BMC Genomics

RESEARCH ARTICLE   **Open Access**

Parallel identification of novel antimicrobial peptide sequences from multiple anuran species by targeted DNA sequencing

CrossMark

Tomislav Rončević[1*] 🆔, Marco Gerdol[2], Francesca Spazzali[2], Fiorella Florian[2], Stjepan Mekinić[3], Alessandro Tossi[2] and Alberto Pallavicini[2]

Abstract

Background: Antimicrobial peptides (AMPs) are multifunctional effector molecules that often combine direct antimicrobial activities with signaling or immunomodulatory functions. The skin secretions of anurans contain a variety of such bioactive peptides. The identification of AMPs from frog species often requires sacrificing several specimens to obtain small quantities of crude peptides, followed by activity based fractionation to identify the active principles.

Results: We report an efficient alternative approach to selectively amplify AMP-coding transcripts from very small amounts of tissue samples, based on RNA extraction and cDNA synthesis, followed by PCR amplification and high-throughput sequencing of size-selected amplicons. This protocol exploits the highly conserved signal peptide region of the AMP precursors from Ranidae, Hylidae and Bombinatoridae for the design of family-specific, forward degenerate primers, coupled with a reverse primer targeting the mRNA poly-A tail.

Conclusions: Analysis of the assembled sequencing output allowed to identify more than a hundred full-length mature peptides, mostly from Ranidae species, including several novel potential AMPs for functional characterization. This (i) confirms the effectiveness of the experimental approach and indicates points for protocol optimization to account for particular cases, and (ii) encourages the application of the same methodology to other multigenic AMP families, also from other genera, sharing common features as in anuran AMPs.

Keywords: Antimicrobial peptides, Anura, Innate immunity, Parallel identification, Signal peptide region

Background

AMPs are endogenous antibiotics present in all organisms, with a direct antimicrobial activity towards pathogens, often also showing immunomodulatory properties, and a regulated gene expression to facilitate and modulate immune responses [1, 2]. The skin secretions of many anurans contain a variety of bioactive peptides encoded by multigenic families [3] that often exhibit antibacterial activity towards multidrug resistant microbial isolates [4]. AMPs have been identified in all anuran families of the phylogenetically more ancient suborder of Archaeobatrachia including

Leiopelmatidae, Alytidae, Bombinatoridae and Pipidae families. In Neobatrachia, AMPs have been identified in Dicroglossidae, Hylidae, Hyperoliidae, Leptodactylidae, Myobatrachidae and, in particular, in Ranidae [5]. The latter family consists of wide-ranging frog species distributed worldwide, except for the polar regions, in which 14 different AMP classes have been identified to date based on the molecular characteristics of the peptides they contain. Ranid frogs, in general, are well known to synthesize and secrete multiple active AMPs (at least 22 are reported in the skin secretions of *Rana palustris*) with rare exceptions such as in *Rana sylvatica*, in which only one antimicrobial peptide has been isolated to date [6]. It is worth noting that production of antimicrobial peptides may be influenced by hormonal and/or environmental factors [7–9], which can

* Correspondence: troncevic@pmfst.hr
[1]Department of Physics, Faculty of Science, University of Split, 21000 Split, Croatia
Full list of author information is available at the end of the article

Rončević *et al. BMC Genomics*   (2018) 19:827

Page 2 of 10

hinder the identification of AMPs under certain experimental conditions.

Identifying novel anuran peptides normally requires handling several individuals, which are either sacrificed or held in captivity and treated with electric shocks/ norepinephrine to obtain small amounts of crude peptide. This is followed by several rounds of purification using different precipitation and chromatographic techniques combined with activity testing of fractions to identify the active principles [10]. This approach however raises problems of animal protection and nature preservation. The International Union for Conservation of Nature (IUCN) reports that 1276 amphibian species worldwide are endangered or critically endangered with 38 having become extinct [11]. In this context, the search for a more efficient and less invasive method that requires minimum amounts of biological samples is highly desirable, such as isolation, amplification and sequencing of the nucleotide sequences coding for the AMPs. Although a few alternative approaches based on the screening of available transcriptomic data have been attempted [12], they did not implement the use of degenerate primers designed on the most conserved regions of AMP precursors [13]. This has resulted in the discovery of a limited number of novel AMPs and has never been applied for large-scale multispecies screening.

We have developed a potentially faster, less invasive and more efficient approach based on the selective amplification and subsequent sequencing of transcripts encoding for antimicrobial peptides, starting from very small amounts of tissue. This method however requires accurate primer design to capture the diversity of AMPs. In general, anuran antimicrobial peptide precursors consist of a highly conserved signal sequence, a negatively charged propeptide and a hypervariable cationic mature region [14–16]. Data on anuran signal sequences pertaining to different families and species are available in a dedicated database, DADP [17]. In many cases, the sequences present in this database were validated by biochemical methods, which also confirmed biological activity. Using this information and combining it with a method based on the 3'-RACE (rapid amplification of cDNA ends) protocol [18] we have developed a methodology for simultaneous identification of novel antimicrobial peptide sequences from multiple anuran species. To this purpose, total RNA was extracted from eight different frog species belonging to three anuran families. cDNA libraries were prepared utilizing a reverse primer based on the mRNA poly-A tail and forward degenerate primers designed based on highly conserved signal regions of the peptide precursors. These were used for selective amplification of the target AMP cDNAs, and the resulting amplicons then size-selected and subjected to Ion Torrent long-read high-throughput sequencing. We present data on the effectiveness of this method in identifying AMPs, including several known sequences and a number of novel sequences, some belonging to known classes. We also indicate deficiencies, discuss the most likely causes and indicate how to possibly overcome them.

Methods

Tissue sampling and RNA extraction
One specimen belonging to each of eight different species from Ranidae, Hylidae and Bombinatoridae family (see Additional file 1) was collected in the Croatian wild during March and April 2017. Frogs were sampled in accordance with applicable EU and Croatian legislation governing animal experimentation (Directive 2010/63/EU and NN 55/2013) and necessary permits were obtained from Croatian Ministry of Environmental and Nature Preservation. All animals were sacrificed by exposure to chloroform 24–48 h after capture to ensure a minimal stay in captivity and suffering. Approximately 200 mg of skin tissue was immediately transferred to RNA*later*® buffer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and stored at − 20 °C according to the manufacturers' instructions. Total RNA was extracted using the TRIzol protocol (TRIzol® Reagent, Life Technologies, Carlsbad, California, USA) from ∼ 50 mg of this tissue, resuspended in RNAase free water, quantified with NanoDrop 2000 (Thermo Fisher Scientific, Waltham, Massachusetts, USA), quality checked using denaturing 1.5% agarose gel electrophoresis, and stored at − 80 °C until further use (see Fig. 1).

Transcriptome assembly and screening
The available RNAseq data in Sequence Read Archive database [19] was retrieved for 16 anuran species belonging to 3 different families, and assembled with the Trinity 2.4.0 software [20] (see Additional file 2) using default parameters and setting a minimal contig length to 200 nucleotides. Signal sequences pertaining to Class-1 (Ranidae and Hylidae) and Class-3 (Bombinatoridae) families were obtained from DADP database taking into account only peptides with reported bioactivity data [17] (see Additional file 3). The protein sequences from DADP were aligned using Muscle [21] and used to generate Hidden Markov Model profiles with the HMMER 3.1b1 hmmbuild module [22]. Trinity assembled transcripts were translated to all six possible reading frames with EMBOSS transeq [23, 24] and then screened for significant matches using the HMMER 3.1b1 hmmsearch module with an E-value cut off < 0.05. Open reading frames (ORFs) encoding

Fig. 1 Schematic representation of the targeted DNA sequencing method
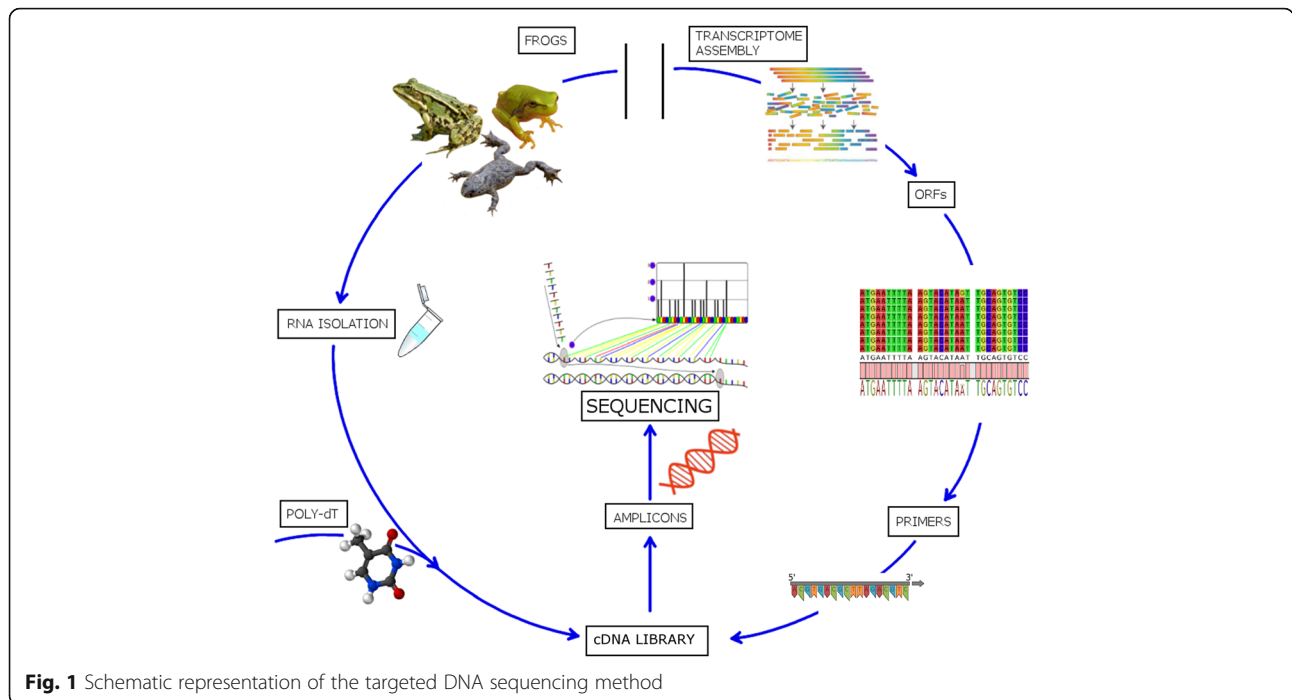
peptides corresponding to positive hits were extracted from each transcriptome with CLC Genomics Workbench 10.1.1 (Qiagen, Hilden, Germany). Incomplete ORFs were also retained.

Primer design

The nucleotide sequences from Ranidae and Hylidae AMP transcripts obtained as described above were separately aligned and only the regions encoding for the signal peptides were kept for further analysis. Redundancies were removed and the remaining sequences were clustered by similarity with the cd-hit software, using a 0.8 identity threshold (i.e. all signal peptide sequences sharing > 50% sequence identity at the nucleotide level were clustered together) [25]. The

same procedure was used for the Bombinatoridae family, but in this case longer highly conserved regions were obtained comprising the propeptide region. The resulting alignments of sequence clusters were used for forward primer design (see Additional file 4). Briefly, the position of forward primers was selected based on the identification of well-conserved 20 nucleotide-long sequence stretches containing a maximum of 3 polymorphic positions, where degenerate nucleotides were inserted (see Table 1). Due to the short length of the signal peptide region of Ranidae and Hylidae AMPs (about 60 nucleotides), the positioning of the forward degenerate primer was not expected to have a significant effect on amplicon size and therefore the maximum allowance of 3

Table 1 List of primers used for cDNA synthesis and amplification

Target family	Primer	Primer sequence (5′-3′)
Forward primers		
Ranidae[a]	TP1	CAGGACCAGGGTACGGTG**ATGTTCACCTTGAAGRAATC**
Ranidae[a]	TP2	CAGGACCAGGGTACGGTG**TTGGGATYGTCTCCTCATCT**
Ranidae[a]	TP3	CAGGACCAGGGTACGGTG**ATGTTCACCWTGARKAAAYC**
Hylidae[b]	TP4	CAGGACCAGGGTACGGTG**AAGAARTCWCTTYTTCTTGT**
Bombinatoridae[c]	TP5	CAGGACCAGGGTACGGTG**TAGAAGAAGAATCACTGAGG**
Outer PCR *Forward*	A-Uni1	CCATCTCATCCCTGCGTGTCTCCGACTCAG(X)$_{10}$CAGGACCAGGGTACGGTG
PCR *Reverse*	trP1	CCTCTCTATGGGCAGTCGGTGAT

AMP specific sequences are in bold
[a]Species pertaining to Ranidae family: *Pelophylax ridibundus, Pelophylax* kl. *esculentus, Rana dalmatina, Rana arvalis* and *Rana temporaria*
[b]Species pertaining to Hylidae family: *Hyla arborea*
[c]Species pertaining to Bombinatoridae family: *Bombina variegata* and *Bombina bombina*

polymorphic positions was aimed at minimizing the chances of non-specific PCR product amplification. In the case of Bombinatoridae, due to the different organization of AMP precursors, the forward primer was designed as close as possible to the 3'end of the propeptide-encoding region, maintaining the maximum limit of three degenerate nucleotides.

The reverse primer (5'-CCTCTCTATGGGCAGTCGG TGATTTTTTTTTTTTTTTTTTTT-3′) contained a poly-dT stretch to match the poly-A tail of the mRNA and was used for cDNA synthesis. It also contained a 5′ tail sequence (trP1) which was required for the cDNA amplification protocol and subsequent parallel sequencing. All forward primers (see Table 1) were also synthesized with a 5'-CAGGACCAGGGTACGGTG-3′ tail required for multiplex sequencing through attachment of the barcodes in a secondary outer amplification. All primers were synthesized by BMR Genomics (Padova, Italy) (see Fig. 1).

Library construction and sequencing
First strand cDNA was synthesized from 1 μg of total RNA using the qScript™ Flex cDNA Synthesis Kit (Quanta Biosciences, Gaithersburg, Maryland, USA) according to the manufacturers' instructions. The reverse transcription reaction was performed at 37 °C for 1 h. The mixture for cDNA amplification contained 0.2 μl of DNA polymerase (5 U/μl), 2.5 μl of 25 mM $MgCl_2$, 2.5 μl of 10 × buffer A (KAPA Taq PCR Kit, Kapa Biosystems, Wilmington, Massachusetts, USA) together with 0.5 μl of 10 μM dNTP solution, 0.5 μl of 10 μM specific forward primer solution, 0.5 μl of 10 μM trP1 primer solution and 1 μl of cDNA template in a total volume of 20 μl. The PCR started with an initial denaturation at 95 °C for 2 min, followed by 10 cycles including 10 s at 95 °C, annealing at 45–50 °C for 20 s (ramping 0.5 °C/cycle) and 20 s elongation at 72 ° C. For the next 25 cycles the annealing temperature was set to 52 °C for 20 s ending with 5 min final elongation at 72 °C (MJ Research PTC-200 Gradient Thermal Cycler, Marshall Scientific, Hampton, New Hampshire, USA).

Outer PCR was performed to attach the barcodes (a sample-specific 10 nucleotide sequence used for de-multiplexing) on the 5' end of the amplicon, which were followed by sequencing adapters. The mixture for outer PCR contained 0.2 μl of DNA polymerase (5 U/μl), 2.5 μl of 10× buffer A (KAPA Taq PCR Kit, Kapa Biosystems, USA) together with 1 μl of 20× EvaGreen™ (Biotium, Fremont, California, USA), 0.5 μl of 10 μM dNTP solution, ~ 20 ng of nucleic acid from the primary PCR and 3 μl of 10 μM primer solution in a total volume of 20 μl. This secondary PCR run was performed for 8 cycles including denaturation at 95 °C for 10 s, annealing at 60 °C

for 10 s and 40 s elongation at 65 °C with 3 min final elongation at 72 °C. The quality of the amplification products was visualized by electrophoresis on 1.5% agarose gel after each amplification run. Based on this analysis, some amplicons were discarded due either to an unsuccessful PCR or out of range size.

The size range and quantity of nucleic acid in each individual library were assessed using a DNA 1000 kit on an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, California, USA) (see Additional file 5). Prior to sequencing, all suitable amplicons were pooled together at equimolar quantities, then purified with E-Gel® SizeSelect™ (Invitrogen, Carlsbad, California, USA) and quantified with a Qubit 2.0 fluorimeter (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Amplicon libraries were concentrated with Omega Cycle Pure Kit (VWR International, Radnor, Pennsylvania, USA) according to the manufacturers' instructions and the subsequent DNA quantification was performed with a Qubit dsDNA Assay Kit (Molecular Probes, Eugene, Oregon, USA) on a Qubit 2.0 fluorimeter (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Sequencing was carried out on an Ion Torrent PGM™ sequencing platform (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and the library prepared using Ion PGM Hi-Q View OT2 Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) according to the manufacturers' instructions. The library was loaded on an Ion 314 chip (Life Technologies, Carlsbad, California, USA), and sequenced for 800 cycles using an Ion PGM Hi-Q View Sequencing Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Post-sequencing analysis
Raw sequencing data from PGM runs were imported into CLC Genomics Workbench 10.1.1 (Qiagen, Hilden, Germany) to perform trimming. Briefly, low quality (trimming limit = 0.05) and ambiguous nucleotides, adapters and short residual reads (< 100 nucleotides) were removed. Filtered reads were then assembled into contigs with an overlap-layout-consensus (OLC) approach. The original reads were re-mapped on the assembled contigs to allow visual inspection of the correctness of the assembly of each contig. Sequences supported by less than 3 reads were discarded without further analysis and in some cases contigs displaying a high amount of polymorphism were re-assembled with more stringent parameters to obtain all the possible sequence variants. Transcripts were blasted (using BLASTx) against a custom sequence database containing all known Class-1 (Ranidae and Hylidae) or Class-3 (Bombinatoridae) nucleotide sequences to detect all positive hits based on E-value threshold of 0.05. The longest ORFs for each of the resulting contigs were translated into amino acid sequences using the ExPASy

Rončević *et al. BMC Genomics* (2018) 19:827

Page 5 of 10

translate tool [26] and grouped together based on common features.

Peptide characterization

Two peptides with amidated C-terminus were obtained from GenicBio Ltd. (Shanghai, China) at > 98% purity as confirmed by RP-HPLC and MS. They were dissolved in doubly distilled water and stock concentration determined as described previously [27, 28]. Minimal inhibitory concentration (MIC) was determined on a Gram-negative and a Gram positive reference laboratory strains, *Escherichia coli* ATCC 25922 and *Staphylococcus aureus* ATCC 29213, obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA), using the serial two-fold microdilution method as described previously [27]. Cytotoxic effects on metabolic activity were determined on human monocytes isolated from buffy coats of informed donors (in accordance with the ethical guidelines and approved from the ethical committee of the University of Trieste) [29], using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay after treatment with the peptides for 24 h, as described previously [30].

Results

Novel peptide sequences

Selective amplification with AMP specific primers of cDNA libraries obtained from seven anuran species (Ranidae, Hylidae, Bombinatoridae families), followed by long-read high-throughput sequencing, resulted in 120.371 reads. Out of the assembled contigs, 5–15% of the sequences from the Ranidae family corresponded to AMPs, depending on the species. The success rate for *Hyla arborea* (1 contig, 0.2%) was much lower (see Table 2). *Bombina variegata* gave no contigs corresponding to AMPs, while *Bombina bombina* amplicons were not pooled with the others for sequencing, due to their excessive average size (573 nucleotides), as evaluated by Agilent Bioanalyzer 2100. This was suggested by the DNA fragment size restrictions (< 450) of the Ion Torrent™ PGM sequencer that was used.

Overall, this approach permitted to identify 128 likely AMP sequences. One hundred twenty seven of these were identified in the five species from the Ranidae family, and only one in *Hyla arborea*, from the Hylidae family (see Table 2). Before proceeding further with the analysis, we extracted the putative mature peptide region by removing the sequence N-terminal to the basic Lys-Arg propeptide cleavage site. The predicted peptides were then checked for identity with known sequences using BLASTp [31] and manually grouped into eight different classes. Seven of these classes were grouped based on the predicted secondary structure (e.g. α-helix, β-sheet), length, frequency of specific

Table 2 Number of raw and trimmed reads, assembled contigs and putative antimicrobial peptides classified per species and the primer of choice

Species	Forward primer	Raw reads	Trimmed reads	Contigs	AMPs (on target)
Pelophylax kl. *esculentus*	TP1	10.278	7.674	399	22 (5%)
	TP2	6.995	3.786		
	TP3	18.067	13.321		
Pelophylax ridibundus	TP1	5.401	4.192	196	31 (15%)
	TP2	3.670	2.446		
	TP3	5.132	3.924		
Rana arvalis	TP1	8.292	6.294	391	31 (8%)
	TP2	12.498	6.272		
	TP3	1.505	1.101		
Rana dalmatina	TP1	4.404	2.697	275	20 (7%)
	TP3	17.884	9.748		
Rana temporaria	TP1	6.751	4.164	187	23 (12%)
Hyla arborea	TP4	11.161	4.690	448	1 (0.2%)
Bombina variegata	TP5	8.333	4.807	577	0

amino acids (e.g. presence of a characteristic rana-box domain) [32] and identity with known AMP classes. The eight class consisted of peptides which could not be classified into any of the other seven groups (see Additional file 6). About half of the peptides had already been identified previously in the same, or in closely related, species (48% with 100% identity), or shared significant similarity with previously described AMPs (27% with 81–99% identity). Thirteen of the peptides (10% of the total) showed significant BLAST e-values but < 80% identity with AMPs deposited in the protein sequences databases (see Table 3 and Additional file 7). Additionally, we could identify 16 entirely novel peptides (12%), here defined as peptides lacking significant similarity with known AMPs, but that may have antimicrobial activities based on their physico-chemical characteristics (charge, overall hydrophobicity) and the conserved secretion signal sequence (see Table 3 and Additional file 7). Four additional peptides (3%) with 100% identity to known peptide sequences were also identified. However, with less than 70% query cover and additional amino acids in their primary structure those were also categorized as novel (see Table 3 and Additional file 7). It is worth noting that some of these peptides were found to be identical (at the amino acid level) in multiple analyzed species, thereby reducing the number of completely unique novel peptide sequences to 14, and to 11 for peptides with less than 80% identity. Altogether, these results validate the reliability of this approach, confirming the effectiveness of the method that has been proven to be fast, accurate and suitable for simultaneous identification of

Table 3 Identified putative AMP sequences with no significant identity to the known proteins, less than 80% identity to known AMP sequences or 100% identity to proteins stored at NCBI database with < 70% query cover and their physico-chemical characteristics

#	Sequence	Species	Charge	H[a]
No identity to proteins stored at NCBI nr database				
1	FLGALGNALSRVLGK	R. temporaria	+3	0.813
2	FIGALVNALTRVLGK	R. temporaria	+3	1.213
3	FIGALVHALTGILGK	R. temporaria	+ 2	2.247
4	LVPFIGRTLGGLLARFGK	R. temporaria	+ 4	1.500
5	VPQLCFKFQKVIYCEINRTLPNEA	R. dalmatina	+ 2	−0.625
6	FSQLFFAWLLRLCRQ	P. kl. esculentus	+3	2.587
7	GIVEAWPLR	P. ridibundus, R. dalmatina	+ 1	1.133
8	NNLRHIVAWCKNRNYSLAVCARFKPQ	R. temporaria	+ 6	−1.612
9	NLLGFLQGAKDILKECEADNYQGWLCESYKPQ	R. dalmatina	−1	−1.250
10	FLPLVLGKTHSEQAEILSWKSSNVEYHLPKCTTDV	P. ridibundus, R. dalmatina	0	−0.763
11	FLPLIAGLWVNCSANNPKMLKLWK	P. kl. esculentus	+ 4	1.363
12	FLPICDKSALRFVGKV	P. ridibundus	+3	0.494
13	EMPMKKKEETIQKKGMLKWKTIFTSHCWSFE	P. ridibundus	+ 4	−1.835
14	RGLLDPITGLVGGLLR	H. arborea	+ 2	1.213
< 80% identity to proteins stored at NCBI database				
15	FLGFVGQALNALLGKLGK	R. dalmatina	+3	1.550
16	FLPAIAGILSQIFGK	P. kl. esculentus	+ 2	2.540
17	FFPAFLKVAAKVVPSIICSITKNVET	P. kl. esculentus	+3	0.573
18	IVPILLGVVPQLVCAITKKC	R. dalmatina, R. temporaria	+3	1.675
19	IIPLLLGKVVCAITKKC	R. dalmatina	+ 4	1.271
20	LVPMFLSKLICFITKKC	R. temporaria	+ 4	1.918
21	GLEVLGKILSGILGK	R. dalmatina, Rana arvalis	+ 2	1.220
22	LLGAALSALSSVIPSVISWFQKG	Rana arvalis	+ 2	1.670
23	LANRAARNTSQNVLNAITCTL	R. dalmatina	+3	−1.662
24	ADFLDKLRNFAAKNLQNKASL	P. ridibundus	+3	−1.595
25	EMLRKKEETIQKKGMLKWKNDFYQSLLEF	P. ridibundus	+3	−1.779
100% identity to proteins stored at NCBI database with < 70% query cover				
26	QKTYNRRPPGWSLYVFHQQISNLELEVI	P. kl. esculentus	+ 2	−0.654
27	FVPLLVSKLVCVVTKNVRIWKLELEII	Rana arvalis	+3	1.663
28	FVPLLVSKLVCVVTKNVRTLET	Rana arvalis	+3	0.455
29	FLPIVTNLLLRFVG	R. dalmatina	+ 2	3.729

[a]Calculated using the CCS consensus hydrophobicity scale [50]

large numbers of AMP sequences from Ranidae. Some potential limitations are discussed in the following section.

Preliminary biological characterization

Two of the novel identified peptides were synthesized (see Table 4) and tested for their in vitro activity against a Gram-negative and a Gram-positive reference bacterial strains. Results are promising, with both peptides active against *S. aureus*, especially rarv_10.1_19 with MIC of 4 μM. On the other hand, peptides don't seem to be selective for Gram-negative strains (see Table 4) with MIC > 64 μM. Another encouraging

aspect was the very low toxicity towards human circulating blood cells, namely monocytes. At the highest concentration used (100 μM) over 80% of cells were fully viable (see Table 4).

Discussion

Selective amplicon sequencing resulted in the simultaneous identification of antimicrobial peptide encoding transcripts from 5 out of 7 different anuran species tested. All the species with a positive result pertain to the Ranidae family, with a single sequence obtained from *Hyla arborea* (Hylidae) and none from *Bombina*

Table 4 Peptides tested for biological activity and their physico-chemical characteristics

Peptide[a]	Coding sequence	Sequence[b]	MW	Charge	MIC (μM)[c]		%Viability[d] (100 μM)
					E. coli	S. aureus	Monocytes
#4	rtemp_4.3_210	LVPFIGRTLGGLLARF-NH2	1729.1	+ 4	> 64	4	> 80
#22	rarv_10.1_19	LLGAALSALSSVIPSVISWFQK-NH$_2$	2286.7	+3	> 64	16	> 80

[a]Peptide number refers to designation in Table 3
[b]Sequences end with a C-terminal amidation signal [51, 52]
[c]Evaluated using microdilution assay in MH medium
[d]Viability assessed in the presence of 100 μM peptide after 24 h exposure by the MTT assay

variegata (Bombinatoridae) (see Table 2). The different success rate of this approach among three anuran families can be explained by several factors. First, the assembled transcriptome data used for primer design comprised several species pertaining to either the Rana or Pelophylax genus (Ranidae), whereas limited data with satisfactory phylogenetic relations was available for Hylidae. The high number of AMP transcripts available in the public databases for the two Ranidae genera enabled to construct three primers suitable for efficiently amplifying the different expressed AMP transcripts in the five target species, improving the chances of correct annealing and amplification in PCR. As the nucleotide sequence data available for other genera increases in the databases, it should become possible to refine primer design and provide suitable primer options also for other more distantly related families.

Another important parameter is transcript size. All known Class-1 Ranidae peptides are shorter than 100 AA long (mostly 70–80). Furthermore, our transcriptome screening revealed that the 3' UTR region of the encoding transcripts was generally shorter than 200 nucleotides. The Ion Torrent™ PGM (Life Technologies, Carlsbad, California, USA) sequencing kit used is suitable for sequencing DNA strands up to about 450 base pairs. The design of degenerate primers based on the Class-1 signal peptide region should therefore have generated amplicons with a size range compatible with this sequencing method. This was confirmed using an Agilent Bioanalyzer 2100 prior to sequencing, which revealed a library size of between 300 and 400 bp for all the Ranidae species (see Additional file 5). This being said, libraries at the top of this size range, at around 400 bp, were very close to the maximal permissible size range, which could create problems. Amplicons successfully obtained from the Bombina species were excluded from sequencing for this reason, as their average size was > 500 nucleotides. Using the new isothermal amplification for Ion Torrent, or a different sequencing platform, may partly counteract this issue.

A third consideration comes from a detailed analysis of the data from Ranidae species. We obtained 22 unique assembled sequences in P. kl. esculentus, 31 in P. ridibundus, 31 in R. arvalis, 20 in R. dalmatina and 23 in R. temporaria. These results highlight the considerable sequence diversity of Class-1 AMPs in Ranidae. Identification of such a high number of variants within a single specimen underlines the fact that these AMPs are encoded by multigenic families [3]. Although the high conservation of the signal peptide region allowed targeted DNA sequencing, this feature can represent a potential obstacle in whole transcriptome sequencing. Indeed, the inefficient assembly of full length transcripts, or the collapse of similar variants within a single contig are well-known issues linked to the use of short reads in the assembly of highly similar transcripts derived from multigenic families [33, 34]. Therefore, we suggest the use of longer reads. In this respect, those obtained by Ion Torrent may represent an optimal balance between high-throughput and reasonable length (up to 450 bp) for the management of this sequence diversity. For the same reason, the use of three different primers is another key factor for successful amplification. This guarantees an efficient pairing with all the possible sequence variants. Indeed, even within the Ranidae family, we noticed substantial differences in the efficiency of the amplification using the 3 primers across species (see Additional file 8).

Considering that only one Class-1 AMP encoding transcript was obtained from H. arborea (Hylidae family), the explanation could thus be a combination of i) the poor representation of transcriptomic datasets from Hyla spp. in the SRA database; ii) the size range of the library, which was close to the maximum capabilities of Ion Torrent sequencing (see Additional file 5), and iii) a different AMP gene organization for this family. With respect to the first consideration, 4 out of 5 transcriptomes used for primer design pertained to a species of a different subfamily (Pelodryadinae) then Hyla (Hylinae) and the only transcriptome available for Hyla arborea was not obtained from AMP-producing tissues. It thus seems likely the designed forward primer did not include all the polymorphisms present in AMPs from the Hyla genus, resulting in the amplification of a single but highly represented sequence (22% of the total sequencing output). Concerning the second consideration, it is likely that AMP amplicons had been removed during the E-gel purification procedure,

Rončević *et al. BMC Genomics*   (2018) 19:827

Page 8 of 10

due to excessive size. The third consideration could be relevant if, unlike Ranidae family peptides, *Hyla* Class-1 AMPs do not have a multi-gene organization, even though this seems to be disproved by previous reports [3]. Based on these observations a similar approach should be undertaken in the future with an improved primer design, based on broader taxonomical sampling, specifically including other Hylinae species.

Unsuccessful results with Bombinatoridae are most certainly linked to the longer AMP precursor and, consequently, longer length of the encoding mRNA molecules. During the initial phases of the experimental design, we tried to overcome this issue by designing more internal primers based on the propeptide rather than on the signal peptide region, thereby reducing the size of the expected amplicons. However, the assessment of the library size range indicated that amplicons were above (*B. bombina*) or very close (*B. variegata*) to maximal input length capabilities of the sequencing technologies used. While the former library was discarded altogether, the latter one was subjected to sequencing but did not produce any positive matches. Despite a similar apparent size range between the libraries obtained from *B. variegata* and some of the longer ones obtained from Ranidae species, the concentration of the former was approximately 7 times lower (see Additional file 5). The most likely cause of unsuccessful sequencing in this species is therefore the removal of AMP amplicons during the E-gel purification procedure due to their borderline size, similarly to *H. arborea*, so that only non-specific amplicons were sequenced. To confirm this hypothesis we carried out purification of a single 573 bp band of *B. bombina* amplification visible on the agarose gel, followed by sequencing on a Sanger ABI 3130 sequencer (Thermo Fischer Scientific, Waltham, Massachusetts, USA). Although the chromatogram was not clean, suggesting that multiple products of the same size were amplified, the consensus sequence clearly confirmed the amplification of a Class-3 AMP precursor. Therefore, while our strategy was not suitable with Bombinatoridae for Ion Torrent PGM, other massive parallel sequencing platforms allowing higher read lengths (such as SMRT PacBio or Oxford Nanopore) could enable its application also in this anuran family.

Overall, the positive results obtained with Ranidae species, with the identification of 127 peptides, including several novel AMPs (i.e. lacking significant sequence similarity with previously characterized anuran sequences) confirm the effectiveness of this experimental design, as long as the degenerate primers are properly designed, and the amplicon size is tailored to the sequencing platform used. Geographical location would have been very important if the selected species were endemic to Croatia.

However, in this case all the species targeted display a relatively broad and partially overlapping area of distribution across Europe and are, in some cases, evolutionarily closely related (e.g. the latest common ancestor of *R. dalmatina*, *R. temporaria* and *R. arvalis* lived in the Miocene [35]). Consequently, the expansion of the taxonomical breadth of sampling to other species adapted to different geographical locations, environmental niches and thereby evolving under different microbial contexts might represent a reliable strategy for novel anuran AMP discovery.

The single result obtained from *H. arborea* indicates that the panel of species analyzed can be quite wide, but this requires a particular effort in collecting as many sequences as possible from species phylogenetically closely related to the target species in the analysis panel to optimize primer design, thereby maximizing the chances of annealing during selective amplification. In this respect, PCR experimental conditions, and the annealing temperature in particular, need to be carefully selected to allow pairing with templates that are not perfectly matching, and allow capturing of as many sequence variants as possible. The lack of success in obtaining Bombinatoridae AMPs instead pinpoints the importance of tailoring the sequencing platform to the expected amplicon length, or alternatively, to identify useable sequences and design primers as close as possible to 3′ end of the mRNAs in order to reduce the amplicon size.

Although the novel identified peptides display physico-chemical features compatible with antimicrobial activity (see Table 3) and clearly possess a well-conserved signal peptide/propeptide region typically found in anuran AMPs, their biological role requires confirmation. A comprehensive evaluation of the antimicrobial activity of six selected novel peptides identified in this study is currently in progress, and preliminary results are presented for two of these peptides (see Table 4). This confirms the reliability of our approach to identify novel, functional antimicrobial peptide sequences (manuscript in preparation).

Conclusions

The approach here presented, with suitable modifications, can be applied also to other gene families sharing a conserved signal peptide and/or propeptide region, an hypervariable mature peptide region, and a limited distance between this region and the mRNA poly(A) tail. These characteristics are well known in many different animal AMPs [36–41], toxins [42–44] and other types of bioactive peptides from other organisms [45]. One should however always keep in mind a key factor, i.e. the detection of sequence variants depends on their being expressed. Therefore, whenever possible, the most appropriate tissue and/or

Rončević et al. BMC Genomics   (2018) 19:827

Page 9 of 10

experimental challenge need to be selected to enhance the expression of the target mRNAs. In our case, the choice of anuran skin was amply supported by abundant literature [5, 46–49], and indeed we could obtain several dozens of different peptides for each species as expected. However, the number of reads obtained for each sequence variant does not necessarily depend only on the level of expression of the transcript itself but is also affected by the efficiency of the amplification, which depends on the match to the primer. For this reason, the results of this type of study can only be considered as qualitative, and not as a proxy to investigate the expression levels of AMP variants. Finally, the small amount of tissue required may permit the identification of novel AMPs from endangered species, potentially without the need for sacrificing any individual. This would permit to fully exploit animal biodiversity in identification of potential novel therapeutic agents, without adding to the threat of reducing it.

Additional files

Additional file 1: Complete list of frog species obtained from Croatian wild. (DOCX 13 kb)

Additional file 2: List of transcriptome data downloaded from SRA database. (DOCX 14 kb)

Additional file 3: List of signal peptides used for transcriptome screening. (XLSX 46 kb)

Additional file 4: Clusters of nucleotide alignments used for forward primer design. (DOCX 7338 kb)

Additional file 5: Size range of each individual library obtained prior to pooling. (DOCX 14 kb)

Additional file 6: Classification of identified peptides. (DOCX 1600 kb)

Additional file 7: BLASTp output for novel identified putative AMP sequences. (XLSX 47 kb)

Additional file 8: Success rate of amplicon synthesis in Ranidae species based on used primer. (DOCX 13 kb)

Abbreviations
AMP: Antimicrobial peptide; OLC: Overlap layout consensus; ORF: Open reading frame; RACE: Rapid amplification of cDNA ends

Availability of data and materials
The datasets generated and/or analyzed during the current study are available in the NCBI SRA database under BioProject accession ID: PRJNA415374.

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

Author details
[1]Department of Physics, Faculty of Science, University of Split, 21000 Split, Croatia. [2]Department of Life Sciences, University of Trieste, 34127 Trieste, Italy. [3]Public Institution for the Management of Protected Areas in the County of Split and Dalmatia – "Sea and karst", 21000 Split, Croatia.

References
1. Cederlund A, Gudmundsson GH, Agerberth B. Antimicrobial peptides important in innate immunity. FEBS J. 2011;278:3942–51.
2. Nijnik A, Hancock R. Host defence peptides: antimicrobial and immunomodulatory activity and potential applications for tackling antibiotic-resistant infections. Emerg Health Threats J. 2009;2. https://doi.org/10.3134/ehtj.09.001.
3. König E, Bininda-Emonds ORP. Evidence for convergent evolution in the antimicrobial peptide system in anuran amphibians. Peptides. 2011;32:20–5.
4. Conlon JM. The potential of frog skin antimicrobial peptides for development into therapeutically valuable anti-infective agents. In: Rajasekaran K, Cary JW, Jaynes JM, Montesinos E, editors. Small wonders: peptides for disease control. Washington, DC: American Chemical Society; 2012. p. 47–60. https://doi.org/10.1021/bk-2012-1095.ch003.
5. Conlon JM. The contribution of skin antimicrobial peptides to the system of innate immunity in anurans. Cell Tissue Res. 2011;343:201–12.
6. Conlon JM, Kolodziejek J, Nowotny N. Antimicrobial peptides from ranid frogs: taxonomic and phylogenetic markers and a potential source of new therapeutic agents. Biochim Biophys Acta BBA Proteins Proteomics. 2004; 1696:1–14.
7. Ohnuma A, Conlon JM, Kawasaki H, Iwamuro S. Developmental and triiodothyronine-induced expression of genes encoding preprotemporins in the skin of Tago's brown frog Rana tagoi. Gen Comp Endocrinol. 2006;146:242–50.
8. Matutte B, Storey KB, Knoop FC, Conlon JM. Induction of synthesis of an antimicrobial peptide in the skin of the freeze-tolerant frog, Rana sylvatica, in response to environmental stimuli. FEBS Lett. 2000;483:135–8.
9. Davidson C, Benard MF, Shaffer HB, Parker JM, O'Leary C, Conlon JM, et al. Effects of chytrid and carbaryl exposure on survival, growth and skin peptide defenses in foothill yellow-legged frogs. Environ Sci Technol. 2007;41:1771–6.
10. Giuliani A, Rinaldi AC, editors. Antimicrobial Peptides. Totowa: Humana Press; 2010. https://doi.org/10.1007/978-1-60761-594-1. Accessed 12 Dec 2016

11. The IUCN Red list of threatened species. 2017. http://www.iucnredlist.org/. Accessed 14 Dec 2017.

12. Reshmy V, Preeji V, Parvin A, Santhoshkumar K, George S. Three novel antimicrobial peptides from the skin of the Indian bronzed frog Hylarana temporalis (Anura: Ranidae). J Pept Sci. 2011;17:342–7.

13. Dong Z, Luo W, Zhong H, Wang M, Song Y, Deng S, et al. Molecular cloning and characterization of antimicrobial peptides from skin of Hylarana guentheri. Acta Biochim Biophys Sin. 2017;49:450–7.

14. Nacif-Marçal L, Pereira GR, Abranches MV, Costa NCS, Cardoso SA, Honda ER, et al. Identification and characterization of an antimicrobial peptide of *Hypsiboas semilineatus* (Spix, 1824) (Amphibia, Hylidae). Toxicon. 2015; 99(Supplement C):16–22.

15. Nicolas P, Vanhoye D, Amiche M. Molecular strategies in biological evolution of antimicrobial peptides. Peptides. 2003;24:1669–80.

16. Vanhoye D, Bruston F, Nicolas P, Amiche M. Antimicrobial peptides from hylid and ranin frogs originated from a 150-million-year-old ancestral precursor with a conserved signal peptide but a hypermutable antimicrobial domain. Eur J Biochem. 2003;270:2068–81.

17. Novkovic M, Simunic J, Bojovic V, Tossi A, Juretic D. DADP: the database of anuran defense peptides. Bioinformatics. 2012;28:1406–7.

18. Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A. 1988;85:8998–9002.

19. Kodama Y, Shumway M, Leinonen R. The sequence read archive: explosive growth of sequencing data. Nucleic Acids Res. 2012; 40(Database issue):D54–6.

20. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011;29:644–52.

21. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

22. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. Nucleic Acids Res. 2015;43:W30–8.

23. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic Acids Res. 2010;38(Web Server issue):W695–9.

24. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000;16:276–7.

25. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

26. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31:3784–8.

27. Rončević T, Vukičević D, Ilić N, Krce L, Gajski G, Tonkić M, et al. Antibacterial activity affected by the conformational flexibility in glycine–lysine based α-helical antimicrobial peptides. J Med Chem. 2018;61:2924–36.

28. Kuipers BJH, Gruppen H. Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography −mass spectrometry analysis. J Agric Food Chem. 2007;55:5445–51.

29. Pacor S, Grillo A, Đorđević L, Zorzet S, Lucafò M, Da Ros T, et al. Effects of two fullerene derivatives on monocytes and macrophages. Biomed Res Int. 2015;2015:1–13.

30. Pelillo C, Benincasa M, Scocchi M, Gennaro R, Tossi A, Pacor S. Cellular internalization and cytotoxicity of the antimicrobial proline-rich peptide Bac7(1-35) in monocytes/macrophages, and its activity against phagocytosed Salmonella typhimurium. Protein Pept Lett. 2014;21:382–90.

31. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.

32. Park S, Park S-H, Ahn H-C, Kim S, Kim SS, Lee BJ, et al. Structural study of novel antimicrobial peptides, nigrocins, isolated from Rana nigromaculata. FEBS Lett. 2001;507:95–100.

33. Phuong MA, Mahardika GN, Alfaro ME. Dietary breadth is positively correlated with venom complexity in cone snails. BMC Genomics. 2016;17:401.

34. Robinson SD, Safavi-Hemami H, McIntosh LD, Purcell AW, Norton RS, Papenfuss AT. Diversity of Conotoxin gene Superfamilies in the venomous snail, Conus victoriae. PLOS ONE. 2014;9:e87648.

35. Voituron Y, Barré H, Ramløv H, Douady CJ. Freeze tolerance evolution among anurans: frequency and timing of appearance. Cryobiology. 2009;58:241–7.

36. Gerdol M, Puillandre N, De Moro G, Guarnaccia C, Lucafò M, Benincasa M, et al. Identification and characterization of a novel family of cysteine-rich peptides (MgCRP-I) from Mytilus galloprovincialis. Genome Biol Evol. 2015;7:2203–19.

37. Kim M, Jeon J-M, Oh C-W, Kim YM, Lee DS, Kang C-K, et al. Molecular characterization of three crustin genes in the morotoge shrimp, Pandalopsis japonica. Comp Biochem Physiol B Biochem Mol Biol. 2012;163:161–71.

38. Leoni G, De Poli A, Mardirossian M, Gambato S, Florian F, Venier P, et al. Myticalins: a novel multigenic family of linear, cationic antimicrobial peptides from marine mussels (Mytilus spp.). Mar Drugs. 2017;15. https://doi.org/10.3390/md15080261.

39. Rosani U, Varotto L, Rossi A, Roch P, Novoa B, Figueras A, et al. Massively parallel amplicon sequencing reveals isotype-specific variability of antimicrobial peptide transcripts in Mytilus galloprovincialis. PLoS One. 2011;6:e26680.

40. Supungul P, Klinbunga S, Pichyangkura R, Hirono I, Aoki T, Tassanakajon A. Antimicrobial peptides discovered in the black tiger shrimp Penaeus monodon using the EST approach. Dis Aquat Org. 2004;61:123–35.

41. Tian C, Gao B, Fang Q, Ye G, Zhu S. Antimicrobial peptide-like genes in Nasonia vitripennis: a genomic perspective. BMC Genomics. 2010;11:187.

42. Pi C, Liu J, Peng C, Liu Y, Jiang X, Zhao Y, et al. Diversity and evolution of conotoxins based on gene expression profiling of Conus litteratus. Genomics. 2006;88:809–19.

43. Pineda SS, Sollod BL, Wilson D, Darling A, Sunagar K, Undheim EAB, et al. Diversification of a single ancestral gene into a successful toxin superfamily in highly venomous Australian funnel-web spiders. BMC Genomics. 2014;15:177.

44. Zhu S, Peigneur S, Gao B, Luo L, Jin D, Zhao Y, et al. Molecular diversity and functional evolution of scorpion Potassium Channel toxins. Mol Cell Proteomics MCP. 2011;10. https://doi.org/10.1074/mcp.M110.002832.

45. Van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, et al. Plant peptides govern terminal differentiation of bacteria in symbiosis. Science. 2010;327:1122–6.

46. Conlon JM. Structural diversity and species distribution of host-defense peptides in frog skin secretions. Cell Mol Life Sci. 2011;68:2303–15.

47. Conlon JM, Kolodziejek J, Nowotny N, Leprince J, Vaudry H, Coquet L, et al. Characterization of antimicrobial peptides from the skin secretions of the Malaysian frogs, Odorrana hosii and Hylarana picturata (Anura:Ranidae). Toxicon. 2008;52:465–73.

48. Simmaco M, Kreil G, Barra D. Bombinins, antimicrobial peptides from Bombina species. Biochim Biophys Acta BBA - Biomembr. 2009;1788:1551–5.

49. Yang X, Lee W-H, Zhang Y. Extremely abundant antimicrobial peptides existed in the skins of nine kinds of Chinese odorous frogs. J Proteome Res. 2012;11:306–19.

50. Tossi A, Sandri L, Giangaspero A. New consensus hydrophobicity scale extended to non-proteinogenic amino acids. Peptides. 2002;27:416.

51. Mangoni ML, Shai Y. Temporins and their synergism against gram-negative bacteria and in lipopolysaccharide detoxification. Biochim Biophys Acta BBA - Biomembr. 2009;1788:1610–9.

52. Solstad RG, Li C, Isaksson J, Johansen J, Svenson J, Stensvåg K, et al. Novel antimicrobial peptides EeCentrocins 1, 2 and EeStrongylocin 2 from the Edible Sea urchin Echinus esculentus have 6-Br-Trp post-translational modifications. PLoS One. 2016;11:e0151820.