



MSc Thesis

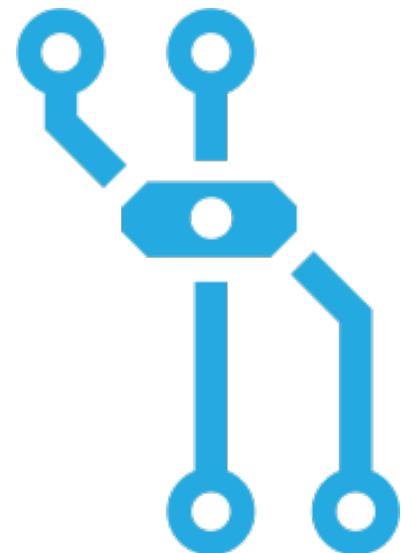


High throughput *in silico* discovery of antimicrobial peptides in amphibian and insect transcriptomes

Diana Lin

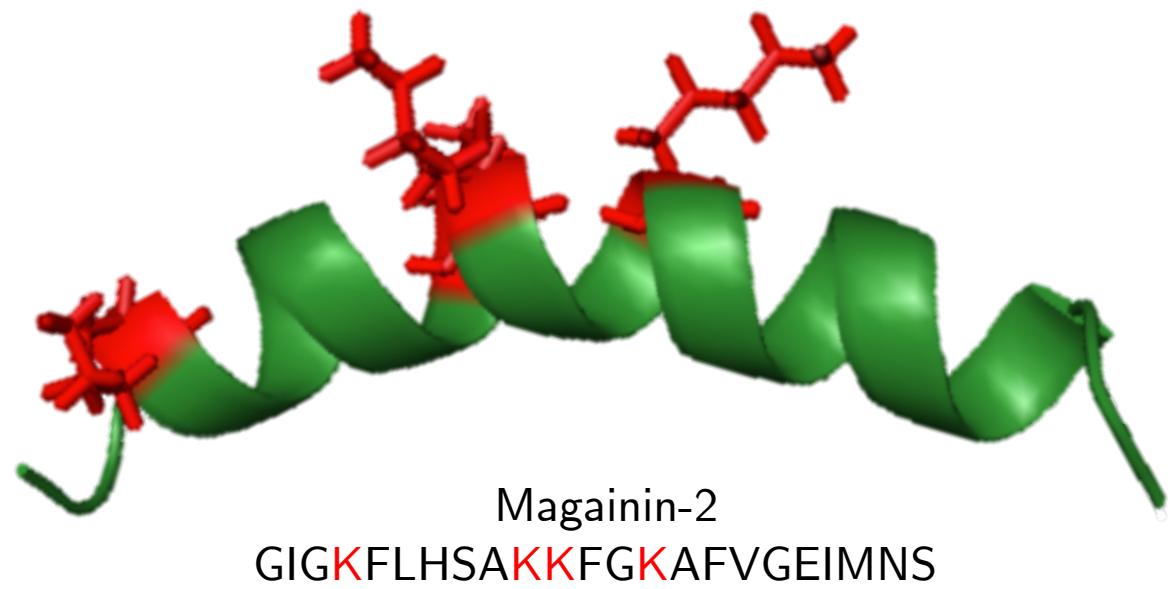
MSc, Bioinformatics, University of British Columbia (UBC)

Dr. Inanc Birol Lab, Genome Sciences Centre (GSC),
BC Cancer

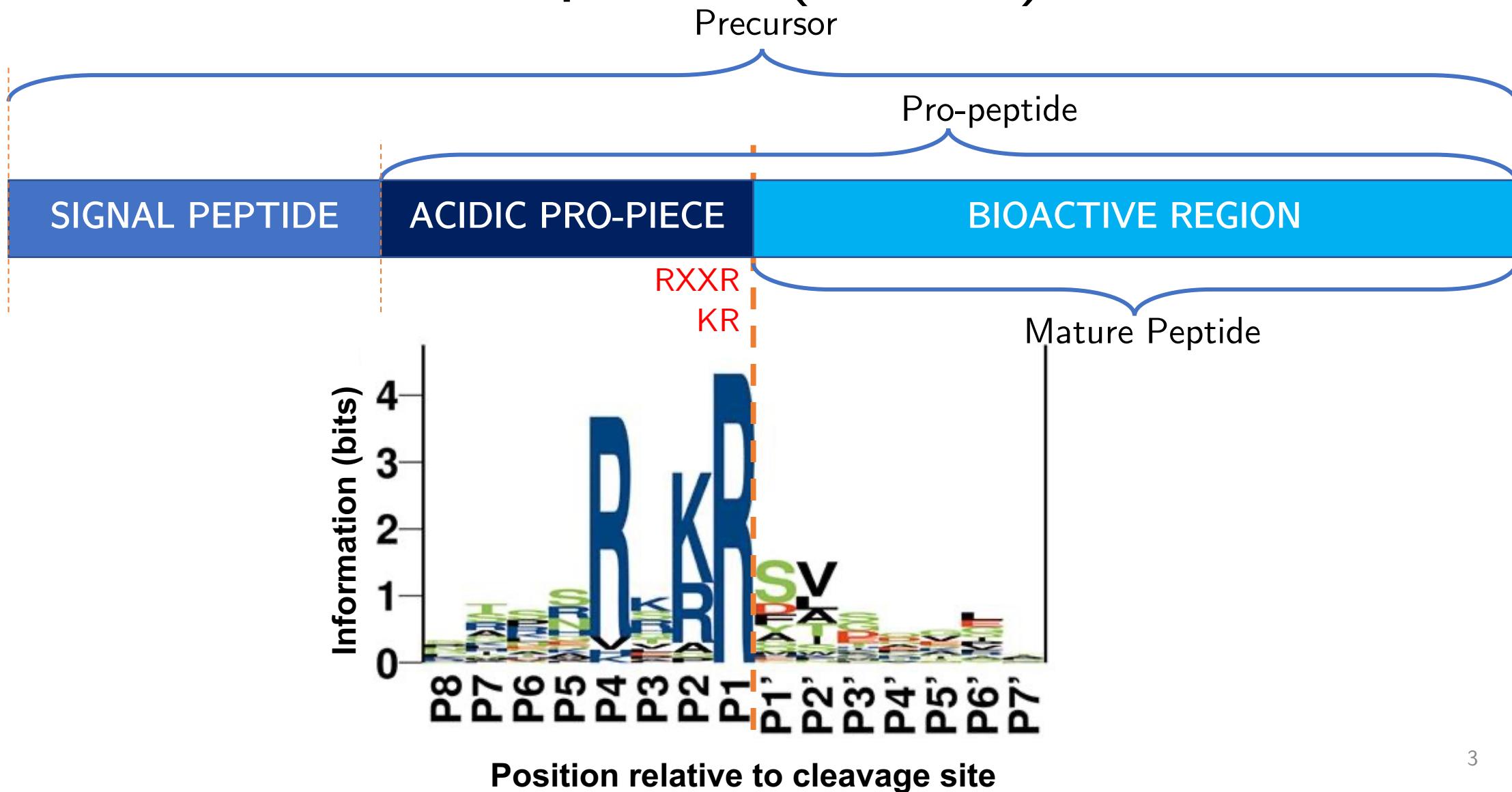


Antimicrobial Peptides (AMPs)

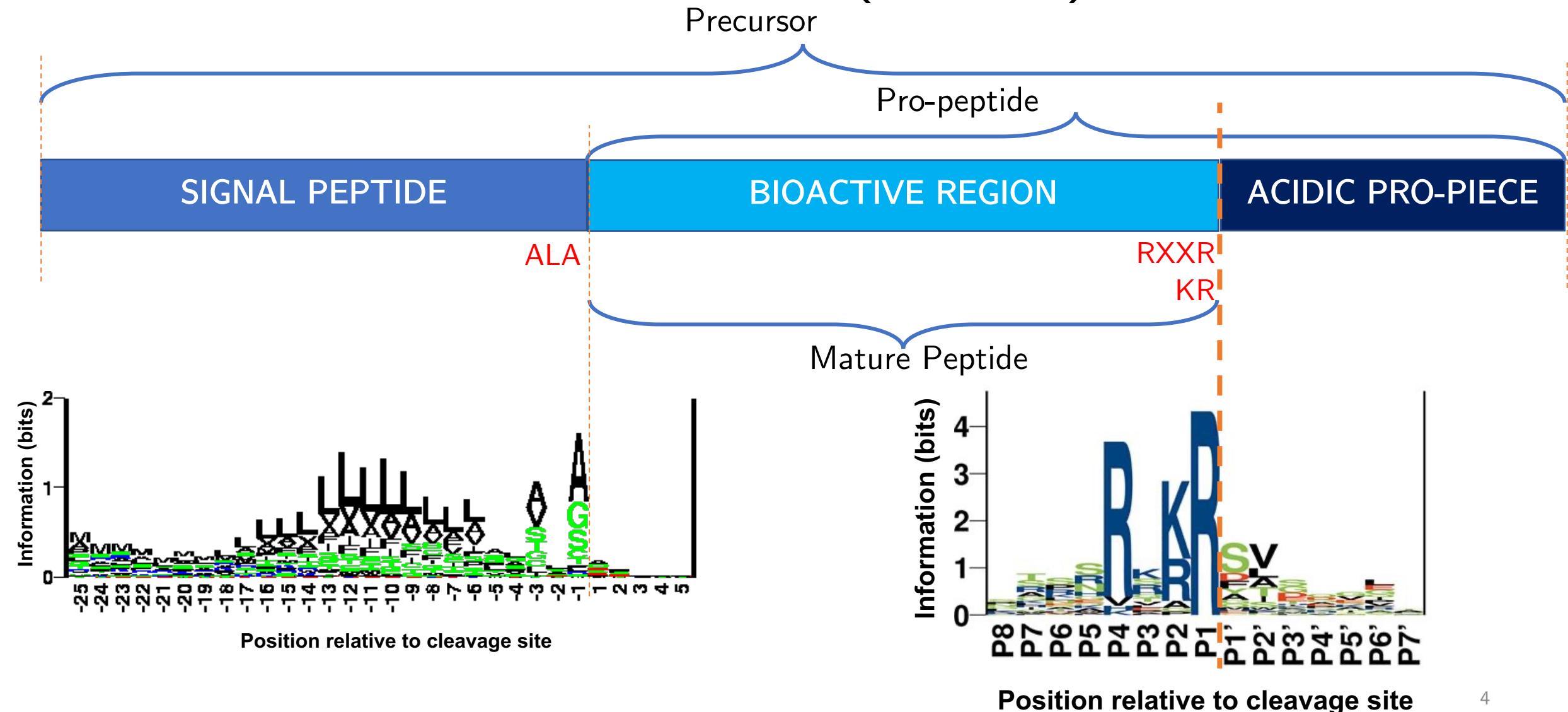
- Short peptide sequences (5 to 50 amino acids)
- Often positively charged
- Amphipathic
- Produced by all multicellular life forms
- Part of the innate immune system



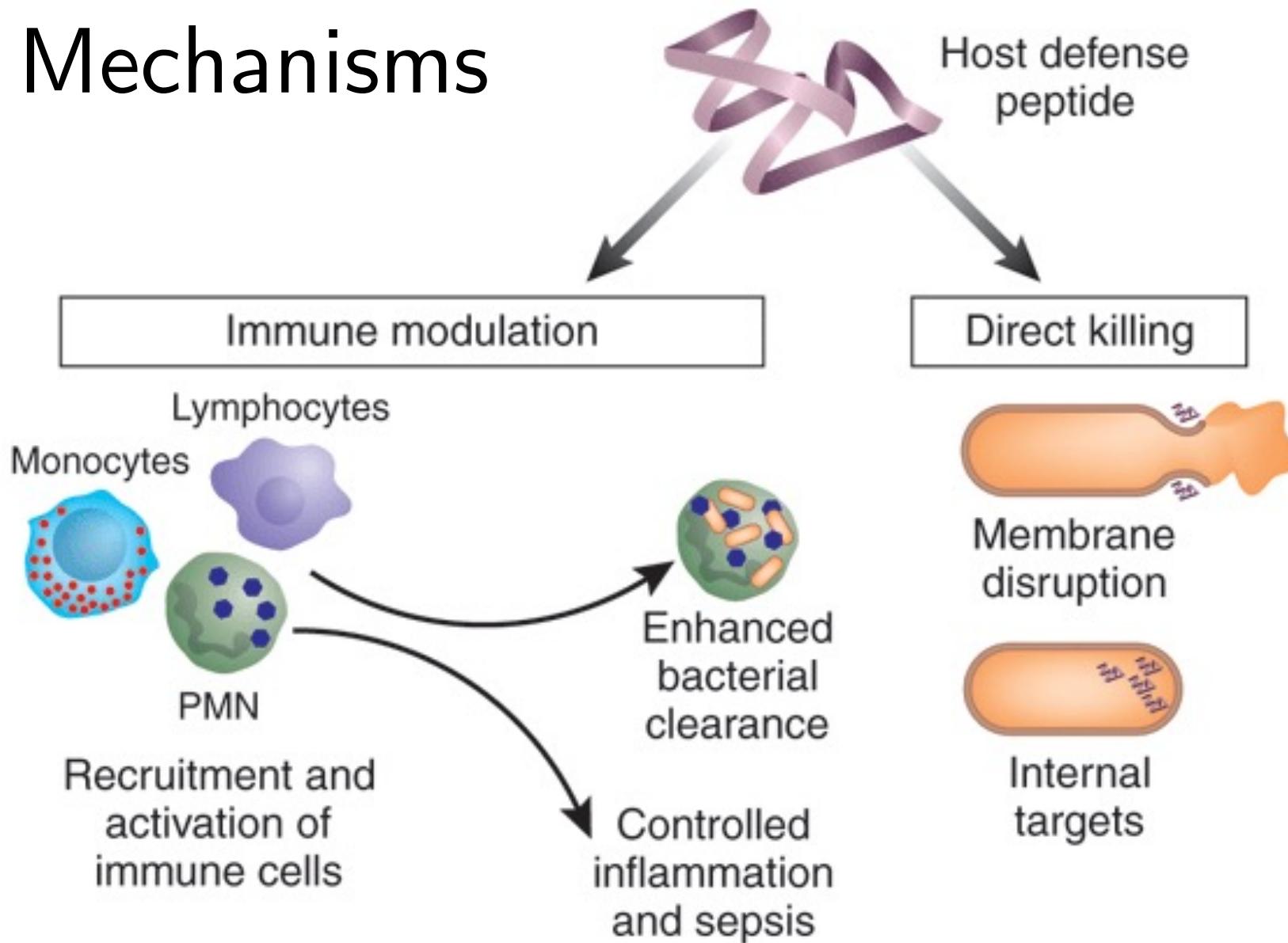
Antimicrobial Peptides (AMPs)



Antimicrobial Peptides (AMPs)



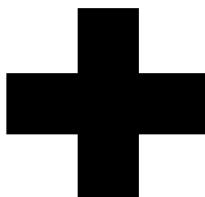
AMP Mechanisms



Motivation

FAST

emergence and spread
of multi-drug
resistance



SLOW

pace of development
of *novel* classes of
antibiotics



THE DISCOVERY VOID

“The [current antibiotic discovery] pipeline is insufficient to counter the rising resistance in these priority bacteria.”

Dr. Peter Beyer, Senior Advisor,
WHO Antimicrobial Resistance Division

rAMPage: Rapid Anti-Microbial Peptide Annotation and Gene Estimation



- ✓ Developed and executed an AMP discovery pipeline
 - ✓ Mined for AMP sequences in publicly available genomic resources

- ✓ Found candidate AMPs suitable for synthesis and wet lab validation





RNA-seq Datasets



Frogs & Toads (38)

Frogs (33)

<i>A. femoralis</i>	<i>P. adspersus</i>	<i>R. sirensis</i>
<i>A. mantzorum</i>	<i>P. amboli</i>	<i>R. sylvatica</i>
<i>A. petersi</i>	<i>P. megacephalus</i>	<i>R. temporaria</i>
<i>C. alboguttata</i>	<i>P. microps</i>	<i>S. ruber</i>
<i>D. auratus</i>	<i>P. nigromaculatus</i>	<i>X. allofraseri</i>
<i>D. leucomelas</i>	<i>P. toftae</i>	<i>X. borealis</i>
<i>D. tinctorius</i>	<i>Q. boulengeri</i>	<i>X. laevis</i>
<i>H. pugnax</i>	<i>R. catesbeiana</i>	<i>X. largeni</i>
<i>L. verreauxii</i>	<i>R. dennysi</i>	<i>X. tropicalis</i>
<i>O. margaretae</i>	<i>R. imitator</i>	
<i>O. sylvatica</i>	<i>R. omeimontis</i>	
<i>O. tormota</i>	<i>R. pipiens</i>	



Toads (5)

<i>A. minuta</i>
<i>B. gargarizans</i>
<i>L. Boringii</i>
<i>M. sangzhiensis</i>
<i>O. rhodostigmatus</i>



Ants, Bees, & Wasps (37)

Ants (8)

<i>A. echinatior</i>
<i>C. castaneus</i>
<i>C. obscurior</i>
<i>M. gulosa</i>
<i>O. monticola</i>
<i>P. barbatus</i>
<i>T. bicarinatum</i>
<i>T. rugulatus</i>

Bees (5)

<i>A. cerana</i>
<i>A. mellifera</i>
<i>B. ardens</i>
<i>B. consobrinus</i>
<i>B. ussurensis</i>



Wasps (24)

<i>A. compressa</i>	<i>P. snelleni</i>
<i>A. flavomarginatum</i>	<i>P. turionellae</i>
<i>B. nigricans</i>	<i>P. varia</i>
<i>C. vestalis</i>	<i>P. vindemmiae</i>
<i>D. collaris</i>	<i>S. deformae</i>
<i>D. longicaudata</i>	<i>S. kj8906</i>
<i>M. demolitor</i>	<i>T. sarcophagae</i>
<i>N. giraulti</i>	<i>U. rufipes</i>
<i>N. vitripennis</i>	<i>V. analis</i>
<i>N. vitripennis x N. giraulti</i>	<i>V. crabro</i>
<i>O. decorates</i>	<i>V. dybowskii</i>
<i>P. rothneyi</i>	<i>V. similis</i>

rAMPage Pipeline: Overview

<https://github.com/bcgsc/rAMPPage>



✓ RNA-Bloom
<https://github.com/bcgsc/RNA-Bloom>

✓ TransDecoder

✓ ProP

- ✓ E_NTAP
- ✓ Exonerate
- ✓ SABLE

Process Reads

✓ fastp

Assemble Transcripts

Expression Filtering

✓ Salmon

in silico Translation

Homology Search

✓ HMMER

Cleave Precursors

Prioritize AMPs

Characterize AMPs

✓ TransDecoder

✓ ProP

- ✓ E_NTAP
- ✓ Exonerate
- ✓ SABLE

✓ AMPlify

<https://github.com/bcgsc/AMPlify>

Assembly: RNA-Bloom



- *De novo* transcriptome assembly with single and paired-end reads
- Reference-guided assembly if reference or draft transcriptome available
- Multi-sample pooled assembly support



Homology Search

- 4,663 (nr) amphibians; 1,204 insect (nr) known “reference” AMPs



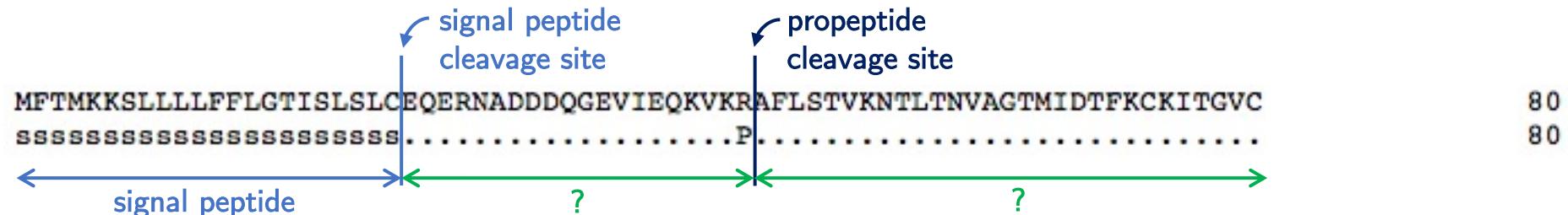
APD3: Antimicrobial Peptide Database 3, <https://aps.unmc.edu>

DADP: Database of Anuran Defense Peptides, <http://split4.pmfst.hr/dadp/>

Cleaving Precursor AMPs

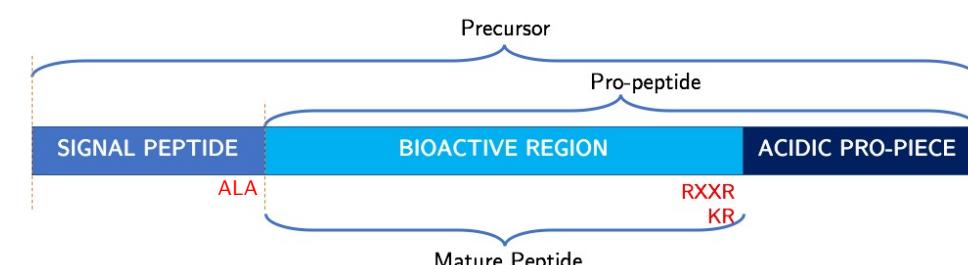
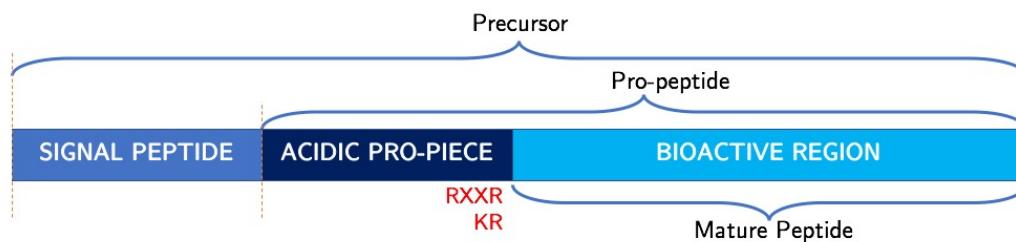
>rcatesbeiana-back_skin-139086.p2

MFTMKKSLLLLFFLGTISLSLCEQERNADDDQGEVIEQKV**KR**AFLSTVKNTLTNVAGTMIDTFKCKITGVC

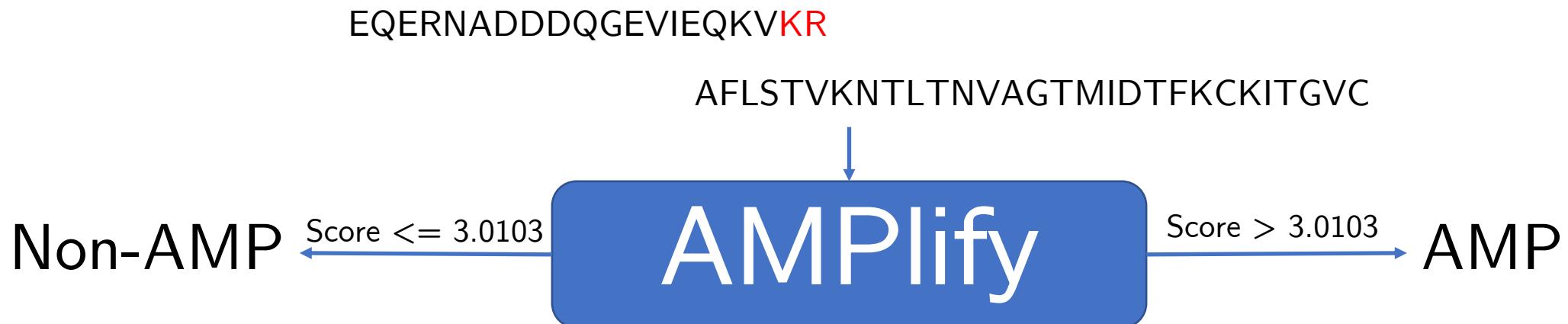


Signal peptide cleavage site predicted: between pos. 22 and 23: SLC-EQ

Propeptide cleavage sites predicted: Arg(R)/Lys(K): 1



Prioritizing AMPs: AMPlify



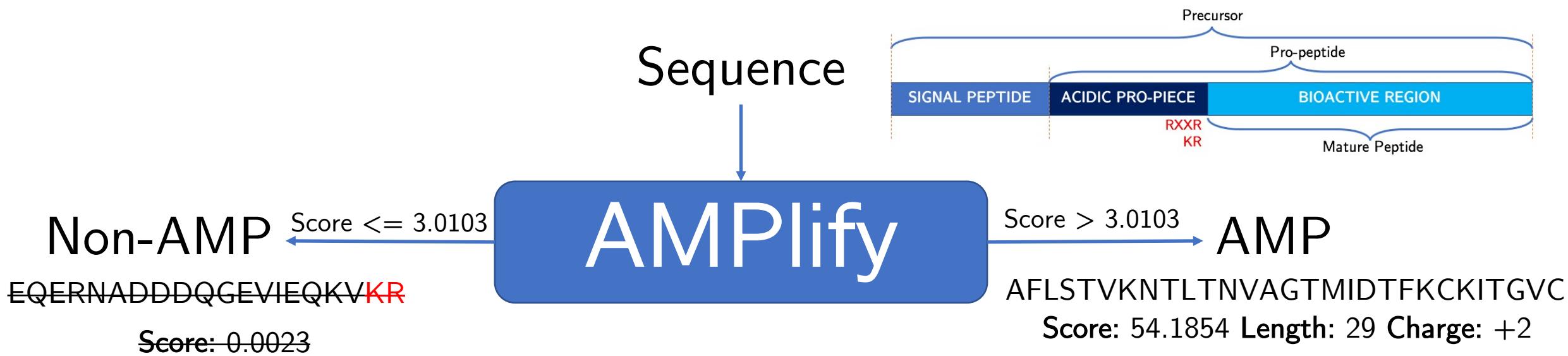
AMPlify Score > 3.0103

Length <= 50aa

Charge >= +1

AMPlify Score = $-10 \log_{10}(1-P)$

Prioritizing AMPs: AMPlify

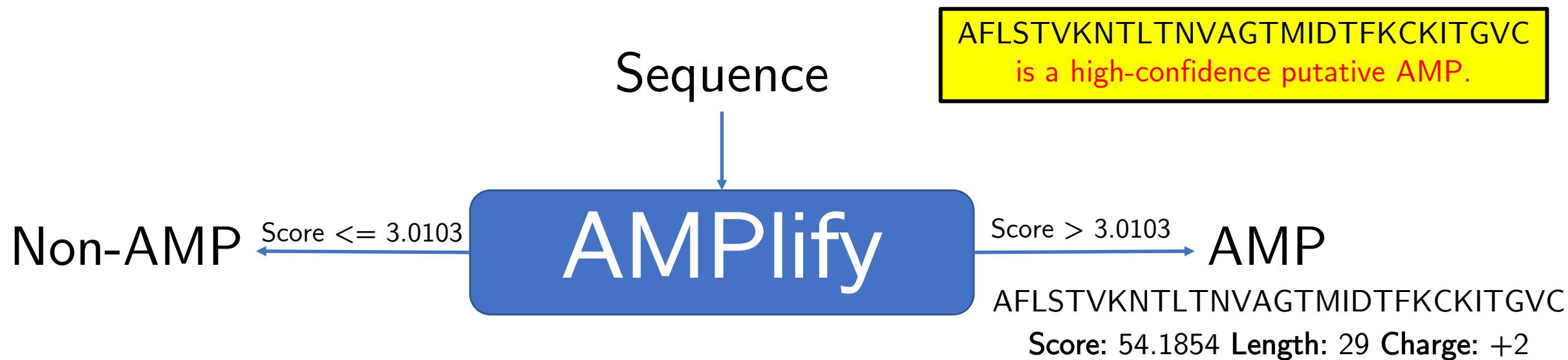


AMPlify Score > 3.0103

Length <= 50aa

Charge >= +1

Prioritizing AMPs: AMPlify



AMPlify Score $> 3.0103 \longrightarrow$ AMPlify Score $\geq 10.0 / 7.0$
Amphibians Insects

Length $\leq 50\text{aa} \longrightarrow$ Length $\leq 30\text{aa}$

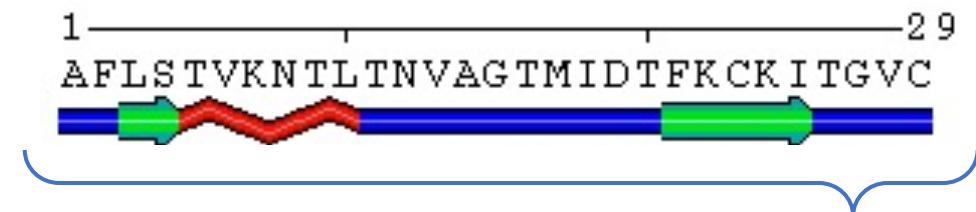
Charge $\geq +1 \longrightarrow$ Charge $\geq +2$

Characterizing AMPs

- Alignment to our “reference” AMPs
- Alignment to orthology, domain, and protein databases
- Secondary structure prediction

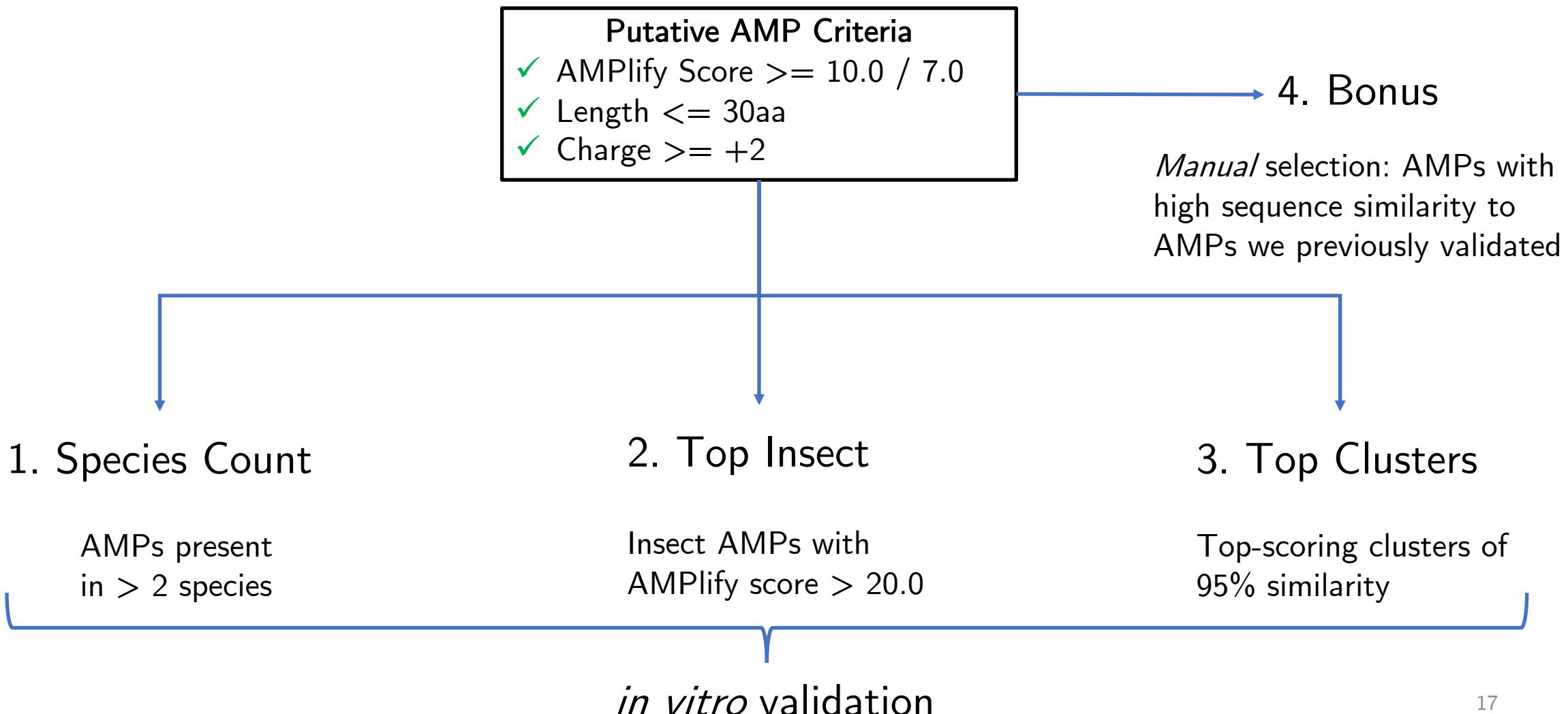
Putative AMP Criteria

- ✓ AMPlify Score $\geq 10.0 / 7.0$
- ✓ Length $\leq 30\text{aa}$
- ✓ Charge $\geq +2$



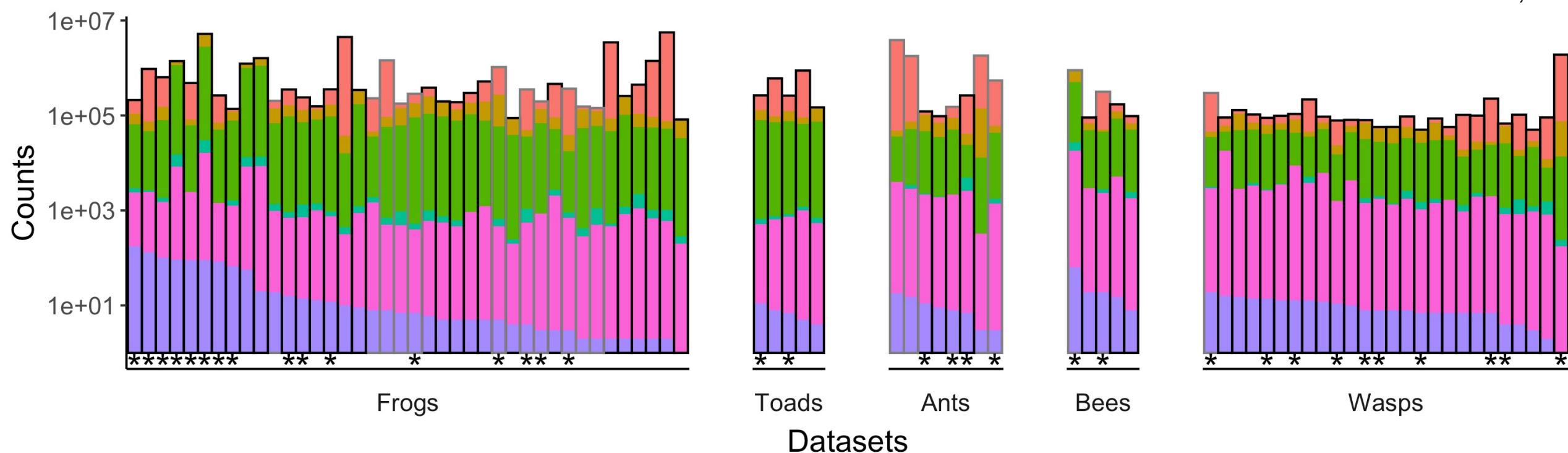
AMPlify Score	Length	Charge	Top Precursor	Top Mature	GO Term	IPScan	Pfam	Secondary Structure
54.1854	29	2	ABB89058.1: ranatuerin 2CHb precursor, partial [Rana chiricahuensis] (75.86%)	AP00617: Palustrin-2 (UCLL1a; XXU; 1S=S, frog, amphibians, animals) (75.86%)	GO:0005576 (extracellular region) GO:0006952 (defense response)	IPR012521 (Frog antimicrobial peptide, brevinin-2/esculentin type)	PF08023 (Frog antimicrobial peptide)	CCEEHHHH HHCCCCCC CCCCEEEE ECCCC

Selecting AMPs for Synthesis

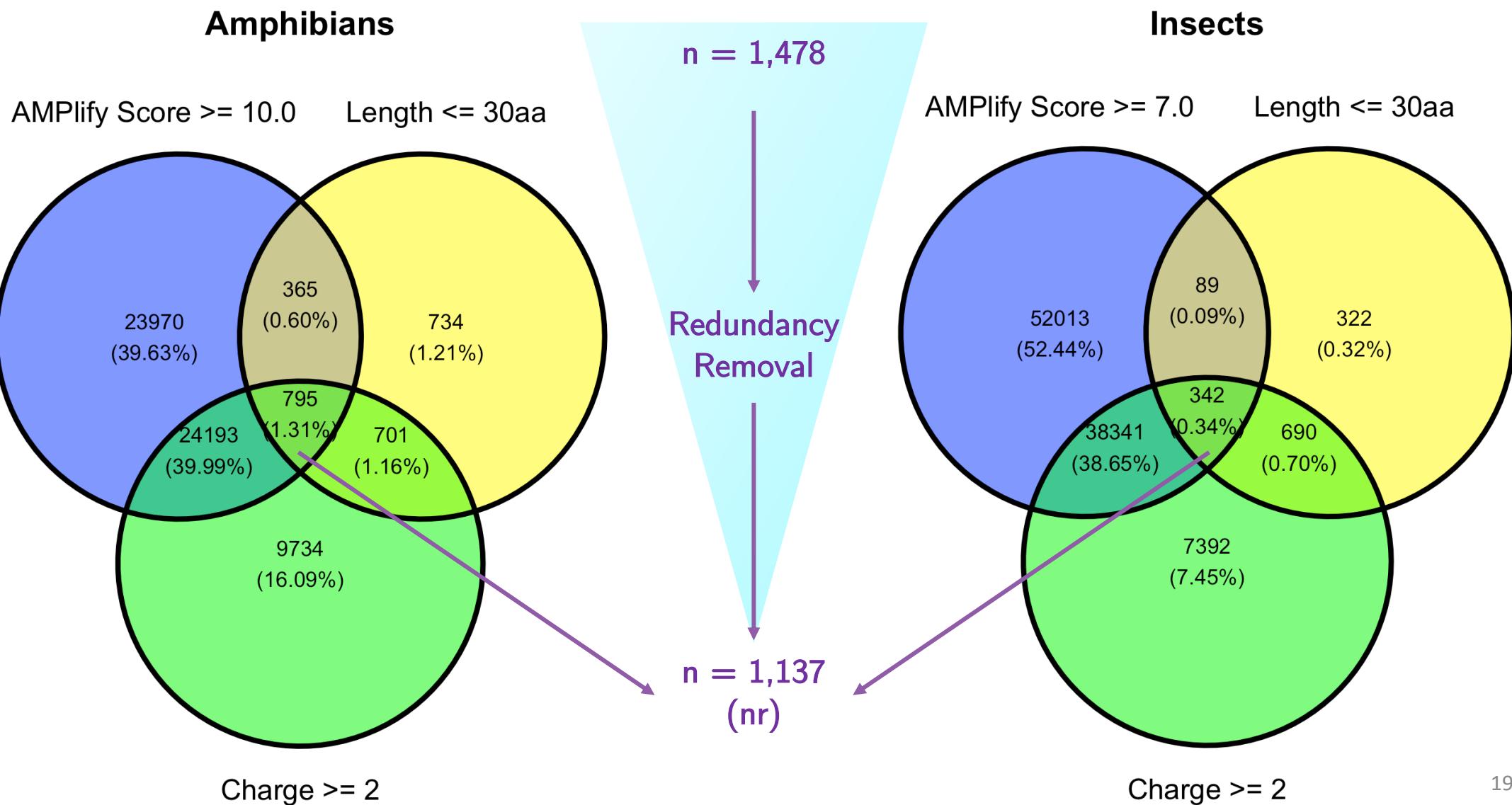


Overview

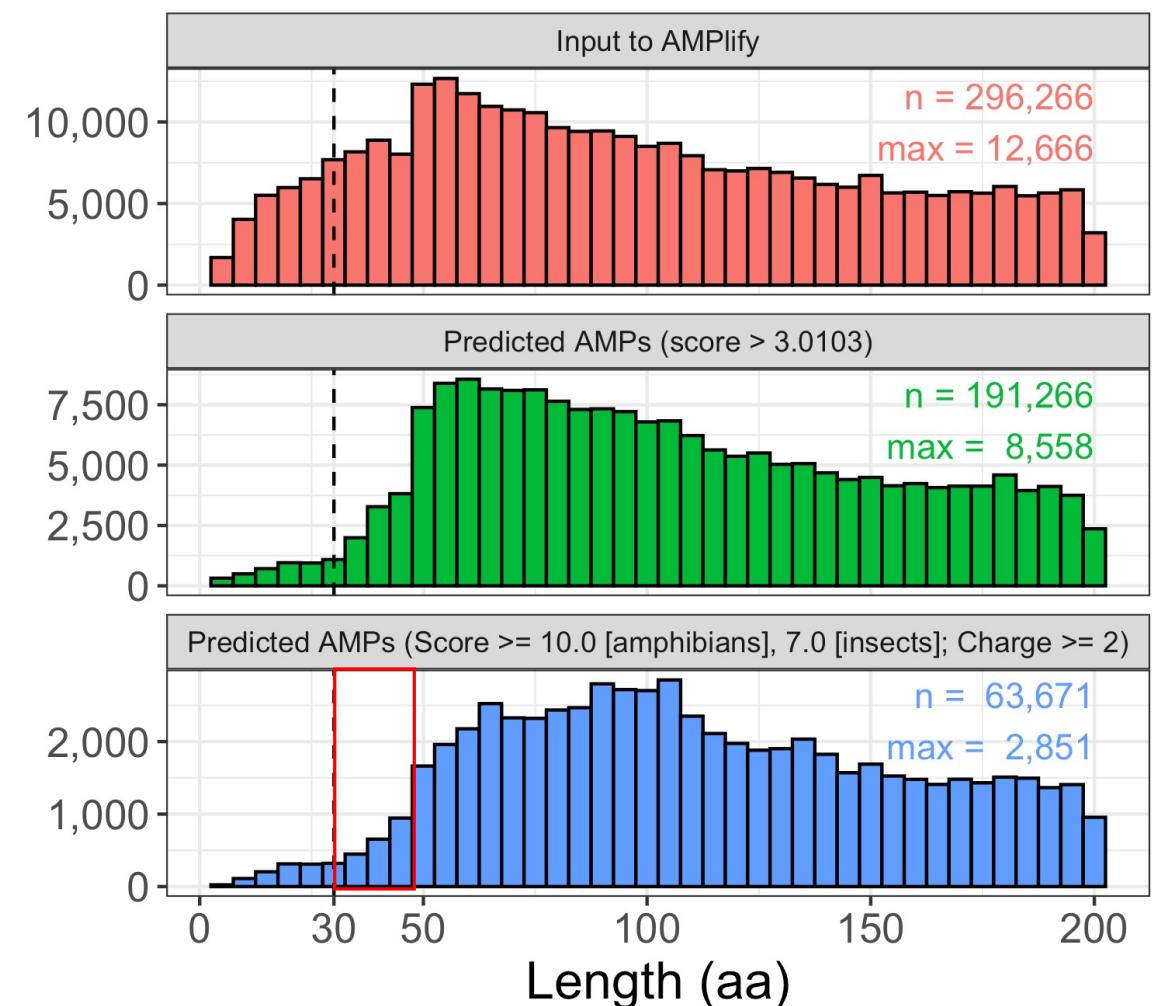
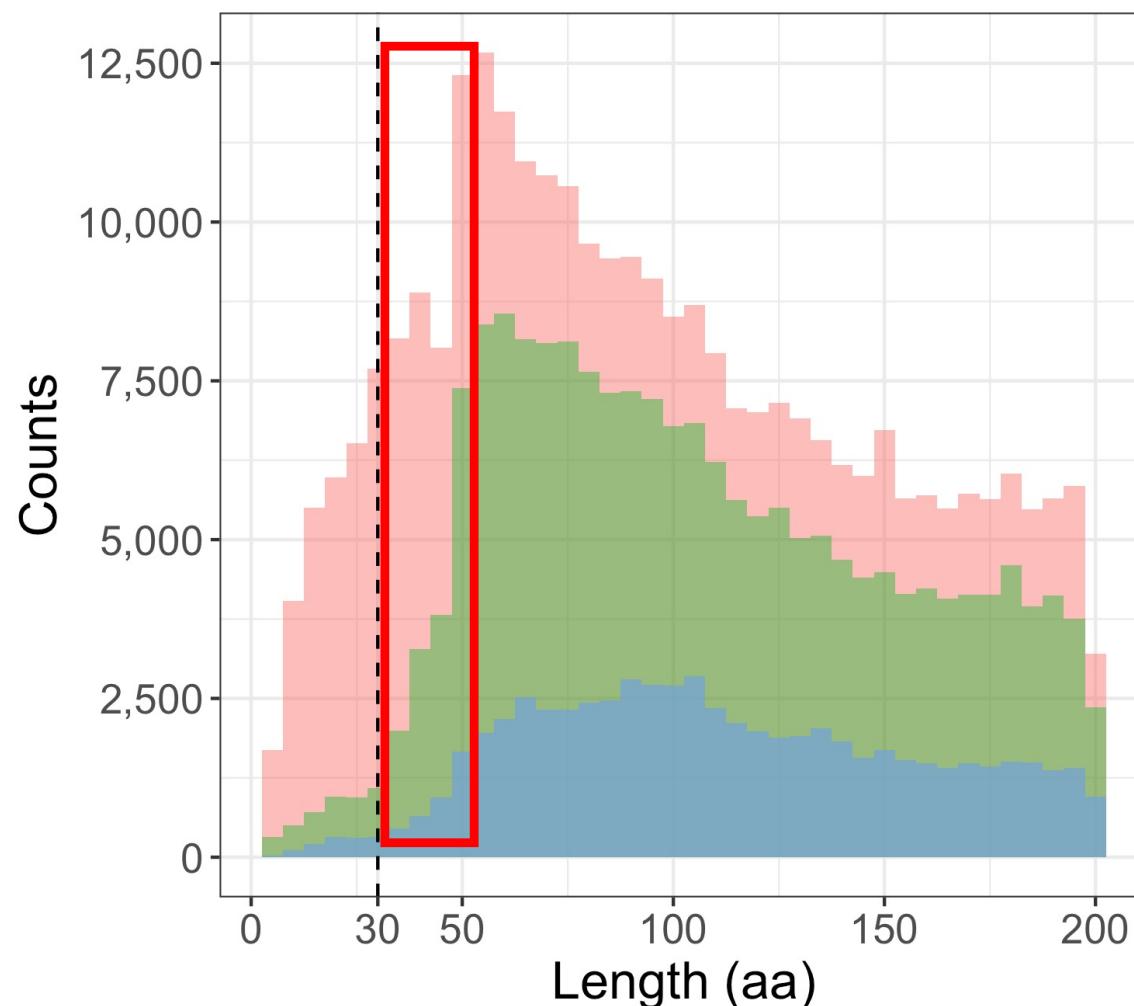
Putative AMP Criteria: AMPlify score ≥ 10.0 (amphibians), 7.0 (insects); Length $\leq 30\text{aa}$; Charge ≥ 2
 Total Transcripts: 53,031,368
 Total Putative AMPs: 1,478



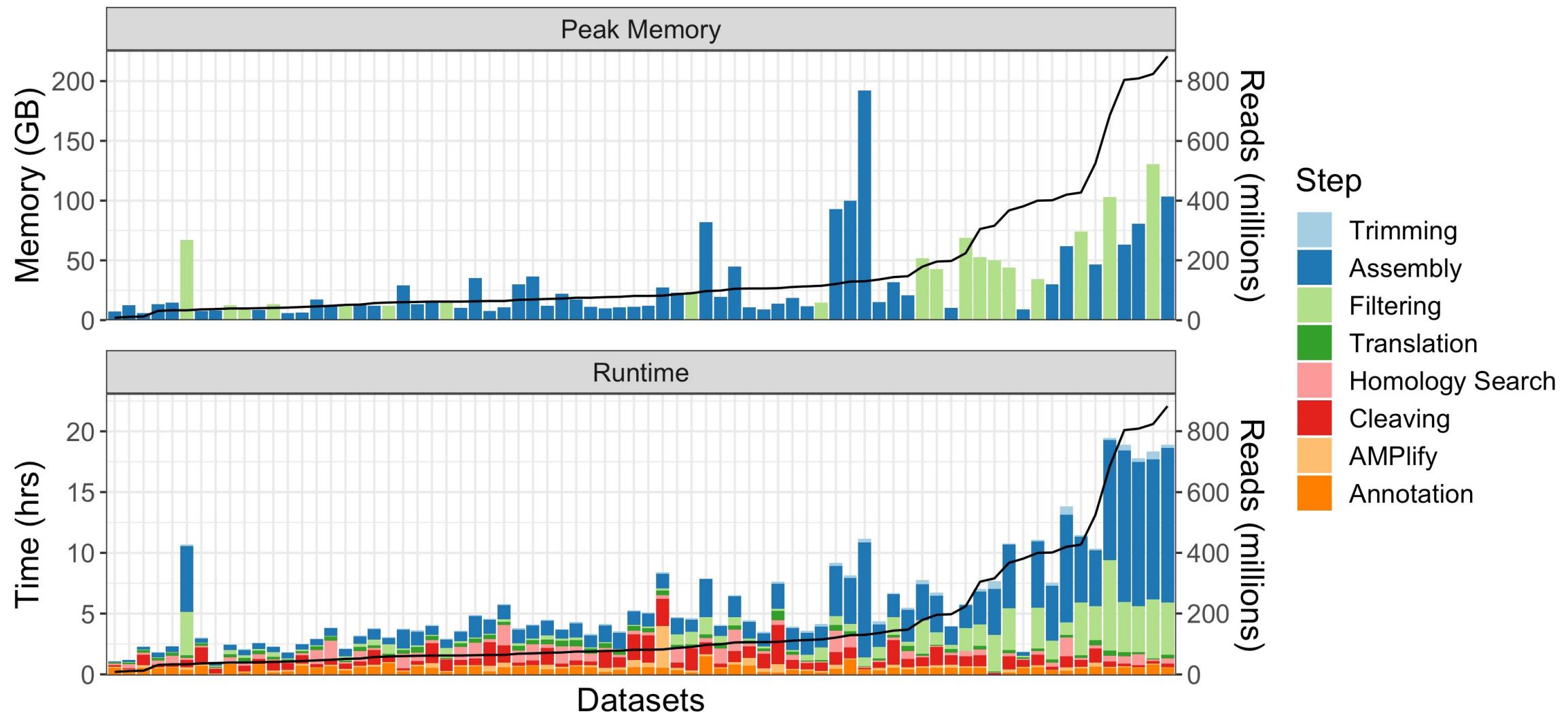
Putative AMP Criteria



Length Distribution



rAMPage Benchmarking



Selected AMPs for Synthesis

n = 1,137

Putative AMP Criteria

- ✓ AMPlify Score >= 10.0 / 7.0
- ✓ Length <= 30aa
- ✓ Charge >= +2

4. Bonus

n = 3

Manual selection: AMPs with high sequence similarity to AMPs we previously validated

1. Species Count

n = 27

AMPs present
in > 2 species

2. Top Insect

n = 30

Insect AMPs with
AMPlify score > 20.0

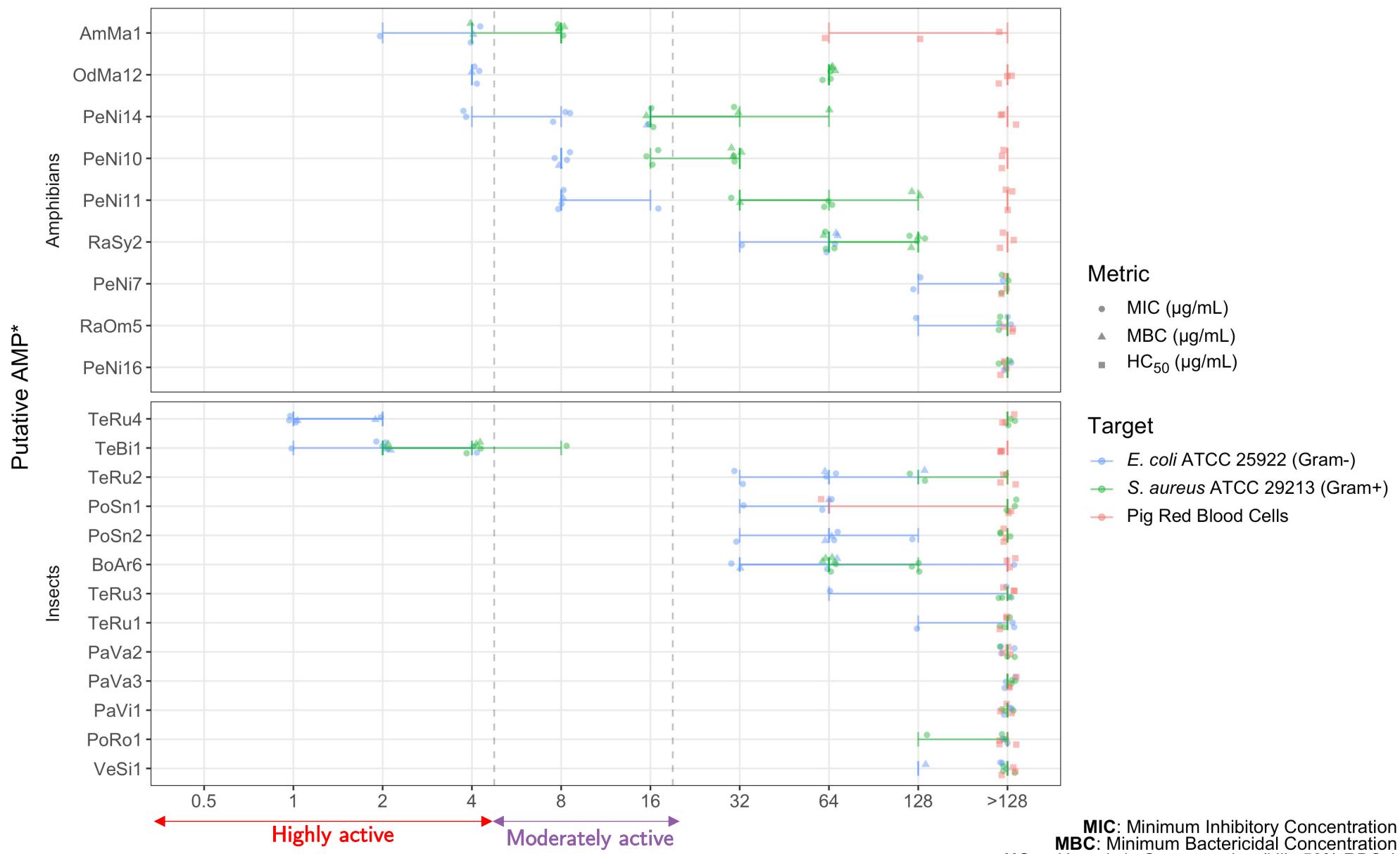
3. Top Clusters

n = 30

Top-scoring clusters of
95% similarity

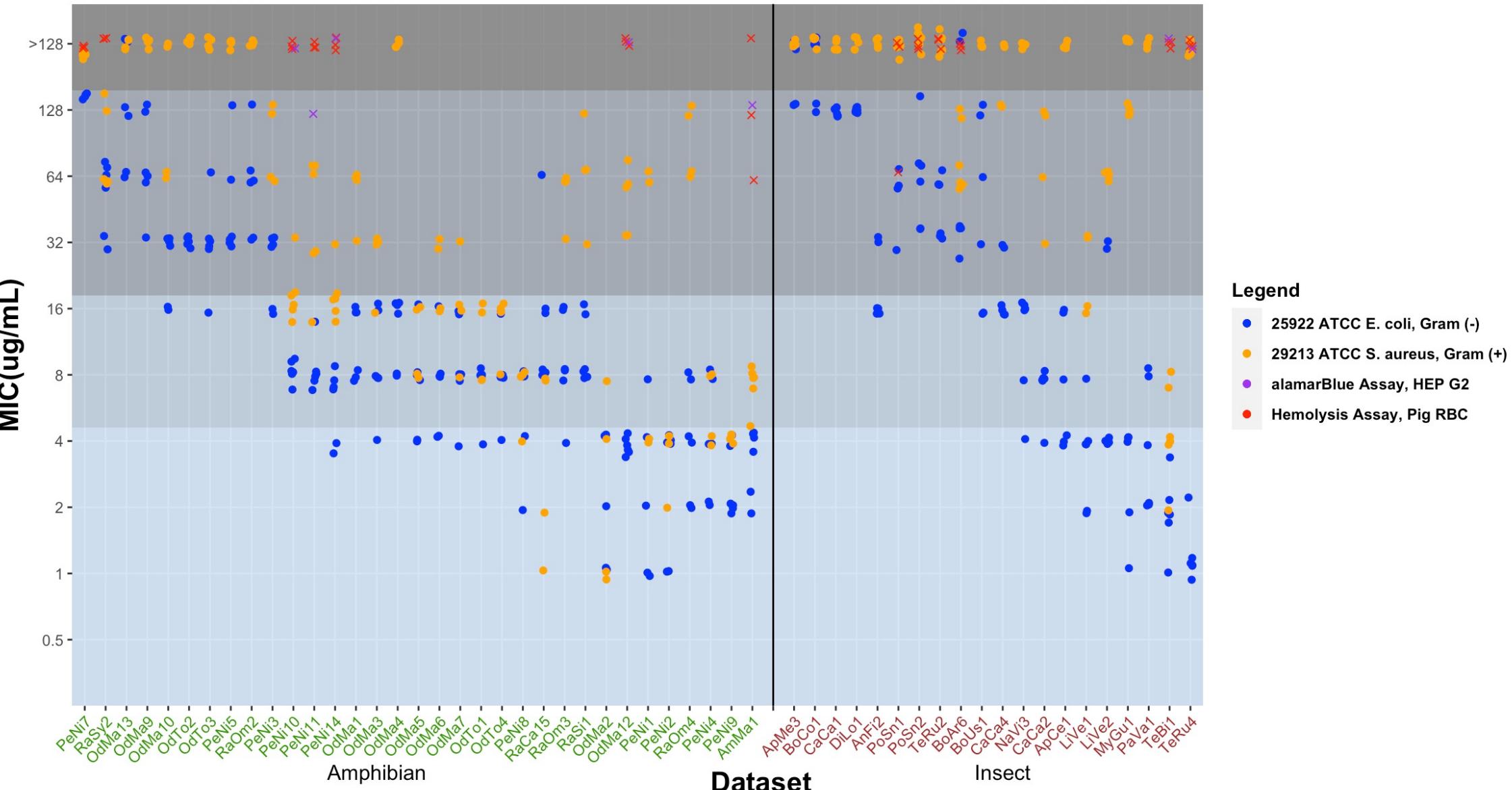
in vitro validation Total = 90

Antimicrobial Susceptibility Testing (AST) and Hemolysis



AST conducted by Darcy Sutherland and Nathan Louie.

rAMPAGE Peptide AST



AST conducted by Darcy Sutherland and Nathan Louie. Hemolysis conducted by Darcy Sutherland.
Cytotoxicity assay conducted by Dr. Anat Yanai. Plot created by Nathan Louie.

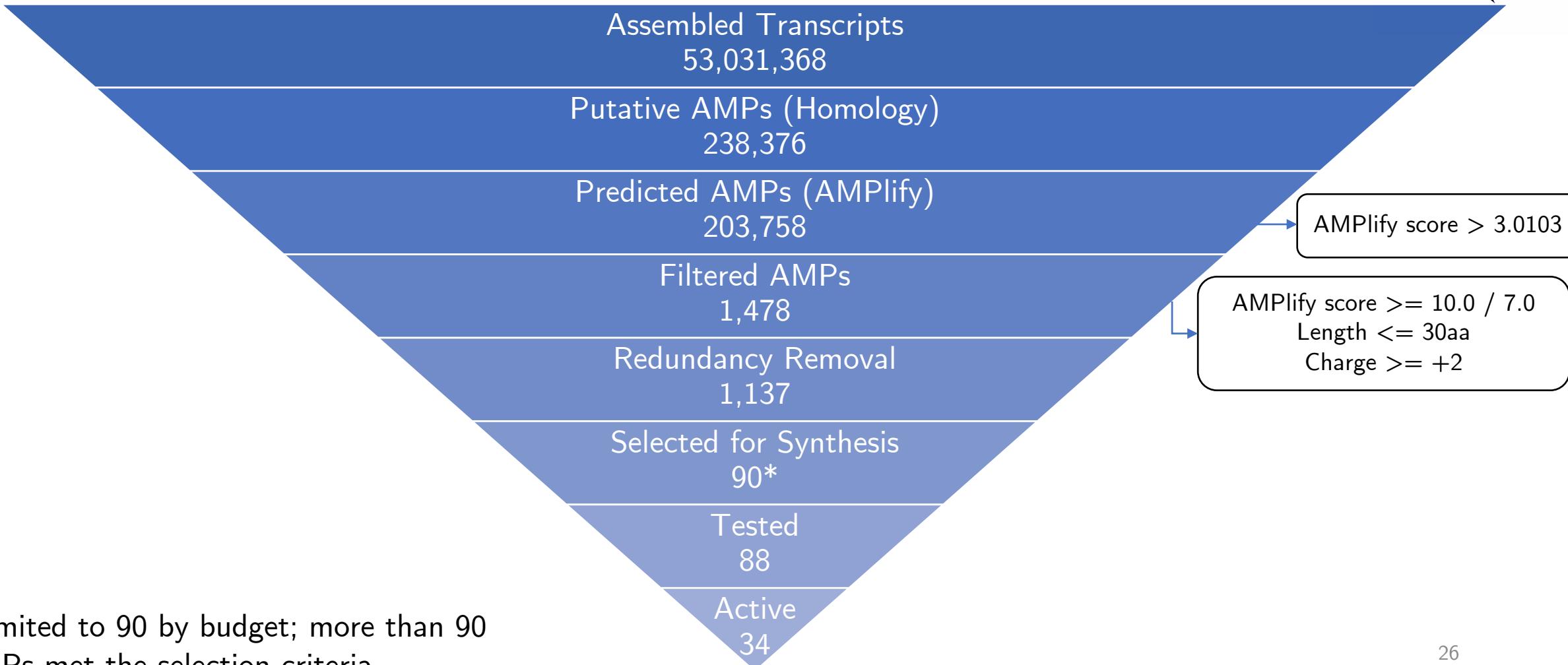
Antimicrobial Susceptibility Testing Results

Activity	<i>E. coli</i> ATCC 29522 (Gram-)	<i>S. aureus</i> ATCC 29213 (Gram+)	TOTAL
Moderately active (MIC <= 4 µg/mL)	20/88 15 amphibians; 5 insects	11/88 10 amphibians; 1 insect	23/88 17 amphibians; 6 insects
Highly active (4 µg/mL < MIC <= 16 µg/mL)	14/88 8 amphibians; 6 insects	5/88 4 amphibians; 1 insect	15/88 9 amphibians; 6 insects
TOTAL	34/88 23 amphibians; 11 insects	16/88 14 amphibians; 2 insects	

- 8 putative AMPs found to be moderately active against both
- 4 putative AMPs found to be highly active against both
- 18 putative AMPs found to be highly active against *E. coli* only

rAMPage: Data Attrition

38 amphibian datasets; 37 insect datasets; ~38 billion RNA-seq reads



Summary



- ✓ *rAMPage* pipeline runs all the tools from beginning to end, from reads to candidate AMPs
- ✓ 90 of 1,137 candidate AMPs detected by *rAMPage* selected for wet lab validation
- ✓ 34 of 88 tested AMPs found to be moderate-to-highly active *in vitro* against *E. coli* or *S. aureus* with low-to-no cytotoxicity or hemolytic activity

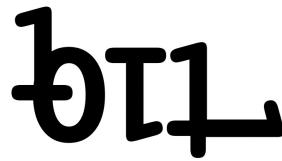
Acknowledgements

Birol Lab

- Darcy Sutherland
- Sambina Islam Aninta
- Nathan Louie
- Anat Yanai
- Ka Ming Nip
- Chenkai Li
- Lauren Coombe
- Monica Kotkoff
- René Warren
- Inanc Birol

Collaborators

- Caren Helbing
- Fraser Hof
- Linda Hoang



BC Centre for Disease Control

