

---

Faculty of Science

Faculty Publications

---

Mining amphibian and insect transcriptomes for antimicrobial peptide sequences with rAMPage

Lin, D., Sutherland, D., Aninta, S. I., Louie, N., Nip, K. M., Li, C., . . . Birol, I.

2022

© 2022 Diana Lin et al. This is an open access article distributed under the terms of the Creative Commons Attribution License. <http://creativecommons.org/licenses/by/4.0/>

This article was originally published at:  
<https://doi.org/10.3390/antibiotics11070952>


---

Citation for this paper:

Lin, D., Sutherland, D., Aninta, S. I., Louie, N., Nip, K. M., Li, C., . . . Birol, I. (2022). "Mining amphibian and insect transcriptomes for antimicrobial peptide sequences with rAMPage." *Antibiotics*, 11(7), 952.  
<https://doi.org/10.3390/antibiotics11070952>

## Article

# Mining Amphibian and Insect Transcriptomes for Antimicrobial Peptide Sequences with rAMPage

Diana Lin <sup>1</sup>, Darcy Sutherland <sup>1,2,3</sup>, Sambina Islam Aninta <sup>1</sup>, Nathan Louie <sup>1</sup>, Ka Ming Nip <sup>1,4</sup>,  
Chenkai Li <sup>1,4</sup>, Anat Yanai <sup>1</sup>, Lauren Coombe <sup>1</sup>, René L. Warren <sup>1</sup>, Caren C. Helbing <sup>5</sup>,  
Linda M. N. Hoang <sup>2,3</sup> and Inanc Birol <sup>1,2,3,\*</sup>

<sup>1</sup> Canada's Michael Smith Genome Sciences Centre at BC Cancer, Vancouver, BC V5Z 4S6, Canada; dlin@bcgsc.ca (D.L.); dsutherland@bcgsc.ca (D.S.); saninta@bcgsc.ca (S.I.A.); nlouie@bcgsc.ca (N.L.); kmnip@bcgsc.ca (K.M.N.); cli@bcgsc.ca (C.L.); ayanai@bcgsc.ca (A.Y.); lcoombe@bcgsc.ca (L.C.); rwarren@bcgsc.ca (R.L.W.)

<sup>2</sup> British Columbia Centre for Disease Control, Public Health Laboratory, Vancouver, BC V6Z R4R, Canada; linda.hoang@bccdc.ca

<sup>3</sup> Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

<sup>4</sup> Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

<sup>5</sup> Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8P 5C2, Canada; chelbing@uvic.ca

\* Correspondence: ibirol@bcgsc.ca



**Citation:** Lin, D.; Sutherland, D.; Aninta, S.I.; Louie, N.; Nip, K.M.; Li, C.; Yanai, A.; Coombe, L.; Warren, R.L.; Helbing, C.C.; et al. Mining Amphibian and Insect Transcriptomes for Antimicrobial Peptide Sequences with rAMPage. *Antibiotics* **2022**, *11*, 952. <https://doi.org/10.3390/antibiotics11070952>

Academic Editors: Guillermin Agüero-Chapin, Agostinho Antunes and Yovani Marrero-Ponce

Received: 9 June 2022

Accepted: 13 July 2022

Published: 15 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Antibiotic resistance is a global health crisis increasing in prevalence every day. To combat this crisis, alternative antimicrobial therapeutics are urgently needed. Antimicrobial peptides (AMPs), a family of short defense proteins, are produced naturally by all organisms and hold great potential as effective alternatives to small molecule antibiotics. Here, we present rAMPage, a scalable bioinformatics discovery platform for identifying AMP sequences from RNA sequencing (RNA-seq) datasets. In our study, we demonstrate the utility and scalability of rAMPage, running it on 84 publicly available RNA-seq datasets from 75 amphibian and insect species—species known to have rich AMP repertoires. Across these datasets, we identified 1137 putative AMPs, 1024 of which were deemed novel by a homology search in cataloged AMPs in public databases. We selected 21 peptide sequences from this set for antimicrobial susceptibility testing against *Escherichia coli* and *Staphylococcus aureus* and observed that seven of them have high antimicrobial activity. Our study illustrates how in silico methods such as rAMPage can enable the fast and efficient discovery of novel antimicrobial peptides as an effective first step in the strenuous process of antimicrobial drug development.

**Keywords:** antimicrobial peptide; AMP discovery; genome mining; antimicrobial resistance

## 1. Introduction

Due in large part to the overuse and misuse of antibiotics, the prevalence of multidrug-resistant bacteria is rapidly growing at a rate that cannot be matched by antibiotic discovery efforts [1]. As a consequence, the world is currently in an arms race and is at the cusp of a post-antibiotic era [1]. The slow pace of new antibiotic drug discovery, development, and regulation, combined with the accelerated emergence of resistance to existing antibiotics creates what is referred to as the “discovery void” [2]. This gap between discovery and emergence of resistance highlights an urgency to develop new antimicrobial therapeutics. One such alternative is formulations based on the antimicrobial peptides (AMPs) [3].

AMPs are short amphipathic host defense peptides that are produced in all multicellular organisms as part of the innate immune system [3]. Many AMPs operate through nonspecific mechanisms [4], such as direct electrostatic interactions with the cell membrane and immunomodulation [3], allowing for a broad spectrum of efficacy against bacteria [5], viruses [6], and fungi [7]. Furthermore, pathogens develop a slower rate of resistance to

AMPs compared to conventional antibiotics [8]. It is these qualities that position AMPs as attractive alternatives to conventional antibiotics [9].

AMPs are often produced as precursor peptides within cells that consist of an N-terminal signal peptide, followed by an acidic pro-sequence, and a C-terminal basic bioactive mature peptide sequence [3]. The acidic pro-sequence neutralizes the basic mature peptide to keep the AMP in its inactive form and the signal peptide and acidic pro-sequence together are referred to as the prepro domain [3]. AMPs are then activated by proteolytic cleavage of the prepro sequence and the release of the mature peptide [3]. While the signal peptide is often highly conserved, the acidic pro-sequence and mature AMP can be quite variable [10]. However, there is evidence that the prepro sequence can vary across different organisms [3] and even within organisms [11].

Past research has shown that amphibians, such as the American bullfrog *Rana [Lithobates] catesbeiana*, possess a rich diversity of AMPs due to their aquatic and terrestrial life cycle, where the species encounter a wide spectrum of pathogens in these two environments [11]. In amphibians, AMPs such as ranatuerin are secreted at the skin surface upon pathogen exposure and can also stimulate an adaptive immune response [12]. In contrast, insects lack a sophisticated adaptive immune system and yet are highly tolerant to bacterial infection [13,14]. This may be due to the production of AMPs by the innate immune system [14]. In insects, AMPs are found in venom or salivary gland secretions. For example, melittin, a 26 amino acid (AA) peptide is the main component of honeybee venom [13]. While there are many known amphibian AMPs, there are far fewer known insect AMPs. Amphibian AMPs have their own designated database of 1923 peptides in the Database of Anuran Defense Peptides (DADP) [15]. Additionally, they also comprise 34% (1128 sequences) of the curated Antimicrobial Peptide Database 3 (APD3) [16]. Insect AMPs, however, only contribute 10% (325 sequences) in APD3, despite being the next largest non-mammalian classification. Better characterization of these AMP arsenals holds great potential in aiding the discovery of novel AMPs.

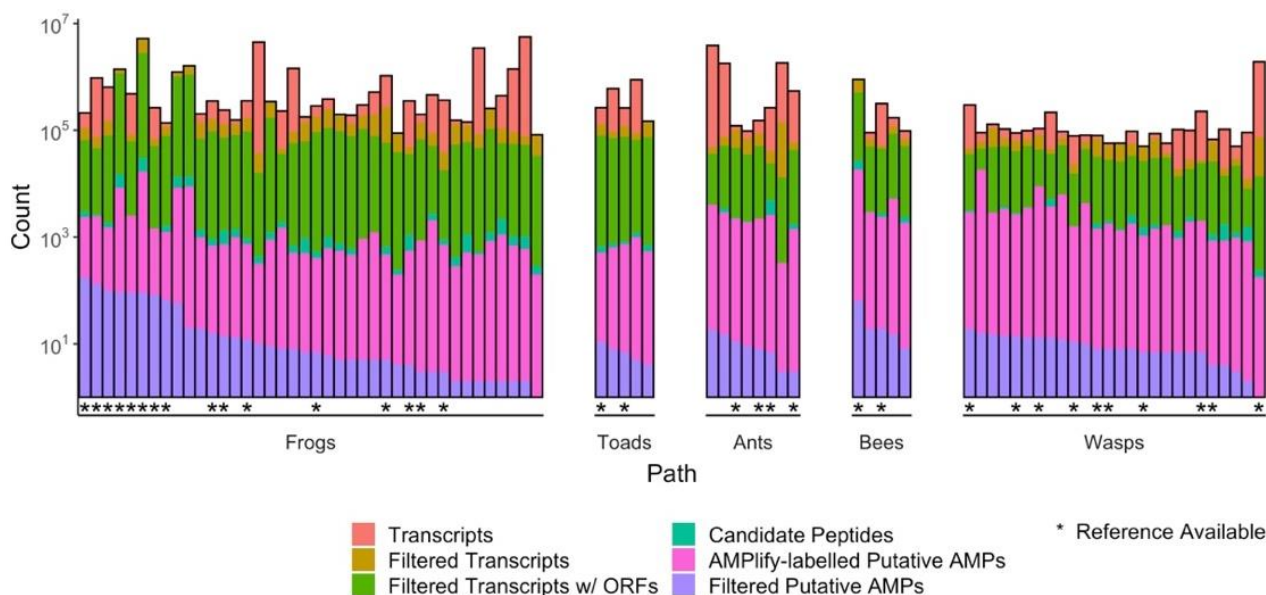
Because most AMPs under therapeutic investigation are derived from naturally occurring AMPs in various organisms [2], effective methods to discover natural AMPs would expand the number of potential candidates. Current wet lab screening protocols consist of extraction, isolation, and purification of AMPs through laborious methods such as the collection of skin secretions followed by liquid chromatography and sequence identification using mass spectrometry [17–21]. However, these protocols are costly, time-consuming, and expertise intensive. To resolve this, a scalable, rapid, high throughput in silico methodology built on genomics technologies and able to mine RNA sequencing (RNA-seq) datasets, would greatly aid in the discovery of AMPs funneling into drug development and enhancement processes. There are in silico AMP discovery methodologies presented in earlier studies [22–26], most of which start with processed data such as assembled genomic or protein sequences. Additionally, there are several state-of-the-art tools that perform AMP prediction [27]. Because AMP precursor genes have conserved sequence characteristics, these properties can be leveraged for filtering, and their inferred mature products can be classified as an AMP or not using machine learning methods. With the current unprecedented expansion of data generation and large amounts of sequencing data available in public repositories [28], there exists a rich untapped resource for AMP discovery.

To help fill the antibiotic discovery void, we offer rAMPage: Rapid Antimicrobial Peptide Annotation and Gene Estimation, a homology-based AMP discovery pipeline to mine for putative AMP sequences in publicly available genomic resources. To classify AMP sequences, rAMPage employs AMPlify [27], an attentive deep learning model. Currently, existing AMP databases, (e.g., APD3, DADP) contain less than 4000 validated nonredundant AMP sequences in total. In comparison, we have found over 1000 putative mature AMPs in the present study, with the potential to discover thousands more. Realizing the full potential of such pipelines would require the synthesis and validation of AMP candidates. Herein, we report our results on a select list of 21 peptides we detected using rAMPage.

## 2. Results

### 2.1. Identification of Putative AMPs

Using rAMPage, we assembled ~53 million transcripts from 84 RNA-seq datasets derived from the transcriptomes of 38 amphibian (33 frogs, five toads; anurans) and 37 hymenopteran insect (eight ants, five bees, 24 wasps) species and flagged 203,758 candidate peptide sequences to be classified (Figure 1). To select a list of high-confidence putative AMPs, we collapsed duplicates from multiple samples and applied three filters: AMPLify prediction score, peptide charge, and peptide length to obtain 1137 peptide sequences. Of these, 795 originate from amphibians, and 342 from insects. Running rAMPage on all 84 datasets took one week, with all datasets (comprising < 1 billion reads) taking less than 24 h (see Supplementary Materials Figure S1 for details on the computational platform and resource usage statistics).



**Figure 1.** Statistics and attrition as the sequencing data are processed by the rAMPage AMP discovery pipeline. rAMPage processes RNA-seq datasets from raw reads to transcripts to putative AMPs. In this case, a putative AMP is defined as a sequence with an AMPLify score  $\geq 10$  for amphibians or  $\geq 7$  for insects, a length  $\leq 30$  AA, and a charge  $\geq 2$ . Datasets with a reference transcriptome used during assembly are indicated with an asterisk. The total number of putative AMPs ( $n = 1478$ , including 341 duplicates) are shown in purple, discovered from a total of ~53 million assembled transcripts.

For each sequence, AMPLify [27] reports a prediction score  $s$  from 0 to 80, where  $s$  is a log-transformation of the AMPLify probability score  $p$

$$s = -10 \log_{10}(1 - p), \quad (1)$$

and 80 represents the highest confidence.

We note that the training data set for the AMPLify model had an over-representation of AMPs from amphibian species [27]; hence, it is biased towards assigning higher scores for amphibian AMPs. To compensate, we have applied separate score cut-offs for the two groups: 10 for amphibians and 7 for insects. Since the majority of AMPs are positively charged, a net charge threshold of  $\geq +2$  was applied. As for length, we filtered for sequences that are  $\leq 30$  AA, because shorter peptides are more cost-effective to synthesize for downstream validation studies. Figure S2 shows that the length filter used is the most restrictive filter of the three, with only 4.28% and 1.45% of the sequences for amphibians and insects, respectively, meeting this criterion.

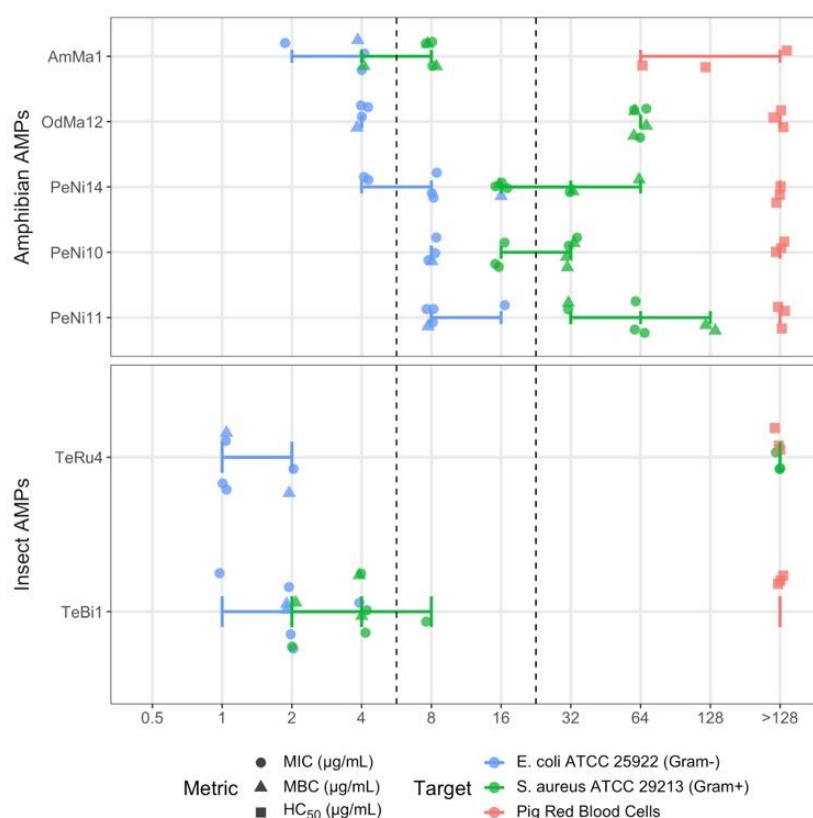
Score, charge, length distributions, and AA compositions of the 1137 putative AMPs are characterized in Figures S1 and S4. From this set, 21 AMPs were selected for synthesis

and validation, using three prioritization strategies: “Species Count”, “Insect Peptide”, and “AMplify Score” (see Section 4). The peptides have been named after the species they were discovered from (Table S1), then numbered in order using their AMplify scores.

## 2.2. Antimicrobial Susceptibility Testing (AST) Results

A total of 21 of the 1137 putative AMPs (Table S2) were synthesized (Genscript Biotech, Piscataway, NJ, USA) and tested for their antimicrobial activity against *Escherichia coli* ATCC 25922 and *Staphylococcus aureus* ATCC 29213 in a minimum of three independent experiments (see Figure S5 for a full set of experimental results). In these antimicrobial susceptibility tests, AMP activity was assessed using two metrics: minimum inhibitory and bactericidal concentrations (MIC and MBC, respectively). Lower MIC and MBC values are desirable as they indicate that lower AMP concentrations are sufficient for inhibitory or bactericidal activity, respectively. AMP toxicity was measured by HC<sub>50</sub> hemolytic concentration values—the concentration required to lyse  $\geq 50\%$  of porcine red blood cells. In contrast to MIC/MBC assays, it is desirable to have higher HC<sub>50</sub> values. All 21 putative AMPs exhibited minimal to no hemolytic activity with HC<sub>50</sub> values of 64  $\mu\text{g}/\text{mL}$  or higher.

Of these 21 putative AMPs, three displayed moderate activity (MIC and MBC in the range 8–16  $\mu\text{g}/\text{mL}$ ) and four displayed high activity ( $\leq 4$   $\mu\text{g}/\text{mL}$ ) against *E. coli* and/or *S. aureus*, all with minimal hemolytic activity, as shown in Figure 2. The characteristics of these seven sequences are described in Table 1. All seven AMPs with moderate to high antimicrobial activity have AMplify scores greater than 25.



**Figure 2.** Antimicrobial susceptibility and hemolysis test results of seven moderately and highly active putative AMPs. AMPs were tested for their bioactivity against *E. coli* and *S. aureus* to determine minimum inhibitory and bactericidal concentrations (MIC and MBC, respectively). AMPs were also tested for their hemolytic activity using pig red blood cells to determine hemolytic concentration (HC<sub>50</sub>) values. Moderate activity (MIC and MBC in the range of 8–16  $\mu\text{g}/\text{mL}$ ) and high activity ( $\leq 4$   $\mu\text{g}/\text{mL}$ ) thresholds indicated by the dashed lines. AMPs are ordered by increasing MIC values against *E. coli* ATCC 25922.

**Table 1.** Characteristics of putative AMP sequences with moderate to high in vitro bioactivity against *E. coli* or *S. aureus*. Each sequence is separated into the prepro sequence and the predicted mature peptide sequence. Conserved proteolytic cleavage sites are underlined in the prepro sequences.

Prepro-Sequence	Putative Mature Peptide						
	Sequence	Length	Charge	AMPLify Score	MIC (µg/mL) *		Peptide ID
					<i>E. coli</i> †	<i>S. aureus</i> †	
MFTMKKSLLVFFLGI-VLSLCEEERNADED-DGEMTEEVKR	GILDTLKQLGKAAVQG-LLSKAACKLAKTC	29	4	80.0	2–4	4–8	AmMa1
LGIVSLSLCQEERSA-DDEEGEVIEEEVKR	GFMDTAKNVAKNV-AVTLLYNLKCKITKAC	29	4	69.2	4	64	OdMa12
MFTMKKSLLVFFLGI-TIALSLCEEERGADEEENGGEITDEEVKR	GLLLD TVKGA AKNVA-GILLNKLKCKVTGDC	30	3	61.8	8	16–32	PeNi10
MFTMKKSLLVFFLGI-TIALSLCEEERGADEEENGGEITDEEIKR	GILDTLKGA AKNVA GVL-LDKLKCKITGGC	30	3	61.8	8–16	32–128	PeNi11
MFTLRKSLLVFFLGM-VLSLCEQERDAD-EDEGEVTEEVKR	GLWTTIKEGVKNFS-VGVLDKIRCKITGGC	29	3	67.5	4–8	16–64	PeNi14
MKLLALVLVLS CVV-AYTARKRGQYWPT-NTKIFTTPYRFRREAD-QGSIVANLKNTPQLPFD-DNENLRLVLFNDPTVDLG-EDDK EIPG PQSQPNALSNN-LHLIDENDYFSSYTSQPGT-YRSFPRNFSGTSGRYRWRREAGGHVEPRLRFD AETQRG-NSFFTFADLQRRANRGI-EPTVSATAGIRFR-QEADQINPLAVRRERR	SWLSKSVK KLVNKKNY-TRLEKLAKK KLFNE	30	8	25.5	1–2	>128	TeRu4
IFLVGCKLFGNFIL-QRMQLLLALADAVA	KIKIPWGKVKDF-LVGGMKAVGKK	23	6	45.0	1–4	2–8	TeBi1

\* MIC: Minimum inhibitory concentration. † *Escherichia coli* ATCC 25922; *Staphylococcus aureus* ATCC 29213.

### 2.3. Novelty of Discovered AMPs

To assess if the putative AMPs discovered using rAMPAGE are novel, a BLASTp [29] (basic local alignment search tool) protein search was performed using the 1137 sequences that met our selection criteria. Of these, 1024 sequences are reported as novel, providing no antimicrobial characterization or exact match (sequence identity = 100%; query coverage = 100%) within the NCBI non-redundant protein database [29]. The novelty analysis results for the seven moderately to highly active AMPs are presented in Table 2. Four of the queried putative AMPs (AmMa1, OdMa12, PeNi10, and PeNi14) are novel in sequence, aligning with high sequence identity ( $\geq 90\%$ ) to existing NCBI annotations [29]. Two putative AMPs (PeNi11 and TeBi1) are known and published AMPs, aligning with 100% sequence identity of the precursor protein and across the prepro and mature regions. One putative AMP (TeRu4) aligns with high sequence identity to an uncharacterized protein in the NCBI non-redundant protein database.

**Table 2.** Comparison of sequence identities (%) of the discovered AMPs with their best-known AMP blastp matches to the NCBI non-redundant (nr) protein database over the entire sequence (precursor), prepro or mature sequences.

Peptide ID	Source Organism	Highest Scoring Blastp Match	Sequence Identity (%)		
			Precursor	Prepro	Mature
AmMa1	<i>Amolops mantzorum</i>	Palustrin-2GN3 (ADM34231.1) [ <i>Amolops granulosis</i> ]	97	100	93
OdMa12	<i>Odorrana margaretae</i>	Odorranain-F2 (ABG76517.1) [ <i>Odorrana grahami</i> ]	98	100	97
PeNi10	<i>Leptobrachium boringii</i>	Pelophylaxin-1 (Q2WCN8.1) [ <i>Pelophylax fukienensis</i> ]	82	86	77
	<i>Polypedates megacephalus</i>	Ranatueringin-2N (AEM68233.1) *	98	97	100
	<i>Pelophylax nigromaculatus</i>	[ <i>Pelophylax nigromaculatus</i> ]			
	<i>Rhacophorus dennysi</i> <i>Rhacophorus omeimontis</i>				
PeNi11	<i>Leptobrachium boringii</i>	Pelophylaxin-1 (Q2WCN8.1) [ <i>Pelophylax fukienensis</i> ]	100	100	100
	<i>Polypedates megacephalus</i>				
	<i>Pelophylax nigromaculatus</i>				
	<i>Rhacophorus dennysi</i> <i>Rhacophorus omeimontis</i>				
PeNi14	<i>Bufo gargarizans</i>	Palustrin-2HB1 (AIU998997.1) [ <i>Pelophylax hubeiensis</i> ]	90	93	86
	<i>Polypedates megacephalus</i>				
	<i>Pelophylax nigromaculatus</i>				
	<i>Rhacophorus omeimontis</i>				
TeRu4	<i>Temnothorax rugatulus</i>	Uncharacterized protein (XP_024884948.1) [ <i>Temnothorax curvoispinosus</i> ]	94	93	97
		Uncharacterized protein (TGZ47385.1) *	91	90	97
		[ <i>Temnothorax longispinosus</i> ]			
TeBi1	<i>Tetramorium bicarinatum</i>	M-myrmecitoxin(01)-Tb1a (W8GNV3.1) [ <i>Tetramorium bicarinatum</i> ]	100	-	100

\* Highest scoring blastp match when query sequence consists of only the mature sequence instead of the whole precursor. -: no significant alignment.

AmMa1, derived from the Moupung sucker frog, *A. mantzorum*, aligned with 97% sequence identity to Palustrin-2GN3 [30] from a species of the same genus, *A. granulosis*, differing only by two AA in the mature region (Figure S3a). Similarly, OdMa12, found in the green odorous frog, *O. margaretae*, aligned with 98% sequence identity to odorranain-F2 [31] from a species of the same genus, *O. grahami*, differing only by one AA in the mature region (Figure S3b). While these two sequences (AmMa1 and OdMa2) are very similar to known sequences, we have additionally discovered each of them in a different species of the same genus.

PeNi10 was detected in the dark-spotted frog *P. nigromaculatus*, and aligned with 82% identity to pelophylaxin-1 [32] from a species of the same genus, *P. fukienensis* (Figure S3c). We also identified PeNi10 in four other species of frogs: *L. boringii*, *P. megacephalus*, *R. dennysi*, *R. omeimontis* (Figure S4a). Although the PeNi10 precursor aligns best to pelophylaxin-1, the mature region aligns with complete sequence identity to ranatueringin-2N (unpublished).

PeNi14, also derived from the dark-spotted frog, *P. nigromaculatus*, aligned with 90% sequence identity to palustrin-2HB1 [33] from a species of the same genus, *P. hubeiensis* (Figure S6d). PeNi14 was also detected in three other species of frogs: *B. gargarizans*, *P. megacephalus*, *R. omeimontis* (Figure S7b).

Originating from the dark-spotted frog, *P. nigromaculatus*, PeNi11 aligned with 100% sequence identity to pelophylaxin-1 [32] from a species of the same genus, *P. fukienensis*, meaning it is identical to a known AMP precursor (Figure S6e). However, in addition to *P. nigromaculatus*, we also detected PeNi11 in four other species of frogs: *L. boringii*, *P. megacephalus*, *R. dennysi*, *R. omeimontis* (Figure S7c).

Found in the venom of tramp ant, *T. bicarinatum*, TeBi1 aligned with 100% sequence identity with bicarinalin [34,35] of the same species (Figure S6f). In the case of TeBi1, its precursor was partial on the 5' end, accounting for no alignments in the prepro sequence.

TeRu4, discovered in the brain of the small myrmicine ant, *T. rugatulus*, aligned with 100% sequence identity to an uncharacterized protein [36] from a species of the same genus, *T. longispinosus* (Figure S6g). While TeRu4 is not a novel protein, it is a novel mature AMP as it has not been previously characterized to have antimicrobial properties.

Additional annotation of the seven bioactive peptides (five amphibians, two insects) can be found in Table S3. The underrepresentation of insect AMPs in the literature, compared to amphibians, is further demonstrated here; while the amphibian peptides have been annotated with “frog antimicrobial peptide” domains in both InterProScan [37] and Pfam [38], the insect sequences have no protein family annotations. Figure S8 illustrates the sequence identity between AMPs identified by rAMPPage and known AMPs for amphibian and insect AMPs. Although the majority of putative AMPs from rAMPPage were novel sequences, previously reported AMP sequences were also identified and are a good demonstration and internal validation of the robustness of this methodology.

### 3. Discussion

Using rAMPPage, we analyzed 84 RNA-seq datasets of 38 amphibian and 37 insect species to discover 1137 putative AMPs, 1024 of which are novel. In the present study we report our validation results on 21 putative AMPs, with over 1000 additional peptide sequences left to investigate. This list is by no means exhaustive; adjusting the described filtering parameters may yield thousands more discoveries (Table S4). Further, the rAMPPage pipeline can be readily used on other transcriptome sequencing datasets, though this might call for modifications in experimental designs. For instance, in the case of bacterial RNA-seq datasets with reduced post-transcriptional polyadenylation, RNA-seq data from rRNA depleted libraries would be recommended as input for the pipeline, as opposed to data from poly(A) enriched libraries [39,40].

While the sensitivity (proportion of reference AMPs captured by the three putative AMP filters) of rAMPPage is <50% (Table S5 and Figure S9) with the default filtering thresholds, the filters are implemented to select for high confidence predictions that are also easier and more cost-effective to synthesize for validation. However, as more putative AMPs are discovered and the number of reference AMPs increase in public databases, the rAMPPage filters can be adjusted accordingly to report more novel AMPs.

Although rAMPPage captures most putative AMPs in their complete mature form, their associated precursor sequences may be incomplete, as shown using multiple sequence alignments with Clustal Omega v1.2.4 [41] (Figure S7). However, most partial transcripts are missing sequence on the 5' end. Therefore, while the AMP precursors may be partial, the mature AMPs at the C-termini are more likely to be complete, thereby still detectable by rAMPPage.

Because progress is rapid in bioinformatics, rAMPPage is designed to be flexible as new technologies are developed. The pipeline is implemented as a Makefile with each step as a separate target, making the pipeline modular and providing analysis checkpoints. The tools for each step can be substituted with newer/improved tools if needed. Similarly, the pipeline is versatile and can be adapted for other sequencing technologies, for instance by assembling RNA/cDNA long reads from Pacific Biosciences of California (Menlo Park, CA, USA) or Oxford Nanopore Technologies Ltd. (Oxford, UK) instruments.

Recently, our group released AMPlify and compared its performance to other state-of-the-art tools for AMP prediction [27]. Other machine learning methods included iAMP-

pred [42], iAMP-2L [43], AMP Scanner Vr. 2 [44], with AMPlify outperforming all previously described AMP prediction tools in metrics of accuracy, sensitivity, and specificity [27]. For this reason, rAMPPage employs AMPlify as its AMP prediction step, and will continue to until it is surpassed in performance. Machine learning in AMP discovery is a dynamic study, ranging from AMP sequence prediction and structure classification to de novo AMP sequence generation and design [45–47]. While there are existing methods to mine protein databases [48,49], rAMPPage is an all-in-one tool to mine next-generation sequencing data directly from reads to AMP prediction.

While rAMPPage can find a substantial number of putative AMPs, its main limitation lies in the fact that it uses homology-based sequence selection and machine learning-based sequence classification steps. These two steps are limited by the quantity and quality of data currently available for training the tools. The homology-based step of rAMPPage would be less sensitive when there are more divergent signal sequences in the precursor genes. Similarly, the sequence classification engine in the pipeline, AMPlify, may be biased by known (and limited) classes of AMPs in the databases. However, this limitation is not restricted to only AMPlify, but all approaches dependent on AMP databases for training data sets [48–50].

Despite these limitations, which are expected to resolve over time as curated AMP sequence databases grow, a sizeable number (>1000 from 84 RNA-seq datasets) of AMPs were reported by the pipeline with the filters described herein. In the tested set of 21 peptides, seven demonstrated antimicrobial activity against a defined set of bacteria in vitro and 15 did not. We note that AST experiments can assess activity against the tested pathogens but cannot rule in or out an activity against other targets. Further, AMPs have multiple modes of action, and the AST protocol used in our study only validates direct action and does not test the putative immunomodulatory effects of these peptides, for instance. Of the seven active putative AMPs, three were moderately active, and all three are expressed in multiple amphibian species, potentially signaling the evolutionary significance of these AMPs.

An AMP of particular interest in the present study is TeRu4, due to its novelty and specificity in bioactivity. The precursor sequence of TeRu4 is 234 AA long, indicating that TeRu4 may be a multi-functional protein, such as a histone whose subsequence includes antimicrobial properties [51]. Additionally, TeRu4 showed a 36.84% sequence similarity to the spaetzle protein from the fruit fly *Drosophila melanogaster*, a protein in the insect Toll pathway, which triggers AMP production [52]. TeRu4 is also the most specific of the active putative AMPs we tested. While all the other active peptides tested are active against both *E. coli* and *S. aureus*, TeRu4 is active only against *E. coli*, a Gram-negative bacterium. This specificity may indicate a unique mechanism of action.

Despite the great promise of discovering putative AMPs with rAMPPage, AMP-based drug development still faces some biological challenges, such as peptide stability and bacterial resistance. AMPs in their mature form are considered more unstable and more easily degraded by proteases. While synthesizing precursors for testing would increase stability, doing so would drive up the cost of synthesis using conventional synthetic chemistry methods. Although resistance to AMPs emerges at a slower rate compared to resistance to antibiotics, bacteria may develop resistance to AMPs through surface remodeling, modulation of AMP gene expression, proteolytic degradation, trapping, efflux pumps, and biofilms [4,53–55]. To combat specific mechanisms of resistance, targeted AMP discovery methods are being developed. A method to discover AMPs with anti-biofilm activity is described in a preprint [26], and a curated 3D structural and functional repository of AMPs relevant to biofilm studies called B-AMP was recently published [26]. Finding solutions to these and other challenges in developing AMPs as replacements for conventional small molecule antibiotics is an active field of research [56–58].

#### 4. Materials and Methods

rAMPPage is an AMP discovery pipeline that takes short RNA-seq reads as input, and outputs candidate putative AMPs for wet lab validation. Since it is a homology-based

method to select a list of candidates for classification, a set of reference AMPs is required. Here, we describe how input datasets and reference AMPs are collated, as well as each step of rAMPage.

#### 4.1. Collating Input RNA-Seq Datasets

The RNA-seq reads from 38 amphibian and 37 insect species were downloaded from the Sequence Read Archive (SRA) [59] using fasterq-dump v2.10.5 (<http://ncbi.github.io/sra-tools/>, accessed on 4 November 2019) from the NCBI SRA Tool Kit. Analyzing RNA-seq (transcriptomic) reads enables the discovery of expressed putative AMPs. Because some RNA-seq experiments were conducted with multiple tissues or treatments, there are 75 species in total, but 84 datasets are shown in Tables S6 and S7.

#### 4.2. Collating Reference AMP Datasets

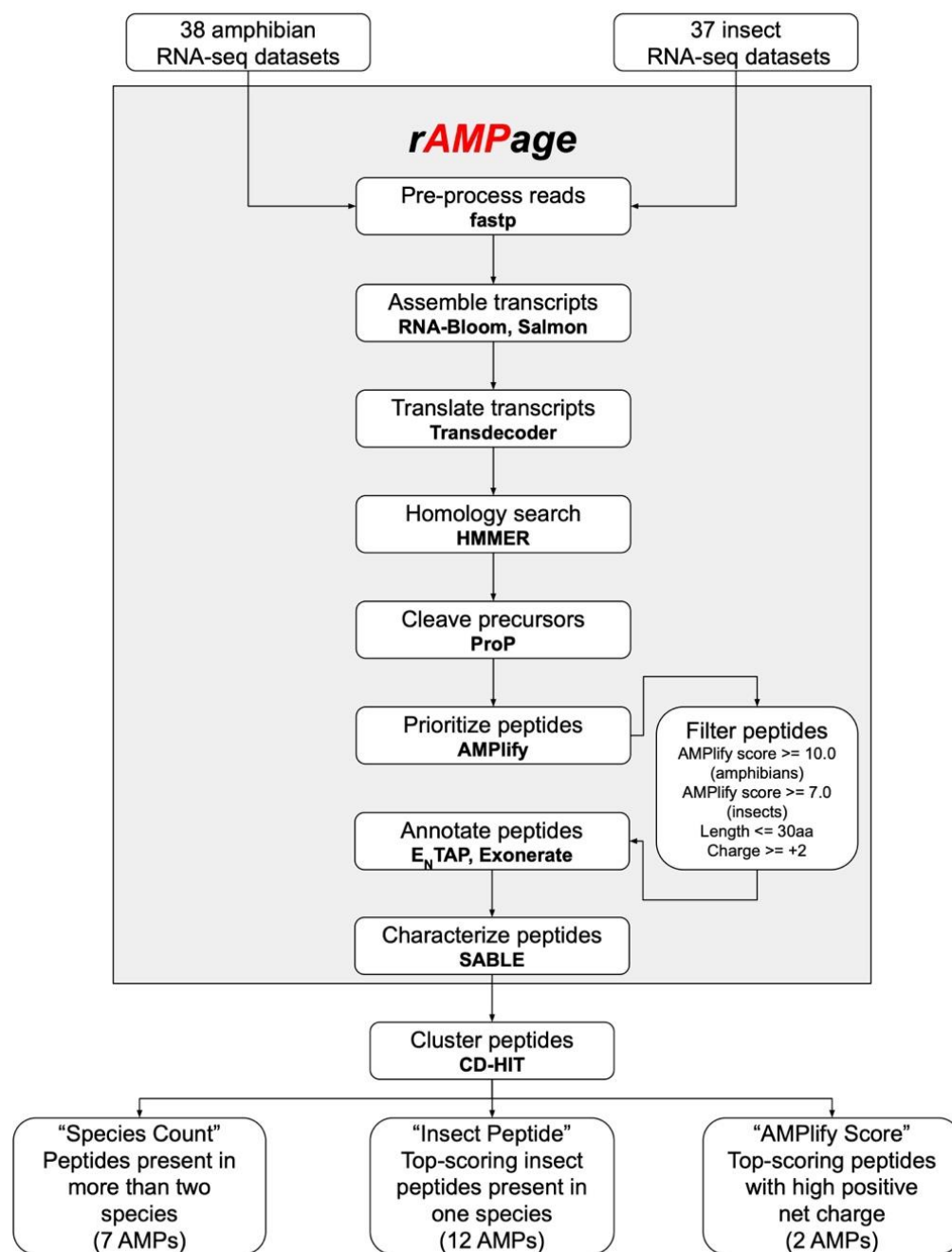
A set of 3306 AMP sequences were collated from two high-quality AMP databases: the Database of Anuran Defense Peptides (DADP; <http://split4.pmfst.hr/dadp/>, accessed on 6 December 2018) [15] and the Antimicrobial Peptide Database 3 (APD3; <https://aps.unmc.edu>, accessed on 14 September 2020) [16]. These databases are highly curated, where sequences have been validated for efficacy. To complement this list, 3835 precursor and mature AMP sequences of amphibian and insect origin were downloaded from the NCBI non-redundant (nr) protein database [29]. These sequences are less curated, including partial sequences and sequences with only in silico prediction, etc., accounting for the difference between numbers from DADP/APD3 and NCBI in Table S8.

#### 4.3. rAMPage Pipeline

rAMPage is implemented as a Makefile and written in bash, Python3, and R. It is publicly available on GitHub (<https://github.com/bcgsc/rAMPage>, v1.0 accessed on 14 February 2021). The pipeline was tested for the dependencies listed in Tables S9 and S10, and is highly customizable, with its major parameter options listed in Table S4. Command and parameters for each step can be found in Table S11. A flowchart of the rAMPage pipeline is shown in Figure 3.

Because the datasets used for rAMPage originate from publicly available genomic resources and we have no control over the experimental design or protocols used, we performed rigorous quality control. The RNA-seq reads were trimmed to remove adapter sequences using fastp v0.20.0 [60], which does not require the adapter sequences to be known, and instead infers adapter sequences from sequence overlaps between reads. This is particularly convenient when dealing with multiple datasets that possibly have different sequencing protocols.

To assemble the RNA-seq reads into transcripts we used RNA-Bloom v1.3.1 [61], a de novo transcriptome assembler that works with single and paired-end reads. RNA-Bloom is able to assemble transcriptomes without a reference but also allows for reference-guided assembly if a reference is available. It also allows for multi-sample pooling, where, for instance, reads describing multiple tissues from the same individual or different treatments for the same species are assembled together while retaining the tissues or treatment specificity of assembled transcripts.



**Figure 3.** rAMPPage workflow. The rAMPPage pipeline and downstream selection of putative AMPs for validation.

We note that the transcripts with a smaller number of reads have less reconstruction evidence; thus, assembled sequences with lower measured expression levels may be enriched for misassemblies. To exclude such sequences from downstream analysis, we used Salmon v1.3.0 [62] to quantify assembled transcript expression levels, and filtered out transcripts with less than 1 TPM (transcripts per million) expression.

To obtain translated peptide sequences from the transcripts, TransDecoder v5.5.0 [63] was used to conduct an in silico six-frame open reading frame (ORF) translation, and ORFs that are at least 50 AA were selected for downstream analysis. In the case of nesting ORFs, the longest ORF was chosen.

To select putative AMP precursors from this vast pool of assembled and translated sequences, we conducted a homology search against our curated reference AMP dataset (Table S8) using HMMER v3.3.1 [64] and assigned an Expect (E) value to every sequence.

The E-value describes the number of hits expected by chance when searching a database of a particular size [65]. Sequences that share a certain degree of identity, with E-values of less than  $10^{-5}$ , were selected as putative AMP precursors.

These putative precursor (or partial precursor) sequences were then cleaved in silico using ProP v1.0c [66] to obtain putative mature AMP sequences, to be further classified. However, cleavage prediction tools only predict where the cleavage occurs, not what each resulting cleaved peptide represents, and the AMP precursor organization shows inter- and intra-species variability [13,67,68]. While amphibian AMPs are typically cleaved at a lysine–arginine (KR) motif and their precursor structure follows a conserved structure (prepro sequence containing acidic AA residues and a mature bioactive AMP) [67], insect AMPs are typically cleaved at an RXXR motif (two arginine residues surrounding two optional AA) and the precursor structure is not always conserved [68]. Insect AMPs are more variable in structure [13], increasing the difficulty in identifying the putative mature peptide. This difficulty is especially present in precursor structures with multiple acidic regions (UniProtKB P54684.1) or multiple bioactive regions (UniProtKB P35581.1). In such multi-peptide precursors, it is unclear whether each bioactive region is its own isoform or part of a larger mature peptide. To account for this and to possibly discover novel but perhaps not naturally occurring putative AMPs, cleaved peptides were also recombined in a manner similar to alternative splicing (Figure S10). In this procedure, the order and orientation of the cleaved peptides were maintained, and cleaved peptides that originally share cleavage sites were not recombined, with a maximum of three cleaved peptides within recombination. This recombination feature can be turned off in rAMPPage's options.

The collected candidate peptide sequences were classified with AMPLify v1.0.3 [27] as AMP or non-AMP sequences. When given a sequence, AMPLify calculates a score between 0 to 80, with the score  $\approx 3.0103$  corresponding to the classification probability cutoff of 50% through Equation (1).

To facilitate AMP synthesis for the validation experiments, we filtered the putative AMPs by length and charge, in addition to the AMPLify score. A maximum length of 30 AA was imposed to control the cost of peptide synthesis and to reduce the number of spurious hits from recombined sequences. A minimum charge of +2 was imposed as a proxy to assess the effectiveness of an AMP, as past evidence indicates that more positively charged AMPs show higher activity, especially when their mechanism of action is membrane disruption [69]. Because AMPLify was trained on mostly amphibian AMPs, different score thresholds were imposed for amphibian ( $\geq 10$ ) and insect ( $\geq 7$ ) datasets to compensate for the dearth of insect training AMPs.

To annotate the final set of filtered putative AMPs, E<sub>N</sub>TAP v0.10.7, Eukaryotic Non-Model Transcriptome Annotation Pipeline [70], were used, along with UniProtKB (release 2020\_06, accessed on 15 December 2020) [71], RefSeq (release 203, accessed on 15 December 2020) [72], and NCBI non-redundant (nr) (v5, accessed on 12 December 2020) [29] protein databases. For AMPs that E<sub>N</sub>TAP failed to annotate, InterProScan 5 v5.30-69.0 [37] was run separately to annotate protein families, functions, and domains. Exonerate v2.4.0 [73] was used to align the filtered putative AMPs against the reference AMPs to assess how many of the labeled AMPs were already known AMPs. Finally, SABLE v4.0 [74] was optionally used to predict secondary structures of the filtered putative AMPs, for visualization.

#### 4.4. Selecting Filtered Putative AMPs for Validation

To select peptides to validate from the filtered putative AMPs, we ranked their sequences using AMPLify and chose peptides based upon three selection criteria (Figure 3): “Species Count” ( $n = 7$ ), “Insect Peptide” ( $n = 12$ ), or “AMPLify Score” ( $n = 2$ ), for a final total of 21 AMPs (Table S2). The sequences were first clustered using CD-HIT [75] v4.8.1 with a sequence similarity cutoff of 100%. We chose the longest sequence for each of these clusters, removing duplicate and subsumed sequences to obtain a non-redundant sequence set.

In the first selection strategy of “Species Count”, sequences that were present in more than two species were chosen. In the “Insect Peptide” strategy, to balance the training bias

of AMPLify towards AMPs of amphibian origin, we specifically selected insect-originating sequences using a reduced AMPLify score cutoff of >20. In the “AMPLify Score” strategy, the two highest-scoring peptides (AMPLify score = 80.0, 69.9) with the highest charge (+4) were chosen for validation.

#### 4.5. Antimicrobial Susceptibility Testing (AST)

Twenty-one putative AMP sequences identified using the rAMPAGE pipeline were validated through a minimum of three AST experiments performed independently on separate days. In these tests, the AMP activity was assessed using two metrics: minimum inhibitory concentration and minimum bactericidal concentration (MIC and MBC, respectively). MIC and MBC values were determined using procedures outlined by the Clinical and Laboratory Standards Institute (CLSI), with the recommended adaptations for the testing of cationic AMPs described previously [76]. “Wild-type” strains of *Escherichia coli* (*E. coli* 25922) and *Staphylococcus aureus* (*S. aureus* 29213) were purchased from the American Type Culture Collection (ATCC; Manassas, VA, USA) and were used for screening of antimicrobial activity. Briefly, putative AMPs were synthesized by Genscript (Piscataway, NJ, USA) and received in lyophilized format. These peptides were suspended using ultrapure water (Life Technologies, Grand Island, NY, USA; Invitrogen cat# 10977-015), and an 11 µL two-fold serial dilution of 1280 to 2.5 µg/mL was prepared in duplicate rows in a 96-well microtiter plate, before being combined with 100 µL standardized bacterial inoculum yielding a final duplicate testing range of 128 to 0.25 µg/mL. The bacterial inoculum was prepared using colonies isolated on non-selective agar and combined with Mueller Hinton Broth. This suspension was measured and adjusted to achieve an optical density of 0.08–0.1, equivalent to a 0.5 McFarland standard ( $1-2 \times 10^8$  cfu/mL). The inoculum was then diluted to a target concentration of  $5 \pm 3 \times 10^5$  cfu/mL; total viable counts from the final inoculum were routinely performed to confirm the target bacterial density was achieved. MIC values were reported at the concentrations in which no visible growth was detected following 20–24-h incubation at 37 °C. The MIC and adjacent wells were plated onto non-selective agar; the concentration in which killed 99.9% of the inoculum following additional overnight incubation was determined to be the MBC.

#### 4.6. Hemolysis Experiments

The twenty-one putative AMPs were evaluated for toxicity using three independent hemolysis experiments performed on separate days. Whole blood from healthy donor pigs was purchased from Lampire Biological Laboratories (Pipersville, PA, USA). Red blood cells (RBCs) were washed and isolated by centrifugation, using Roswell Park Memorial Institute medium (RPMI) (Life Technologies, Grand Island, NY, USA; Gibco cat# 11835-030). Lyophilized AMPs were suspended and serially diluted from 128–1 µg/mL using RPMI in a 96-well plate, before being combined with 100 µL of a 1% RBC solution. Following a minimum 30 min incubation at 37 °C, plates were centrifuged and  $\frac{1}{2}$  volume from each supernatant was transferred to a new 96-well plate. The absorbance of these wells was measured at 415 nm. To quantify hemolytic activity and determine the AMP concentration that kills 50% of the RBCs ( $HC_{50}$ ), absorbance readings from wells containing RBCs treated with 11 µL of a 2% Triton-X100 solution or RPMI (AMP solvent-only) were used to define 100% and 0% hemolysis, respectively. All centrifugation steps were performed at  $500 \times g$  for five minutes in an Allegra-6R centrifuge (Beckman Coulter, CA, USA).

## 5. Conclusions

rAMPage is a bioinformatics pipeline for high throughput identification of putative AMPs in RNA-seq datasets. It fills a current void in the AMP discovery process, bridging the gap between in silico and in vitro methods. The pipeline has the potential to accelerate the discovery of novel antibiotics, with the possibility to enrich existing AMP sequence repositories. The easy-to-run pipeline design with various checkpoints and the low computational resources required to run rAMPage increase its accessibility to users. By executing rAMPage on publicly available amphibian and insect transcriptome sequencing data, we have identified over 1000 putative AMPs. Of those, we performed functional tests on twenty-one putative AMPs and demonstrated that seven have moderate to high activity against *E. coli* ATCC 25922 and/or *S. aureus* ATCC 29213. As the number of tested peptides increases, the wet lab validation results can feed back into rAMPage by augmenting the reference AMP datasets, helping refine the underlying homology and machine learning approaches. We expect rAMPage to have broad utility in the discovery of novel antimicrobials from a wide variety of transcriptome sequencing datasets.

## 6. Patents

Patent applications pending on the reported novel peptides.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/antibiotics11070952/s1>, Figure S1: Putative AMP filters used for amphibians and insects; Figure S2: Runtime and memory of each dataset through rAMPage; Figure S3: Score, length, and charge distribution of filtered putative AMPs; Figure S4: Amino acid composition of filtered putative AMPs; Figure S5: Antimicrobial susceptibility and hemolysis testing of 21 putative AMPs; Figure S6: Multiple sequence alignments of moderately to highly active AMPs; Figure S7: Multiple sequence alignments of moderately to highly active AMP precursors; Figure S8: Distribution of alignment of filtered putative AMPs to mature reference AMPs; Figure S9: Distribution of reference mature AMPs; Figure S10: Approach for peptides with multiple cleavage sites; Table S1: Peptide naming convention; Table S2: Subset of 21 putative AMPs synthesized and validated against *E. coli* and *S. aureus*; Table S3: Annotation of moderately to highly active putative mature AMPs; Table S4: Major options for rAMPage; Table S5: Sensitivity of all putative AMP filter combinations; Table S6: Amphibian RNA-seq datasets; Table S7: Insect RNA-seq datasets; Table S8: Breakdown of AMP sequences in AMP databases; Table S9: Shell scripting dependencies of rAMPage; Table S10: Bioinformatic tool dependencies of rAMPage; Table S11: Command and parameters for each step of rAMPage. References [11,27,29,30,41,59–64,73–75,77–117] are cited in the supplementary materials.

**Author Contributions:** Conceptualization, I.B., C.C.H. and L.M.N.H.; methodology, I.B., R.L.W., L.C., D.L. and D.S.; software, D.L., S.I.A., K.M.N. and C.L.; validation, D.S., A.Y. and N.L.; data curation, D.L. and C.L.; writing—original draft preparation, D.L.; writing—review and editing, I.B., R.L.W., L.C., D.S., A.Y., C.L., D.L. and C.C.H.; funding acquisition, I.B., C.C.H. and L.M.N.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by funds from Genome Canada, and Genome BC as part of the PeptAid (291PEP) and AnnoVis (281ANV) projects. Additional support was provided by the Canadian Agricultural Partnership, a federal–provincial–territorial initiative. The program is delivered by the Investment Agriculture Foundation of BC (INV106). Further funds were received from the Office of the Vice President, Research and Innovation of the University of British Columbia. Opinions expressed in this document are those of the author and not necessarily those of the Governments of Canada and British Columbia or the Investment Agriculture Foundation of BC. The Governments of Canada and British Columbia, and the Investment Agriculture Foundation of BC, and their directors, agents, employees, or contractors will not be liable for any claims, damages, or losses of any kind whatsoever arising out of the use of, or reliance upon, this information.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Accessions for input RNA-seq datasets can be found in Tables S6 and S7. rAMPPage code is publicly available at <https://github.com/bcgsc/rAMPPage> (v1.0, accessed on 14 February 2021).

**Acknowledgments:** This work was based on a prototype pipeline created by S. Austin Hammond.

**Conflicts of Interest:** I.B. is the founder of, and a shareholder in, Amphorax Life Sciences Inc.

## References

1. Hede, K. Antibiotic Resistance: An Infectious Arms Race. *Nature* **2014**, *509*, S2–S3. [[CrossRef](#)] [[PubMed](#)]
2. Koo, H.B.; Seo, J. Antimicrobial Peptides under Clinical Investigation. *Pept. Sci.* **2019**, *111*, e24122. [[CrossRef](#)]
3. Zhang, L.; Gallo, R.L. Antimicrobial Peptides. *Curr. Biol.* **2016**, *26*, R14–R19. [[CrossRef](#)] [[PubMed](#)]
4. Andersson, D.I.; Hughes, D.; Kubicek-Sutherland, J.Z. Mechanisms and Consequences of Bacterial Resistance to Antimicrobial Peptides. *Drug Resist. Updat.* **2016**, *26*, 43–57. [[CrossRef](#)] [[PubMed](#)]
5. Brandenburg, K.; Heinbockel, L.; Correa, W.; Lohner, K. Peptides with Dual Mode of Action: Killing Bacteria and Preventing Endotoxin-Induced Sepsis. *Biochim. Biophys. Acta BBA-Biomembr.* **2016**, *1858*, 971–979. [[CrossRef](#)]
6. Klotman, M.E.; Chang, T.L. Defensins in Innate Antiviral Immunity. *Nat. Rev. Immunol.* **2006**, *6*, 447–456. [[CrossRef](#)] [[PubMed](#)]
7. De Lucca, A.J.; Walsh, T.J. Antifungal Peptides: Novel Therapeutic Compounds against Emerging Pathogens. *Antimicrob. Agents Chemother.* **1999**, *43*, 1–11. [[CrossRef](#)]
8. Hancock, R.E.W.; Sahl, H.-G. Antimicrobial and Host-Defense Peptides as New Anti-Infective Therapeutic Strategies. *Nat. Biotechnol.* **2006**, *24*, 1551–1557. [[CrossRef](#)]
9. Moravej, H.; Moravej, Z.; Yazdanparast, M.; Heiat, M.; Mirhosseini, A.; Moosazadeh Moghaddam, M.; Mirnejad, R. Antimicrobial Peptides: Features, Action, and Their Resistance Mechanisms in Bacteria. *Microb. Drug Resist.* **2018**, *24*, 747–767. [[CrossRef](#)]
10. 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*, 2068–2081. [[CrossRef](#)]
11. Helbing, C.C.; Hammond, S.A.; Jackman, S.H.; Houston, S.; Warren, R.L.; Cameron, C.E.; Birol, I. Antimicrobial Peptides from Rana [Lithobates] Catesbeiana: Gene Structure and Bioinformatic Identification of Novel Forms from Tadpoles. *Sci. Rep.* **2019**, *9*, 1529. [[CrossRef](#)] [[PubMed](#)]
12. Conlon, J.M.; Mechkaraska, M. Host-Defense Peptides with Therapeutic Potential from Skin Secretions of Frogs from the Family Pipidae. *Pharmaceuticals* **2014**, *7*, 58–77. [[CrossRef](#)]
13. Wu, Q.; Patočka, J.; Kuča, K. Insect Antimicrobial Peptides, a Mini Review. *Toxins* **2018**, *10*, 461. [[CrossRef](#)]
14. Sheehan, G.; Farrell, G.; Kavanagh, K. Immune Priming: The Secret Weapon of the Insect World. *Virulence* **2020**, *11*, 238–246. [[CrossRef](#)] [[PubMed](#)]
15. Novković, M.; Simunić, J.; Bojović, V.; Tossi, A.; Juretić, D. DADP: The Database of Anuran Defense Peptides. *Bioinformatics* **2012**, *28*, 1406–1407. [[CrossRef](#)] [[PubMed](#)]
16. Wang, G.; Li, X.; Wang, Z. APD3: The Antimicrobial Peptide Database as a Tool for Research and Education. *Nucleic Acids Res.* **2016**, *44*, D1087–D1093. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, R.-W.; Liu, W.-T.; Geng, L.-L.; Chen, X.-H.; Bi, K.-S. Quantitative Analysis of a Novel Antimicrobial Peptide in Rat Plasma by Ultra Performance Liquid Chromatography–Tandem Mass Spectrometry. *J. Pharm. Anal.* **2011**, *1*, 191–196. [[CrossRef](#)] [[PubMed](#)]
18. Shen, W.; Chen, Y.; Yao, H.; Du, C.; Luan, N.; Yan, X. A Novel Defensin-like Antimicrobial Peptide from the Skin Secretions of the Tree Frog, *Theloderma Kwangsiensis*. *Gene* **2016**, *576*, 136–140. [[CrossRef](#)]
19. Pei, J.; Feng, Z.; Ren, T.; Sun, H.; Han, H.; Jin, W.; Dang, J.; Tao, Y. Purification, Characterization and Application of a Novel Antimicrobial Peptide from *Andrias Davidianus* Blood. *Lett. Appl. Microbiol.* **2018**, *66*, 38–43. [[CrossRef](#)]
20. Chen, W.; Hwang, Y.Y.; Gleaton, J.W.; Titus, J.K.; Hamlin, N.J. Optimization of a Peptide Extraction and LC–MS Protocol for Quantitative Analysis of Antimicrobial Peptides. *Future Sci. OA* **2019**, *5*, FSO348. [[CrossRef](#)]
21. Chowdhury, T.; Mandal, S.M.; Kumari, R.; Ghosh, A.K. Purification and Characterization of a Novel Antimicrobial Peptide (QAK) from the Hemolymph of *Antheraea Mylitta*. *Biochem. Biophys. Res. Commun.* **2020**, *527*, 411–417. [[CrossRef](#)] [[PubMed](#)]
22. Amaral, A.C.; Silva, O.N.; Mundim, N.C.C.R.; de Carvalho, M.J.A.; Migliolo, L.; Leite, J.R.S.A.; Prates, M.V.; Bocca, A.L.; Franco, O.L.; Felipe, M.S.S. Predicting Antimicrobial Peptides from Eukaryotic Genomes: In Silico Strategies to Develop Antibiotics. *Peptides* **2012**, *37*, 301–308. [[CrossRef](#)]
23. Prichula, J.; Primon-Barros, M.; Luz, R.C.Z.; Castro, Í.M.S.; Paim, T.G.S.; Tavares, M.; Ligabue-Braun, R.; d’Azevedo, P.A.; Frazzon, J.; Frazzon, A.P.G.; et al. Genome Mining for Antimicrobial Compounds in Wild Marine Animals-Associated Enterococci. *Mar. Drugs* **2021**, *19*, 328. [[CrossRef](#)]
24. De la Lastra, J.M.P.; Garrido-Orduña, C.; Borges, A.A.; Jiménez-Arias, D.; García-Machado, F.J.; Hernández, M.; González, C.; Boto, A. Bioinformatics discovery of vertebrate cathelicidins from the mining of available genomes. In *Drug Discovery—Concepts to Market*; Bobbarala, V., Ed.; InTech: London, UK, 2018; ISBN 978-1-78923-696-5.
25. Tomazou, M.; Oulas, A.; Anagnostopoulos, A.K.; Tsangaris, G.T.; Spyrou, G.M. In Silico Identification of Antimicrobial Peptides in the Proteomes of Goat and Sheep Milk and Feta Cheese. *Proteomes* **2019**, *7*, 32. [[CrossRef](#)]

26. Mhade, S.; Panse, S.; Tendulkar, G.; Awate, R.; Kadam, S.; Kaushik, K.S. AMPing Up the Search: A Structural and Functional Repository of Antimicrobial Peptides for Biofilm Studies, and a Case Study of Its Application to *Corynebacterium striatum*, an Emerging Pathogen. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 803774. [[CrossRef](#)]
27. Li, C.; Sutherland, D.; Hammond, S.A.; Yang, C.; Taho, F.; Bergman, L.; Houston, S.; Warren, R.L.; Wong, T.; Hoang, L.M.N.; et al. AMPLify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens. *BMC Genom.* **2022**, *23*, 77. [[CrossRef](#)]
28. Muir, P.; Li, S.; Lou, S.; Wang, D.; Spakowicz, D.J.; Salichos, L.; Zhang, J.; Weinstock, G.M.; Isaacs, F.; Rozowsky, J.; et al. The Real Cost of Sequencing: Scaling Computation to Keep Pace with Data Generation. *Genome Biol.* **2016**, *17*, 53. [[CrossRef](#)]
29. NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44*, D7–D19. [[CrossRef](#)]
30. Guo, R.; Chen, D.; Diao, Q.; Xiong, C.; Zheng, Y.; Hou, C. Transcriptomic Investigation of Immune Responses of the Apis Cerana Larval Gut Infected by *Ascosphaera Apis*. *J. Invertebr. Pathol.* **2019**, *166*, 107210. [[CrossRef](#)]
31. Li, J.; Xu, X.; Xu, C.; Zhou, W.; Zhang, K.; Yu, H.; Zhang, Y.; Zheng, Y.; Rees, H.H.; Lai, R.; et al. Anti-Infection Peptidomics of Amphibian Skin. *Mol. Cell. Proteom.* **2007**, *6*, 882–894. [[CrossRef](#)]
32. Song, Y.; Ji, S.; Liu, W.; Yu, X.; Meng, Q.; Lai, R. Different Expression Profiles of Bioactive Peptides in *Pelophylax nigromaculatus* from Distinct Regions. *Biosci. Biotechnol. Biochem.* **2013**, *77*, 1075–1079. [[CrossRef](#)] [[PubMed](#)]
33. Wang, X.; Ren, S.; Guo, C.; Zhang, W.; Zhang, X.; Zhang, B.; Li, S.; Ren, J.; Hu, Y.; Wang, H. Identification and Functional Analyses of Novel Antioxidant Peptides and Antimicrobial Peptides from Skin Secretions of Four East Asian Frog Species. *Acta Biochim. Biophys. Sin.* **2017**, *49*, 550–559. [[CrossRef](#)] [[PubMed](#)]
34. Rifflet, A.; Gavalda, S.; Téné, N.; Orivel, J.; Leprince, J.; Guilhaudis, L.; Génin, E.; Vétillard, A.; Treilhou, M. Identification and Characterization of a Novel Antimicrobial Peptide from the Venom of the Ant *Tetramorium bicarinatum*. *Peptides* **2012**, *38*, 363–370. [[CrossRef](#)] [[PubMed](#)]
35. Téné, N.; Bonnafé, E.; Berger, F.; Rifflet, A.; Guilhaudis, L.; Ségalas-Milazzo, I.; Pipy, B.; Coste, A.; Leprince, J.; Treilhou, M. Biochemical and Biophysical Combined Study of Bicarinalin, an Ant Venom Antimicrobial Peptide. *Peptides* **2016**, *79*, 103–113. [[CrossRef](#)] [[PubMed](#)]
36. Kaur, R.; Stoldt, M.; Jongepier, E.; Feldmeyer, B.; Menzel, F.; Bornberg-Bauer, E.; Foitzik, S. Ant Behaviour and Brain Gene Expression of Defending Hosts Depend on the Ecological Success of the Intruding Social Parasite. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2019**, *374*, 20180192. [[CrossRef](#)] [[PubMed](#)]
37. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)]
38. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The Protein Families Database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [[CrossRef](#)]
39. Sarkar, N. Polyadenylation of mRNA in Prokaryotes. *Annu. Rev. Biochem.* **1997**, *66*, 173–197. [[CrossRef](#)]
40. Wangsanuwat, C.; Heom, K.A.; Liu, E.; O'Malley, M.A.; Dey, S.S. Efficient and Cost-Effective Bacterial mRNA Sequencing from Low Input Samples through Ribosomal RNA Depletion. *BMC Genom.* **2020**, *21*, 717. [[CrossRef](#)]
41. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [[CrossRef](#)]
42. Meher, P.K.; Sahu, T.K.; Saini, V.; Rao, A.R. Predicting Antimicrobial Peptides with Improved Accuracy by Incorporating the Compositional, Physico-Chemical and Structural Features into Chou's General PseAAC. *Sci. Rep.* **2017**, *7*, 42362. [[CrossRef](#)] [[PubMed](#)]
43. Xiao, X.; Wang, P.; Lin, W.-Z.; Jia, J.-H.; Chou, K.-C. IAMP-2L: A Two-Level Multi-Label Classifier for Identifying Antimicrobial Peptides and Their Functional Types. *Anal. Biochem.* **2013**, *436*, 168–177. [[CrossRef](#)] [[PubMed](#)]
44. Veltri, D.; Kamath, U.; Shehu, A. Deep Learning Improves Antimicrobial Peptide Recognition. *Bioinformatics* **2018**, *34*, 2740–2747. [[CrossRef](#)]
45. Das, P.; Wadhawan, K.; Chang, O.; Sercu, T.; Santos, C.D.; Riemer, M.; Chenthamarakshan, V.; Padhi, I.; Mojsilovic, A. PepCVAE: Semi-Supervised Targeted Design of Antimicrobial Peptide Sequences. *arXiv* **2018**, arXiv:1810.07743. [[CrossRef](#)]
46. Dean, S.N.; Alvarez, J.A.E.; Zabetakis, D.; Walper, S.A.; Malanoski, A.P. PepVAE: Variational Autoencoder Framework for Antimicrobial Peptide Generation and Activity Prediction. *Front. Microbiol.* **2021**, *12*, 725727. [[CrossRef](#)] [[PubMed](#)]
47. Szymczak, P.; Możejko, M.; Grzegorzec, T.; Bauer, M.; Neubauer, D.; Michalski, M.; Sroka, J.; Setny, P.; Kamysz, W.; Szczurek, E. HydrAMP: A Deep Generative Model for Antimicrobial Peptide Discovery. *bioRxiv* **2022**. [[CrossRef](#)]
48. Porto, W.F.; Pires, A.S.; Franco, O.L. Computational Tools for Exploring Sequence Databases as a Resource for Antimicrobial Peptides. *Biotechnol. Adv.* **2017**, *35*, 337–349. [[CrossRef](#)]
49. Ramazi, S.; Mohammadi, N.; Allahverdi, A.; Khalili, E.; Abdolmaleki, P. A Review on Antimicrobial Peptides Databases and the Computational Tools. *Database* **2022**, *2022*, baac011. [[CrossRef](#)]
50. Aronica, P.G.A.; Reid, L.M.; Desai, N.; Li, J.; Fox, S.J.; Yadahalli, S.; Essex, J.W.; Verma, C.S. Computational Methods and Tools in Antimicrobial Peptide Research. *J. Chem. Inf. Model.* **2021**, *61*, 3172–3196. [[CrossRef](#)]
51. Cho, J.H.; Sung, B.H.; Kim, S.C. Buforins: Histone H2A-Derived Antimicrobial Peptides from Toad Stomach. *Biochim. Biophys. Acta BBA-Biomembr.* **2009**, *1788*, 1564–1569. [[CrossRef](#)]

52. De Gregorio, E.; Spellman, P.T.; Tzou, P.; Rubin, G.M.; Lemaitre, B. The Toll and Imd Pathways Are the Major Regulators of the Immune Response in *Drosophila*. *EMBO J.* **2002**, *21*, 2568–2579. [[CrossRef](#)] [[PubMed](#)]
53. Guilhelmelli, F.; Vilela, N.; Albuquerque, P.; Derengowski, L.D.S.; Silva-Pereira, I.; Kyaw, C.M. Antibiotic Development Challenges: The Various Mechanisms of Action of Antimicrobial Peptides and of Bacterial Resistance. *Front. Microbiol.* **2013**, *4*, 353. [[CrossRef](#)] [[PubMed](#)]
54. Rodríguez-Rojas, A.; Baeder, D.Y.; Johnston, P.; Regoes, R.R.; Rolff, J. Bacteria Primed by Antimicrobial Peptides Develop Tolerance and Persist. *PLoS Pathog.* **2021**, *17*, e1009443. [[CrossRef](#)]
55. da Cunha, N.B.; Cobacho, N.B.; Viana, J.F.C.; Lima, L.A.; Sampaio, K.B.O.; Dohms, S.S.M.; Ferreira, A.C.R.; de la Fuente-Núñez, C.; Costa, F.F.; Franco, O.L.; et al. The next Generation of Antimicrobial Peptides (AMPs) as Molecular Therapeutic Tools for the Treatment of Diseases with Social and Economic Impacts. *Drug Discov. Today* **2017**, *22*, 234–248. [[CrossRef](#)]
56. Cao, J.; de la Fuente-Núñez, C.; Ou, R.W.; Torres, M.D.T.; Pande, S.G.; Sinskey, A.J.; Lu, T.K. Yeast-Based Synthetic Biology Platform for Antimicrobial Peptide Production. *ACS Synth. Biol.* **2018**, *7*, 896–902. [[CrossRef](#)] [[PubMed](#)]
57. Hazam, P.K.; Goyal, R.; Ramakrishnan, V. Peptide Based Antimicrobials: Design Strategies and Therapeutic Potential. *Prog. Biophys. Mol. Biol.* **2019**, *142*, 10–22. [[CrossRef](#)] [[PubMed](#)]
58. Hirano, M.; Saito, C.; Goto, C.; Yokoo, H.; Kawano, R.; Misawa, T.; Demizu, Y. Rational Design of Helix-Stabilized Antimicrobial Peptide Foldamers Containing  $\alpha,\alpha$ -Disubstituted AAs or Side-Chain Stapling. *ChemPlusChem* **2020**, *85*, 2731–2736. [[CrossRef](#)]
59. Leinonen, R.; Sugawara, H.; Shumway, M.; On behalf of the International Nucleotide Sequence Database Collaboration. Sequence Read Archive. *Nucleic Acids Res.* **2011**, *39*, D19–D21. [[CrossRef](#)]
60. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)]
61. Nip, K.M.; Chiu, R.; Yang, C.; Chu, J.; Mohamadi, H.; Warren, R.L.; Birol, I. RNA-Bloom Enables Reference-Free and Reference-Guided Sequence Assembly for Single-Cell Transcriptomes. *Genome Res.* **2020**, *30*, 1191–1200. [[CrossRef](#)]
62. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods* **2017**, *14*, 417–419. [[CrossRef](#)] [[PubMed](#)]
63. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)] [[PubMed](#)]
64. Johnson, L.S.; Eddy, S.R.; Portugaly, E. Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure. *BMC Bioinform.* **2010**, *11*, 431. [[CrossRef](#)] [[PubMed](#)]
65. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [[CrossRef](#)]
66. Duckert, P.; Brunak, S.; Blom, N. Prediction of Proprotein Convertase Cleavage Sites. *Protein Eng. Des. Sel.* **2004**, *17*, 107–112. [[CrossRef](#)]
67. Wang, X.; Song, Y.; Li, J.; Liu, H.; Xu, X.; Lai, R.; Zhang, K. A New Family of Antimicrobial Peptides from Skin Secretions of *Rana Pleuraden*. *Peptides* **2007**, *28*, 2069–2074. [[CrossRef](#)]
68. Yi, H.-Y.; Chowdhury, M.; Huang, Y.-D.; Yu, X.-Q. Insect Antimicrobial Peptides and Their Applications. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 5807–5822. [[CrossRef](#)]
69. Jiang, Z.; Vasil, A.I.; Hale, J.D.; Hancock, R.E.W.; Vasil, M.L.; Hodges, R.S. Effects of Net Charge and the Number of Positively Charged Residues on the Biological Activity of Amphipathic Alpha-Helical Cationic Antimicrobial Peptides. *Biopolymers* **2008**, *90*, 369–383. [[CrossRef](#)]
70. Hart, A.J.; Ginzburg, S.; Xu, M.; Fisher, C.R.; Rahmatpour, N.; Mitton, J.B.; Paul, R.; Wegryzn, J.L. EnTAP: Bringing Faster and Smarter Functional Annotation to Non-model Eukaryotic Transcriptomes. *Mol. Ecol. Resour.* **2020**, *20*, 591–604. [[CrossRef](#)]
71. The UniProt Consortium; Bateman, A.; Martin, M.-J.; Orchard, S.; Magrane, M.; Agivetova, R.; Ahmad, S.; Alpi, E.; Bowler-Barnett, E.H.; Britto, R.; et al. UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489. [[CrossRef](#)]
72. O’Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciuffo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [[CrossRef](#)] [[PubMed](#)]
73. Slater, G.; Birney, E. Automated Generation of Heuristics for Biological Sequence Comparison. *BMC Bioinform.* **2005**, *6*, 31. [[CrossRef](#)]
74. Adamczak, R.; Porollo, A.; Meller, J. Combining Prediction of Secondary Structure and Solvent Accessibility in Proteins. *Proteins* **2005**, *59*, 467–475. [[CrossRef](#)] [[PubMed](#)]
75. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)]
76. Wiegand, I.; Hilpert, K.; Hancock, R.E.W. Agar and Broth Dilution Methods to Determine the Minimal Inhibitory Concentration (MIC) of Antimicrobial Substances. *Nat. Protoc.* **2008**, *3*, 163–175. [[CrossRef](#)] [[PubMed](#)]
77. Sanchez, E.; Rodríguez, A.; Grau, J.H.; Lötters, S.; Künzel, S.; Saporito, R.A.; Ringler, E.; Schulz, S.; Wollenberg Valero, K.C.; Vences, M. Transcriptomic Signatures of Experimental Alkaloid Consumption in a Poison Frog. *Genes* **2019**, *10*, 733. [[CrossRef](#)]

78. Siu-Ting, K.; Torres-Sánchez, M.; San Mauro, D.; Wilcockson, D.; Wilkinson, M.; Pisani, D.; O'Connell, M.J.; Creevey, C.J. Inadvertent Paralog Inclusion Drives Artifactual Topologies and Timetree Estimates in Phylogenomics. *Mol. Biol. Evol.* **2019**, *36*, 1344–1356. [[CrossRef](#)]
79. Xia, Y.; Luo, W.; Yuan, S.; Zheng, Y.; Zeng, X. Microsatellite Development from Genome Skimming and Transcriptome Sequencing: Comparison of Strategies and Lessons from Frog Species. *BMC Genom.* **2018**, *19*, 886. [[CrossRef](#)]
80. Fan, W.; Jiang, Y.; Zhang, M.; Yang, D.; Chen, Z.; Sun, H.; Lan, X.; Yan, F.; Xu, J.; Yuan, W. Comparative Transcriptome Analyses Reveal the Genetic Basis Underlying the Immune Function of Three Amphibians' Skin. *PLoS ONE* **2017**, *12*, e0190023. [[CrossRef](#)]
81. Reilly, B.D.; Schlipalius, D.I.; Cramp, R.L.; Ebert, P.R.; Franklin, C.E. Frogs and Estivation: Transcriptional Insights into Metabolism and Cell Survival in a Natural Model of Extended Muscle Disuse. *Physiol. Genom.* **2013**, *45*, 377–388. [[CrossRef](#)]
82. Liscano Martinez, Y.; Arenas Gómez, C.M.; Smith, J.; Delgado, J.P. A Tree Frog (Boana Pugnax) Dataset of Skin Transcriptome for the Identification of Biomolecules with Potential Antimicrobial Activities. *Data Brief* **2020**, *32*, 106084. [[CrossRef](#)] [[PubMed](#)]
83. Grogan, L.F.; Mulvenna, J.; Gummer, J.P.A.; Scheele, B.C.; Berger, L.; Cashins, S.D.; McFadden, M.S.; Harlow, P.; Hunter, D.A.; Tregrove, R.D.; et al. Survival, Gene and Metabolite Responses of *Litoria Verreauxii* Alpina Frogs to Fungal Disease Chytridiomycosis. *Sci. Data* **2018**, *5*, 180033. [[CrossRef](#)] [[PubMed](#)]
84. Qiao, L.; Yang, W.; Fu, J.; Song, Z. Transcriptome Profile of the Green Odorous Frog (*Odorrana Margaretae*). *PLoS ONE* **2013**, *8*, e75211. [[CrossRef](#)] [[PubMed](#)]
85. Chang, L.; Zhu, W.; Shi, S.; Zhang, M.; Jiang, J.; Li, C.; Xie, F.; Wang, B. Plateau Grass and Greenhouse Flower? Distinct Genetic Basis of Closely Related Toad Tadpoles Respectively Adapted to High Altitude and Karst Caves. *Genes* **2020**, *11*, 123. [[CrossRef](#)]
86. Caty, S.N.; Alvarez-Buylla, A.; Byrd, G.D.; Vidoudez, C.; Roland, A.B.; Tapia, E.E.; Budnik, B.; Trauger, S.A.; Coloma, L.A.; O'Connell, L.A. Molecular Physiology of Chemical Defenses in a Poison Frog. *J. Exp. Biol.* **2019**, *222*, jeb.204149. [[CrossRef](#)]
87. Shu, Y.; Xia, J.; Yu, Q.; Wang, G.; Zhang, J.; He, J.; Wang, H.; Zhang, L.; Wu, H. Integrated Analysis of mRNA and miRNA Expression Profiles Reveals Muscle Growth Differences between Adult Female and Male Chinese Concave-Eared Frogs (*Odorrana Tormota*). *Gene* **2018**, *678*, 241–251. [[CrossRef](#)]
88. Yoshida, N.; Kaito, C. Dataset for de Novo Transcriptome Assembly of the African Bullfrog *Pyxicephalus Adspersus*. *Data Brief* **2020**, *30*, 105388. [[CrossRef](#)]
89. Bossuyt, F.; Schulte, L.M.; Maex, M.; Janssenswillen, S.; Novikova, P.Y.; Biju, S.D.; Van de Peer, Y.; Matthijs, S.; Roelants, K.; Martel, A.; et al. Multiple Independent Recruitment of Sodefrin Precursor-Like Factors in Anuran Sexually Dimorphic Glands. *Mol. Biol. Evol.* **2019**, *36*, 1921–1930. [[CrossRef](#)]
90. Zhang, Y.; Li, Y.; Qin, Z.; Wang, H.; Li, J. A Screening Assay for Thyroid Hormone Signaling Disruption Based on Thyroid Hormone-Response Gene Expression Analysis in the Frog *Pelophylax Nigromaculatus*. *J. Environ. Sci.* **2015**, *34*, 143–154. [[CrossRef](#)]
91. Eskew, E.A.; Shock, B.C.; LaDouceur, E.E.B.; Keel, K.; Miller, M.R.; Foley, J.E.; Todd, B.D. Gene Expression Differs in Susceptible and Resistant Amphibians Exposed to *Batrachochytrium Dendrobatidis*. *R. Soc. Open sci.* **2018**, *5*, 170910. [[CrossRef](#)]
92. Stuckert, A.M.M.; Chouteau, M.; McClure, M.; LaPolice, T.M.; Linderoth, T.; Nielsen, R.; Summers, K.; MacManes, M.D. The Genomics of Mimicry: Gene Expression throughout Development Provides Insights into Convergent and Divergent Phenotypes in a Müllerian Mimicry System. *Mol. Ecol.* **2021**, *30*, 4039–4061. [[CrossRef](#)] [[PubMed](#)]
93. Christenson, M.K.; Trease, A.J.; Potluri, L.-P.; Jezewski, A.J.; Davis, V.M.; Knight, L.A.; Kolok, A.S.; Davis, P.H. De Novo Assembly and Analysis of the Northern Leopard Frog *Rana Pipiens* Transcriptome. *J. Genom.* **2014**, *2*, 141–149. [[CrossRef](#)] [[PubMed](#)]
94. Price, S.J.; Garner, T.W.J.; Balloux, F.; Ruis, C.; Paszkiewicz, K.H.; Moore, K.; Griffiths, A.G.F. A de Novo Assembly of the Common Frog (*Rana Temporaria*) Transcriptome and Comparison of Transcription Following Exposure to Ranavirus and *Batrachochytrium Dendrobatidis*. *PLoS ONE* **2015**, *10*, e0130500. [[CrossRef](#)]
95. Furman, B.L.S.; Evans, B.J. Sequential Turnovers of Sex Chromosomes in African Clawed Frogs (*Xenopus*) Suggest Some Genomic Regions Are Good at Sex Determination. *G3 (Bethesda)* **2016**, *6*, 3625–3633. [[CrossRef](#)] [[PubMed](#)]
96. Birol, I.; Behsaz, B.; Hammond, S.A.; Kucuk, E.; Veldhoen, N.; Helbing, C.C. De Novo Transcriptome Assemblies of *Rana (Lithobates) Catesbeiana* and *Xenopus Laevis* Tadpole Livers for Comparative Genomics without Reference Genomes. *PLoS ONE* **2015**, *10*, e0130720. [[CrossRef](#)]
97. Barbosa-Morais, N.L.; Irimia, M.; Pan, Q.; Xiong, H.Y.; Gueroussov, S.; Lee, L.J.; Slobodeniuc, V.; Kutter, C.; Watt, S.; Colak, R.; et al. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science* **2012**, *338*, 1587–1593. [[CrossRef](#)]
98. Arvidson, R.; Kaiser, M.; Lee, S.S.; Urenda, J.-P.; Dail, C.; Mohammed, H.; Nolan, C.; Pan, S.; Stajich, J.E.; Libersat, F.; et al. Parasitoid Jewel Wasp Mounts Multipronged Neurochemical Attack to Hijack a Host Brain. *Mol. Cell. Proteom.* **2019**, *18*, 99–114. [[CrossRef](#)]
99. Yek, S.H.; Boomsma, J.J.; Schiøtt, M. Differential Gene Expression in *Acromyrmex* Leaf-Cutting Ants after Challenges with Two Fungal Pathogens. *Mol. Ecol.* **2013**, *22*, 2173–2187. [[CrossRef](#)]
100. Yoon, K.A.; Kim, K.; Kim, W.-J.; Bang, W.Y.; Ahn, N.-H.; Bae, C.-H.; Yeo, J.-H.; Lee, S.H. Characterization of Venom Components and Their Phylogenetic Properties in Some Aculeate Bumblebees and Wasps. *Toxins* **2020**, *12*, 47. [[CrossRef](#)]
101. McNamara-Bordewick, N.K.; McKinstry, M.; Snow, J.W. Robust Transcriptional Response to Heat Shock Impacting Diverse Cellular Processes despite Lack of Heat Shock Factor in Microsporidia. *mSphere* **2019**, *4*, e00219-19. [[CrossRef](#)]
102. Becchimanzi, A.; Avolio, M.; Bostan, H.; Colantuono, C.; Cozzolino, F.; Mancini, D.; Chiusano, M.L.; Pucci, P.; Caccia, S.; Pennacchio, F. Venomics of the Ectoparasitoid Wasp *Bracon Nigricans*. *BMC Genom.* **2020**, *21*, 34. [[CrossRef](#)]

103. de Bekker, C.; Ohm, R.A.; Loreto, R.G.; Sebastian, A.; Albert, I.; Meroow, M.; Brachmann, A.; Hughes, D.P. Gene Expression during Zombie Ant Biting Behavior Reflects the Complexity Underlying Fungal Parasitic Behavioral Manipulation. *BMC Genom.* **2015**, *16*, 620. [[CrossRef](#)] [[PubMed](#)]
104. von Wyszczetki, K.; Lowack, H.; Heinze, J. Transcriptomic Response to Injury Sheds Light on the Physiological Costs of Reproduction in Ant Queens. *Mol. Ecol.* **2016**, *25*, 1972–1985. [[CrossRef](#)]
105. Zhao, W.; Shi, M.; Ye, X.; Li, F.; Wang, X.; Chen, X. Comparative Transcriptome Analysis of Venom Glands from *Cotesia Vestalis* and *Diadromus Collaris*, Two Endoparasitoids of the Host *Plutella Xylostella*. *Sci. Rep.* **2017**, *7*, 1298. [[CrossRef](#)]
106. Coffman, K.A.; Harrell, T.C.; Burke, G.R. A Mutualistic Poxvirus Exhibits Convergent Evolution with Other Heritable Viruses in Parasitoid Wasps. *J. Virol.* **2020**, *94*, e02059-19. [[CrossRef](#)] [[PubMed](#)]
107. Burke, G.R.; Strand, M.R. Systematic Analysis of a Wasp Parasitism Arsenal. *Mol. Ecol.* **2014**, *23*, 890–901. [[CrossRef](#)] [[PubMed](#)]
108. Robinson, S.D.; Mueller, A.; Clayton, D.; Starobova, H.; Hamilton, B.R.; Payne, R.J.; Vetter, I.; King, G.F.; Undheim, E.A.B. A Comprehensive Portrait of the Venom of the Giant Red Bull Ant, *Myrmecia Gulosa*, Reveals a Hyperdiverse Hymenopteran Toxin Gene Family. *Sci. Adv.* **2018**, *4*, eaau4640. [[CrossRef](#)]
109. Martinson, E.O.; Mrinalini; Kelkar, Y.D.; Chang, C.-H.; Werren, J.H. The Evolution of Venom by Co-Option of Single-Copy Genes. *Curr. Biol.* **2017**, *27*, 2007–2013. [[CrossRef](#)]
110. Cook, N.; Boulton, R.A.; Green, J.; Trivedi, U.; Tauber, E.; Pannebakker, B.A.; Ritchie, M.G.; Shuker, D.M. Differential Gene Expression Is Not Required for Facultative Sex Allocation: A Transcriptome Analysis of Brain Tissue in the Parasitoid Wasp *Nasonia vitripennis*. *R. Soc. Open sci.* **2018**, *5*, 171718. [[CrossRef](#)]
111. Sim, A.D.; Wheeler, D. The Venom Gland Transcriptome of the Parasitoid Wasp *Nasonia Vitripennis* Highlights the Importance of Novel Genes in Venom Function. *BMC Genom.* **2016**, *17*, 571. [[CrossRef](#)]
112. Kazuma, K.; Masuko, K.; Konno, K.; Inagaki, H. Combined Venom Gland Transcriptomic and Venom Peptidomic Analysis of the Predatory Ant *Odontomachus Monticola*. *Toxins* **2017**, *9*, 323. [[CrossRef](#)] [[PubMed](#)]
113. Smith, C.R.; Helms Cahan, S.; Kemena, C.; Brady, S.G.; Yang, W.; Bornberg-Bauer, E.; Eriksson, T.; Gadau, J.; Helmkampf, M.; Gotzek, D.; et al. How Do Genomes Create Novel Phenotypes? Insights from the Loss of the Worker Caste in Ant Social Parasites. *Mol. Biol. Evol.* **2015**, *32*, 2919–2931. [[CrossRef](#)] [[PubMed](#)]
114. Özbek, R.; Wielsch, N.; Vogel, H.; Lochnit, G.; Foerster, F.; Vilcinskas, A.; von Reumont, B.M. Proteo-Transcriptomic Characterization of the Venom from the Endoparasitoid Wasp *Pimpla Turionellae* with Aspects on Its Biology and Evolution. *Toxins* **2019**, *11*, 721. [[CrossRef](#)] [[PubMed](#)]
115. Yang, L.; Yang, Y.; Liu, M.-M.; Yan, Z.-C.; Qiu, L.-M.; Fang, Q.; Wang, F.; Werren, J.H.; Ye, G.-Y. Identification and Comparative Analysis of Venom Proteins in a Pupal Ectoparasitoid, *Pachycrepoideus Vindemniae*. *Front. Physiol.* **2020**, *11*, 9. [[CrossRef](#)]
116. Bouzid, W.; Verdenaud, M.; Klopp, C.; Ducancel, F.; Noirot, C.; Vétillard, A. De Novo Sequencing and Transcriptome Analysis for *Tetramorium Bicarinatum*: A Comprehensive Venom Gland Transcriptome Analysis from an Ant Species. *BMC Genom.* **2014**, *15*, 987. [[CrossRef](#)]
117. Negrone, M.A.; Foitzik, S.; Feldmeyer, B. Long-Lived *Temnothorax* Ant Queens Switch from Investment in Immunity to Antioxidant Production with Age. *Sci. Rep.* **2019**, *9*, 7270. [[CrossRef](#)]