peigneur, S., Yu, P.Y., Thapa, P., Milisen, J.W., Tytgat, J., Livett, B.G., Bingham, J.P., 2013. A 'conovenomic' analysis of the milked venom from the mollusk-hunting cone snail *Conus textile*--the pharmacological importance of post-translational modifications. Peptides 49, 145-158.

Binford, G., 2001. Differences in venom composition between orb-weaving and wandering Hawaiian *Tetragnatha* (Araneae). Biological Journal of the Linnean Society 74, 581-595.

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

Chang, D., Duda, T.F., Jr., 2012. Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. Molecular Biology and Evolution29, 2019-2029.

Chang, D., Duda, T.F., Jr., 2016. Age-related association of venom gene expression and diet of predatory gastropods. BMC Evolutionary Biology 16, 27.

Conticello, S.G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z., Fainzilber, M., 2001. Mechanisms for evolving hypervariability: the case of conopeptides. Molecular Biology and Evolution 18, 120- 131.

Cunha, R.L., Castilho, R., Rüber, L., Zardoya, R., 2005. Patterns of cladogenesis in the venomous marine gastropod genus *Conus* from the Cape Verde islands. Systematic Biology 54, 634-650.

Cunha, R.L., Lima, F.P., Tenorio, M.J., Ramos, A.A., Castilho, R., Williams, S.T., 2014. Evolution at a different pace: distinctive phylogenetic patterns of cone snails from two ancient oceanic archipelagos. Systematic Biology 63, 971-987.

Duda, T.F., Jr., 2008. Differentiation of venoms of predatory marine gastropods: divergence of orthologous toxin genes of closely related *Conus* species with different dietary specializations. Journal of Molecular Evolution 67, 315-321.

Duda, T.F., Jr., Chang, D., Lewis, B.D., Lee, T., 2009. Geographic variation in venom allelic composition and diets of the widespread predatory marine gastropod *Conus ebraeus*. PLoS One 4, e6245.

Duda, T.F., Jr., Palumbi, S.R., 2004. Gene expression and feeding ecology: evolution of piscivory in the venomous gastropod genus *Conus*. Proceedings of the Royal Society series B 271, 1165-1174.

Duda, T.F., Jr., Remigio, E.A., 2008. Variation and evolution of toxin gene expression patterns of six closely related venomous marine snails. Molecular Ecology 17, 3018-3032.

Duda, T.F., Lee, T., 2009. Isolation and population divergence of a widespread Indo-West Pacific marine gastropod at Easter Island. Marine Biology 156, 1193-1202.

Duda, T.F., Palumbi, S.R., 1999. Molecular genetics of ecological diversification: Duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. Proceedings of the National Academy of Sciences USA 96, 6820-6823.

Duda, T.F., Rolán, E., 2005. Explosive radiation of Cape Verde *Conus*, a marine species flock. Molecular Ecology 14, 267-272.

Dutertre, S., Jin, A.-H., Vetter, I., Hamilton, B., Sunagar, K., Lavergne, V., Dutertre, V., Fry, B.G., Antunes, A., Venter, D.J., Alewood, P.F., Lewis, R.J., 2014. Evolution of separate predation- and defence-evoked venoms in carnivorous cone snails. Nature Communications 5, 3521.

Dutertre, S., Jin, A.H., Kaas, Q., Jones, A., Alewood, P.F., Lewis, R.J., 2013. Deep venomics reveals the mechanism for expanded peptide diversity in cone snail venom. Molecular Cell Proteomics 12, 312-329.

Elliger, C.A., Richmond, T.A., Lebaric, Z.N., Pierce, N.T., Sweedler, J.V., Gilly, W.F., 2011. Diversity of conotoxin types from *Conus californicus* reflects a diversity of prey types and a novel evolutionary history. Toxicon 57, 311-322.

England, L.J., Imperial, J.S., Jacobsen, R.B., Craig, A.G., Gulyas, J., Akhtar, M., Rivier, J., Julius, J., Olivera, B.M., 1998. Inactivation of a serotonin-gated ion channel by a polypeptide toxin from marine snails. Science 281, 575 - 578.

Espiritu, D.J.D., Watkins, M., Dia-Monje, V., Cartier, G.E., Cruz, L.J., Olivera, B.M., 2001. Venomous cone snails: molecular phylogeny and the generation of toxin diversity. Toxicon 39, 1899-1916.

Francis, W.R., Christianson, L., M., Kiko, R., Powers, M.L., Shaner, N.C., Haddock, S.H.D., 2013. A comparison across non-model animals suggests an optimal sequencing depth for *de novo* transcriptome assembly. BMC Genomics 14, 167.

Gao, B., Peng, C., Yang, J., Yi, Y., Zhang, J., Shi, Q., 2017. Cone snails: a big store of conotoxins for novel drug discovery. Toxins 9, E397.

Gibbs, H.L., Sanz, L., Sovic, M.G., Calvete, J.J., 2013. Phylogeny-based comparative analysis of venom proteome variation in a clade of rattlesnakes (Sistrurus sp.). PLoS One 8, e67220.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29, 644-652.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology 59, 307 - 321.

Hansson, K., Thämlitz, A.M., Furie, B., Furie, B.C., 2006. A single γ-carboxyglutamic acid residue in a novel cysteine-rich secretory protein without propeptide. Biochemistry 45, 12828 - 12839.

Hoggard, M.F., Rodriguez, A.M., Cano, H., Clark, E., Tae, H.S., Adams, D.J., Godenschwege, T.A., Mari, F., 2017. *In vivo* and *in vitro* testing of native alpha-conotoxins from the injected venom of *Conus purpurascens*. Neuropharmacology 127, 253-259.

Hu, H., Bandyopadhyay, P.K., Olivera, B.M., Yandell, M., 2011. Characterization of the *Conus bullatus* genome and its venom-duct transcriptome. BMC Genomics 12, 60.

Jin, A.H., Dutertre, S., Dutt, M., Lavergne, V., Jones, A., Lewis, R.J., Alewood, P.F., 2019. Transcriptomic-proteomic correlation in the predation-evoked venom of the cone snail, *Conus imperialis*. Marine Drugs 17, 177.

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

Kaas, Q., Yu, R., Jin, A.H., Dutertre, S., Craik, D.J., 2012. ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. Nucleic Acids Research 40, D325-330.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics bbx108.

Kohn, A.J., 1959. The ecology of *Conus* in Hawaii. Ecological Monographs 29, 47-90.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357-359.

Lavergne, V., Dutertre, S., Jin, A.-H., Lewis, R.J., Taft, R.J., Alewood, P.F., 2013. Systematic interrogation of the *Conus marmoreus* venom duct transcriptome with ConoSorter reveals 158 novel conotoxins and 13 new gene superfamilies. BMC Genomics 14, 708.

Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M., Haag, J.D., Gould, M.N., Stewart, R.M., Kendziorski, C., 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29, 1035-1043.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

Li, Q., Barghi, N., Lu, A., Fedosov, A., E., Bandyopadhyay, P.K., Lluisma, A., O., Concepcion, G.P., Yandell, M., Olivera, B.M., Safavi-Hemami, H., 2017. Divergence of the venom exogene repertoire in two sister species of *Turriconus.* Genome Biology and Evolution 9, 2211-2225.

Lomonte, B., Tsai, W.C., Urena-Diaz, J.M., Sanz, L., Mora-Obando, D., Sanchez, E.E., Fry, B.G., Gutierrez, J.M., Gibbs, H.L., Sovic, M.G., Calvete, J.J., 2014. Venomics of New World pit vipers: genus-wide comparisons of venom proteomes across *Agkistrodon*. Journal of Proteomics 96, 103-116.

Lopez-Vera, E., Jacobsen, R.B., Ellison, M., Olivera, B.M., Teichert, R.W., 2007. A novel alpha conotoxin (alpha-PIB) isolated from *C. purpurascens* is selective for skeletal muscle nicotinic acetylcholine receptors. Toxicon 49, 1193-1199.

Lu, A., Yang, L., Xu, S., Wang, C., 2014. Various conotoxin diversifications revealed by a venomic study of *Conus flavidus*. Molecular Cell Proteomics 13, 105-118.

Maddison, W.P., Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. Version 3.51.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10-12.

Milne, T.J., Abbenante, G., Tyndall, J.D., Halliday, J., Lewis, R.J., 2003. Isolation and characterization of a cone snail protease with homology to CRISP proteins of the pathogenesis-related protein superfamily. Journal of Biological Chemistry 278, 31105-31110.

Olivera, B.M., 2006. Conus peptides: biodiversity-based discovery and exogenomics. Journal of Biological Chemistry 281, 31173-31177.

Olivera, B.M., McIntosh, J.M., Clark, C., Middlemas, D., Gray, W.R., Cruz, L.J., 1985. A sleep-inducing peptide from *Conus geographus* venom. Toxicon 23, 277 - 282.

Olivera, B.M., Rivier, J., Clark, C., Ramilo, C.A., Corpuz, G.P., Abogadie, F.C., Mena, E.E., Woodward, S.R., Hillyard, D.R., Cruz, L.J., 1990. Diversity of *Conus* neuropeptides. Science 249, 257 - 263.

Olivera, B.M., Seger, J., Horvath, M.P., Fedosov, A.E., 2015. Prey-capture strategies of fish-hunting cone snails: behavior, neurobiology and evolution. Brain Behavior and Evolution 86, 58-74.

Pahari, S., Bickford, D., Fry, B.G., Kini, R.M., 2007. Expression pattern of three-finger toxin and phospholipase A2 genes in the venom glands of two sea snakes, *Lapemis curtus* and *Acalyptophis peronii*: comparison of evolution of these toxins in land snakes, sea kraits and sea snakes. BMC Evolutionary Biology 7, 175.

Pekar, S., Bocanek, O., Michalek, O., Petrakova, L., Haddad, C.R., Sedo, O., Zdrahal, Z., 2018. Venom gland size and venom complexity-essential trophic adaptations of venomous predators: A case study using spiders. Molecular Ecology 27, 4257-4269.

Peng, C., Liu, L., Shao, X., Chi, C., Wang, C., 2008. Identification of a novel class of conotoxins defined as V-conotoxins with a unique cysteine pattern and signal peptide sequence. Peptides 29, 985-991.

Peng, C., Yao, G., Gao, B.-M., Fan, C.-X., Bian, C., Wang, J., Cao, y., Wen, B., zhu, Y., Ruan, Z., Zhao, X., You, X., Bai, J., Li, J., Lin, Z., Zou, S., Zhang, X., Qiu, Y., Chen, J., Coon, S.L., Yang, J., Chen, J.-

S., Shi, Q., 2016. High-throughput identification of novel conotoxins from the Chinese tubular cone snail (*Conus betulinus*) by multitranscriptome sequencing. GigaScience 5, 17.
Phuong, M.A., Alfaro, M.E., Mahardika, G.N., Marwoto, R.M., Prabowo, R.E., von Rintelen, T., Wh Vogt, P., Hendricks, J.R., Puillandre, N., 2019. Lack of signal for the impact of conotoxin gene diversity on speciation rates in cone snails. Systematic Biology syz016.
Phuong, M.A., Mahardika, G.N., 2018. Targeted sequencing of venom genes from cone snail genomes improves understanding of conotoxin molecular evolution. Molecular Biology and Evolution 35, 1210-1224.
Phuong, M.A., Mahardika, G.N., Alfaro, M.E., 2016. Dietary breadth is positively correlated with venom complexity in cone snails. BMC Genomics 17, 401.
Pin, M., Leung Tack, K.D., 1995. The Cones of Senegal. Conchiglia.
Prashanth, J.R., Dutertre, S., Jin, A.H., Lavergne, V., Hamilton, B., Cardoso, F.C., Griffin, J., Venter, D.J., Alewood, P.F., Lewis, R.J., 2016. The role of defensive ecological interactions in the evolution of conotoxins. Molecular Ecology 25, 598-615.
Prashanth, J.R., Dutertre, S., Lewis, R.J., 2017. Pharmacology of predatory and defensive venom peptides in cone snails. Molecular Biosystems 13, 2453-2465.
Prator, C.A., Murayama, K.M., Schulz, J., R., 2014. Venom variation during prey capture by the cone snail, *Conus textile*. PLoS One 9, e98991.
Ranasinghe, S., McManus, D.P., 2013. Structure and function of invertebrate Kunitz serine protease inhibitors. Developmental and Comparative Immunology 39, 219-227.
R Core Team, 2013. R: A language and environment for statistical computing. In: computing, R.F.f.S. (Ed.), Vienna, Austria.
Remigio, E.A., Duda, T.F., Jr., 2008. Evolution of ecological specialization and venom of a predatory marine gastropod. Molecular Ecology 17, 1156-1162.
Robinson, S.D., Li, Q., Lu, A., Bandyopadhyay, P.K., Yandell, M., Olivera, B.M., Safavi-Hemami, H., 2017. The venom repertoire of *Conus gloriamaris* (Chemnitz, 1777), the glory of the sea. Marine Drugs 15, 145.
Robinson, S.D., Norton, R.S., 2014. Conotoxin gene superfamilies. Marine Drugs 12, 6058-6101.
Robinson, S.D., Safavi-Hemami, H., McIntosh, L.D., Purcell, A.W., Norton, R., S., Papenfuss, A.P., 2014. Diversity of conotoxin gene superfamilies in the venomous snail, *Conus victoriae*. PLoS One 9, e87648.
Safavi-Hemami, H., Gajewiak, J., Karanth, S., Robinson, S.D., Ueberheide, B., Douglass, A.D., Schlegel, A., Imperial, J.S., Watkins, M., Bandyopadhyay, P.K., Yandell, M., Li, Q., Purcell, A.W., Norton, R.S., Ellgaard, L., Olivera, B.M., 2015. Specialized insulin is used for chemical warfare by fish-hunting cone snails. Proceedings of the National Academy of Sciences USA 112, 1743 - 1748.
Salzberg, S.L., 2019. Next-generation genome annotation: we still struggle to get it right. Genome Biology 20, 92.
Terrat, Y., Biass, D., Dutertre, S., Favreau, P., Remm, M., Stöcklin, R., Piquemal, D., Ducancel, F., 2012. High-resolution picture of a venom gland transcriptome: Case study with the marine snail *Conus consors*. Toxicon 59, 34-46.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673 - 4680.
Tucker, J.K., Tenorio, M.J., 2009. Systematic classification of recent and fossil conoidean gastropods, with keys to the genera of cone shells. ConchBooks.
Tucker, J.K., Tenorio, M.J., 2013. Illustrated catalog of the living cone shells. MdM Publishing.
UniProt, C., 2015. UniProt: a hub for protein information. Nucleic Acids Research 43, D204-212.