



# MSc Thesis Defence



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

Wednesday, September 29, 2021

Diana Lin

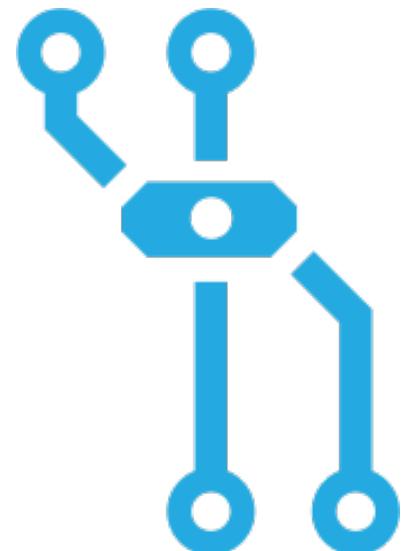
MSc Student, Bioinformatics Graduate Program, UBC

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

**Supervisor:** Dr. Inanc Birol

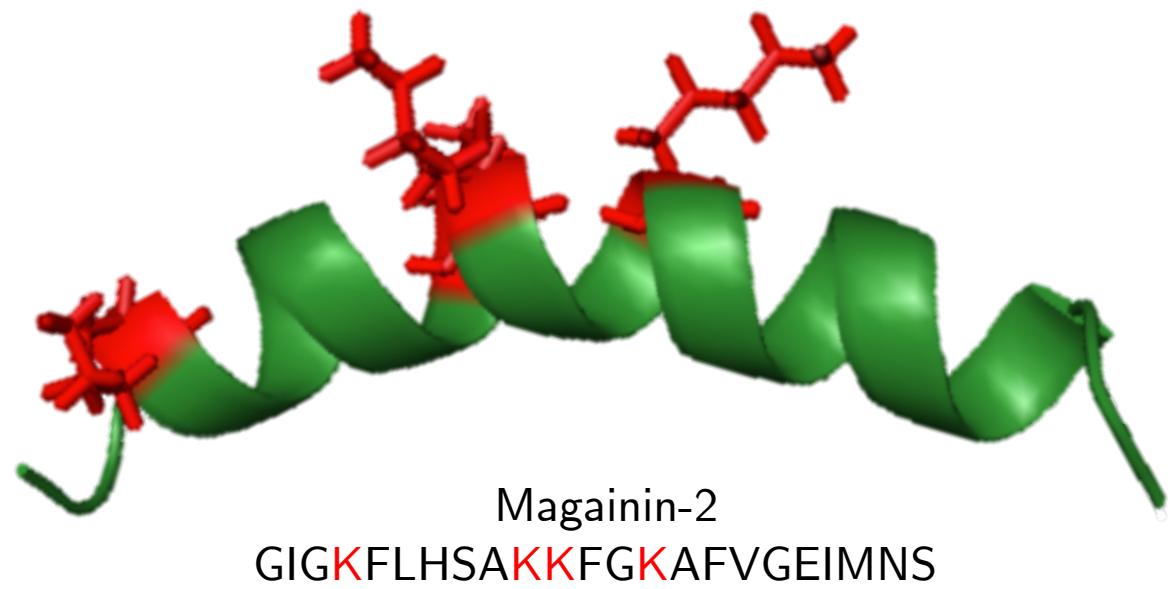
**Committee Members:** Dr. Michael Murphy, Dr. Phil Hieter

**Examining Chair:** Dr. Will Hsiao

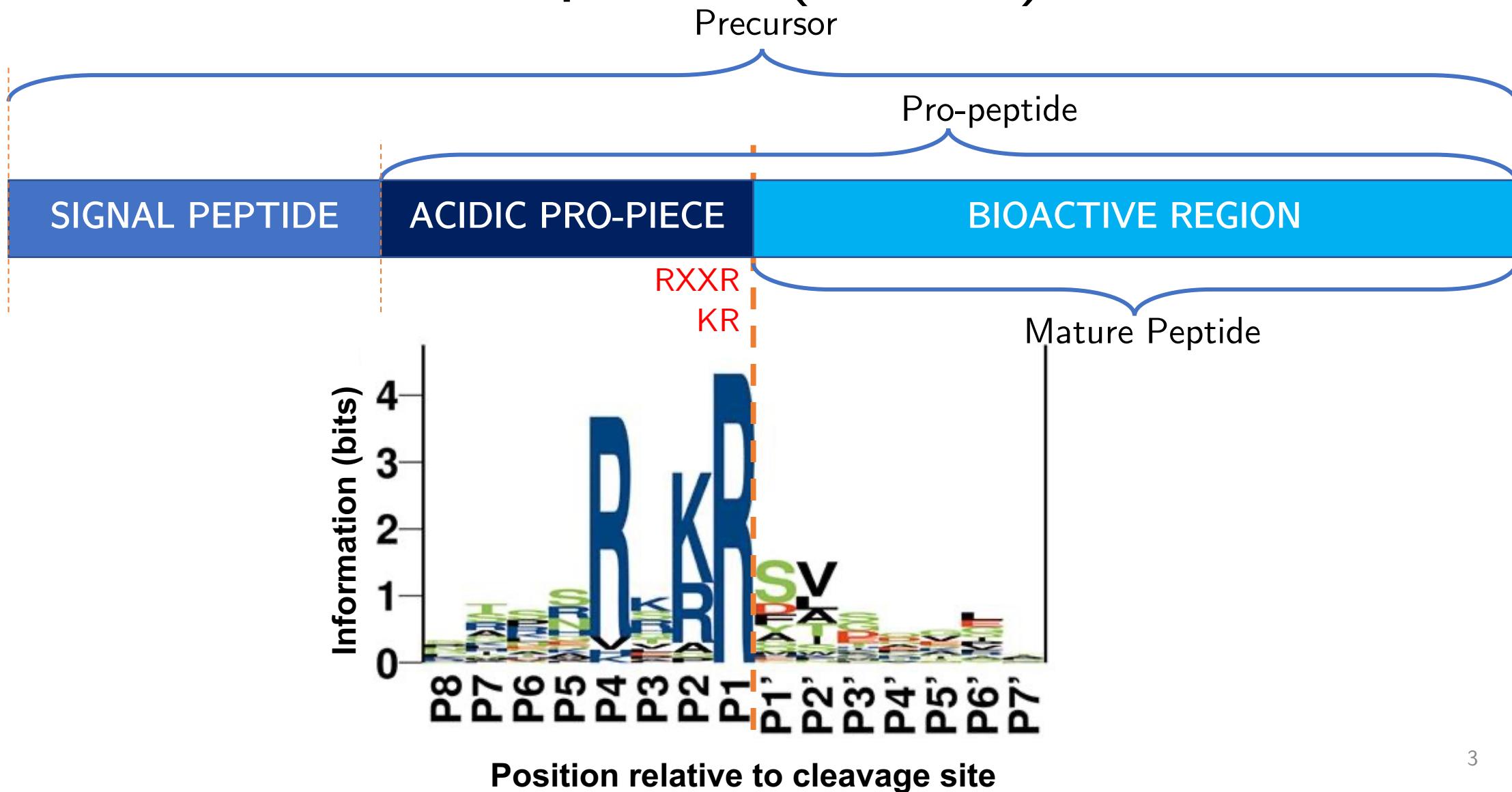


# Antimicrobial Peptides (AMPs)

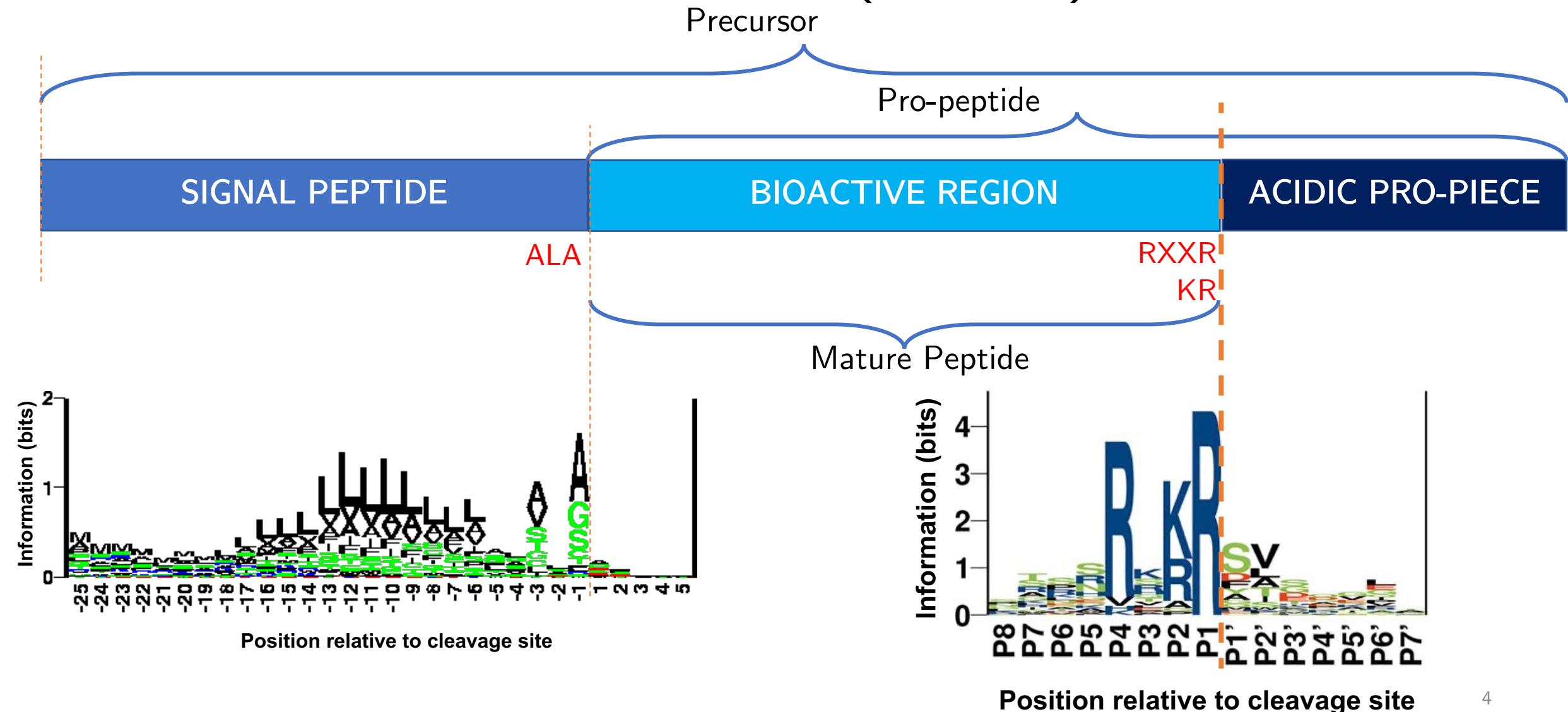
- Short peptide sequences (5 to 50 amino acids)
- Often positively charged
- Amphipathic
- Produced by all multicellular life forms
- Part of the innate immune system



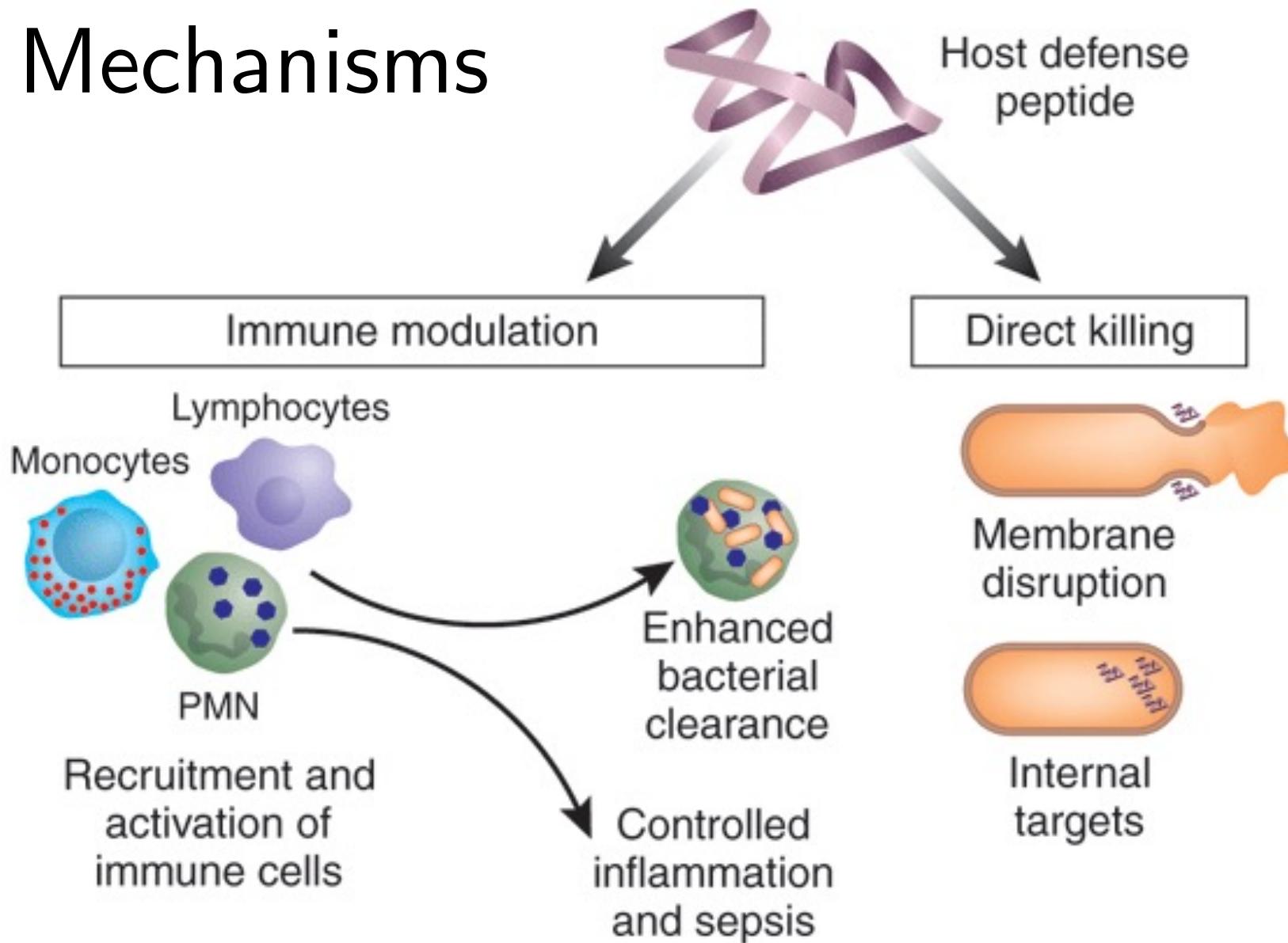
# Antimicrobial Peptides (AMPs)



# Antimicrobial Peptides (AMPs)



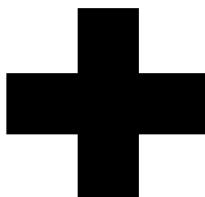
# AMP Mechanisms



# 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

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



# Assembly: RNA-Bloom



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



# Homology Search

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



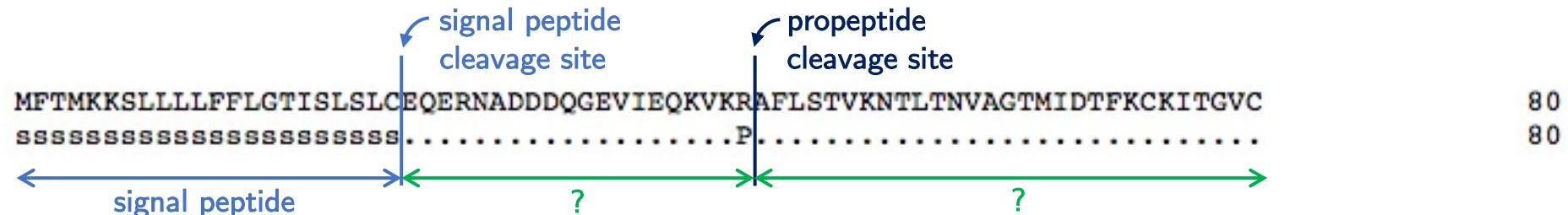
APD3: Antimicrobial Peptide Database 3, <https://aps.unmc.edu>

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

# Cleaving Precursor AMPs

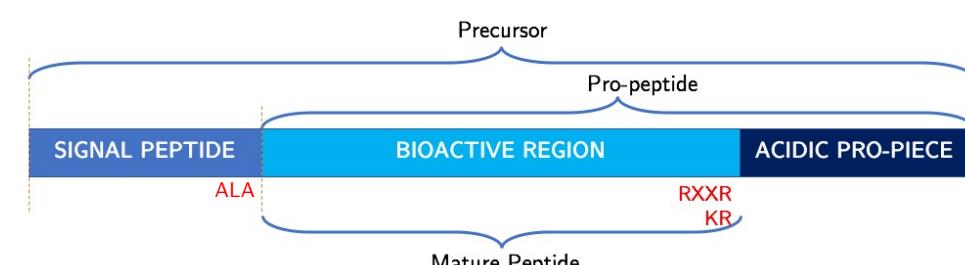
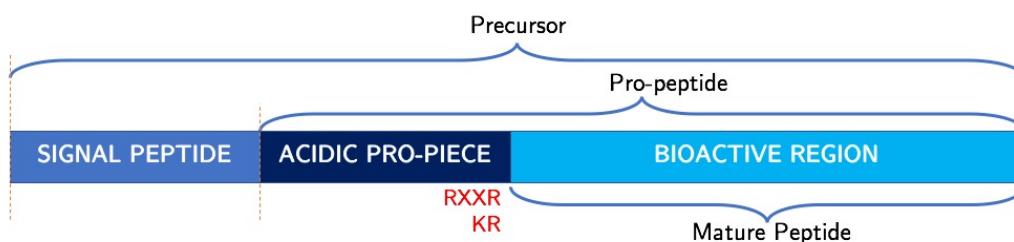
>rcatesbeiana-back\_skin-139086.p2

MFTMKKSLLLLFFLGTISLSLCEQERNADDDQGEVIEQKV**KR**AFLSTVKNTLTNVAGTMIDTFKCKITGVC

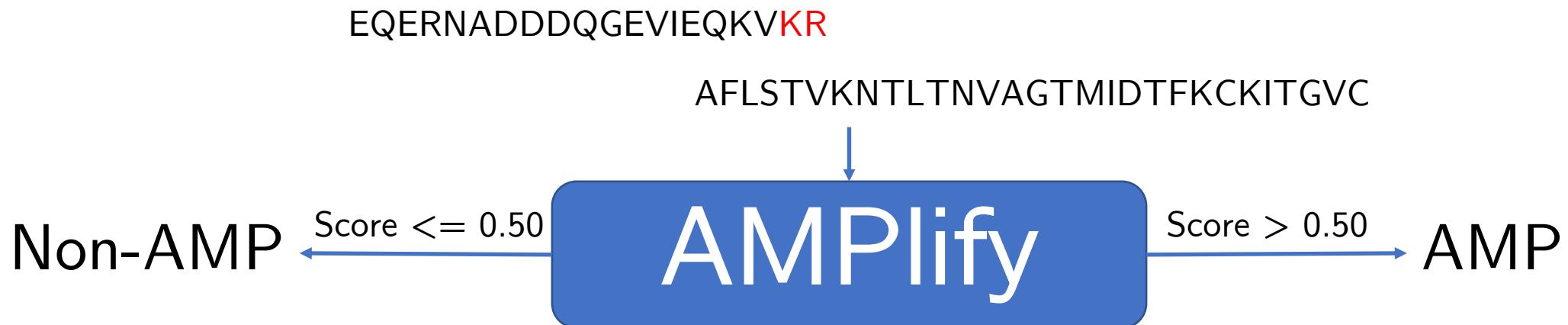


Signal peptide cleavage site predicted: between pos. 22 and 23: SLC-EQ

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



# Prioritizing AMPs: AMPlify

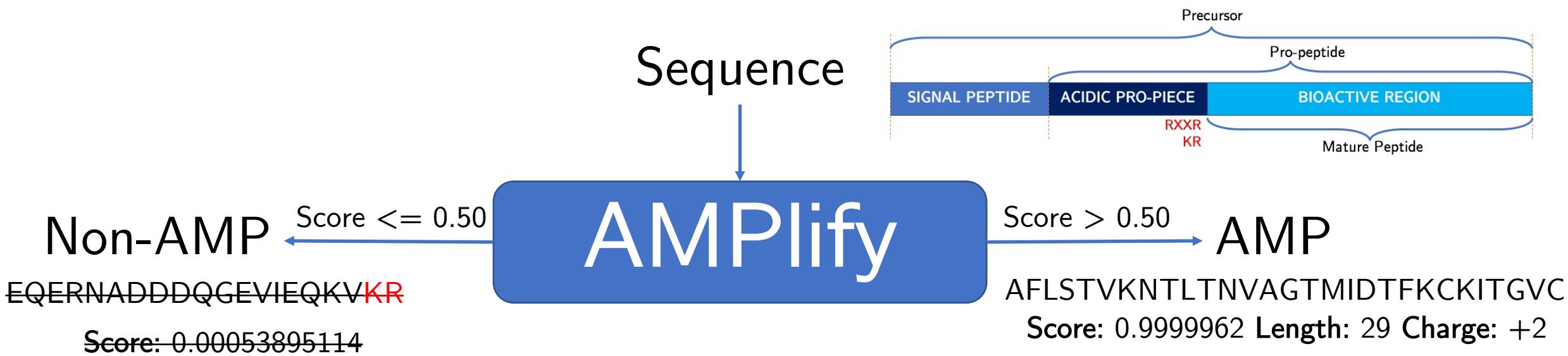


AMPlify Score  $> 0.50$

Length  $\leq 50\text{aa}$

Charge  $\geq +1$

# Prioritizing AMPs: AMPlify

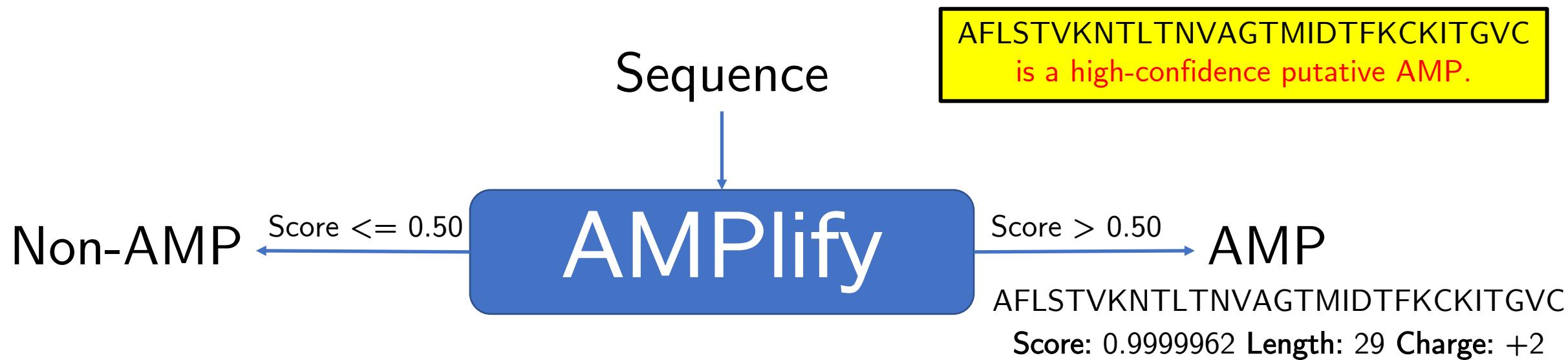


AMPlify Score > 0.50

Length <= 50aa

Charge >= +1

# Prioritizing AMPs: AMPlify



AMPlify Score  $> \cancel{0.50}$   $\longrightarrow$  AMPlify Score  $\geq \cancel{0.50}$  /  $0.90 / 0.80$   
Amphibians Insects

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

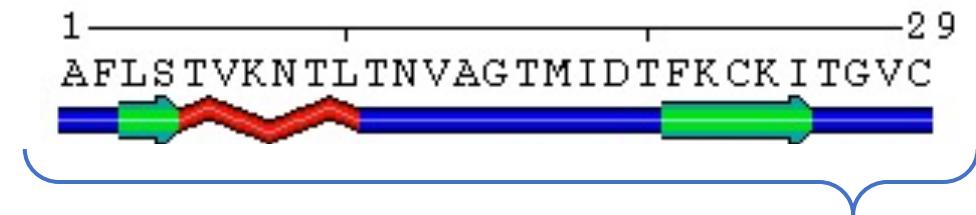
Charge  $\geq \cancel{+1}$   $\longrightarrow$  Charge  $\geq +2$

# Characterizing AMPs

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

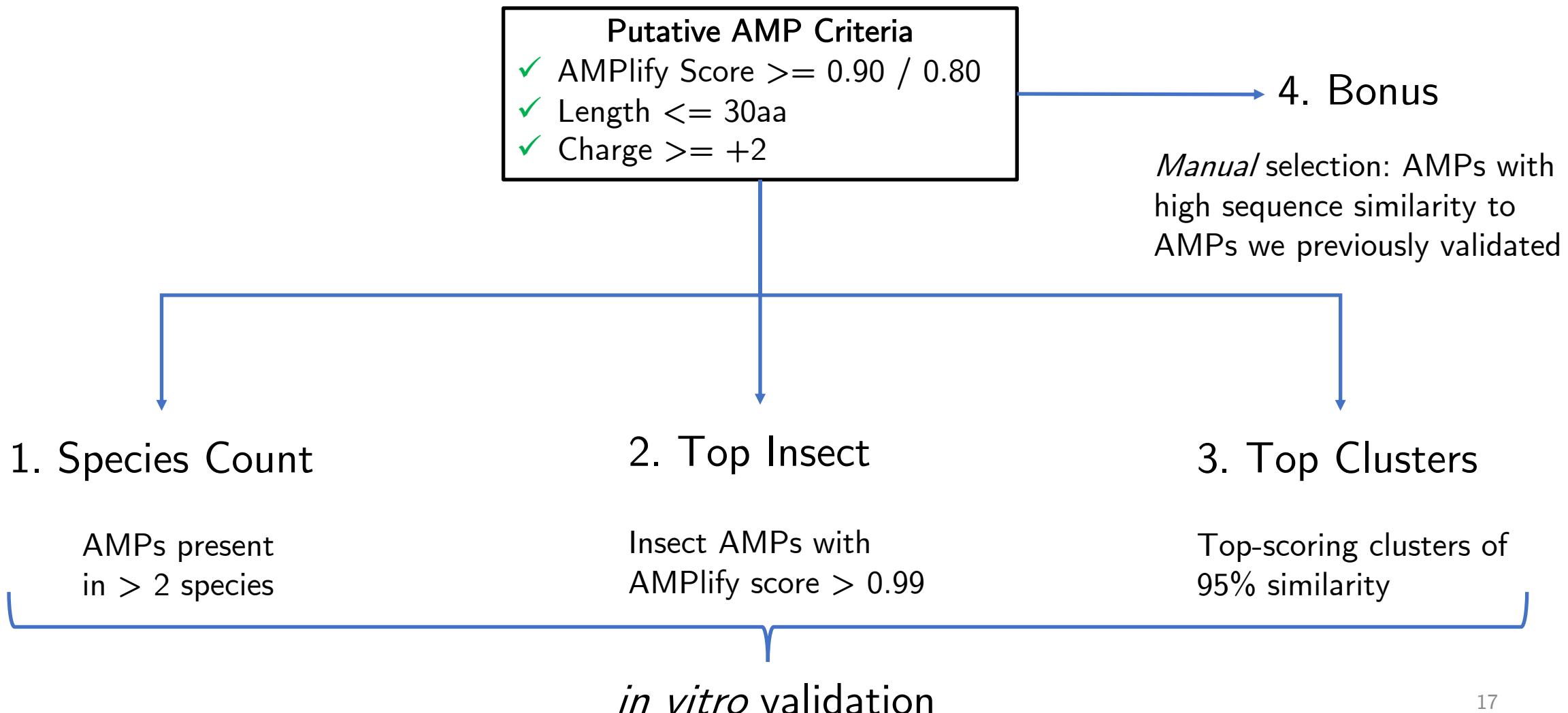
**Putative AMP Criteria**

- ✓ AMPlify Score  $\geq 0.90 / 0.80$
- ✓ Length  $\leq 30\text{aa}$
- ✓ Charge  $\geq +2$



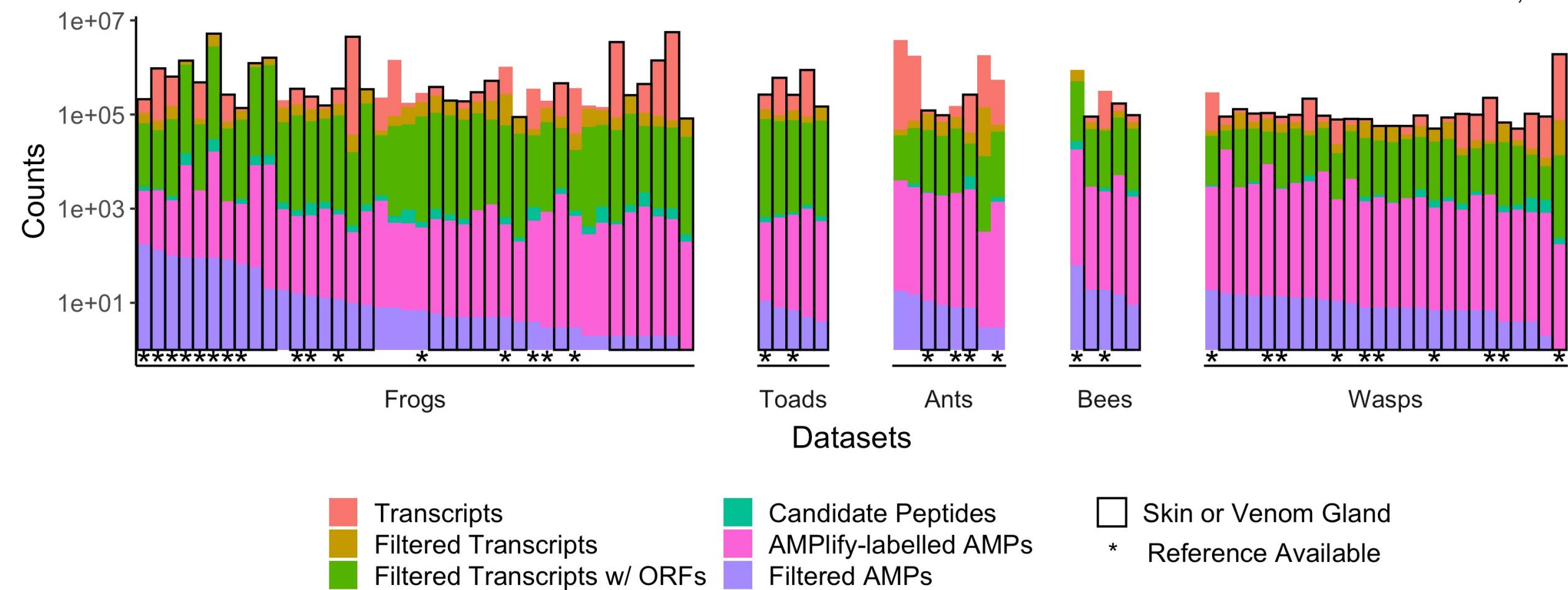
AMPlify Score	Length	Charge	Top Precursor	Top Mature	GO Term	IPScan	Pfam	Secondary Structure
0.9999962	29	2	ABB89058.1: ranatuerin 2CHb precursor, partial [Rana chiricahuensis] (75.86%)	AP00617: Palustrin-2 (UCLL1a; XXU; 1S=S, frog, amphibians, animals) (75.86%)	GO:0005576 (extracellular region) GO:0006952 (defense response)	IPR012521 (Frog antimicrobial peptide, brevinin-2/esculentin type)	PF08023 (Frog antimicrobial peptide)	CCEEHHHH HHCCCCCC CCCCEEEE ECCCC

# Selecting AMPs for Synthesis

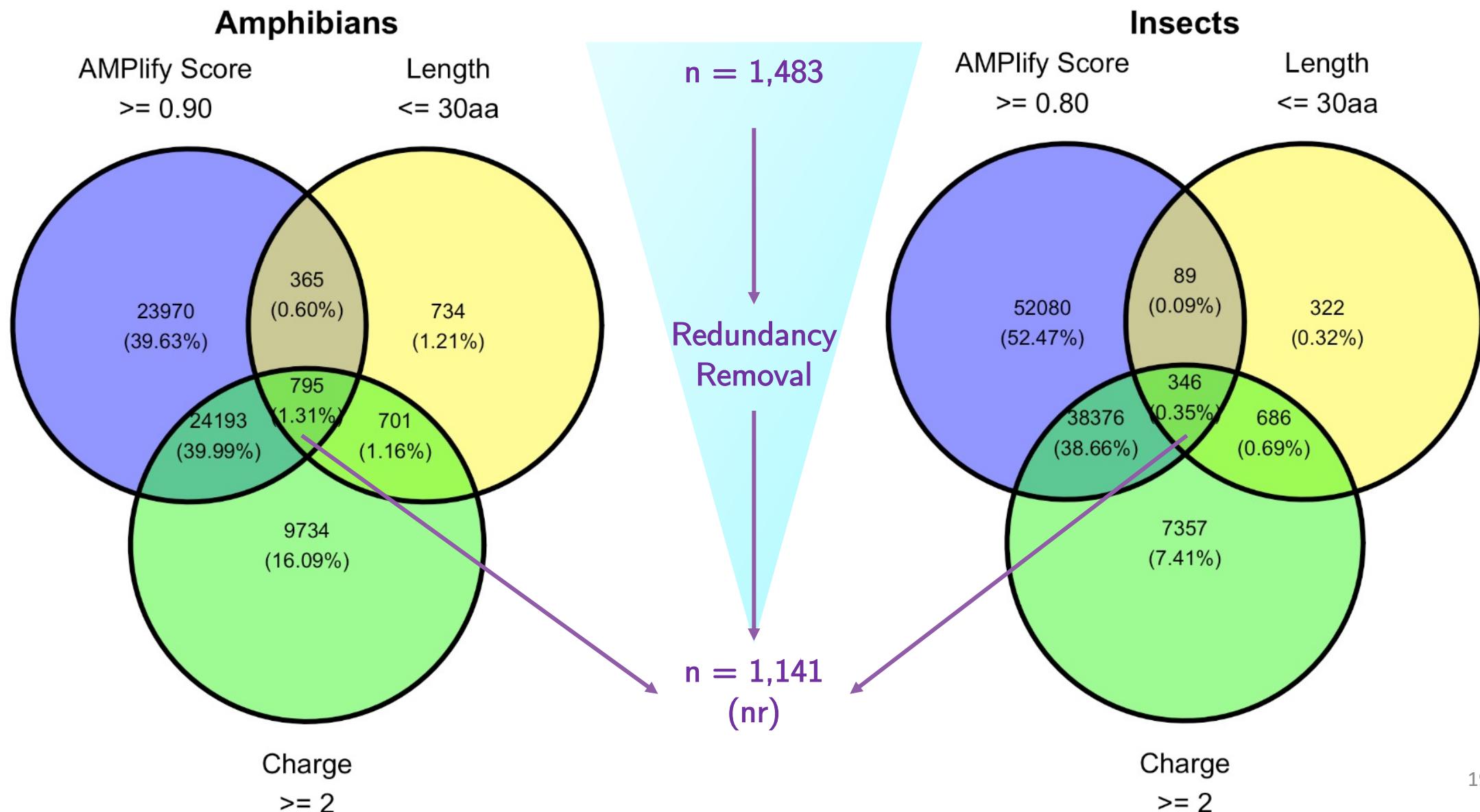


# Overview

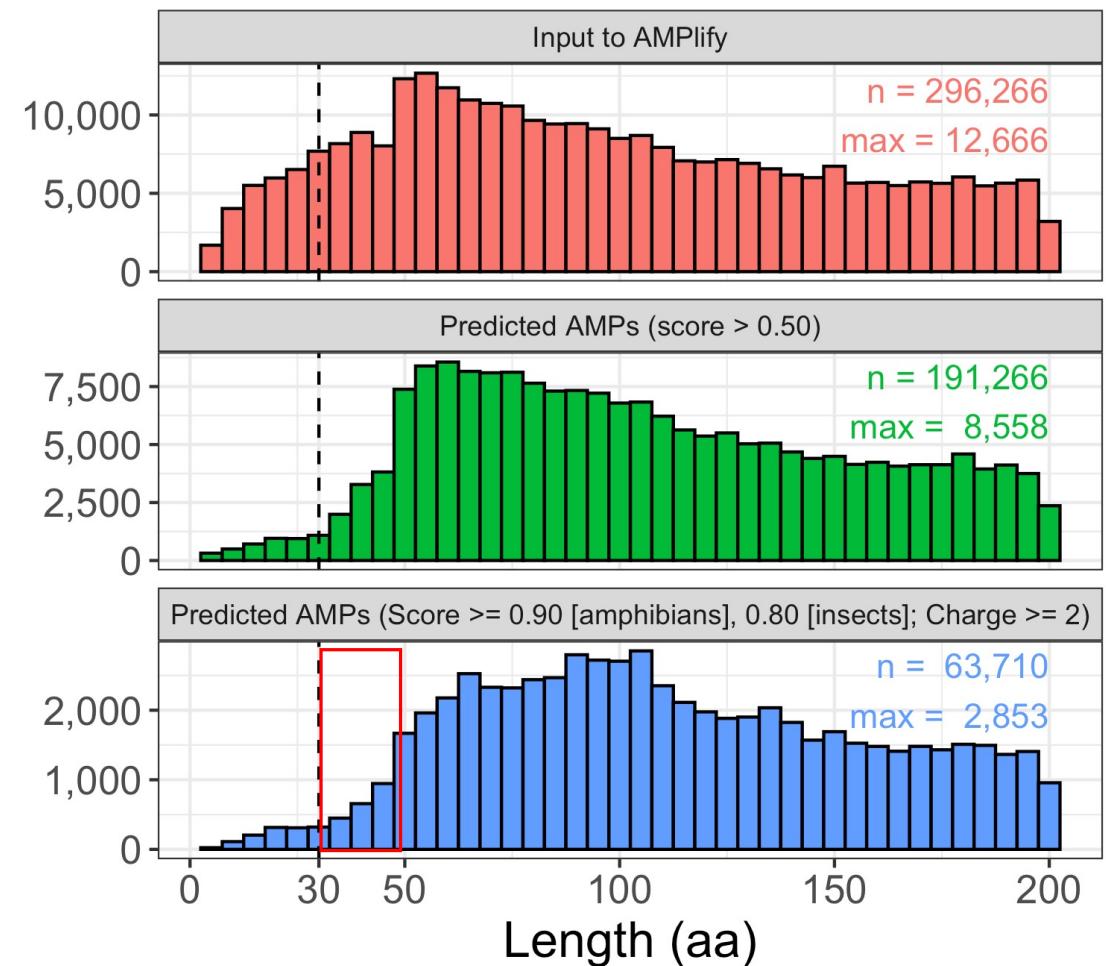
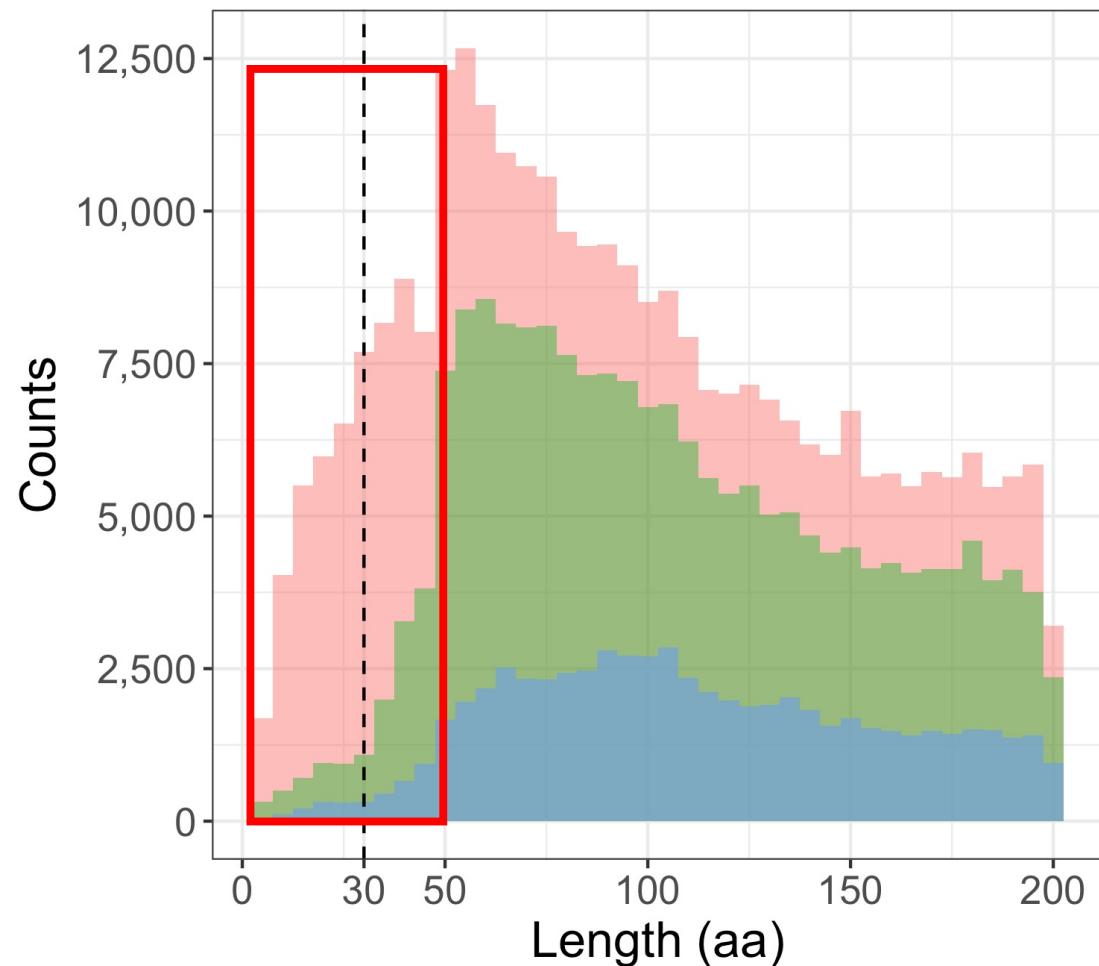
Criteria: AMPlify score  $\geq 0.90$  (amphibians), 0.80 (insects); Length  $\leq 30\text{aa}$ ; Charge  $\geq 2$   
 Total Transcripts: 53,031,368  
 Total AMPs: 1,483



# Putative AMP Criteria

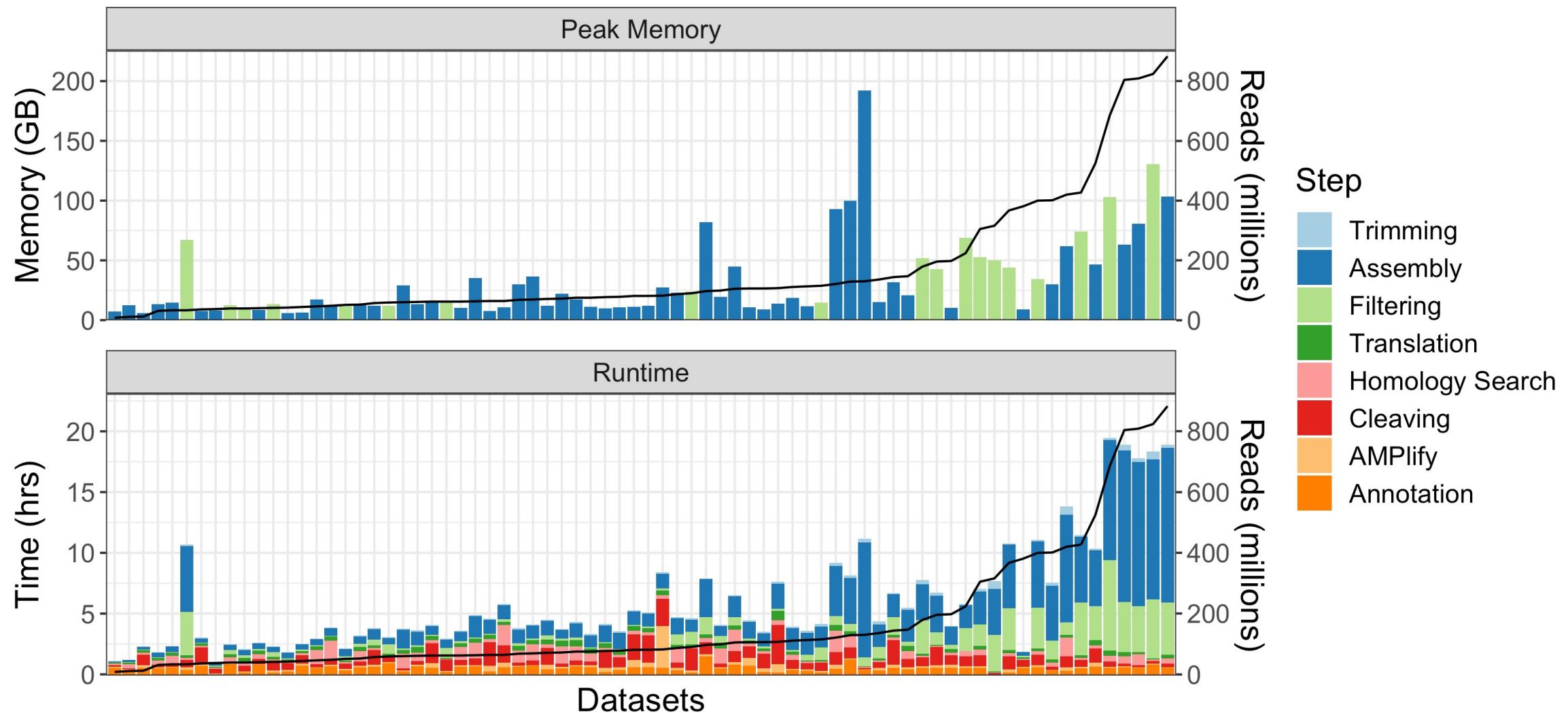


# Length Distribution



binwidth = 5

# *rAMPage* Benchmarking



# Selected AMPs for Synthesis

n = 1,141

## Putative AMP Criteria

- ✓ AMPlify Score >= 0.90 / 0.80
- ✓ 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 > 0.99

## 3. Top Clusters

n = 30

Top-scoring clusters of  
95% similarity

*in vitro* validation Total = 90

# *in vitro validation*

MIC: Minimum Inhibitory Concentration

- MIC <= 16 µg/mL → ACTIVE
- MIC <= 4 µg/mL → HIGHLY ACTIVE

Name	Sequence	MIC (µg/mL)	Score	Length	Charge	Selection Method
AmMa1	GILDTLKQLGKAAVQGLLSKAACKLAKTC	2	1.0	29	4	Top Cluster
OdMa12	GFMDTAKNVAKNVAVTLLYNLKCKITKAC	4	0.9999999	29	4	Top Cluster
PeNi7	VIPFVASVAAEMMHVYCAASKRCKN	128	0.9999522	26	2	Species Count
PeNi10	GLLLDTVKGAAKNVAGILLNKLKCKVTGDC	8	0.9999993	30	3	Species Count
PeNi11	GILTDLKGAAKNVAGVLLDKLKCKITGGC	8	0.9999993	30	3	Species Count
PeNi14	GLWTTIKEGVKNFSVGVLVDKIRCKITGGC	8	0.9999998	29	3	Species Count
PeNi16	ATAWKVPPGLQPIRPIRPLCGNDKS	>128	0.9978241	27	4	Species Count
RaOm5	AGYSRMIRRPPGFSPFRVAPASSLKR	>128	0.9433209	26	6	Species Count
RaSy2	EEQRFLPVVAGLAALKVLPsiICAVTKKC	32	0.9935637	28	3	Bonus
BoAr6	GILRLVTRRFRFSPTNLRYTVARLVSGVP	>128	0.9938134	30	6	Top Insect
TeRu3	AVLSFVHKLFLNFLHVDTSKGKCRATLQ	64	0.9976805	28	3	Top Insect
TeRu4	SWLSKSVKKLVNKKNYTRLEKLAKKLFNE	1	0.9971965	30	8	Top Insect

# *in vitro validation*

MIC/MBC: Minimum Inhibitory/Bactericidal Concentration  
 ➤ MIC/MBC <= 16 µg/mL → ACTIVE  
 ➤ MIC/MBC <= 4 µg/mL → HIGHLY ACTIVE

Name	Sequence	MIC (µg/mL)	MBC (µg/mL)	Score	Length	Charge	Selection Method
AmMa1	GILDTLKQLGKAAVQGLLSKAACKLAKTC	2-4	4	1.0	29	4	Top Cluster
OdMa12	GFMDTAKNVAKNVAVTLLYNLKCKITKAC	4	4	0.9999999	29	4	Top Cluster
PeNi7	VIPFVASVAAEMMHHVYCAASKRCKN	128	-	0.9999522	26	2	Species Count
PeNi10	GLLLDTVKGAAKNVAGILLNKLKCKVTGDC	8	8	0.9999993	30	3	Species Count
PeNi11	GILTDTLKGAAKNVAGVLLDKLKCKITGGC	8-16	8	0.9999993	30	3	Species Count
PeNi14	GLWTTIKEGVKNFSVGVLVDKIRCKITGGC	4-8	4-8	0.9999998	29	3	Species Count
PeNi16	ATAWKVPPGLQPIRPIRPLCGNDKS	>128	-	0.9978241	27	4	Species Count
RaOm5	AGYSRMIRRPPGFSPFRVAPASSLKR	>128	-	0.9433209	26	6	Species Count
RaSy2	EEQRFLPVVAGLAALKVLPsiICAVTKKC	32	-	0.9935637	28	3	Bonus
BoAr6	GILRLVTRRFRFSPTNLNRYTVARLVSGVP	>128	-	0.9938134	30	6	Top Insect
TeRu3	AVLSFVHKLFLNFLHVDTSKGKCRATLQ	64	-	0.9976805	28	3	Top Insect
TeRu4	SWLSKSVKKLVNKNYTRLEKLAKKLFNE	1-2	1	0.9971965	30	8	Top Insect

Active AMPs from Batch #1 tested for MIC (n = 3) and MBC (n = 1). Tested in duplicate against *E. coli* ATCC 25922 for each experiment by Darcy Sutherland.

# *in vitro validation*

MIC/MBC: Minimum Inhibitory/Bactericidal Concentration  
 ➤ MIC/MBC <= 16 µg/mL → ACTIVE  
 ➤ MIC/MBC <= 4 µg/mL → HIGHLY ACTIVE

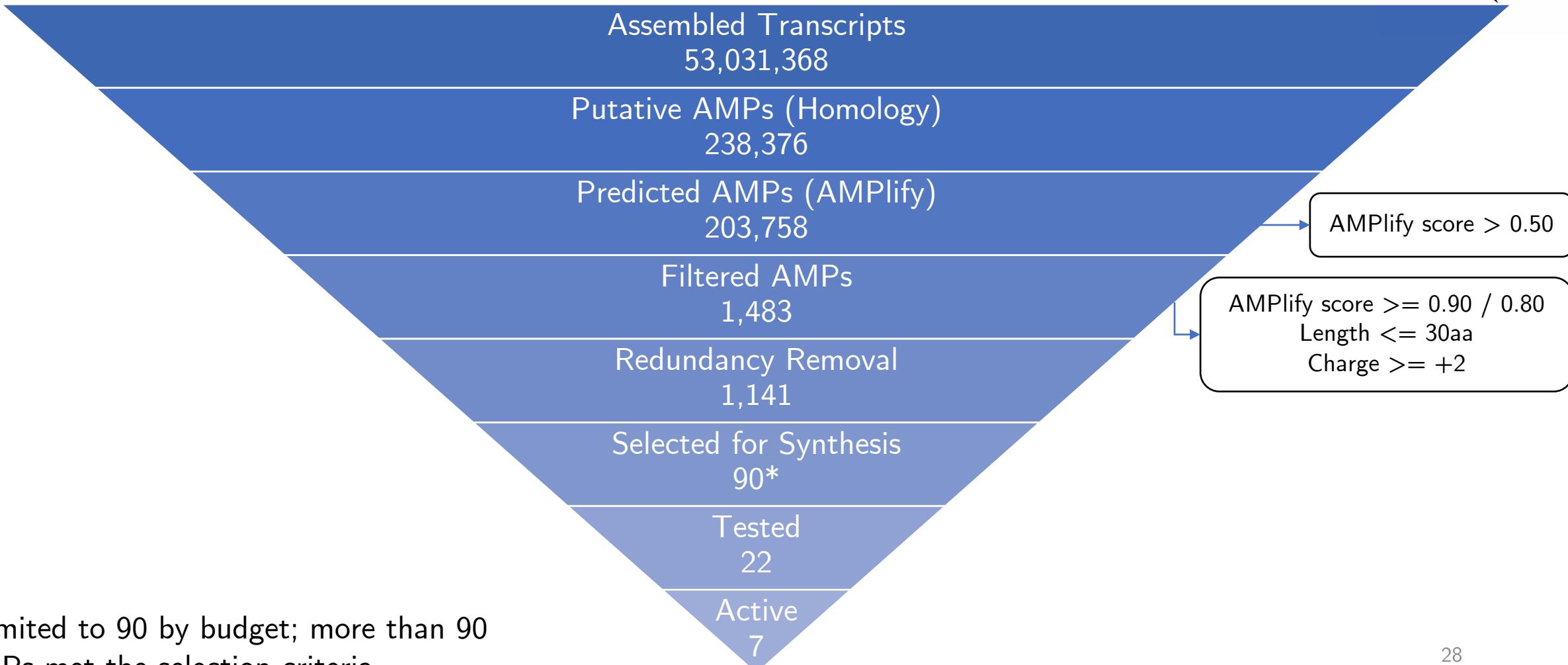
Name	Sequence	MIC (µg/mL)	MBC (µg/mL)	Score	Length	Charge	Selection Method
PaVa2	KYHHIKLRHGRHRRTIH	NI	NI	0.997491	17	6	Top Insect
PaVa3	ITEPVGKAPFTSELRGGWLKRR	NI	NI	0.995996	24	3	Top Insect
PaVi1	WALRWKTR	NI	NI	0.996768	8	3	Top Insect
PoRo1	VAAFAIIGCLCCRRPRR	NI	NI	0.995010	17	4	Top Insect
PoSn1	ISIKEALEHSFFHTVPRKWCKKH	32-64	32-64	0.999081	23	3	Top Insect
PoSn2	TALKSLSILKKLAKLM	32	32	0.995694	17	4	Top Insect
TeBi1	KIKIPWGKVKDFLVGGMKAVGKK	1-2	1-2	0.999968	23	6	Top Insect
TeRu1	VPFGLKPR	NI	NI	0.997391	8	2	Species Count
TeRu2	AFVRILCYCCPRRIKRR	32-64	128	0.999356	17	6	Top Insect
VeSi1	FILHAKKTRSAK	NI	NI	0.994341	12	4	Top Insect

NI = No inhibition

Batch #2, tested in duplicate against *E. coli* ATCC 25922 for MIC (n = 2) and MBC (n = 1) in each experiment by Darcy Sutherland.

# rAMPage: Data Attrition

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



\*Limited to 90 by budget; more than 90 AMPs met the selection criteria

# Summary



- ✓ *rAMPage* pipeline runs all the tools from beginning to end, from reads to candidate AMPs
- ✓ 90 of 1,141 candidate AMPs detected by *rAMPage* selected for wet lab validation
- ✓ 7 of 22 tested AMPs found to be active *in vitro* against *E. coli* with MIC/MBC  $\leq 16 \mu\text{g/mL}$

# 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

