MSc Thesis Committee Meeting #2

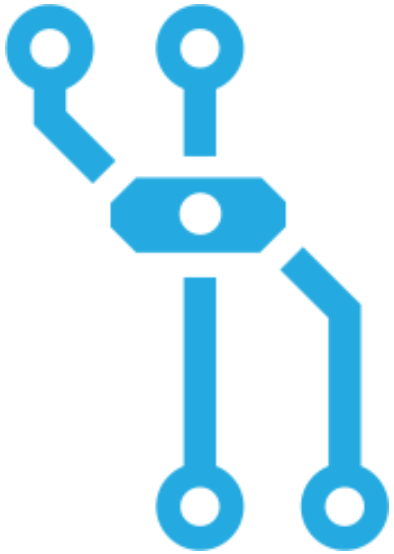
Thursday, May 20, 2021

Diana Lin

MSc Student, Bioinformatics Graduate Program, UBC

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

Supervisor: Dr. Inanc Birol



Outline

1. Recap and Progress Update (15 min.)
2. Thesis Outline (5 min.)
3. Open Discussion and Feedback (30 min.)

Timeline

LAST MEETING

TODAY

July 2019

Started MSc Bioinformatics at UBC

April 2020

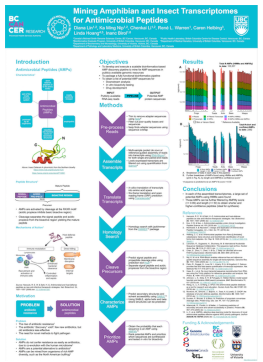
Completed required coursework

June 2020

MSc Thesis committee meeting #1

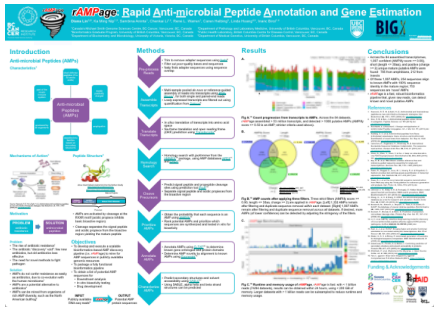
September 2020

Presented poster at BC Cancer Research Day



March 2021

Presented poster at BIG Research Day



April 2021

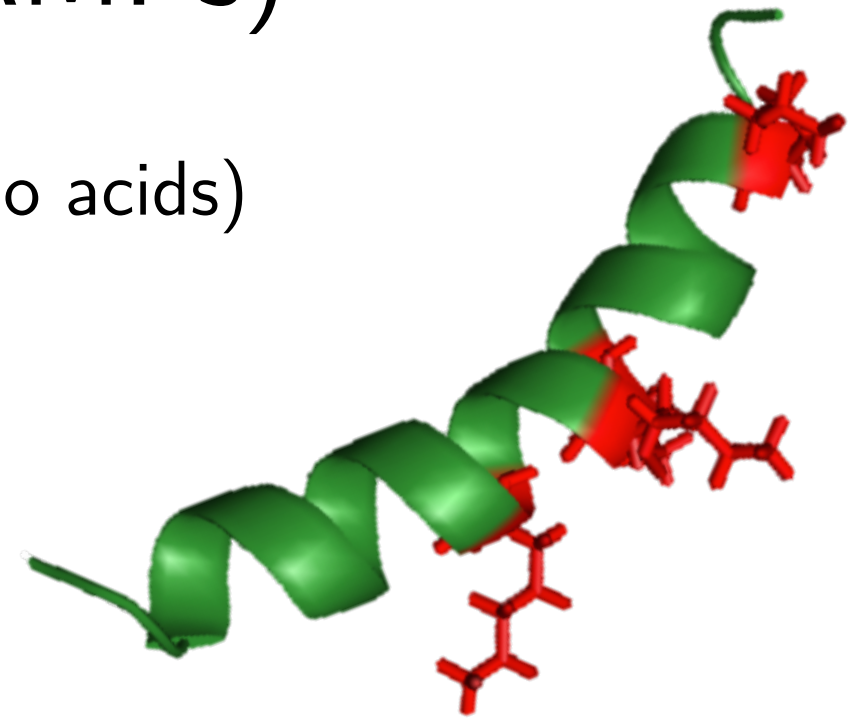
Presented thesis work at PeptAID Science meeting

May 2021

MSc Thesis committee meeting #2

Antimicrobial Peptides (AMPs)

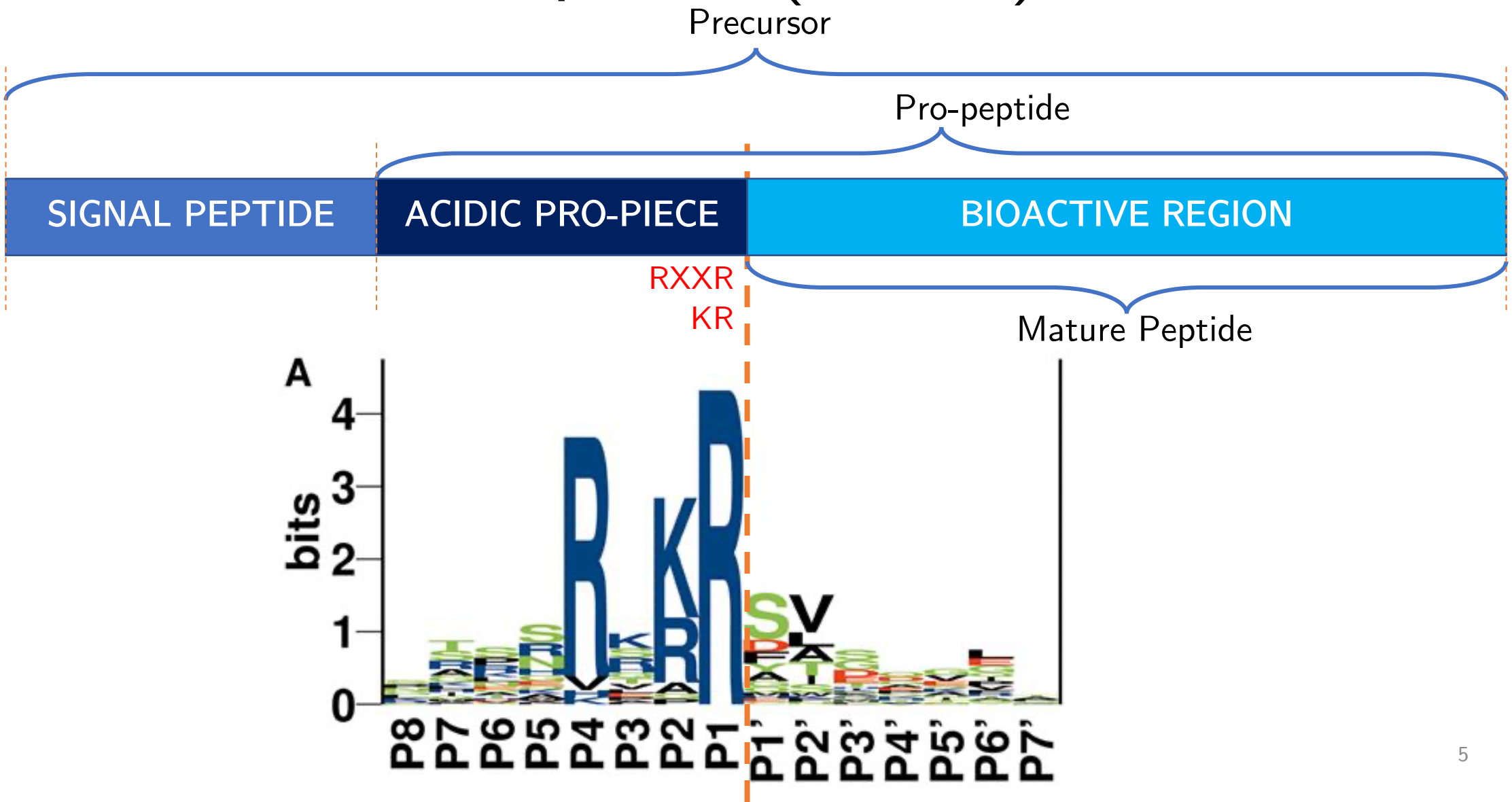
- Short peptide sequences (5 to 50 amino acids)
- Often positively charged
- Amphipathic
- Produced by all life forms
- Part of the innate immune system
- Mechanisms of action:
  1. Direct interaction
  2. Modulation of host immunity



Magainin-2

Source: Darcy Sutherland, PeptAID AIM 1 Science Meeting

Antimicrobial Peptides (AMPs)



Motivation

- Rise of antibiotic resistance creates a problem that requires a novel method to fight pathogens

PROBLEM

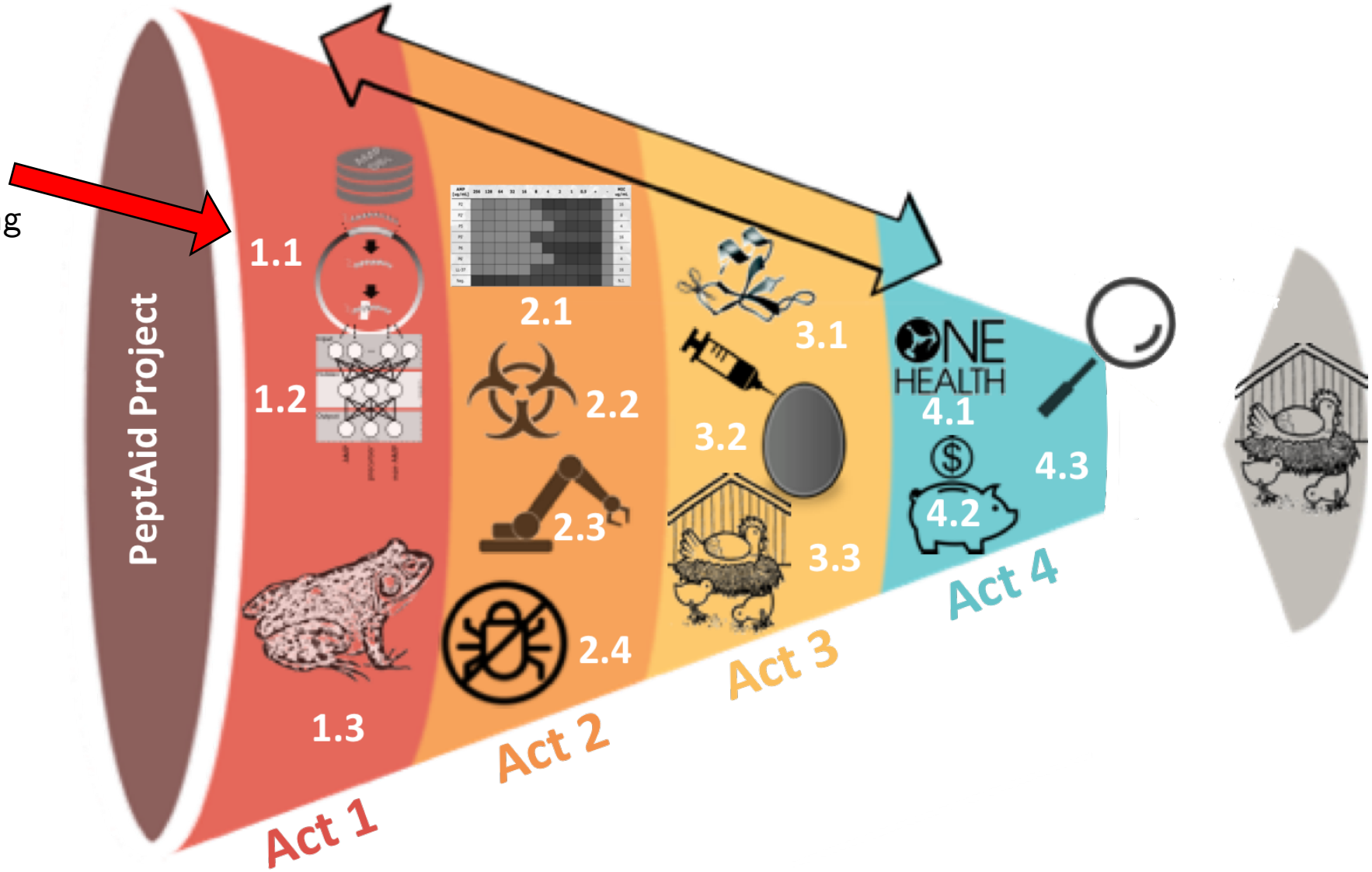
Antibiotic Resistance

SOLUTION

Antimicrobial Peptides!

Antimicrobial Peptides to Replace Antibiotics in Farm Veterinary Practice

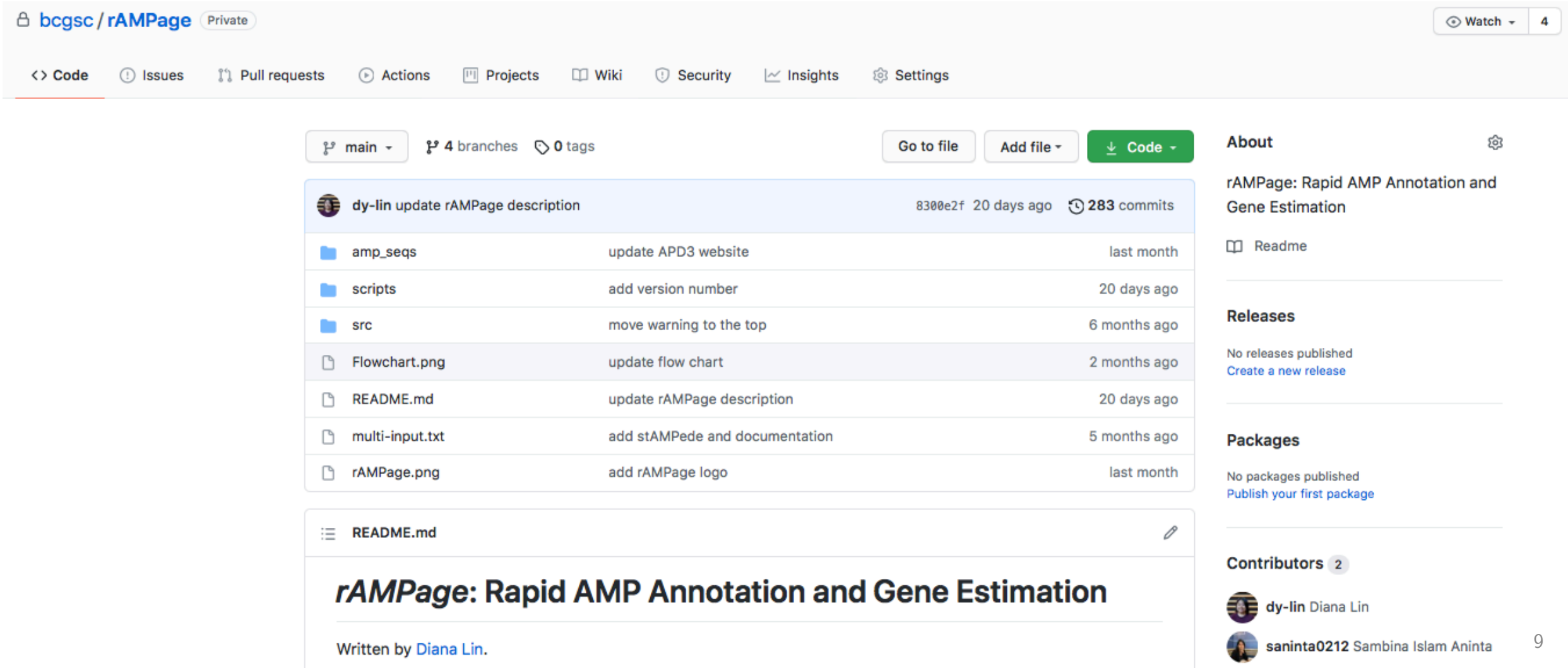
Activity 1.1:
Genome Mining



Objectives

- To develop **and** execute an AMP discovery pipeline to mine for AMP sequences in publicly available genomic resources
- To find candidate AMPs suitable for synthesis and wet lab validation
- To improve existing AMP annotation in AMP databases

rAMPage: Rapid Anti-microbial Peptide Annotation and Gene Estimation



bcgsc / rAMPage Private

Watch 4

Code | Issues | Pull requests | Actions | Projects | Wiki | Security | Insights | Settings

main | 4 branches | 0 tags | Go to file | Add file | Code

dy-lin update rAMPage description — 8300e2f 20 days ago — 283 commits

amp_seqs	update APD3 website	last month
scripts	add version number	20 days ago
src	move warning to the top	6 months ago
Flowchart.png	update flow chart	2 months ago
README.md	update rAMPage description	20 days ago
multi-input.txt	add stAMPede and documentation	5 months ago
rAMPage.png	add rAMPage logo	last month

README.md

rAMPage: Rapid AMP Annotation and Gene Estimation

Written by Diana Lin.

About

rAMPage: Rapid AMP Annotation and Gene Estimation

Readme

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package

Contributors 2

dy-lin Diana Lin

saninta0212 Sambina Islam Aninta

RNA-seq Datasets

Sequence Read Archive

Frogs & Toads (38)

Frogs (33)

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

Toads (5)

A. minuta
B. gargarizans
L. Boringii
M. sangzhiensis
O. rhodostigmatus

Ants, Bees, & Wasps (37)

Ants (8)

A. echinatior
C. castaneus
C. obscurior
M. gulosa
O. monticola
P. barbatus
T. bicarinatum
T. rugulatus

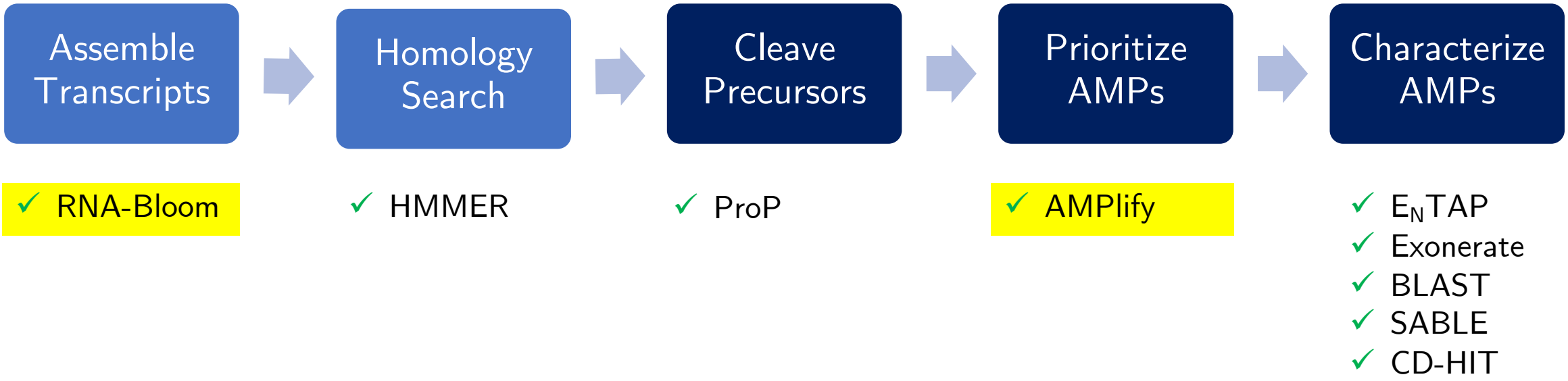
Bees (5)

A. cerana
A. mellifera
B. ardens
B. consobrinus
B. ussurensis

Wasps (24)

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

rAMPage Pipeline



Homology Search: Databases

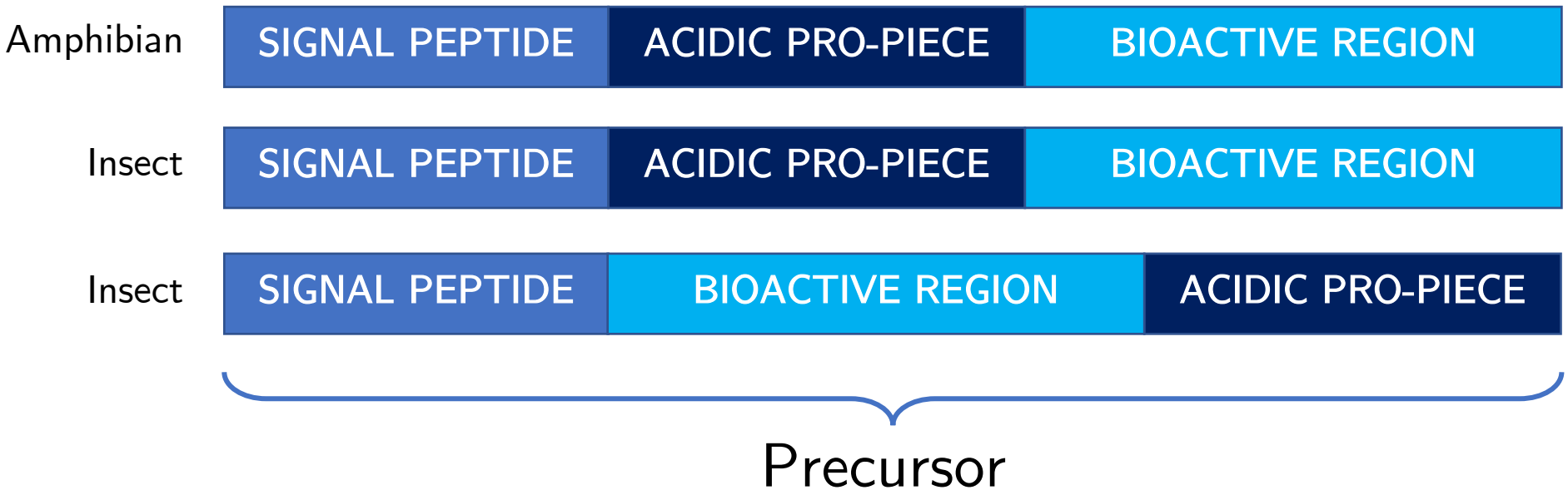
May 10, 2021

Database	Amphibians (nr)	Insects (nr)	All
APD3 Sep. 2020	1,075 +34	310 +13	3,125 +97
DADP Sep. 2020	1,921	0	1,921
NCBI Jul. 2020	2,850 +41	985 +57	185,967
"Reference" AMPs	4,663 +73	1,204 +62	N/A

APD3: Antimicrobial Peptide Database 3, https://wangapd3.com/main.php
DADP: Database of Anuran Defense Peptides, http://split4.pmfst.hr/dadp/

Homology Search

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



Johnson, L. S., Eddy, S. R. & Portugaly, E. **Hidden Markov model speed heuristic and iterative HMM search procedure.** BMC Bioinformatics 11, 431 (2010).

Cleaving Precursor AMPs: Amphibians

R. catesbeiana
American bullfrog

>rcatesbeiana-back_skin-139086.p2

MFTMKKSLLLLFFLGTISLSLCEQERNADDDQGEVIEQKVKRAFLSTVKNTLTNVAGTMIDTFKCKITGVC

SIGNAL PEPTIDE	ACIDIC PRO-PIECE	BIOACTIVE REGION

signal peptide cleavage site

propeptide cleavage site

```
MFTMKKSLLLLFFLGTISLSLCEQERNADDDQGEVIEQKVKRAFLSTVKNTLTNVAGTMIDTFKCKITGVC  80
sssssssssssssssssssss.....................P.............................  80
```

signal peptide

acidic pro-piece

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

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

Duckert, P., Brunak, S. & Blom, N. **Prediction of proprotein convertase cleavage sites.** Protein Eng. Des. Sel. 17, 107–112 (2004).

Cleaving Precursor AMPs: Insects

M. gulosa
red bulldog ant

>mgulosa-venom-325.p1
MRTFIVLIALAIICAVIEA GRPNPVNTKPTPYPRLRR EAEAEAETEPGRPQNWNNKPTPNPRGRR
EAEANAVAASPLGTIQTRPTGPPRPRREAEPHAEALMPGAPSVHVVKPTYAPPPRGRREAEPGRLNPVNNKPTPHPRLRR
EAEAEAEPGRPNPVNNKPTPHPRLRR EAETEAEAEAEPGRPQNWNNKPTPNPR

SIGNAL PEPTIDE	ACIDIC PRO-PIECE?	BIOACTIVE REGION?

SIGNAL PEPTIDE	BIOACTIVE REGION?	ACIDIC PRO-PIECE?

signal peptide
signal peptide cleavage site
propeptide cleavage site

```
MRTFIVLIALAIICAVIEAGRPNPVNTKPTPYPRLRREAEAEAETEPGRPQNWNNKPTPNPRGRREAEANAVAASPLGTI    80
QTRPTGPPRPRREAEPHAEALMPGAPSVHVVKPTYAPPPRGRREAEPGRLNPVNNKPTPHPRLRREAEAEAEPGRPNPVN   160
NKPTPHPRLRREAETEAEAEAEPGRPQNWNNKPTPNPR                                              240
sssssssssssssssss.................P.....................................            80
.....................................P.....................P..............            160
..........P..........................                                               240
```

propeptide cleavage site
acidic pro-piece?

```
Signal peptide cleavage site predicted: between pos. 19 and 20: IEA-GR

Propeptide cleavage sites predicted:    Arg(R)/Lys(K): 4
```

Duckert, P., Brunak, S. & Blom, N. **Prediction of proprotein convertase cleavage sites.** Protein Eng. Des. Sel. 17, 107–112 (2004).

Prioritizing AMPs: AMPlify

Sequence → AMPlify → score >= 0.50 → AMP non-AMP

Criteria
- ✓ AMPlify score >= 0.50
- ✓ Length <= 50aa
- ✓ Charge >= +2

Sequence ID	Sequence	Prediction	Score	Length	Charge	Criteria Met?
rcatesbeiana-back_skin-139086.p2-sig_mature-2	AFLSTVKNTLTNVAGTMIDTFKCKITGVC	AMP	0.9999962	29	2	
mgulosa-venom-325.p1-sig_mature-1	GRPNPVNTKPTPYPRLRR	AMP	0.9907686	18	5	
mgulosa-venom-325.p1-sig_mature-2	EAEAEAETEPGRPQNWNNKPTPNPRGRREA EANAVAASPLGTIQTRPTGPPRPRREAEPHA EALMPGAPSVHVVKPTYAPPPRGRR	AMP	0.9987987	86	3	
mgulosa-venom-325.p1-sig_mature-3	EAEPGRLNPVNNKPTPHPRLRR	AMP	0.50080854	22	3	
mgulosa-venom-325.p1-sig_mature-4	EAEAEAEPGRPNPVNNKPTPHPRLRR	non-AMP	0.25241083	26	1	
mgulosa-venom-325.p1-sig_mature-5	EAETEAEAEAEPGRPQNWNNKPTPNPR	non-AMP	0.017407548	27	-3	

Li, C. et al. **AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens.** bioRxiv (2020).

Prioritizing AMPs: AMPlify

Sequence → AMPlify → score >= 0.50 → **AMP** non-AMP

Criteria
- ✓ AMPlify score >= 0.50
- ✓ Length <= 50aa
- ✓ Charge >= +2

Sequence ID	Sequence	Prediction	Score	Length	Charge	Criteria Met?
rcatesbeiana-back_skin-139086.p2-sig_mature-2	AFLSTVKNTLTNVAGTMIDTFKCKITGVC	AMP	0.9999962	29	2	
mgulosa-venom-325.p1-sig_mature-1	GRPNPVNTKPTPYPRLRR	AMP	0.9907686	18	5	
mgulosa-venom-325.p1-sig_mature-2	EAEAEAETEPGRPQNWNNKPTPNPRGRREA EANAVAASPLGTIQTRPTGPPRPRREAEPHA EALMPGAPSVHVVKPTYAPPPRGRR	AMP	0.9987987	86	3	
mgulosa-venom-325.p1-sig_mature-3	EAEPGRLNPVNNKPTPHPRLRR	AMP	0.50080854	22	3	
mgulosa-venom-325.p1-sig_mature-4	EAEAEAEPGRPNPVNNKPTPHPRLRR	non-AMP	0.25241083	26	1	
mgulosa-venom-325.p1-sig_mature-5	EAETEAEAEAEPGRPQNWNNKPTPNPR	non-AMP	0.017407548	27	-3	

Li, C. et al. **AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens.** bioRxiv (2020).

Prioritizing AMPs: AMPlify

Sequence → AMPlify → score >= 0.50 → **AMP** non-AMP

Criteria
- ✓ AMPlify score >= 0.50
- ✓ Length <= 50aa
- ✓ Charge >= +2

Sequence ID	Sequence	Prediction	Score	Length	Charge	Criteria Met?
rcatesbeiana-back_skin-139086.p2-sig_mature-2	AFLSTVKNTLTNVAGTMIDTFKCKITGVC	AMP	0.9999962	29	2	
mgulosa-venom-325.p1-sig_mature-**1**	GRPNPVNTKPTPYPRLRR	AMP	0.9907686	18	5	
mgulosa-venom-325.p1-sig_mature-**2**	EAEAEAETEPGRPQNWNNKPTPNPRGRREA EANAVAASPLGTIQTRPTGPPRPRREAEPHA EALMPGAPSVHVVKPTYAPPPRGRR	AMP	0.9987987	86	3	
mgulosa-venom-325.p1-sig_mature-**3**	EAEPGRLNPVNNKPTPHPRLRR	AMP	0.50080854	22	3	
mgulosa-venom-325.p1-sig_mature-**4**	EAEAEAEPGRPNPVNNKPTPHPRLRR	non-AMP	0.25241083	26	1	
mgulosa-venom-325.p1-sig_mature-**5**	EAETEAEAEAEPGRPQNWNNKPTPNPR	non-AMP	0.017407548	27	-3	

Li, C. et al. **AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens.** bioRxiv (2020).

Prioritizing AMPs: AMPlify

Sequence → AMPlify → score >= 0.50 → **AMP** non-AMP

Criteria
- ✓ AMPlify score >= 0.50
- ✓ Length <= 50aa
- ✓ Charge >= +2

Sequence ID	Sequence	Prediction	Score	Length	Charge	Criteria Met?
rcatesbeiana-back_skin-139086.p2-sig_mature-2	AFLSTVKNTLTNVAGTMIDTFKCKITGVC	AMP	0.9999962	29	2	
mgulosa-venom-325.p1-sig_mature-1	GRPNPVNTKPTPYPRLRR	AMP	0.9907686	18	5	
mgulosa-venom-325.p1-sig_mature-2	EAEAEAETEPGRPQNWNNKPTPNPRGRREA EANAVAASPLGTIQTRPTGPPRPRREAEPHA EALMPGAPSVHVVKPTYAPPPRGRR	AMP	0.9987987	86	3	
mgulosa-venom-325.p1-sig_mature-3	EAEPGRLNPVNNKPTPHPRLRR	AMP	0.50080854	22	3	
mgulosa-venom-325.p1-sig_mature-4	EAEAEAEPGRPNPVNNKPTPHPRLRR	non-AMP	0.25241083	26	1	
mgulosa-venom-325.p1-sig_mature-5	EAETEAEAEAEPGRPQNWNNKPTPNPR	non-AMP	0.017407548	27	-3	

Li, C. et al. **AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens.** bioRxiv (2020).

Prioritizing AMPs: AMPlify

Sequence → AMPlify → score >= 0.50 → **AMP** non-AMP

Criteria
- ✓ AMPlify score >= ~~0.50~~ 0.90
- ✓ Length <= 50aa
- ✓ Charge >= +2

Sequence ID	Sequence	Prediction	Score	Length	Charge	Criteria Met?
rcatesbeiana-back_skin-139086.p2-sig_mature-2	AFLSTVKNTLTNVAGTMIDTFKCKITGVC	AMP	0.9999962	29	2	
mgulosa-venom-325.p1-sig_mature-1	GRPNPVNTKPTPYPRLRR	AMP	0.9907686	18	5	
mgulosa-venom-325.p1-sig_mature-2	EAEAEAETEPGRPQNWNNKPTPNPRGRREA EANAVAASPLGTIQTRPTGPPRPRREAEPHA EALMPGAPSVHVVKPTYAPPPRGRR	AMP	0.9987987	86	3	
mgulosa-venom-325.p1-sig_mature-3	EAEPGRLNPVNNKPTPHPRLRR	AMP	0.50080854	22	3	
mgulosa-venom-325.p1-sig_mature-4	EAEAEAEPGRPNPVNNKPTPHPRLRR	non-AMP	0.25241083	26	1	
mgulosa-venom-325.p1-sig_mature-5	EAETEAEAEAEPGRPQNWNNKPTPNPR	non-AMP	0.017407548	27	-3	

Li, C. et al. **AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens.** bioRxiv (2020).

Prioritizing AMPs: AMPlify

Sequence → AMPlify → score >= 0.50 → **AMP** non-AMP

Criteria
- ✓ AMPlify score >= ~~0.50~~ 0.90
- ✓ Length <= 50aa
- ✓ Charge >= +2

Sequence ID	Sequence	Prediction	Score	Length	Charge	Criteria Met?
rcatesbeiana-back_skin-139086.p2-sig_mature-2	AFLSTVKNTLTNVAGTMIDTFKCKITGVC	AMP	0.9999962	29	2	
mgulosa-venom-325.p1-sig_mature-1	GRPNPVNTKPTPYPRLRR	AMP	0.9907686	18	5	
mgulosa-venom-325.p1-sig_mature-2	EAEAEAETEPGRPQNWNNKPTPNPRGRREA EANAVAASPLGTIQTRPTGPPRPRREAEPHA EALMPGAPSVHVVKPTYAPPPRGRR	AMP	0.9987987	86	3	
mgulosa-venom-325.p1-sig_mature-3	EAEPGRLNPVNNKPTPHPRLRR	AMP	0.50080854	22	3	
mgulosa-venom-325.p1-sig_mature-4	EAEAEAEPGRPNPVNNKPTPHPRLRR	non-AMP	0.25241083	26	1	
mgulosa-venom-325.p1-sig_mature-5	EAETEAEAEAEPGRPQNWNNKPTPNPR	non-AMP	0.017407548	27	-3	

Li, C. et al. **AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens.** bioRxiv (2020).

Prioritizing AMPs: AMPlify

Sequence → AMPlify → score >= 0.50 → **AMP** non-AMP

Criteria

- ✓ AMPlify score >= ~~0.50~~ 0.90
- ✓ Length <= ~~50aa~~ 30aa
- ✓ Charge >= +2

Sequence ID	Sequence	Prediction	Score	Length	Charge	Criteria Met?
rcatesbeiana-back_skin-139086.p2-sig_mature-2	AFLSTVKNTLTNVAGTMIDTFKCKITGVC	AMP	0.9999962	29	2	
mgulosa-venom-325.p1-sig_mature-1	GRPNPVNTKPTPYPRLRR	AMP	0.9907686	18	5	
mgulosa-venom-325.p1-sig_mature-2	EAEAEAETEPGRPQNWNNKPTPNPRGRREA EANAVAASPLGTIQTRPTGPPRPRREAEPHA EALMPGAPSVHVVKPTYAPPPRGRR	AMP	0.9987987	86	3	
mgulosa-venom-325.p1-sig_mature-3	EAEPGRLNPVNNKPTPHPRLRR	AMP	0.50080854	22	3	
mgulosa-venom-325.p1-sig_mature-4	EAEAEAEPGRPNPVNNKPTPHPRLRR	non-AMP	0.25241083	26	1	
mgulosa-venom-325.p1-sig_mature-5	EAETEAEAEAEPGRPQNWNNKPTPNPR	non-AMP	0.017407548	27	-3	

Li, C. et al. **AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens.** bioRxiv (2020).

Prioritizing AMPs: AMPlify

Sequence → AMPlify → score >= 0.50 → **AMP** non-AMP

Criteria

- ✓ AMPlify score >= ~~0.50~~ 0.90
- ✓ Length <= ~~50aa~~ 30aa
- ✓ Charge >= +2

Sequence ID	Sequence	Prediction	Score	Length	Charge	Criteria Met?
rcatesbeiana-back_skin-139086.p2-sig_mature-2	AFLSTVKNTLTNVAGTMIDTFKCKITGVC	AMP	0.9999962	29	2	✅
mgulosa-venom-325.p1-sig_mature-1	GRPNPVNTKPTPYPRLRR	AMP	0.9907686	18	5	✅
mgulosa-venom-325.p1-sig_mature-2	EAEAEAETEPGRPQNWNNKPTPNPRGRREA EANAVAASPLGTIQTRPTGPPRPRREAEPHA EALMPGAPSVHVVKPTYAPPPRGRR	AMP	0.9987987	86	3	❌
mgulosa-venom-325.p1-sig_mature-3	EAEPGRLNPVNNKPTPHPRLRR	AMP	0.50080854	22	3	❌
mgulosa-venom-325.p1-sig_mature-4	EAEAEAEPGRPNPVNNKPTPHPRLRR	non-AMP	0.25241083	26	1	❌
mgulosa-venom-325.p1-sig_mature-5	EAETEAEAEAEPGRPQNWNNKPTPNPR	non-AMP	0.017407548	27	-3	❌

Li, C. et al. **AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens.** bioRxiv (2020).

Characterizing AMPs

Novelty Criteria
✓ Alignment < 100%

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

Species Count	Sequence_ID	AMPlify Score	Length	Charge	Top Precursor	Top Mature	GO Term	IPScan	Pfam	Secondary Structure	Criteria Met?
1	rcatesbeiana-back_skin-139086.p2-sig_mature-2	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**)	CCEEHHHHHH CCCCCCCCCC EEEEECCCC	✅
1	mgulosa-venom-325.p1-sig_mature-1	0.9907686	18	5	NA	P81437.1: **Formaecin**-2 [Myrmecia gulosa] (100%)	GO:0042381 (hemolymph coagulation) GO:0042742 (**defense response to bacterium**)	IPR012514 (Formaecin)	PF08106 (Formaecin family)	CCCCCCCCCC CCCCCCCC	❌

Final Output

- FASTA file

>rcatesbeiana-back_skin-139086.p2-sig_mature-2 length=29 score=0.9999962 prediction=AMP charge=2
AFLSTVKNTLTNVAGTMIDTFKCKITGVC

>mgulosa-venom-325.p1-sig_mature-1 length=18 score=0.9907686 prediction=AMP charge=5
GRPNPVNTKPTPYPRLRR

>rsylvatica-skin-357168.p1-no_sig_mature-2 length=28 score=0.9976821 prediction=AMP charge=10
SSKKKKCKFFCKLKKKINSIPIISIPFK

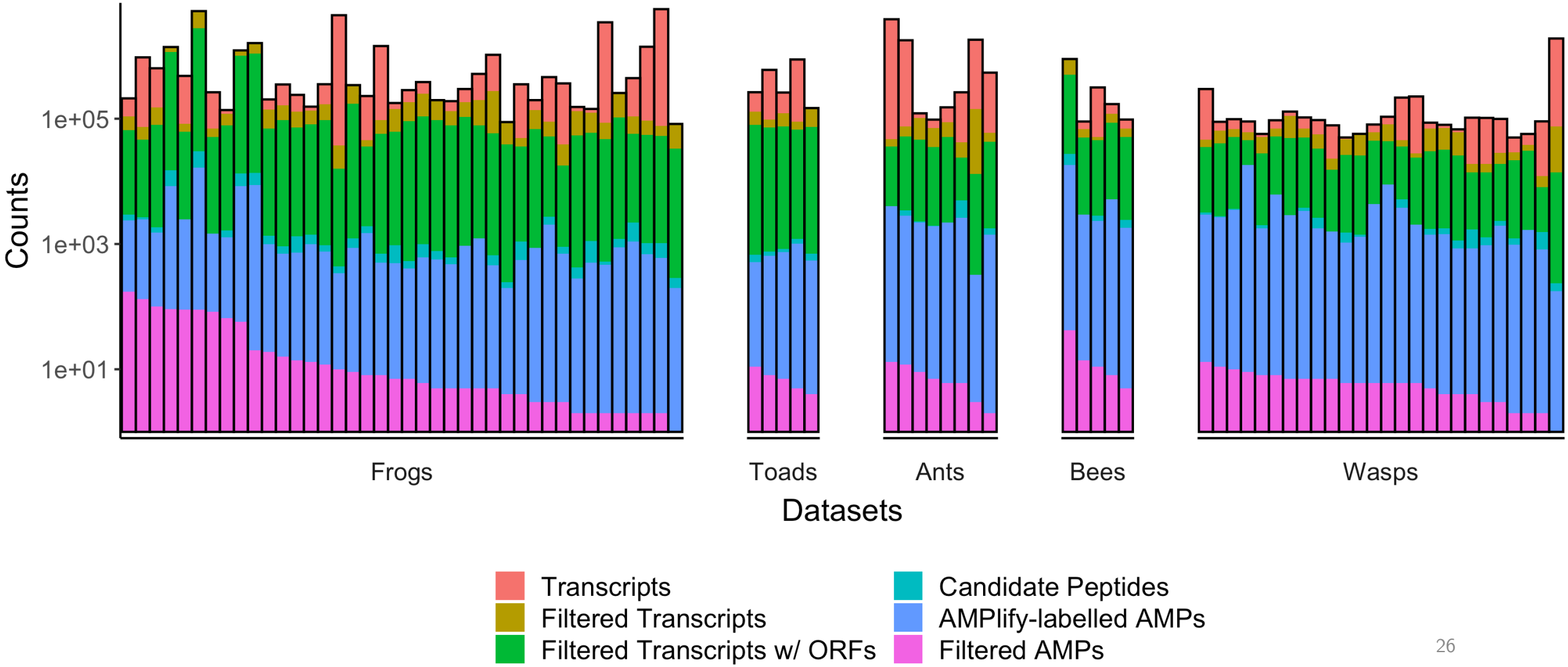
- TSV file

Criteria

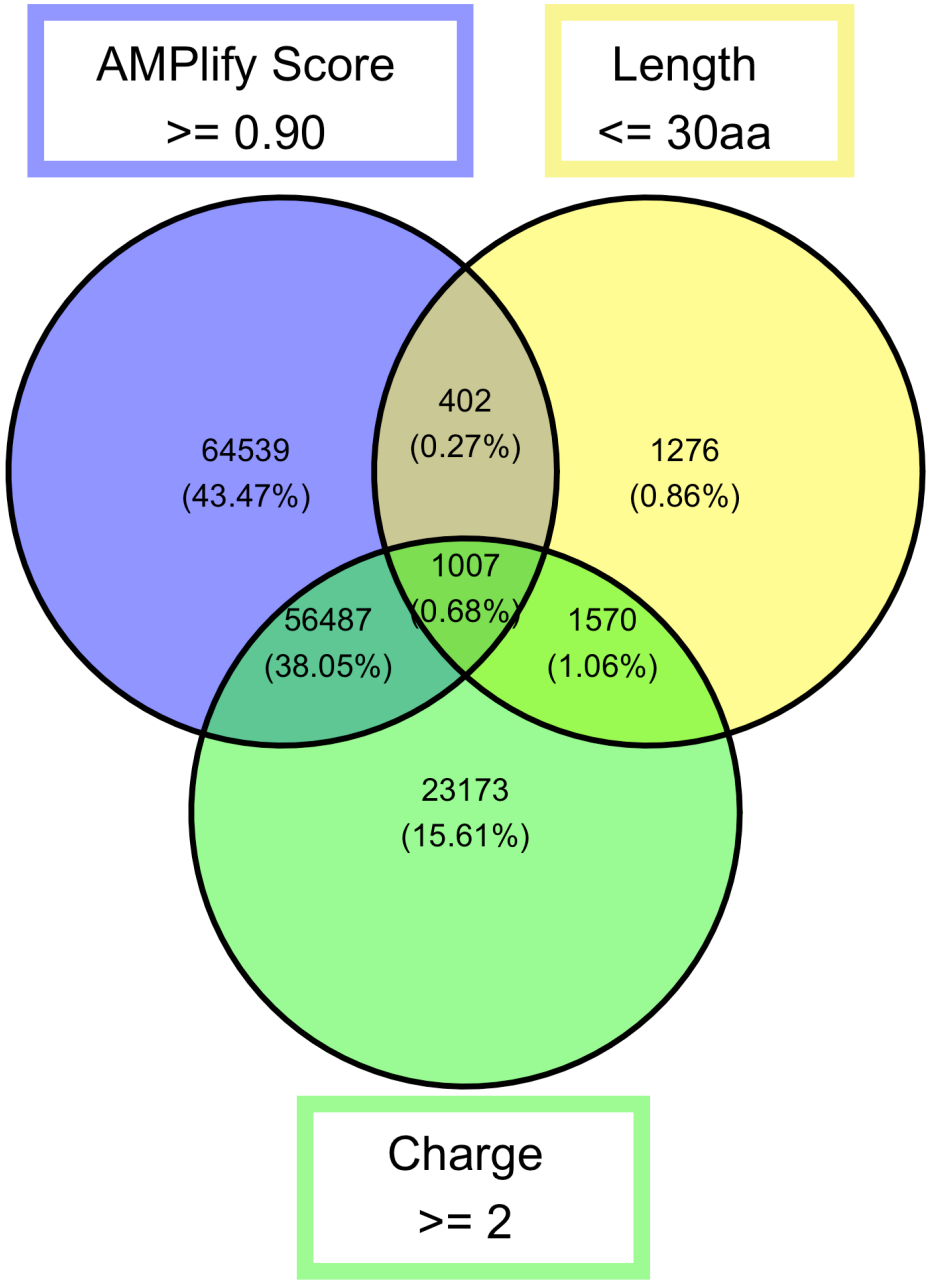
- ✓ AMPlify score >= 0.90
- ✓ Length <= 30aa
- ✓ Charge >= +2

Results

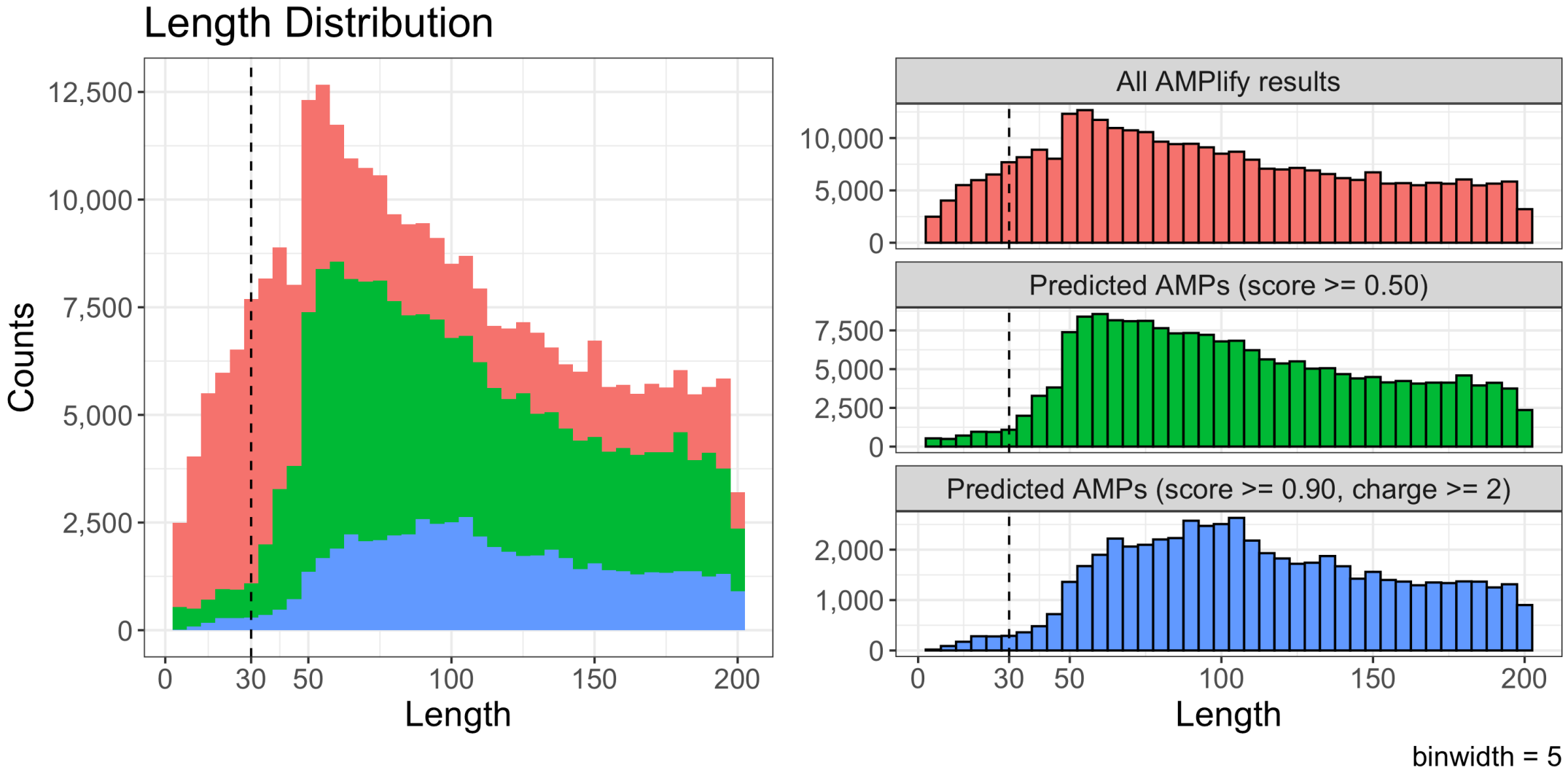
Filtering criteria: AMPlify score >= 0.90, Length <= 30aa, Charge >= 2
Total Transcripts: 53,031,368
Total AMPs: 1,332



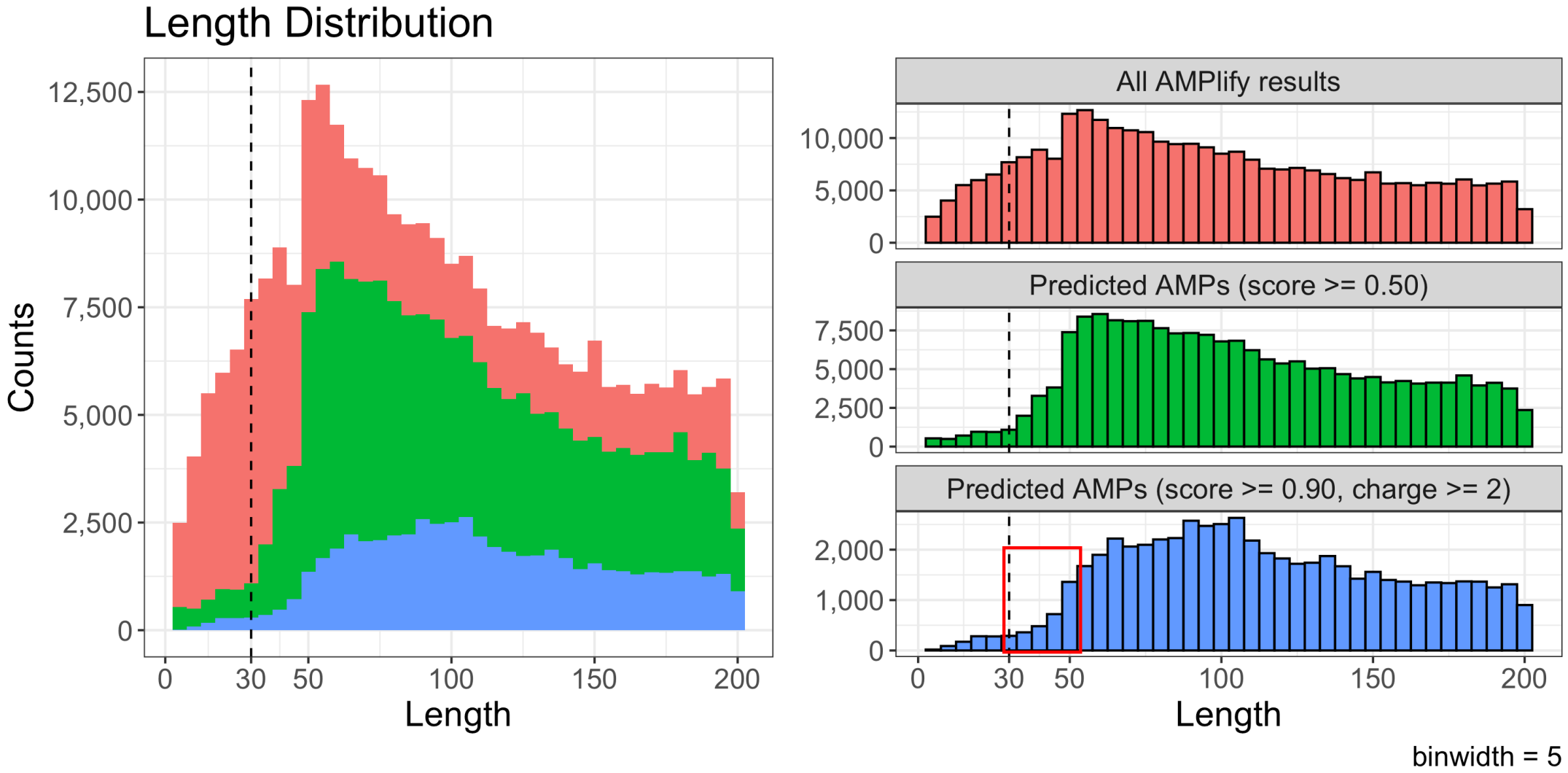
Results


AMPlify Score
>= 0.90
Length
<= 30aa
64539
(43.47%)
402
(0.27%)
1276
(0.86%)
1007
(0.68%)
56487
(38.05%)
1570
(1.06%)
23173
(15.61%)
Charge
>= 2

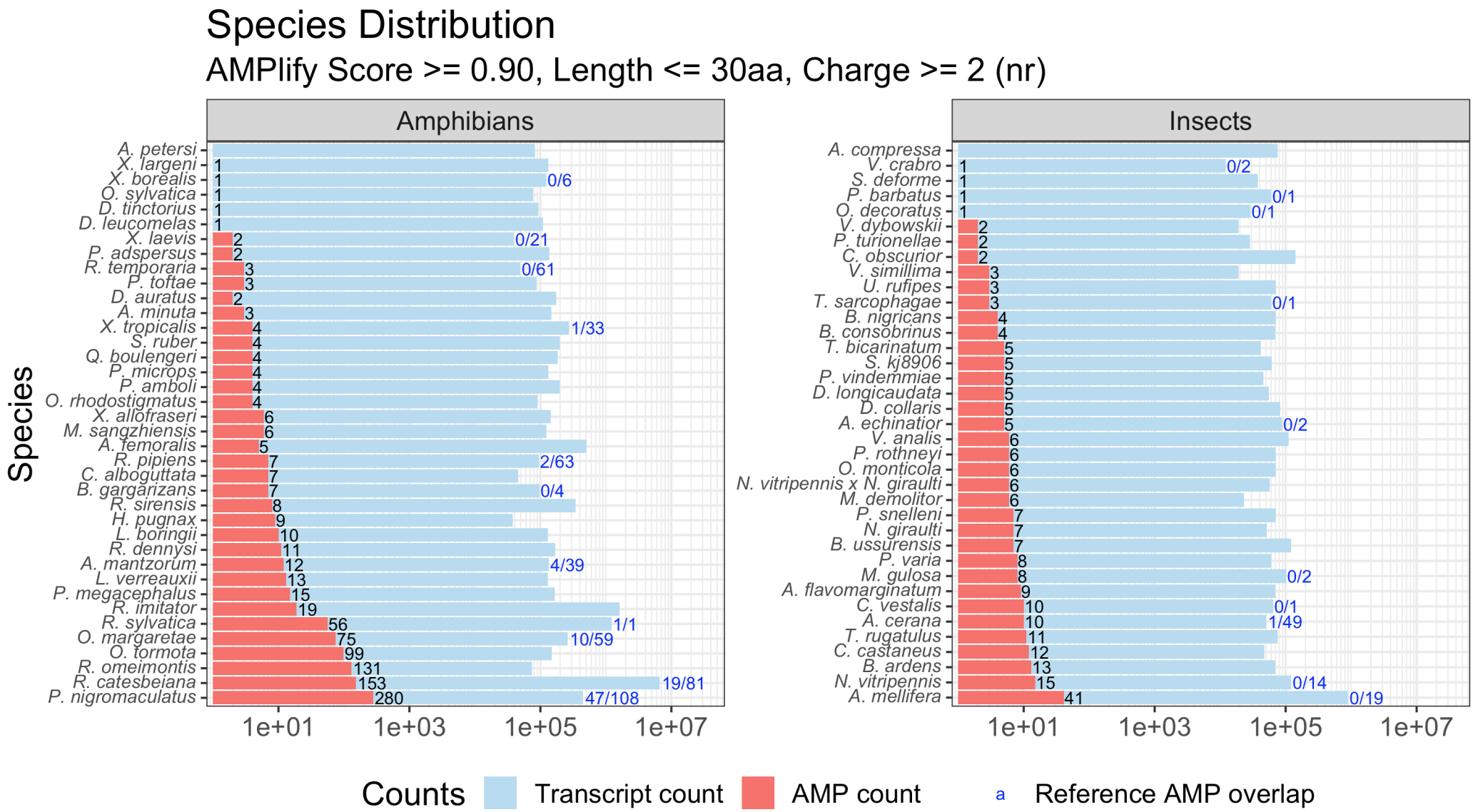
Results


Length Distribution
Counts
12,500
10,000
7,500
5,000
2,500
0
0
30
50
100
150
200
Length
All AMPlify results
10,000
5,000
0
Predicted AMPs (score >= 0.50)
7,500
5,000
2,500
0
Predicted AMPs (score >= 0.90, charge >= 2)
2,000
1,000
0
0
30
50
100
150
200
Length
binwidth = 5

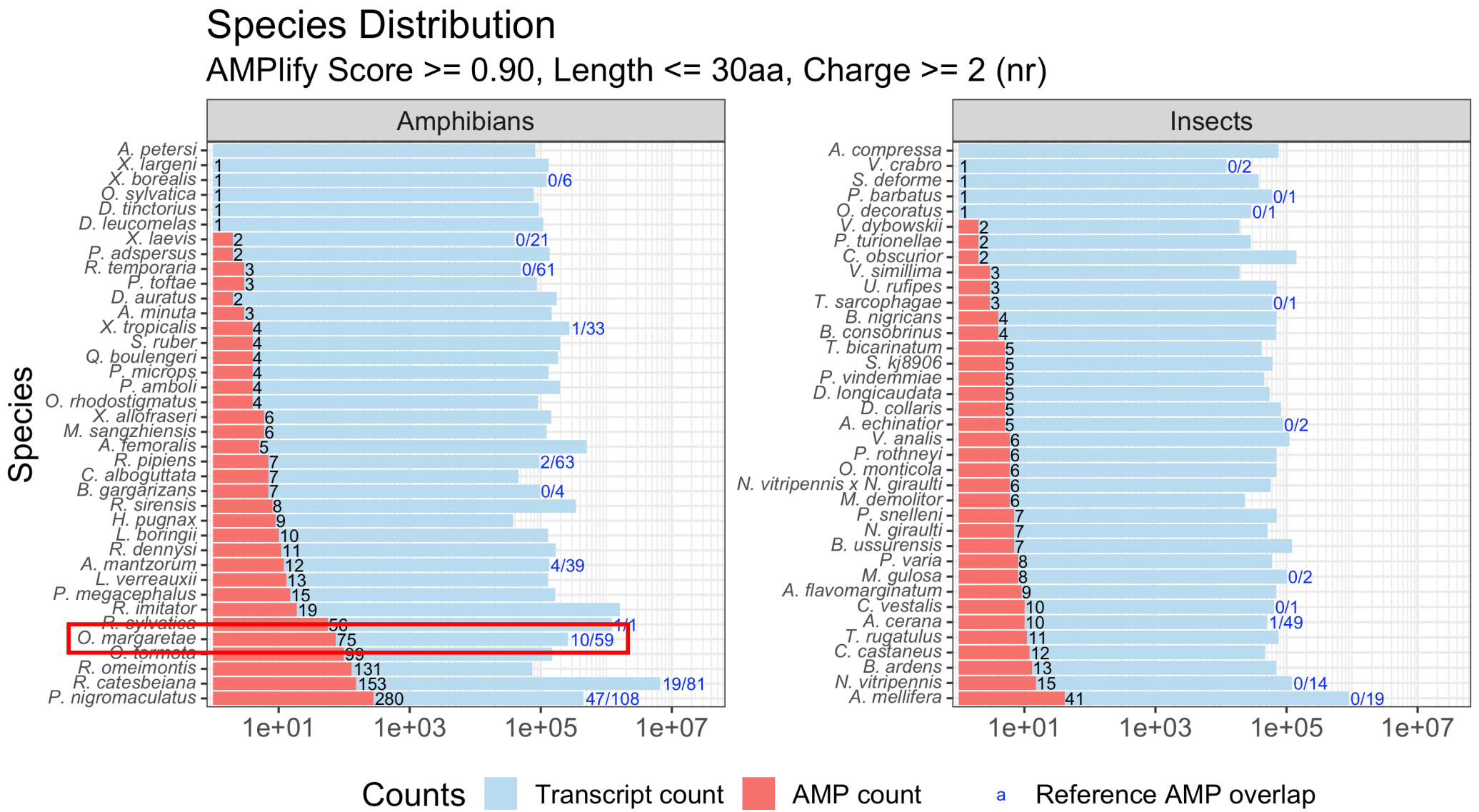
Results


Length Distribution
Counts
12,500
10,000
7,500
5,000
2,500
0
0
30
50
100
150
200
Length
All AMPlify results
10,000
5,000
0
Predicted AMPs (score >= 0.50)
7,500
5,000
2,500
0
Predicted AMPs (score >= 0.90, charge >= 2)
2,000
1,000
0
0
30
50
100
150
200
Length
binwidth = 5

Results


Species Distribution
AMPlify Score >= 0.90, Length <= 30aa, Charge >= 2 (nr)
Amphibians
Insects
Species
A. petersi
X. largeni
X. borealis
O. sylvatica
D. tinctorius
D. leucomelas
X. laevis
P. adspersus
R. temporaria
P. toftae
D. auratus
A. minuta
X. tropicalis
S. ruber
Q. boulengeri
P. microps
P. amboli
O. rhodostigmatus
X. allofraseri
M. sangzhiensis
A. femoralis
R. pipiens
C. alboguttata
B. gargarizans
R. sirensis
H. pugnax
L. boringii
R. dennysi
A. mantzorum
L. verreauxii
P. megacephalus
R. imitator
R. sylvatica
O. margaretae
O. tormota
R. omeimontis
R. catesbeiana
P. nigromaculatus
1 1 1 1 1 2 2 3 3 2 3 4 4 4 4 4 4 6 6 5 7 7 7 8 9 10 11 12 13 15 19 56 75 99 131 153 280
0/6 0/21 0/61 1/33 2/63 0/4 4/39 1/1 10/59 19/81 47/108
A. compressa
V. crabro
S. deforme
P. barbatus
O. decoratus
V. dybowskii
P. turionellae
C. obscurior
V. simillima
U. rufipes
T. sarcophagae
B. nigricans
B. consobrinus
T. bicarinatum
S. kj8906
P. vindemmiae
D. longicaudata
D. collaris
A. echinatior
V. analis
P. rothneyi
O. monticola
N. vitripennis x N. giraulti
M. demolitor
P. snelleni
N. giraulti
B. ussurensis
P. varia
M. gulosa
A. flavomarginatum
C. vestalis
A. cerana
T. rugatulus
C. castaneus
B. ardens
N. vitripennis
A. mellifera
1 1 1 1 1 2 2 2 3 3 3 4 4 5 5 5 5 5 5 6 6 6 6 6 7 7 7 8 8 9 10 10 11 12 13 15 41
0/2 0/1 0/1 0/1 0/2 0/2 0/1 1/49 0/14 0/19
1e+01 1e+03 1e+05 1e+07
Counts
Transcript count
AMP count
a
Reference AMP overlap

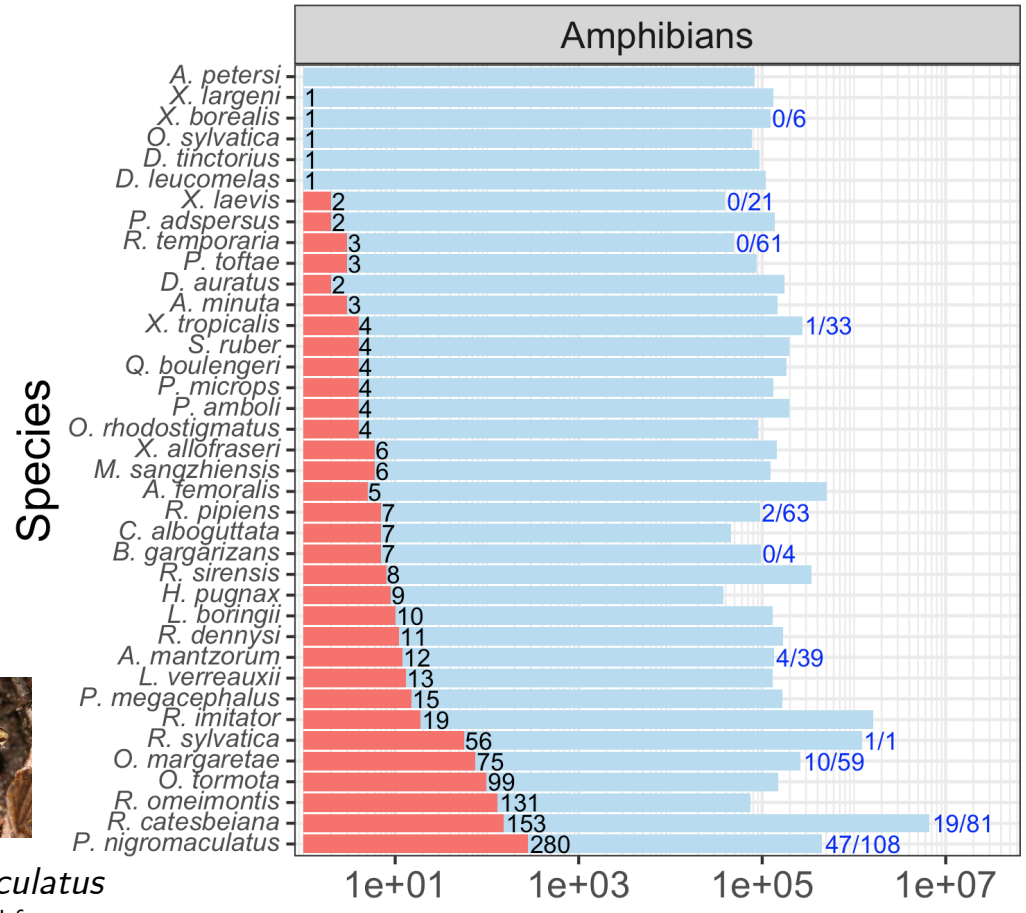
Results

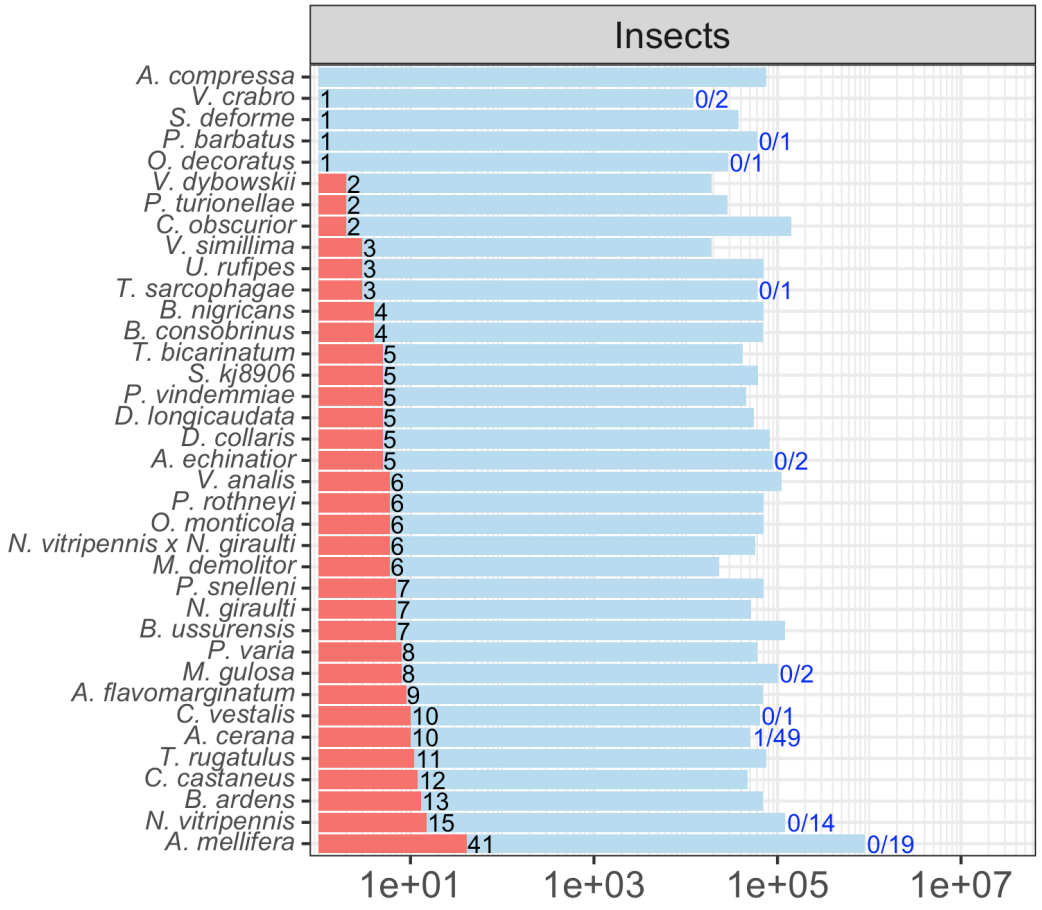

Species Distribution
AMPlify Score >= 0.90, Length <= 30aa, Charge >= 2 (nr)
Amphibians
Insects
Species
A. petersi
X. largeni
X. borealis
O. sylvatica
D. tinctorius
D. leucomelas
X. laevis
P. adspersus
R. temporaria
P. toftae
D. auratus
A. minuta
X. tropicalis
S. ruber
Q. boulengeri
P. microps
P. amboli
O. rhodostigmatus
X. allofraseri
M. sangzhiensis
A. femoralis
R. pipiens
C. alboguttata
B. gargarizans
R. sirensis
H. pugnax
L. boringii
R. dennysi
A. mantzorum
L. verreauxii
P. megacephalus
R. imitator
R. sylvatica
O. margaretae
O. tormota
R. omeimontis
R. catesbeiana
P. nigromaculatus
1 1 1 1 1 2 2 3 3 2 3 4 4 4 4 4 6 6 5 7 7 7 8 9 10 11 12 13 15 19 56 75 99 131 153 280
0/6 0/21 0/61 1/33 4/39 2/63 0/4 1/1 10/59 19/81 47/108
A. compressa
V. crabro
S. deforme
P. barbatus
O. decoratus
V. dybowskii
P. turionellae
C. obscurior
V. simillima
U. rufipes
T. sarcophagae
B. nigricans
B. consobrinus
T. bicarinatum
S. kj8906
P. vindemmiae
D. longicaudata
D. collaris
A. echinatior
V. analis
P. rothneyi
O. monticola
N. vitripennis x N. giraulti
M. demolitor
P. snelleni
N. giraulti
B. ussurensis
P. varia
M. gulosa
A. flavomarginatum
C. vestalis
A. cerana
T. rugatulus
C. castaneus
B. ardens
N. vitripennis
A. mellifera
1 1 1 1 1 2 2 2 3 3 3 4 4 5 5 5 5 5 5 6 6 6 6 6 7 7 7 8 8 9 10 10 11 12 13 15 41
0/2 0/1 0/1 0/1 0/2 0/2 0/1 1/49 0/14 0/19
1e+01 1e+03 1e+05 1e+07
Counts
Transcript count
AMP count
a
Reference AMP overlap

Results

Species Distribution

AMPlify Score >= 0.90, Length <= 30aa, Charge >= 2 (nr)



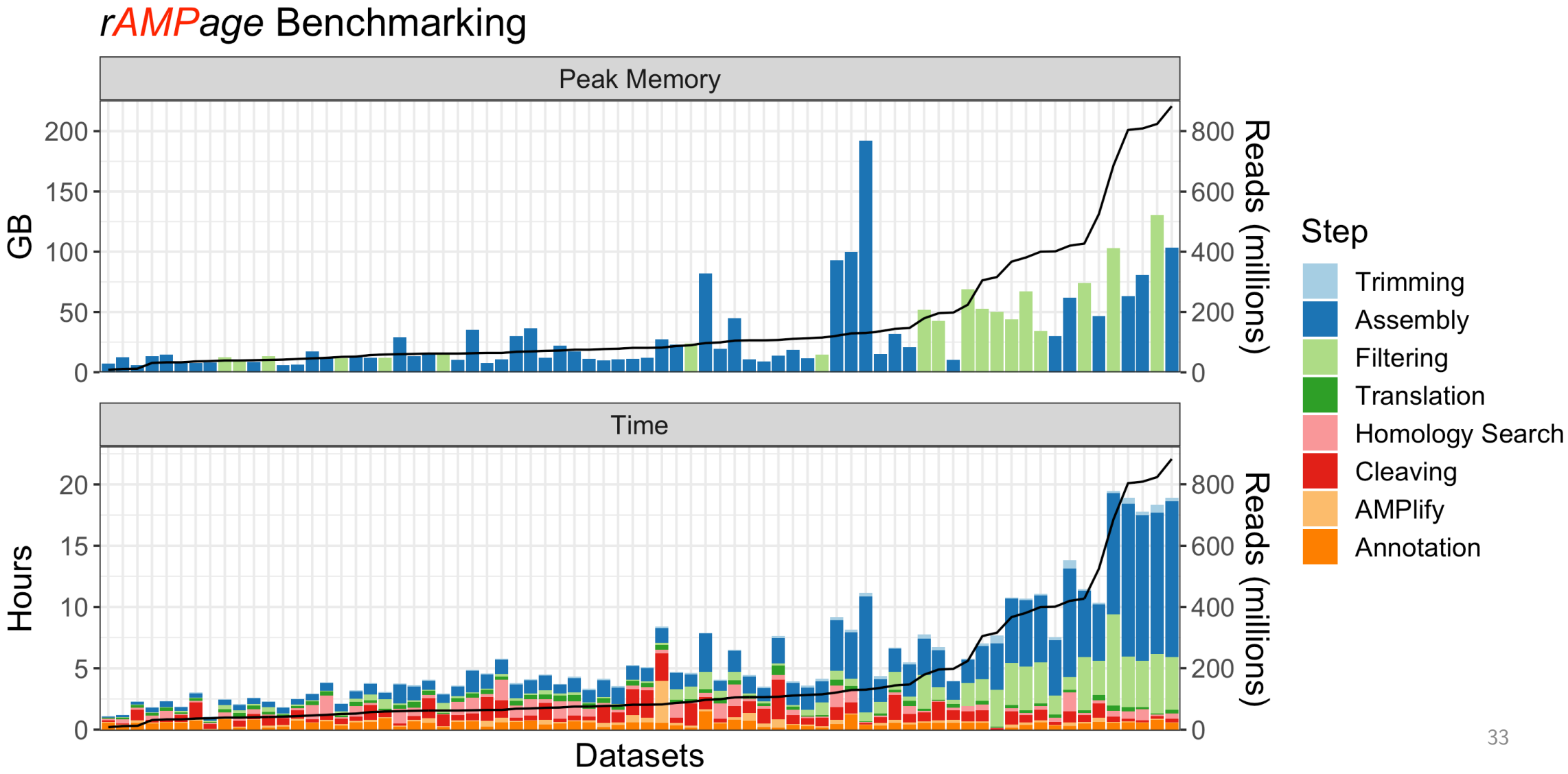


P. nigromaculatus
dark-spotted frog

A. mellifera
honeybee

Counts Transcript count AMP count a Reference AMP overlap

Results


rAMPage Benchmarking
Peak Memory
GB
200
150
100
50
0
800
600
400
200
0
Reads (millions)
Time
Hours
20
15
10
5
0
Datasets
Step
Trimming
Assembly
Filtering
Translation
Homology Search
Cleaving
AMPlify
Annotation

Summary

- ✓ ***rAMPage*** pipeline runs all the tools from beginning to end, from reads to candidate AMPs

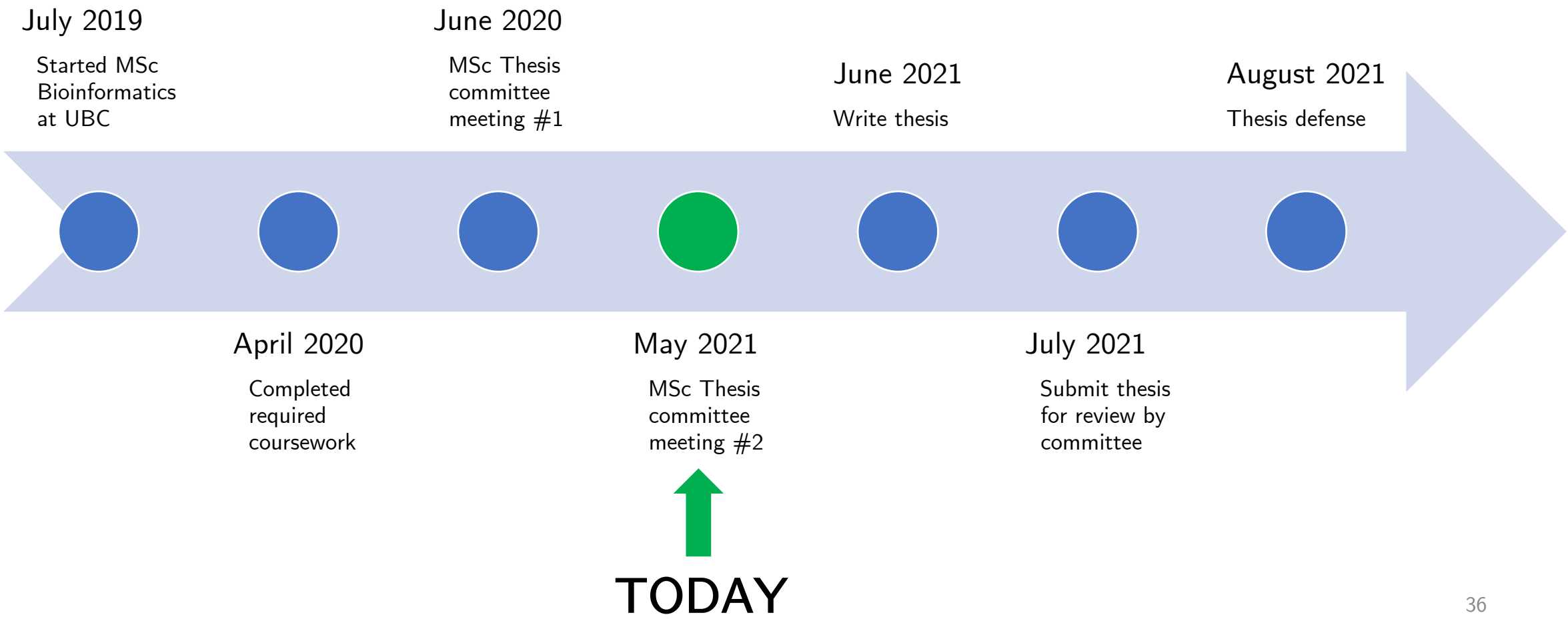
- ✓ 28 candidate AMPs from ***rAMPage*** selected for synthesis and wet lab validation

Thesis Outline

Working Title: Discovering anti-microbial peptides in amphibian and insect transcriptomes

1. Introduction
   i. Anti-microbial Peptides (AMPs)
   ii. Characteristics of AMPs
   iii. Structure of AMPs
   iv. AMPs in Therapeutics and AMR
   v. Thesis Objectives and Motivations
2. Methods
   i. Collation of Input RNA-seq Datasets
   ii. Curation of Reference AMP Datasets
   iii. rAMPage Pipeline
3. Results
   i. AMPs Discovered
   ii. Novelty of AMPs
   iii. rAMPage Benchmarks
4. Discussion and Conclusions

Timeline


July 2019
Started MSc Bioinformatics at UBC
April 2020
Completed required coursework
June 2020
MSc Thesis committee meeting #1
May 2021
MSc Thesis committee meeting #2
TODAY
June 2021
Write thesis
July 2021
Submit thesis for review by committee
August 2021
Thesis defense

Acknowledgements

Birol Lab

- Sambina Islam Aninta
- Ka Ming Nip
- Kristina Gagalova
- Darcy Sutherland
- Chenkai Li
- René Warren
- Inanc Birol

PeptAID Collaborators

- Caren Helbing

PeptAID Funding

Genome BritishColumbia

GenomeCanada

BC CANCER

Investment Agriculture Foundation of British Columbia

UBC

University of Victoria

Questions?