Research Article

Genetic Analysis of Signal Peptides in Amphibian Antimicrobial Secretions.

Running title: Signal Peptide Diversity from Frogs

L.O. Pérez [a]; N.L. Cancelarich [b]; S. Aguilar [b]; N.G. Basso [c] and M.M. Marani [b*]

[a]Instituto Patagónico de Ciencias Sociales y Humanas (IPCSH), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), 2915 Brown Boulevard, Puerto Madryn, Argentina.

[b]Instituto Patagónico para el Estudio de Ecosistemas Continentales (IPEEC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), 2915 Brown Boulevard, Puerto Madryn, Argentina. mmarani@cenpat-conicet.gob.ar; lncancelarich@cenpat-conicet.gob.ar

[c] Instituto de Diversidad y Evolución Austral (IDEAus), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), 2915 Brown Boulevard, Puerto Madryn, Argentina.

* *Correspond*ing *author*. Instituto Patagónico para el Estudio de Ecosistemas Continentales (IPEEC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), 2915 Brown Boulevard, Puerto Madryn, Argentina. Email address: mmarani@cenpat-conicet.gob.ar.

*Corresponding author at: IPEEC–CONICET, Laboratorio de Bioprospección y Aplicaciones Biotecnológicas de Péptidos (BIAPEP), Bvd. Brown 2915, CP U9120ACD, Puerto Madryn, Chubut, Argentina.

E-mail address: mmarani@cenpat-conicet.gob.ar (M.M. Marani)

Summary

Amphibian secretion is an important source of bioactive molecules that naturally protect the skin against noxious microorganisms. Collectively called Antimicrobial Peptides (AMPs), these molecules have a wide spectrum of action, targeting viruses, bacteria and fungi. Like many membrane and secreted proteins, AMPs have cleavable signal sequences that mediate and translocate the nascent polypeptide chains into the endoplasmic reticulum. Although it is accepted that the signal peptides are simple and interchangeable, there is neither sequence nor structure that is conserved among all gene families. They derived from a common ancestor but developed different traits as they adapt to distinct environmental pressures. The aim of this study was to provide an overview of the diversity of signal peptides of the frog, taking into account reported cDNA sequences and the evolutionary relationship among them. We analysed more than two thousand records, reported the relative abundance, diversity and evolutionary divergence based on the peptide signals of frog antimicrobial peptides. We conclude that the physical properties of the sequence is more important than the specific peptides in AMP signal peptides. Since there is significant overlapping among related genera, differences in secretion from different peptides types should be regulated by additional levels, like post-transcriptional modifications or 5´UTR sequences.

Keywords: Antimicrobial peptides; Signal Peptide; Diversity

Introduction

Antimicrobial peptides (AMPs) are an essential part of the innate immunity of organisms and display activity against a wide range of microorganisms. Based on their potential use in the pharmaceutical and biotechnology industry, there has been a rise in medicinal peptide patent applications and developmental work (Uhlig et al. 2014). Particularly, many amphibian species were studied and more than a thousand skin-frog-derived peptides have been screened for a wide diversity of applications, demonstrating activities such as anticancer (Wu et al. 2014), insecticidal (Smith et al. 2011), chemotactic (Köning et al. 2015), antioxidant (Guo et al. 2014), wound healing (Gallo and Huttner 1998), and protease inhibitor (Miller et al 1998).

AMPs typically display a tripartite structure. In modern frogs, the N-terminal region has a 22-residue conserved signal peptide (SP), followed by an acidic propeptide domain (16-27 residue) and a hyper variable C-terminal domain containing the biologically active peptide. It is encoded in a single locus, spanning two exons separated by an intron that can range from several hundred to over a thousand bases. The first exon encodes the SP, while the second encodes the acidic region and the mature peptide. Depending on the genus, the whole structure could be duplicated in tandem or scattered throughout the genome (Amiche et al. 1999).

SPs are found in the N-terminal end of secreted and membrane protein precursors of eukaryotes and prokaryotes and they mediate the targeting of the nascent polypeptide chains to the endoplasmic reticulum (ER). A protein complex, the translocon, identifies the SP and transports the polypeptide to the ER. The SP is cleaved in the process and the peptide is exported outside the cell by the secretory pathway (Lodish et al. 2000). While SPs of different proteins share low sequence identity, they also have a tripartite structure: a) A hydrophilic N-terminus, b) a hydrophilic core and c) a more conserved region that ends with cysteine (Hon et al. 2009). In frog AMPs, there is no certainty whether the SPs have a specific peptidase for secretion or whether they follow an ER independent pathway.

From a practical point of view, a better understanding of the diversity of SPs would lead to a more efficient screening by molecular techniques. These complement conventional methodologies like liquid chromatography and others that infer the peptide sequence from the mature peptide (Csordas and Michl 1969). Cloning and DNA sequencing allow the identification of the complete AMP prepropeptide by directly examining the mRNA, providing unambiguous sequence assignments. A common practice is to perform the 3'RACE amplification with primers that match the 5'UTR or the SP region (Nicolas et al. 2003). In this case, the presence of polymorphic regions is a concern for researcher when designing specific primers. Degenerate primers are needed to recover the diversity of products at expenses of sensitivity.

In this study, we performed an overview of the SP regions of amphibian skin antimicrobial peptides. The main objective of this study was to analyse their sequences and report frequent motifs in order to provide useful information for bioprospecting AMPs.

Materials and Methods

Signal peptide data set

Complementary DNA sequences corresponding to complete prepropeptides of skin antimicrobial peptides from amphibian were downloaded from the NCBI (http://www.ncbi.nlm.nih.gov/pubmed/). In both cases, the words "cDNA", "antimicrobial peptides" and "frog" were used as search keywords. The search comprised a period that extended from1988 to the beginning of 2017. Curation of sequences was performed according to the following criteria: exclusion of sequences that were obtained with primers that overlapped the SP region (the source article was examined when available), data redundancy, and incomplete sequences. Sequence alignment was performed at the nucleotide level using the Muscle algorithm of the *mass* package of R (Edgar 2004).

Diversity description

Total compiled and manually curated SP sequences were analysed. The SP region was defined as the first 22 amino acids residues that usually ends with a cysteine (Neobatrachia). In the case of sequence descriptions that stated otherwise (primitive frogs), or sequences that did not end with a Cys, the length was based on published data (if available) or the SignalP 4.1 algorithm (Petersen et al. 2011). The variation of SP nucleotides was described with basic statistics using the *Poppr* packages from the R software (Kamvar et al. 2014). This included the Simpson index, a common diversity measure that estimates the probability that two entities taken at random represent a same type in a specific database. We used the Shannon Weaver index for SP genotype variability, a measure of uncertainty of information (H) (Morris et al. 2014). The analysis of diversity amongst individuals, genera and families was performed by the analysis of molecular variance (AMOVA) using the *amove.ppoppr* function with the following parameters: positive clone correcting, distance matrix by raw pairwise distances and farthest neighbour clustering. For test significance of the results, we performed a randomization test that permutes the samples (Excoffier et al. 1992). After that, we performed discriminant analysis of principal components (DAPC), a method specifically designed for genetic data to find the best discrimination of individuals in pre-defined groups in order to provide a visual assessment of genetic differentiation and allele contribution. We used the *ade4* package for cross-validation of DAPC, using varying numbers of principal components (PC), setting 300 to the maximum number of PCs, training set to 0.9 of the sample and 30 repetitions.

Signal peptide alignment and consensuses

Sequence alignment was performed by the MUSCLE algorithm from the *mass* package (R) (Edgar 2004). DNA consensuses of families were obtained by the *Biostrings* package (Pagès et al. 2017). The output was uploaded to the CIPRES (https://www.phylo.org/) (Miller et al. 2010) portal and the best tree was built using the maximum likelihood method (RAxML v8 algorithm using a GTR+Gamma nucleotide substitution model) and Bombanitoridae as the rooted branch. Analysis of records of the

whole data set was performed by Maximum likelihood analysis and the model selection was calculated by Neighbour Joining analysis. Graphical representation of trees was done using the MEGA software.

Results

Signal peptide data set

We compiled 2852 individual records of the AMP precursors (prepropeptides) from the NCBI nucleotide database. After processing the genotypes that did not meet the criteria, there were 2165 sequences left (1125 genotypes). These AMPs belonged to 14 Amphibian families, 32 genera, and 118 species. Frog families were not equally represented. Ranidae was the largest family, comprising 86% of the total. A summary of the data is shown in Table 1.

The SP region of AMP prepropeptides was defined as the first 66 nucleotides in modern frogs (Neobatrachia). A few exceptions did not show the common Cys residue ending (see Materials and Methods). In order to avoid information redundancy, we removed duplicated SP sequences within species (but not among them), leaving 1368 sequences out of 2165. This is not a clone correction *per se* because each sequence is part of different complete mRNAs.

Overall, sequence similarity remained high among groups. However, if we consider genotypes differentiated by at least one or more nucleotides, and ignore changes below 2% of the population, 1125 genotypes were unique sequences. Nine hundred and ninety sequences were species specific SPs and 135 of them were shared by two or more species: 8.7% were shared by only two species, 1.2% by three species, 0.8% by four species, and 0.6% by six or more species.

According to the record descriptions of NCBI, 90% of sequences had a label that identified the type of mature peptide and 10% had the generic description "mRNA antimicrobial precursor". Figure 1 shows an example of the peptide diversity uploaded for the Ranidae family. In total, eighty-six different peptide designations were found, but only 30% belonged to well established AMP families, taking

(König et al. 2015) as the reference. A first revision of the peptide names revealed that many designations did not follow the nomenclature proposed by Simmaco (Simmaco et al. 1994). The lack of general agreement in nomenclature, along with a number of peptides that have single or few exemplars ("orphan peptides"), discouraged us from analysing SPs according to the type of mature peptide.

Diversity description

Overall, the AMP SP was a conserved sequence. At the nucleotide level, genetic variability remained relatively uniform across the sequence. The most variable sites belonged to two bases (52 and 53), near the C-terminal codon 17, Simpson´s index > 1.2 (Figure 2). On the other hand, codons 35, 38, 41 and 50 were highly conserved amongst species. They were all located in the middle region and the triplets coded for hydrophobic residues.

The overall diversity, measured by the Shannon-Wiener Index H, was 6.383 with Standard Error of 0.47. In general, variability increased with increasing sample size within each genus. Although they were correlated, it was not always the case, where *Amolops* (n=321), *Odorrana* (n=721), *Rana* (n=226), and *Bombina* (n=113) were diverse groups that had SW indexes above four. However, notable exceptions were *Silvyrana* (H = 3.9, n = 238) and *Babina* (H = 3.5, n= 161). The SPs coming from primitive taxons (Archaeobatrachia and Mesobatrachia) were shorter, and although they carry the typical structure of a hydrophobic core and the semi polar ending, the nucleotide sequence was markedly different.

The AMOVA revealed high genetic variation within genera (63.1%, Phi 0.36). There was an overlapping of the sequences among genera, resulting in low genetic differentiation between families (18.8%, Phi 0.22) and between genera (18.1%, Phi 0.18). The randomization test showed that within and between genus variance were significantly different from expected ($p > 0.05$). However, variation between families was not statistically significant ($p = 0.04$). We applied Discriminant Analysis of Principal Components in an attempt to summarize the genetic differences among genera while overlooking the variance within them. There were three main clusters (Figure 3) according to the two

first discriminant functions. Archaeobatrachia genera like *Bombina* and *Alytes* clustered with Myobatrachidae, *Xenopus* and *Kassina*. In the top right quadrant, *Pithecopus, Phyllomedusa, Agalychnis* and Ranoidea clustered together. All the rest grouped in the middle. Interestingly, the discriminant functions showed that the polymorphisms that most differentiated the SPgenera were located in the first 24 bases (8 amino acids).

Signal peptide alignment and consensuses

We performed the consensus for each family using the available data by multiple alignment. Pipidae, Alytidae, and Bombinatoridae had distinctive SPs. The Pipidae family, in particular, exhibited the shortest SP sequence, encoding for 15 amino acids instead of the typical 22. The translated signal revealed that Bombinatoridae had a conserved serine in the middle of the sequence, an uncommon feature in any other family of our database. This was an exception to the rule of having apolar residues at the core. Another interesting feature of Archaeobatrachia was a SAYA motif at the C-terminal region or a similar variation. This might indicate the presence of enzyme specific recognition by the peptidase process.

Consensus dendrogram by family

We built a dendrogram based on DNA alignments of family consensus. The dendrogram is shown in Figure 4, indicating that clusters at the super family level were consistent with the diversification found in phylogenetic studies (Pyron and Wiens 2011). We observed that main divisions were maintained: Archaeobatrachia, Mesobatrachia and Neobatrachia. Inside Neobatrachia, all members of the Ranoidea superfamily stayed together (squares in Fig 4). The Hyloidea family members (circles in Fig 4) clustered with Myobatrachidae and Ranoidea. Representative AMP signal sequences were described at the right of the Figure 4.

DISCUSSION

Classical methods for skin peptide identification are based on purification of complex frog secretions followed by sequencing methods, such as mass spectrometry and Edman degradation. The

emergence of high-throughput molecular technologies involving *de novo* peptide sequencing via cDNA cloning, opened the field to fast data acquisition and the generation of extensive databases. Some of these techniques rely on previous knowledge of contiguous sequence, usually from the SP region or 5´UTR.

There are several databases available to search for AMP information: APD (Antimicrobial Peptide Database) (Wang and Wang 2004; Wang et al. 2016), AMPer (Fjell et al. 2007), ANTIMI (Brahmachary et al. 2004) and CAMPR3 (Collection of Anti-Microbial Peptides), among others. APD is the largest specific database on AMP with more than 2884 peptides manually curated from literature (PubMed, PDB, Google and Swiss-Prot), of which 1050 records are from amphibians. In general, most of the information is based on peptides and there is not a direct link to their DNA sequences. For raw data, the NCBI nucleotide database or the original studies are the main sources.

Most of the uploaded records belonged to Ranidae, the diverse anuran family known as true frogs. This result was expected due to the wide distribution around the world, comprising several genera and more than 600 species. It has at least 13 different peptide families with limited structural similarities and diverse properties (Conlon 2008). Another abundant family was the South American Phyllomedusidae (formerly Hylidae), a group that pioneered the search of new bioactive peptides. Most families of modern frogs have, in general, SPs with 22 aminoacids long, hydrophobic core with a subtle polar end, and typically one or two serines in the terminal region. This pattern belongs to the old World or tropical regions. Geographically separated families, such as the Australian Myobatrachidae, showed patterns so distinctive that they were grouped near primitive frogs.

The most studied family in primitive frogs was Bombinatoridae, whose products showed interesting biological activities (bombinins and bombininH) in at least three species (Simmaco et al. 2009). They are represented by two patterns with some resemblance to SPs from Alytidae.

It is worth noting that many important families do not have skin SPs in the database. An example is Bufonidae, whose AMPs have been recovered from internal organs (Park et al. 1996) and the skin

(Gupta et al. 2017). The authors who collected these secretions did not report the pre propeptide DNA. Another example is a particular sequence derived from the South African frog *Xenopus laevis*, a well-studied experimental model. Its skin has a complex composition of secretion with tens of peptides (Mechkarska et al. 2013). The SP is typically 20 aminoacids long, with the exception of magainins (18 aminoacids long). All of them share a typical polar end, composed by serine, threonine or asparagine before the last two or three aminoacids.

The SP region has sufficient variability to differentiate between large groups, like superfamily. At the family class, the within group variation decreases noticeably and it is very frequent that the same SP is shared between genera. This observation is opposite to the great diversity seen in the mature region of the gene. Since this region is the only one that interacts with microorganisms, it is likely under great selective pressure. It is not clear whether all antimicrobial peptides are homologous, but they have probably arisen by a process of loci replication and focal hyper-mutations, similar to what is seen in venoms peptides or toxins from other species (Tang et al. 2010; König and Bininda-Emonds 2011).

The SP and the acidic region, on the other hand, seem to have a relevant function inside the cell. They are more conserved than the mature region and critical to the regulation of intracellular transportation. The variation pattern observed in this study suggests that the physical chemical properties are more important than the sequence *per se*. The cationic N-terminus and the hydrophobic core are usually present. The hydrophobic portion was totally missing in just a few exceptions, like in brevinin and temporin prepropeptides from *Rana saharica*, although we do not know whether those SP were completely functional. Other exceptions were a few exemplars from *Rana suchinae* that apparently have lost their cationic amino acids from the N-terminal part. The *Kassina* genus had a shorter SP, although the sequence was very similar to the Neobatrachia consensus in all its extension. On the other hand, the AMPs from the *Crinia* genus, native from Australia, had a unique SP. It encoded for riparins, a series of host defensins with neuropeptide, antioxidant and anti-inflammatory effects (Olivera and Teichert 2007).

Nucleotides microdeletions (one to three sites) could be detected through all the families. Given its regularity, we hypothesise that this event does not alter the functionality. It could be a regulation mechanism. However, we presume that some of these deletions could have been experimental artefacts of the sequencing technique. At the nucleotide level, four sites were remarkable conserved in SPs of almost every species. Interestingly, they were located in the middle of the signal peptide and were part of hydrophobic codons, suggesting that a hydrophobic core is physiological important in transport signalling, presumably for translocating the lipid bilayer during synthesis. The first eight residues were one of the most important sources of variation according to the DAPC analysis. Although the specific function of this subregion is largely unknown, other SPs, like the one from the *Escherichia coli* Autotransporter (EspP), has a large N terminal region that can influence the behaviour of downstream events of translocation, such as glycosylation sites (Maselli et al. 2006).

Regarding the literature, the number of studies on AMP identification has grown steadily over the last ten years in parallel to technological advances. For each peptide showing a possible useful effect, several others show little or no effect and sometimes not even mentioned in the study (i.e. direct submission to the database). Conversely, we found published AMP sequences that were absent from the nucleotide database. This is problematic since discarded peptides could not be useful in the future. In special those from endangered species, whose biodiversity is in danger.

We conclude that the physical properties of the sequence is more important than the specific peptides in AMP SPs. Since there is significant overlapping among related genera, differences in secretion from different peptides types should be regulated by additional levels, like post-trancriptional modifications or 5´UTR sequences.

Acknowledgments

This work was supported by grants from the National Council for Scientific and Technological Research (CONICET) (PIP No. 11220120100050OC). This research is framed within the P-UE CONICET N° 22920160100044. N.L.C. is grateful to CONICET by her grant. L.O.P. and M.M.M. are researchers of the CONICET. The authors state that there is no conflict of interest regarding the contents of this article.

References

Amiche M., Seon A. A., Pierre T. N. and Nicolas P. 1999 The dermaseptin precursors: a protein family with a common preproregion and a variable C-terminal antimicrobial domain. *FEBS Lett.* **456**, 352−356.

Brahmachary M., Krishnan S. P. T., Koh J. L. Y., Khan A. M., Seah S. H., Tan T. W. *et al*. 2004 ANTIMIC: a database of antimicrobial sequences. *Nucleic Acids Res.* **32**, D586–D589.

Conlon J. M. 2008 Reflections on a systematic nomenclature for antimicrobial peptides from the skins of frogs of the family Ranidae. *Peptides* **29,** 1815-9.

Csordas A. and Michl H. 1969 Primary structure of two oligopeptides of the toxin of *Bombina variegate*. *L. Toxicon* **7**, 103–108.

Edgar R. C. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-97.

Excoffier L., Smouse P. E. and Quattro J. M. 1992 Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* **131**, 479-91.

Fjell C. D., Hancock R. E. W. and Cherkasov A. 2007 AMPer: a database and an automated discovery tool for antimicrobial peptides. *Bioinformatics* **23**, 1148-1155.

Gallo R. L. and Huttner K. M. 1998 Antimicrobial peptides: an emerging concept in cutaneous biology. *J. Invest. Dermatol*. **111**, 739−743.

Gupta G. K., Joshi T., Madhudiya K., Biswas K., Bhatanagar S., Ranawat B. et al. 2017 Antibacterial activity of skin secretion of toad *Bufo melanostictus and* frog *Rana cyanophylictitus. Int. J. Acad. Res. Dev.* **2,** 684-686.

Guo C., Hu Y., Li J., Liu Y., Li S. et al. 2014 Identification of multiple peptides with antioxidant and antimicrobial activities from the skin and its secretions of *Hylarana taipehensis*, *Amolops lifanensis*, and *Amolops granulosus*. *Biochimie* **105**, 192−201.

Hon L. S., Zhang Y., Kaminker J. S. and Zhang Z. 2009 Computational prediction of the functional effects of amino acid substitutions in signal peptides using a model-based approach. *Hum. Mutat.* **30**, 99-106.

Kamvar Z. N., Tabima J. F. and Grünwald N. J. 2014 Poppr: an R package for genetic analysis of populations with clonal, parti ally clonal, and/or sexual reproduction. *PeerJ.* **2**, e281.

König E. and Bininda-Emonds O. R. P. 2011 Evidence for convergent evolution in the antimicrobial peptide system in anuran amphibians. *Peptides* **32**, 20–25.

König E., Bininda-Emonds O. R. P., Shaw C. 2015 The diversity and evolution of anuran skin peptides. *Peptides* **63**, 96–117.

Lodish H., Berk A., Zipursky S. L., Matsudaira P., Baltimore D. and Darnell J. 2000 Molecular Cell Biology, 4th edition. W. H. Freeman & Company, New York.

Maselli V. M., Bilusich D., Bowie J. H. and Tyler M. J. 2006 Host-defence skin peptides of the Australian Streambank Froglet *Crinia riparia*: isolation and sequence determination by positive and negative ion electrospray mass spectrometry. *Rapid. Commun. Mass Spectrom* **20**, 797-803.

Mechkarska M., Prajeep M., Leprince J., Vaudry H., Meetani M. A., Evans B. J. and Conlon J.M. 2013 A comparison of host-defense peptides in skin secretions of female *Xenopus laevis* × *Xenopus borealis* and *X. borealis* × *X. laevis* F1 hybrids. *Peptides* **45**:1-8.

Miller K. W., Evans E. J., Eisenberg S. P. and Thompson R. C. 1989 Secretory leukocyte protease inhibitor binding to mRNA and DNA as a possible cause of toxicity to *Escherichia coli*. *J. Bacteriol.* **171**, 2166–2172.

Miller M. A., Pfeiffer W., Schwartz T. 2010 Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans..

Morris E. K., Caruso T., Buscot F., Fischer M., Hancock C., Maier T. S. et al. 2014 Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution* **4**, 3514-3524.

Nicolas P., Vanhoye D. and Amiche M. 2003 Molecular strategies in biological evolution of antimicrobial peptides. *Peptides* **24**, 1669-80.

Olivera B. M. and Teichert R. W. 2007 Diversity of the neurotoxic Conus peptides: a model for concerted pharmacological discovery. *Mol. Interv.* **7**, 251–260.

Pagès H., Aboyoun P., Gentleman R. and DebRoy S. 2017. *Biostrings: Efficient manipulation of biological strings*. R package version 2.46.0.

Park C. B., Kim M. S. and Kim S. C. 1996 A novel antimicrobial peptide from *Bufo bufo gargarizans*. *Biochem. Biophys. Res. Commun*. 1996 **218**, 408-13.

Petersen N. T., Brunak S., von Heijne G. and Nielsen H 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786.

Pyron R. A. and Wiens J. J. A. 2011 Large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution* **61**, 543–583.

Simmaco M., Kreil G. and Barra D. 2009 Bombinins, antimicrobial peptides from Bombina species. *Biochim. Biophys Acta* **1788**, 1551-5.

Simmaco M., Mignona G., Barra D. and Bossa F. 1994 Antimicrobial peptides from skin secretions of *Rana esculenta* - molecular cloning of cDNA encoding and isolation of new active peptides. *J. Biol. Chem.* **269**, 11956-11961.

Smith A. B., Daly N. L. and Craik D. J. 2011 Cyclotides: a patent review. *Expert Opin. Ther. Pat.* **21**, 1657−1672.

Tang X., Zhang Y., Hu W., Xu D., Tao H., Yang X. et al. 2010 Molecular diversification of peptide toxins from the tarantula *Haplopelma hainanum* (Ornithoctonus hainana) venom based on transcriptomic, peptidomic, and genomic analyses. *J. Proteome Res.* **9**, 2550-64.

Uhlig T., Kyprianou T., Martinelli F. G., Oppici, C.A., Heiligers D.. Hills, D. R. J. et al. 2014 The emergence of peptides in the pharmaceutical business: From exploration to explotation. *EuPA. Open Proteomics* **4**, 58–69.

Wang G., Li X. and Wang Z. 2016 APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research* **44**, D1087-D1093.

Wang Z. and Wang G. 2004 APD: the antimicrobial peptide database. *Nucleic Acids Research* **32**, D590-D592.

Wu D., Gao Y., Qi Y., Chen L., Ma Y. and Li Y. 2014 Peptide-based cancer therapy: opportunity and challenge. *Cancer Lett.* **351**, 13−22.

Received 22 February 2018; accepted 1 May 2018

Table 1. Description of signal peptide nucleotide sequences, families and species.

Family	Nº of records		Unique Signal Peptide Sequences (DNA)	Nº ofspecies
Alytidae	**2**	**2**	**1**	
Bombinatoridae	**113**	**85**	**3**	
Ceratophryidae	**1**	**1**	**1**	
Dicroglossidae	**34**	**26**	**7**	
Hylidae	**16**	**16**	**4**	
Hyperoliidae	**10**	**10**	**2**	
Leptodactylidae	**10**	**9**	**2**	
Myobatrachidae	**4**	**4**	**1**	
Pelodryadidae	**22**	**20**	**4**	
Phyllomedusidae	**75**	**72**	**10**	
Pipidae	**6**	**6**	**1**	
Ranidae	**1864**	**1144**	**78**	
Rhacophoridae	**9**	**9**	**4**	

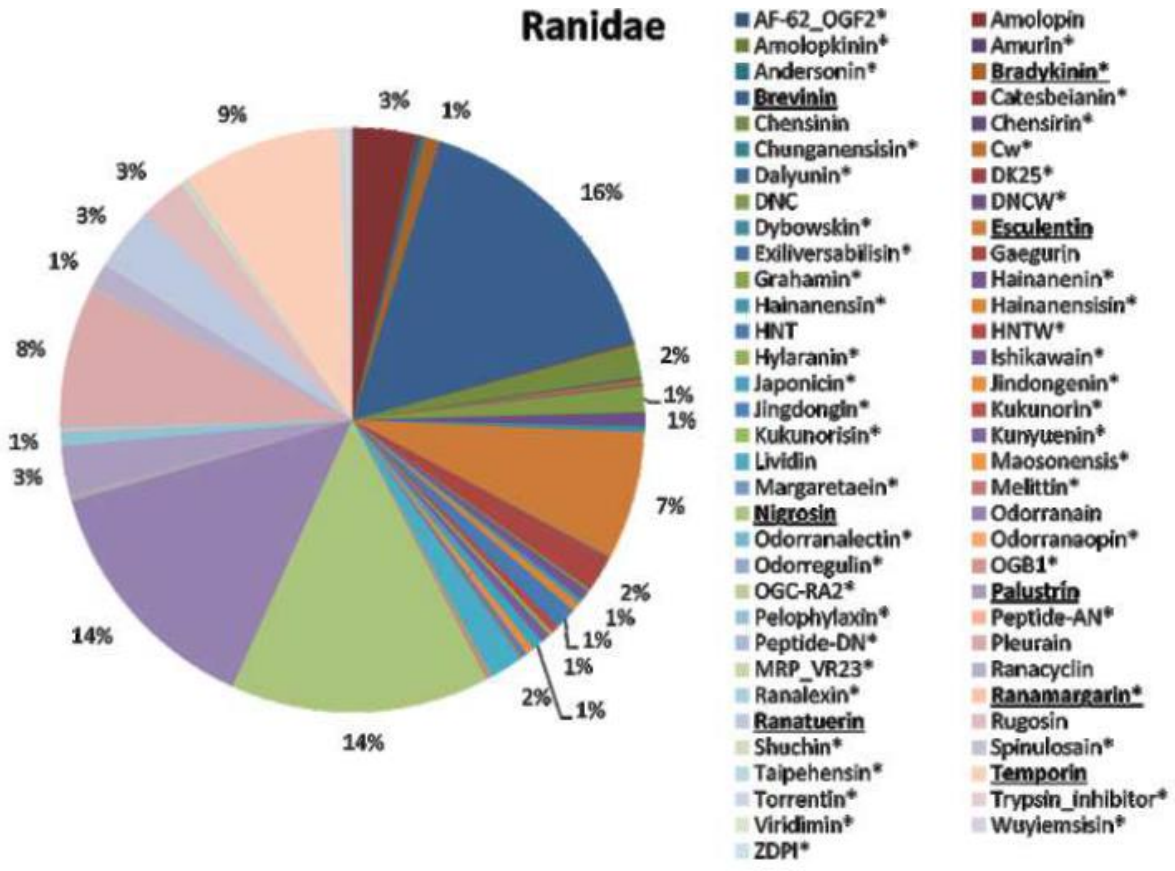

Figure 1: Peptide family diversity in Ranidae. The pie chart represents the relative abundance of mature peptides. Underlined AMPs belonged to frequent peptide families. Peptides that contributed less than 1% of the total were marked with an asterisk.

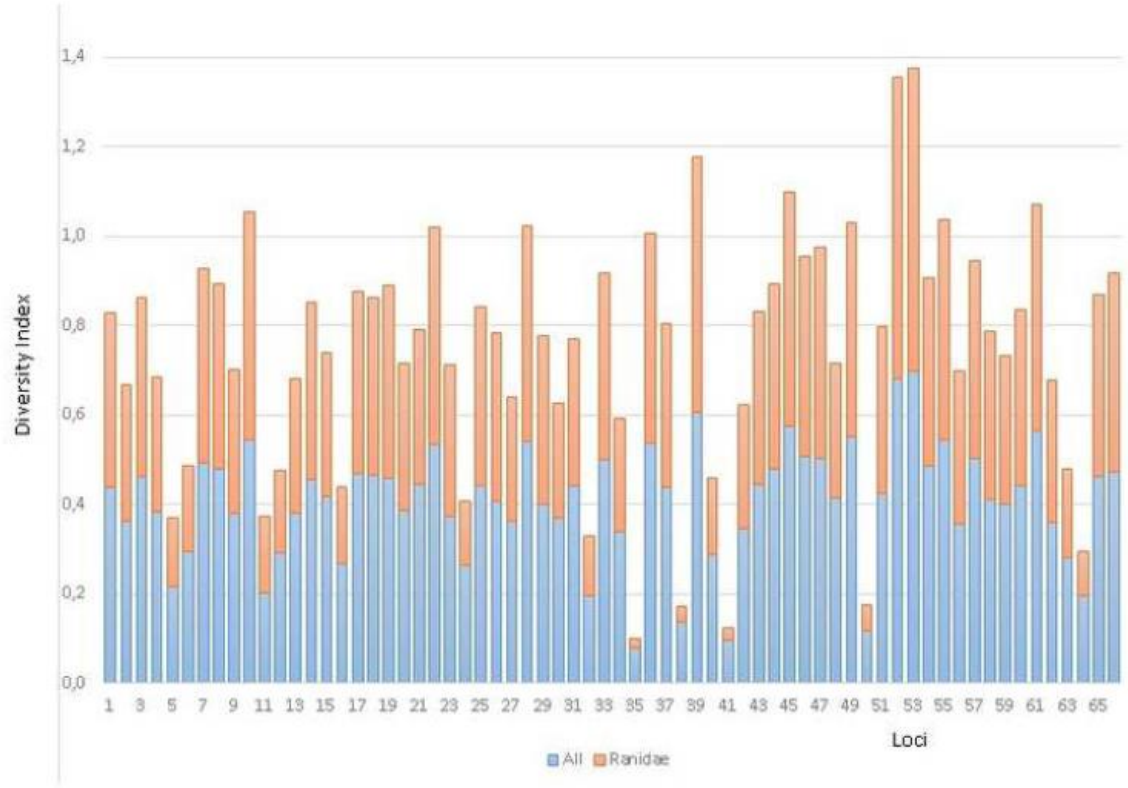

Figure 2. Nucleotide diversity of the SP sequences measured by the inverse of the Simpson index. Cut off value for SNP detection was set at 2 %.

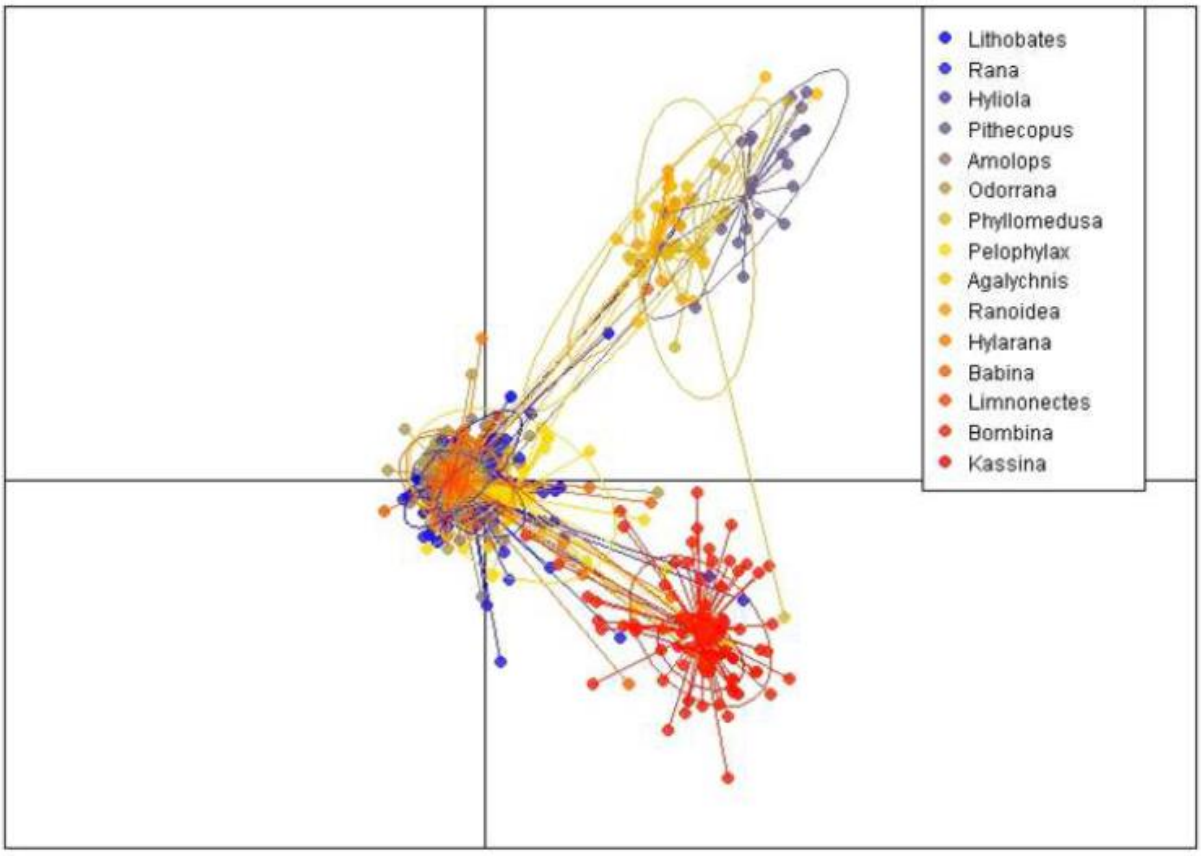

Figure 3. Discriminant Analysis of Principal Components of frog genera. The axes represents the first (horizontal) and second (vertical) discriminant functions.

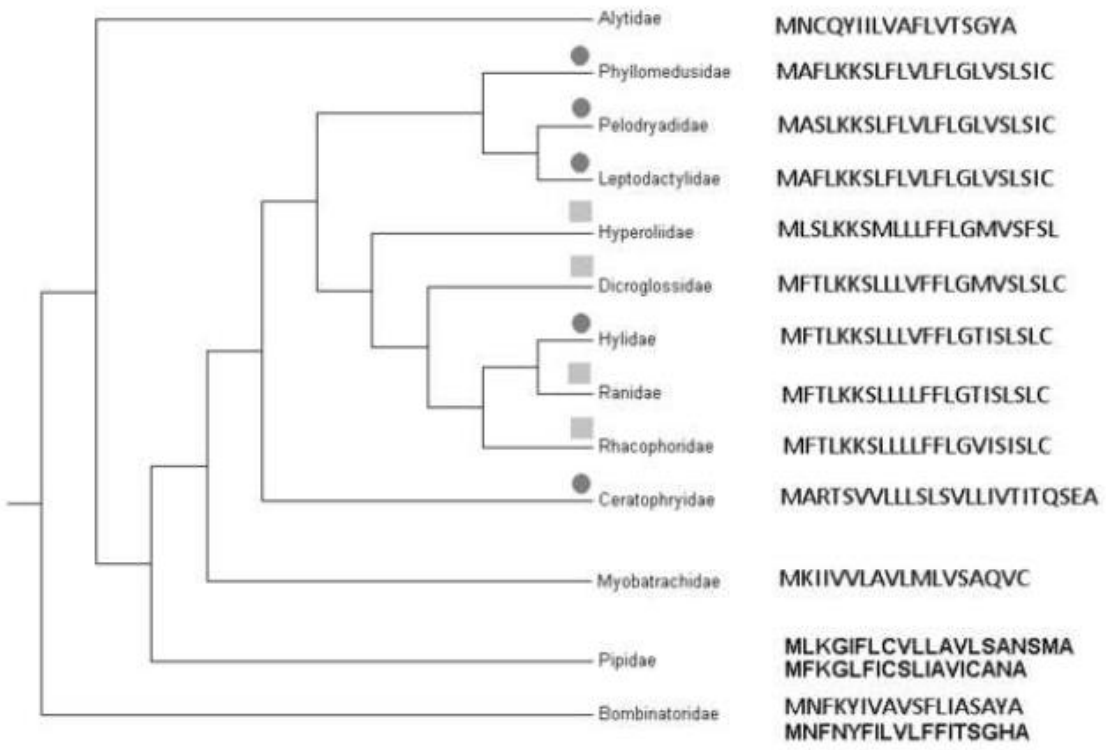

Figure 4. Dendrogram of the signal peptide sequence at the family level. The square indicates families included in the Ranoidea superfamily. The circle indicates families included in Hyloidea superfamily. Representative AMP signal sequences were provided next to the branches.