



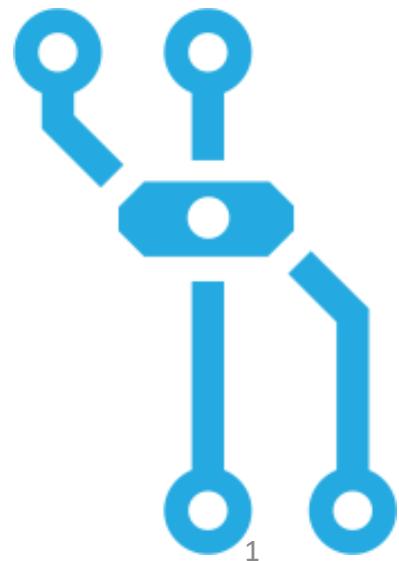
Mining Amphibian and Insect Transcriptomes for Antimicrobial Peptides



BioTalk
Friday, July 3, 2020

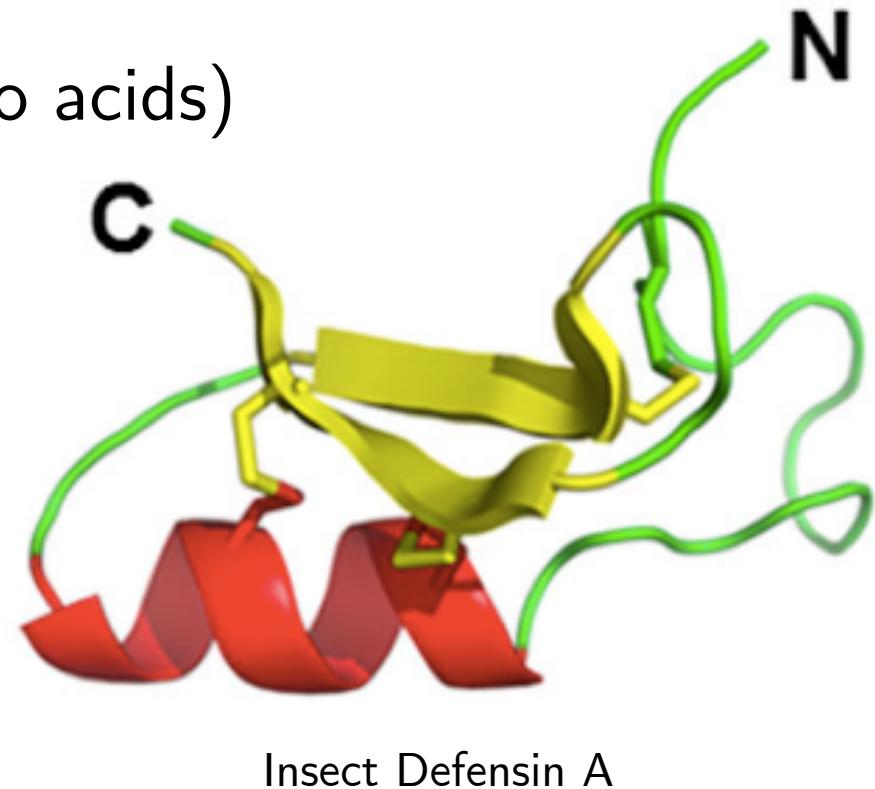
Diana Lin

MSc Student, Bioinformatics Graduate Program, UBC
Dr. Inanc Birol Lab, Genome Sciences Centre (GSC), BC Cancer

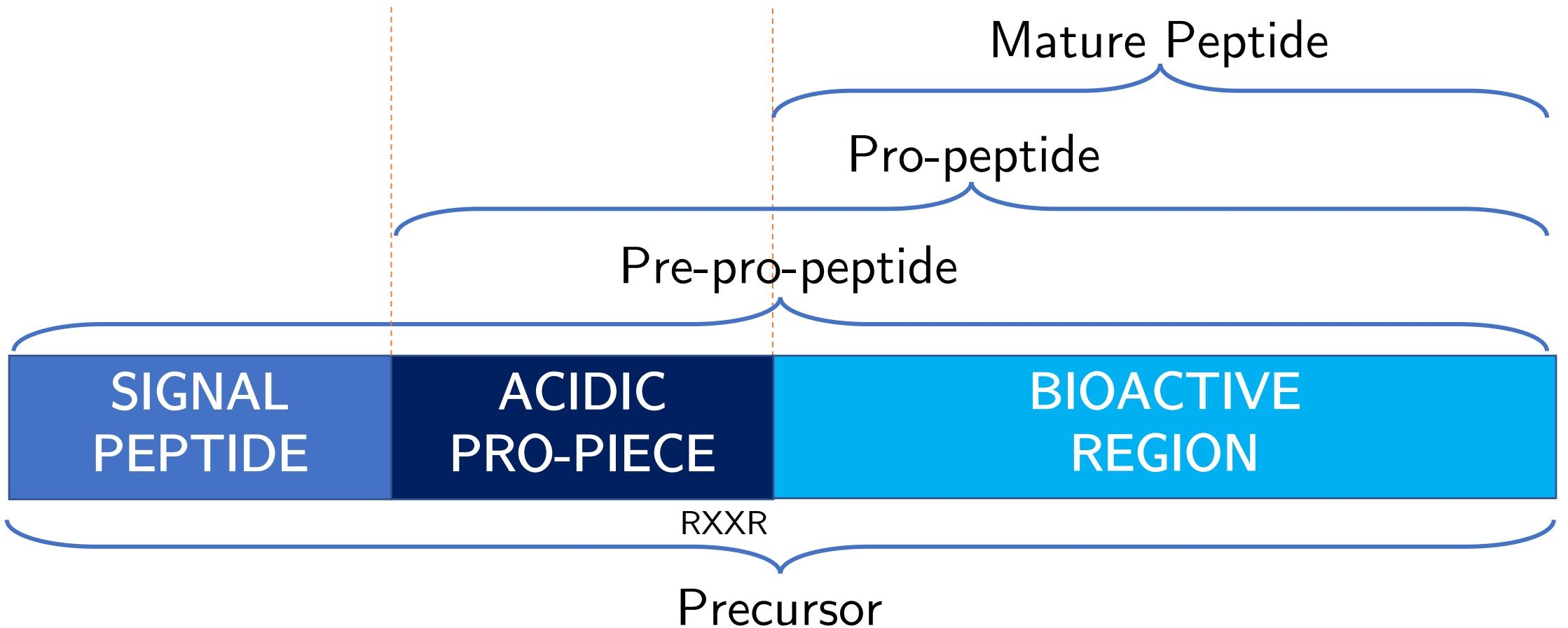


Antimicrobial Peptides (AMPs)

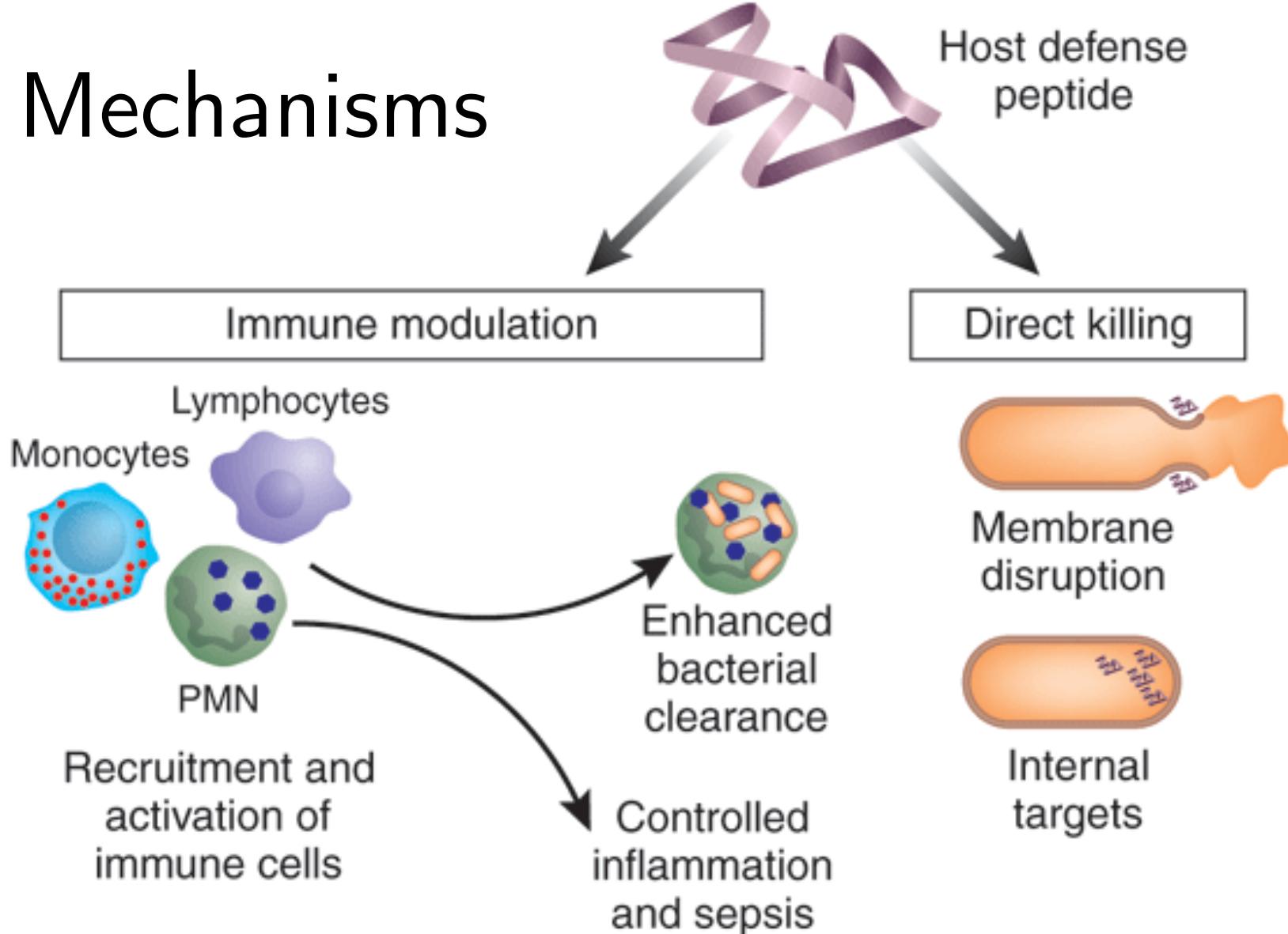
- Short peptide sequences (5 to 50 amino acids)
- Produced by all life forms
- Part of the innate immune system
- Often positively charged
- Amphipathic



AMP Structure



AMP Mechanisms



Source: Hancock, R. E. W. & Sahl, H.-G. Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat. Biotechnol. 24, 1551–1557 (2006).

AMPs Under Clinical Investigation (-2016)

Table 3

Antimicrobial peptides in clinical development.

Peptide	AMP source (host)	Status	Administration	Indication	Company
OP-145 hLF1-11 (Lactoferrin)	LL-37 (human) Lactoferrin 1-11 (human)	Phase I/II Not specified	Ear drops Intravenous	Chronic bacterial ear infection Neutropenic stem cell transplantation patients	OctoPlus Inc. AM-Pharma B.V.
Pexiganan (MSI-78)	Magainin (frog)	Phase III	Topical cream	Diabetic foot infection	Dipexium Pharmaceuticals, Inc.
Iseganan (IB-367)	Protegrin-1 (porcine leukocytes)	Phase III	Topical cream	Diabetic foot ulcers	MacroChem Corporation
			Mouth wash	Prevention of chemotherapy-induced mucositis	National Cancer Institute (NCI)
		Phase II/III	Mouth wash	Prevention of ventilator-associated pneumonia	IntraBiotics Pharmaceuticals
Omiganan (MBI 226, CLS001)	Indolicidin (bovine neutrophils)	Phase III	Topical cream	Topical skin antisepsis, prevention of catheter infections	Mallinckrodt
		Phase III	Topical cream	Rosacea	Cutanea Life Sciences, Inc.
		Phase II	Topical cream	Usual type vulvar intraepithelial neoplasia (uVIN)	Cutanea Life Sciences, Inc.
		Phase II	Topical cream	Moderate to severe inflammatory acne vulgaris	Cutanea Life Sciences, Inc.
		Phase II	Topical cream	Mild to moderate atopic dermatitis	Cutanea Life Sciences, Inc.
Lytixar (LTX-109)	Synthetic antimicrobial peptidomimetic	Phase II	Topical cream	Uncomplicated Gram-positive skin infections	Lytix Biopharma AS
		Phase I/IIa	Nasal	Nasal carriers of <i>Staphylococcus aureus</i>	Lytix Biopharma AS
C16G2	Synthetic specifically targeted antimicrobial peptide	Phase II	Mouth wash	Prevent tooth decay caused by <i>Streptococcus mutans</i>	C3 Jian, Inc.

7 AMPs reported

Source: Andersson, D. I., Hughes, D. & Kubicek-Sutherland, J. Z. Mechanisms and consequences of bacterial resistance to antimicrobial peptides. Drug Resist. Updat. 26, 43–57 (2016).

AMPs Under Clinical Investigation (-2019)

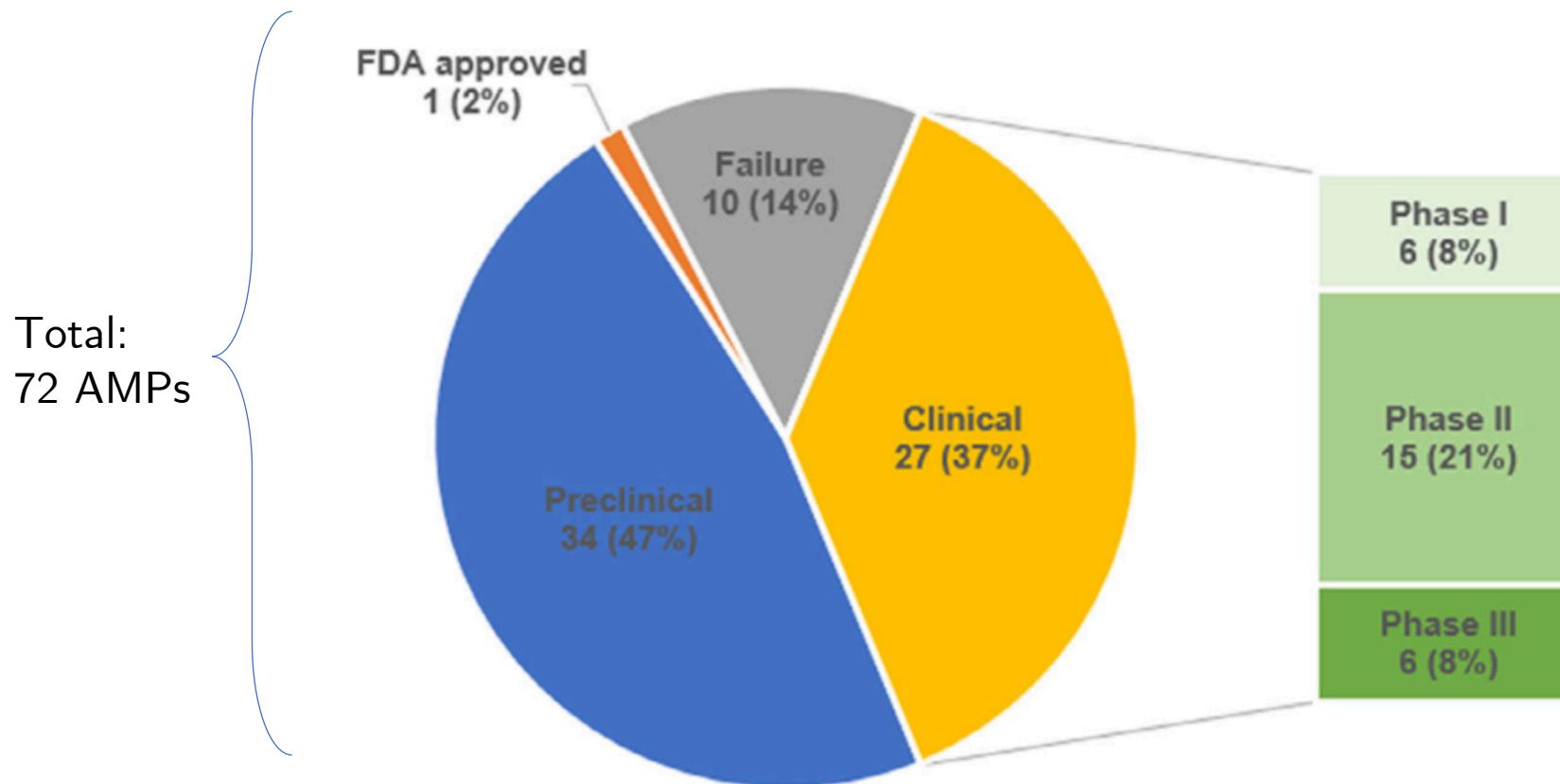
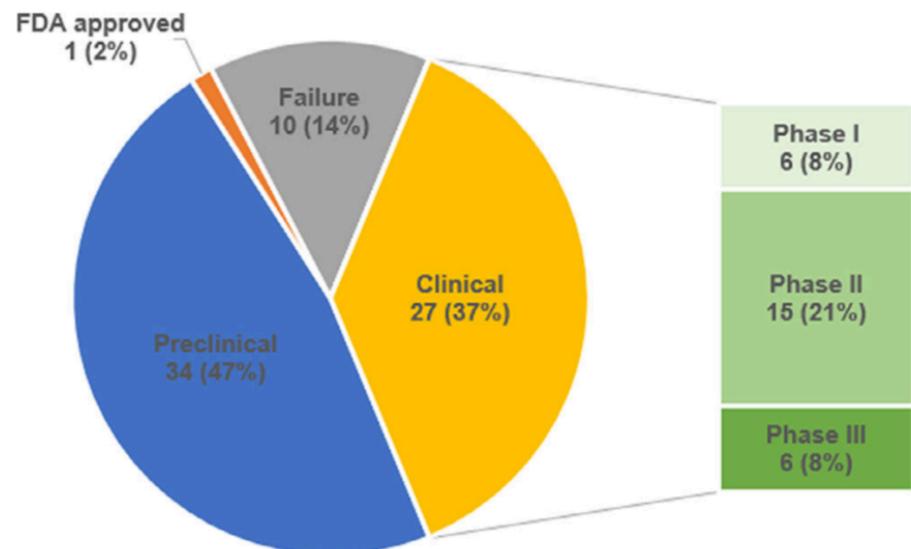


FIGURE 4 Current developmental status of antimicrobial peptides

Source: Koo, H. B. & Seo, J. Antimicrobial peptides under clinical investigation. Peptide Science vol. 111 (2019).

TABLE 3 AMPs in clinical trials (phase I-III)



Entry	AMP	AMP source	Target ^c	Phase	Company	Admin. ^d
1	EA-230	hCG derivative	Sepsis and renal failure protection	II	Exponential Biotherapies	i.v.
2	CZEN-002	α -MSH derivative	Anti-fungal	IIb	Zengen	top.
3	D2A21	Synthetic	Burn wound infections	III	Demegen	top.
4	XMP-629	BPI derivative	Impetigo and acne rosacea	III ^b	Xoma Ltd.	top.
5	Neuprex(rBPI21)	BPR derivative	Pediatric meningococcemia	III ^b	Xoma Ltd.	i.v.
6	Delmitide(RDP58)	HLA class I derivative	Inflammatory bowel disease	II ^a	Genzyme	top.
7	Ghrelin	Endogenous HDP	Chronic respiratory failure	II ^a	University of Miyazaki; Papworth Hospital	i.v.
8	NVB-302	Lantibiotic	<i>C. difficile</i>	I ^b	Novacta	oral
9	hLF1-11	Lactoferricin derivative	MRSA, <i>K. pneumoniae</i> , <i>L. monocytogenes</i>	I/II	AM-Pharma	i.v.
10	Wap-8294A2 (Lotillicin)	Lysobactor spp.	G(+) bacteria(VRE, MRSA)	I/II	aRigen	top.
11	C16G2	Synthetic	Tooth decay by <i>Streptococcus mutans</i>	II	C3 Jian Inc.	Mouth wash
12	SGX942(Dusquertide)	Synthetic	Oral mucositis	III	Soligenix	Oral rinse
13	DPK-060	Kininogen derivative	Acute external otitis	II	ProMore Pharma	Ear drops
14	PXL01	Lactoferrin analog	Postsurgical adhesions	III	ProMore Pharma	top.
15	PAC113	Histatin 5 analog	Oral candidiasis	II ^a	Pacgen Biopharmaceuticals	Mouth rinse
16	POL7080	Protegrin analog	<i>P. aeruginosa</i> <i>K. pneumoniae</i>	III	Polyphor Ltd.	i.v.
17	LTX-109 (Lytxiar)	Synthetic	G(+) MRSA skin infections; impetigo	IIa ^a	Lytx Biopharma	Nasal top.
18	OP-145	LL-37 derivative	Chronic middle ear infection	II ^a	Dr. Reddy's Research	Ear drops
19	LL-37	Human cathelicidin	Leg ulcer	IIb	ProMore Pharma	top.
20	Novexatin (NP213)	Cyclic cationic peptide	Fungal nail infection	IIb	Novabiotics	top.
21	p2TA (AB103)	Synthetic	Necrotizing soft tissue infections	III	Atox Bio Ltd.	i.v.
22	Iseganan (IB-367)	Protegrin analog	Pneumonia, stomatitis	III ^b	IntraBiotics Pharmaceuticals	top.
23	Pexiganan (MSI-78)	Magainin analog	Diabetic foot ulcers	III ^b	Dipexium Pharmaceuticals	top.
24	Omiganan (CLS001)	Indolicidin derivative	Rosacea	III	Cutanea Life Sciences	top.
25	Surotomycin	Cyclic lipopeptide	<i>C. difficile</i> (diarrhea)	III ^b	Cubist Pharmaceuticals/Merck	oral
26	Ramoplanin (NTI-851)	Actinoplanes spp	G(+) (VRE, <i>C. difficile</i>)	III	Nano-therapeutics	oral
27	Friulimicin B	Cyclic lipopeptide	Pneumonia, MRSA	I ^b	MerLion Pharmaceuticals	i.v.

Source: Koo, H. B. & Seo, J. Antimicrobial peptides under clinical investigation. Peptide Science vol. 111 (2019).

Motivation

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

PROBLEM

Antibiotic
Resistance



SOLUTION

Antimicrobial
Peptides?

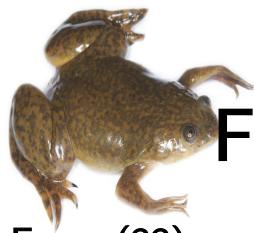


Objective

- To develop **and** execute an AMP discovery pipeline to mine for AMP precursor sequences in publicly available genomic resources



Datasets



Frogs & Toads (38)



Frogs (33)

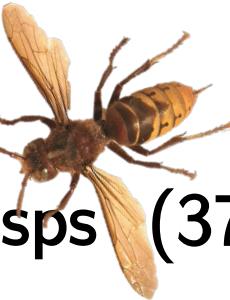
<i>A. femoralis</i>	<i>P. adspersus</i>	<i>R. sirensis</i>
<i>A. mantzorum</i>	<i>P. amboli</i>	<i>R. sylvatica</i>
<i>A. petersi</i>	<i>P. megacephalus</i>	<i>R. temporaria</i>
<i>C. alboguttata</i>	<i>P. microps</i>	<i>S. ruber</i>
<i>D. auratus</i>	<i>P. nigromaculatus</i>	<i>X. allofraseri</i>
<i>D. 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)

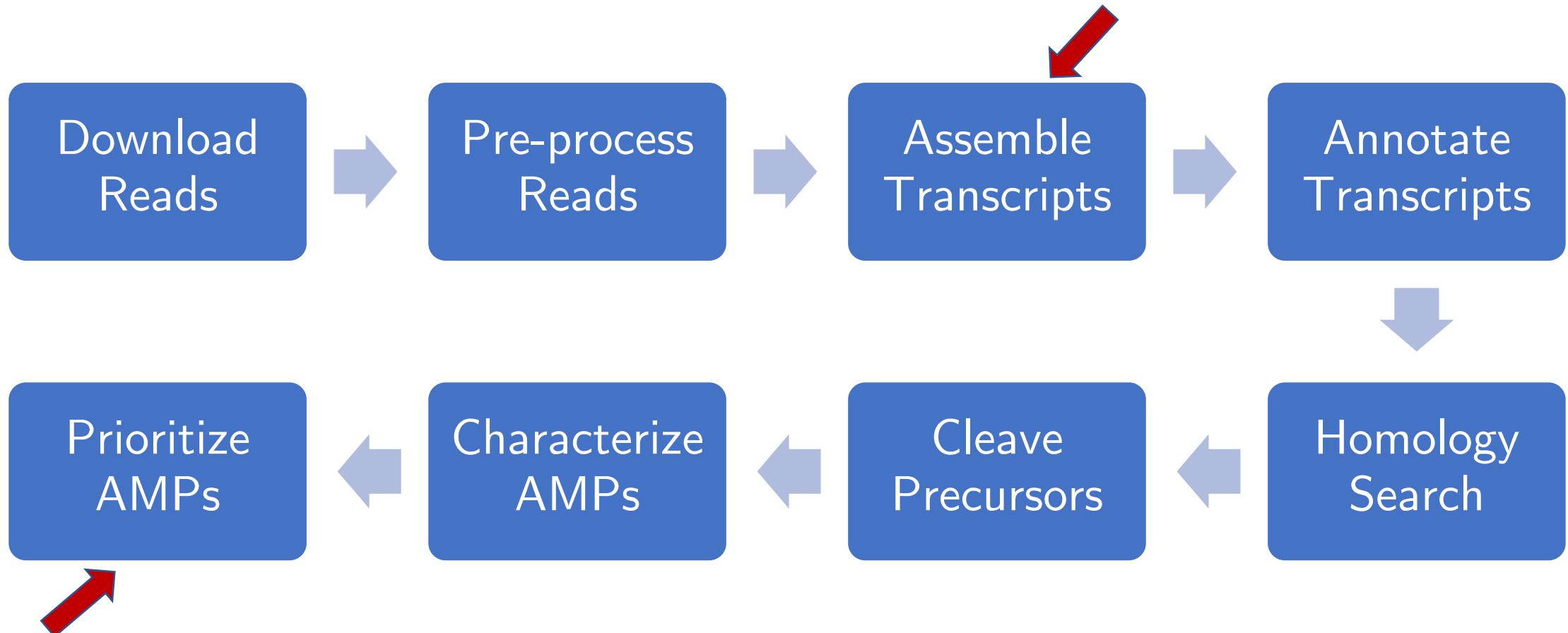
<i>A. compressa</i>	<i>P. snelleni</i>
<i>A. flavomarginatum</i>	<i>P. turionellae</i>
<i>B. nigricans</i>	<i>P. varia</i>
<i>C. vestalis</i>	<i>P. vindemmiae</i>
<i>D. collaris</i>	<i>S. deformae</i>
<i>D. longicaudata</i>	<i>S. kj8906</i>
<i>M. demolitor</i>	<i>T. sarcophagae</i>
<i>N. giraulti</i>	<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>

Why Amphibians and Insects?



Source: Darcy Sutherland, PeptAID AIM 1 Science Meeting, dkfindout.com

Methods Overview



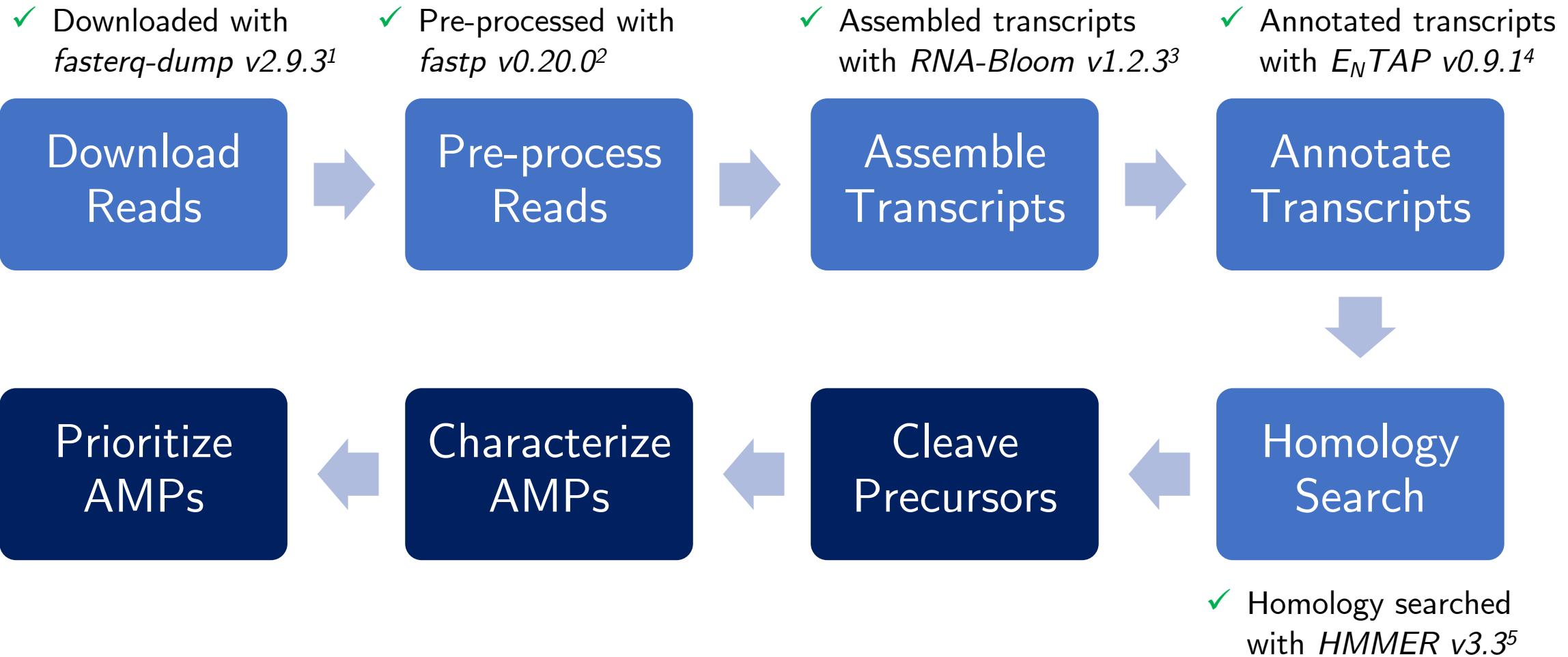
AMPlify

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

RNA-Bloom

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

Completed Steps





Pre-processing: fastp

Chen, S., Zhou, Y., Chen, Y. & Gu, J. **fastp: an ultra-fast all-in-one FASTQ preprocessor**. Bioinformatics 34, i884–i890 (2018).

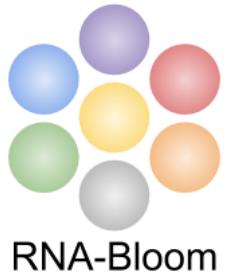
- Quality control - FASTQC
- Adapter trimming - Cutadapt
- Quality filtering - Trimmomatic



fastp **replaces** the need
to use *three separate tools*

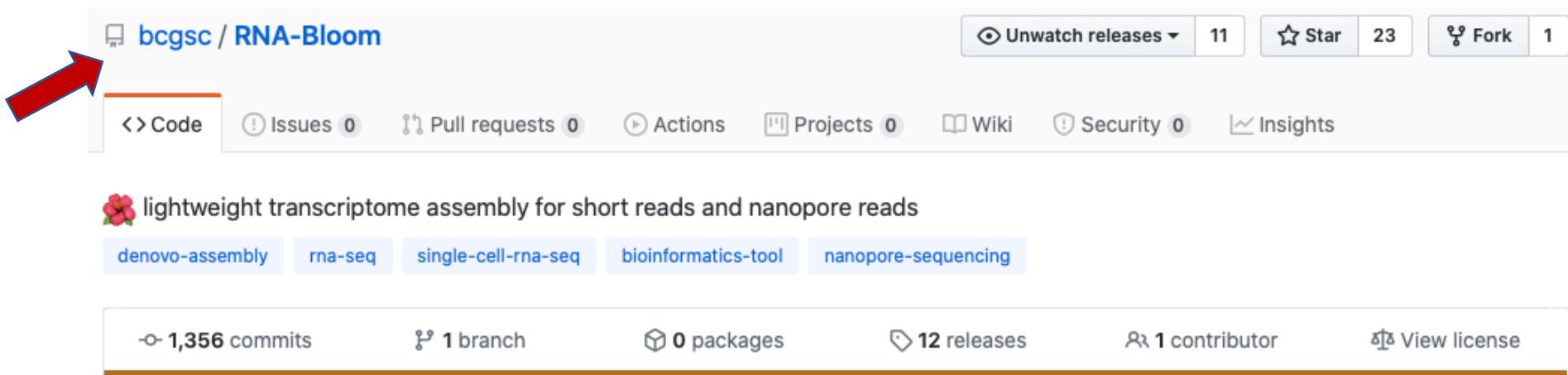


Assembly: RNA-Bloom



Nip, K. M. et al. RNA-Bloom provides lightweight reference-free transcriptome assembly for single cells. Genome Research (under revision).

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



bcgsc / RNA-Bloom

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights

lightweight transcriptome assembly for short reads and nanopore reads

denovo-assembly rna-seq single-cell-rna-seq bioinformatics-tool nanopore-sequencing

1,356 commits 1 branch 0 packages 12 releases 1 contributor View license

Assembly: RNA-Bloom

Stage 1

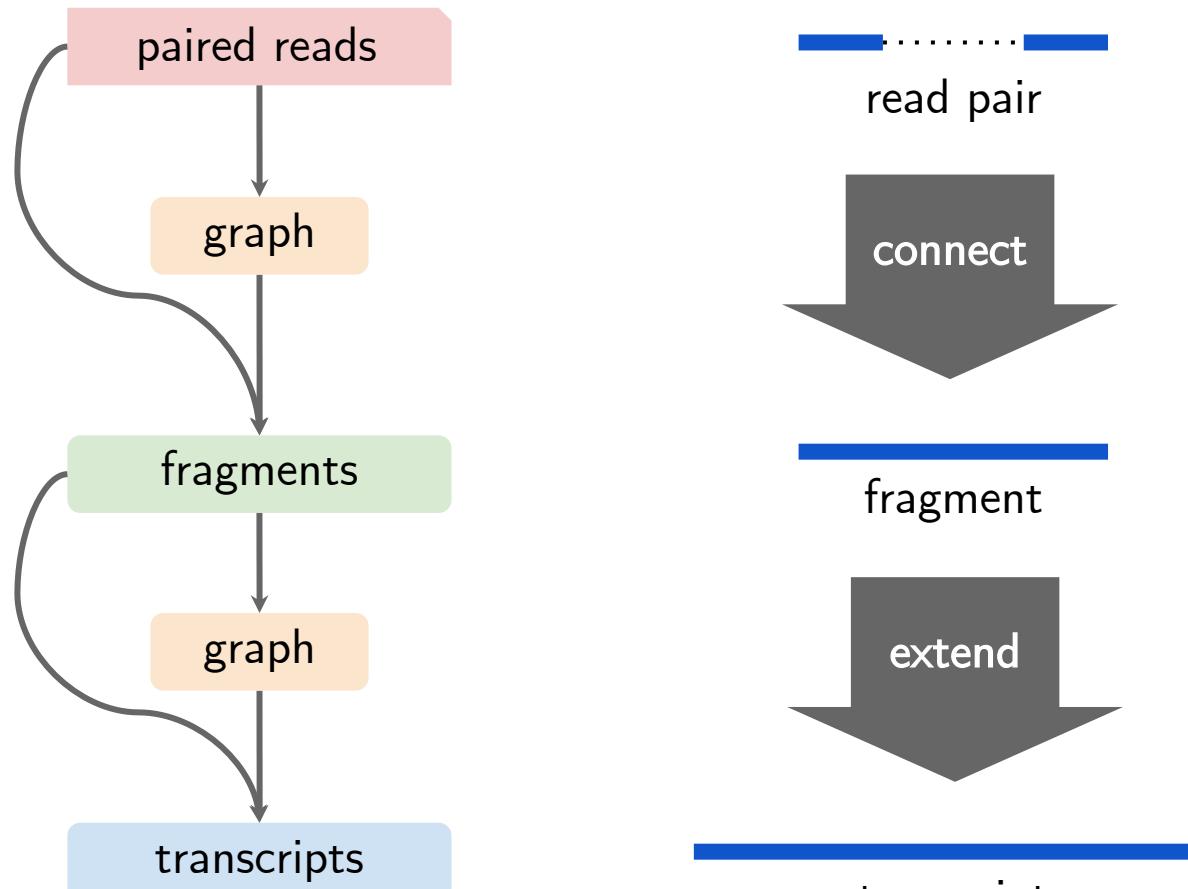
Construction of
Bloom filter de Bruijn graph

Stage 2

Reconstruction of
read fragments

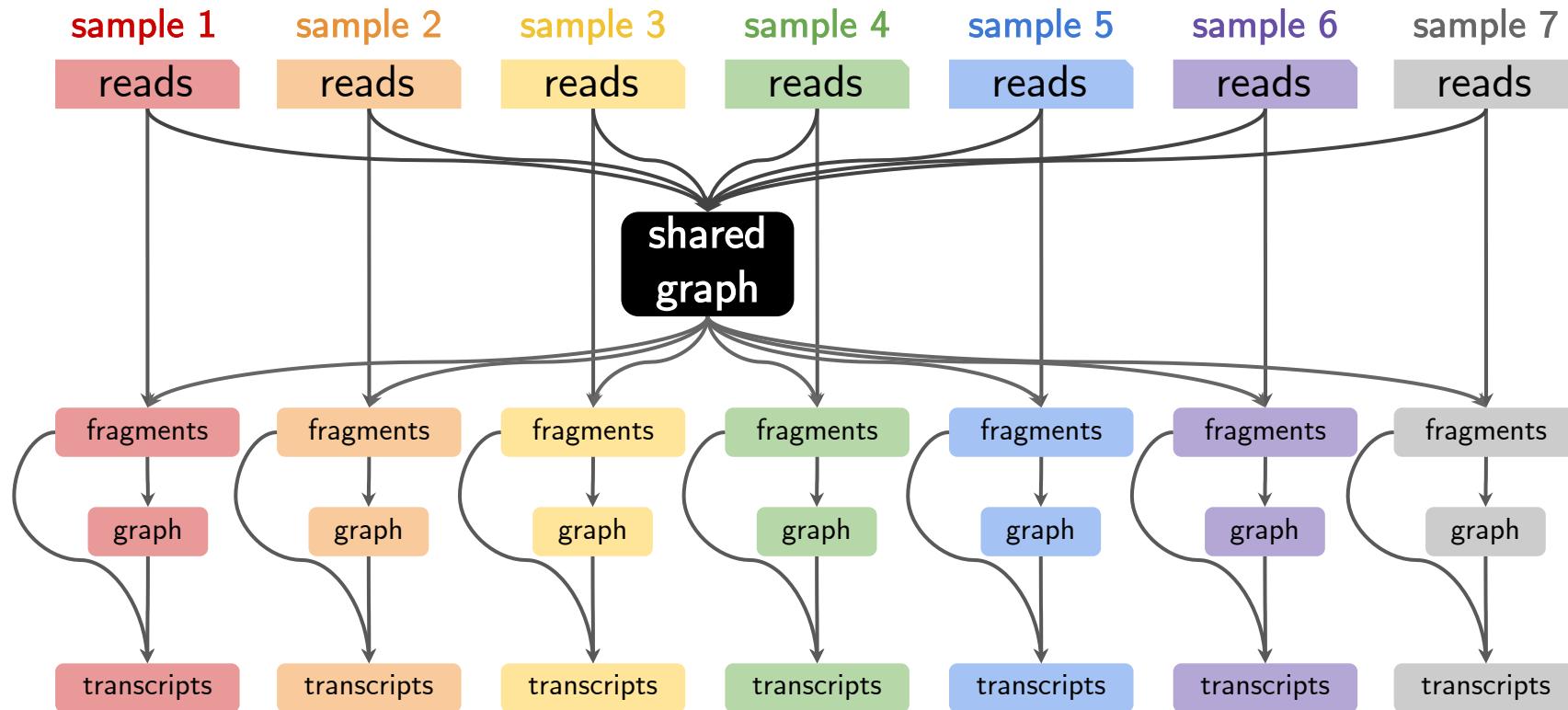
Stage 3

Reconstruction of
transcript sequence



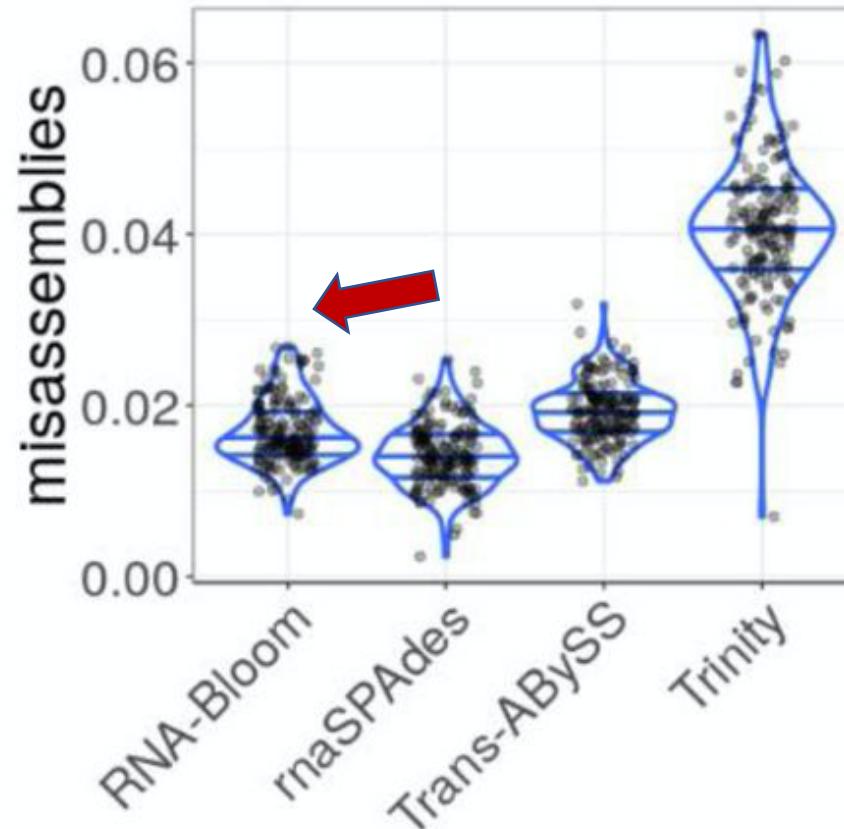
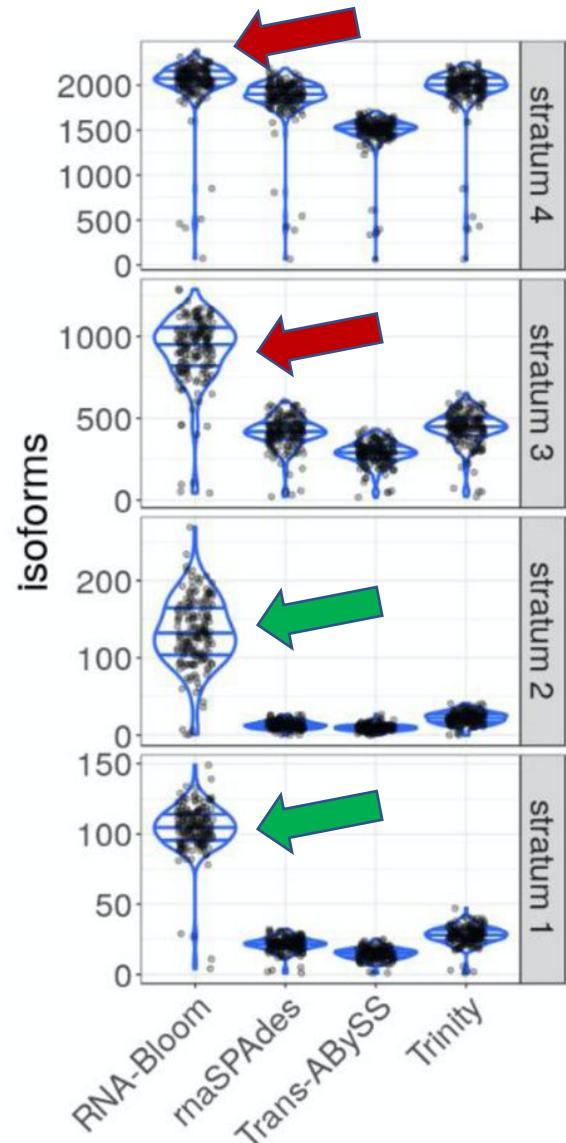
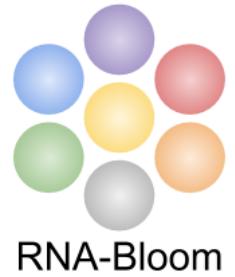
Source: Nip, K. M. et al. RNA-Bloom provides lightweight reference-free transcriptome assembly for single cells. *Genome Research* (under revision).

RNA-Bloom



Source: Nip, K. M. et al. RNA-Bloom provides lightweight reference-free transcriptome assembly for single cells. *Genome Research* (under revision).

Assembly: RNA-Bloom



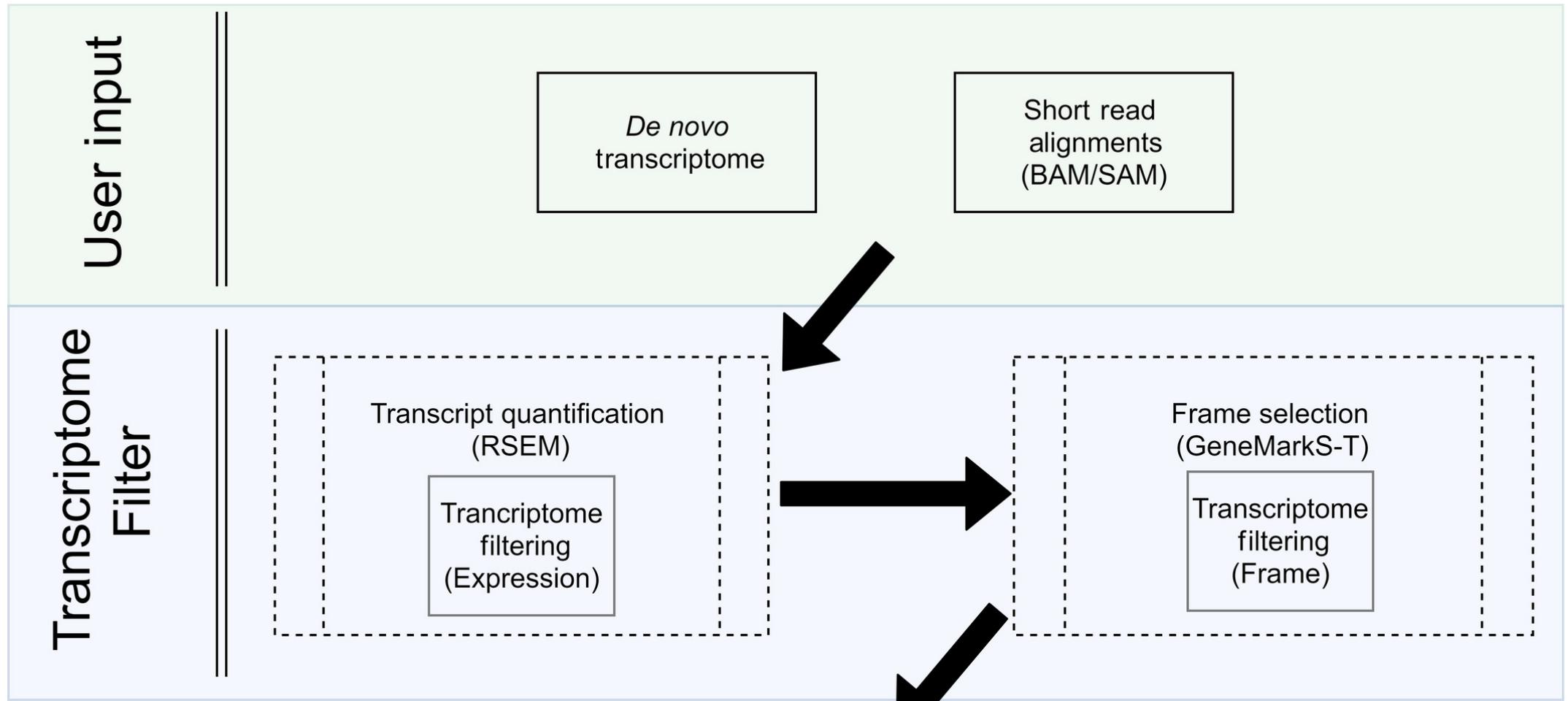
Source: Nip, K. M. et al. RNA-Bloom provides lightweight reference-free transcriptome assembly for single cells. *Genome Research* (under revision).

Annotation: E_NTAP

Hart, A. J. et al. E_NTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20, 591–604 (2020).

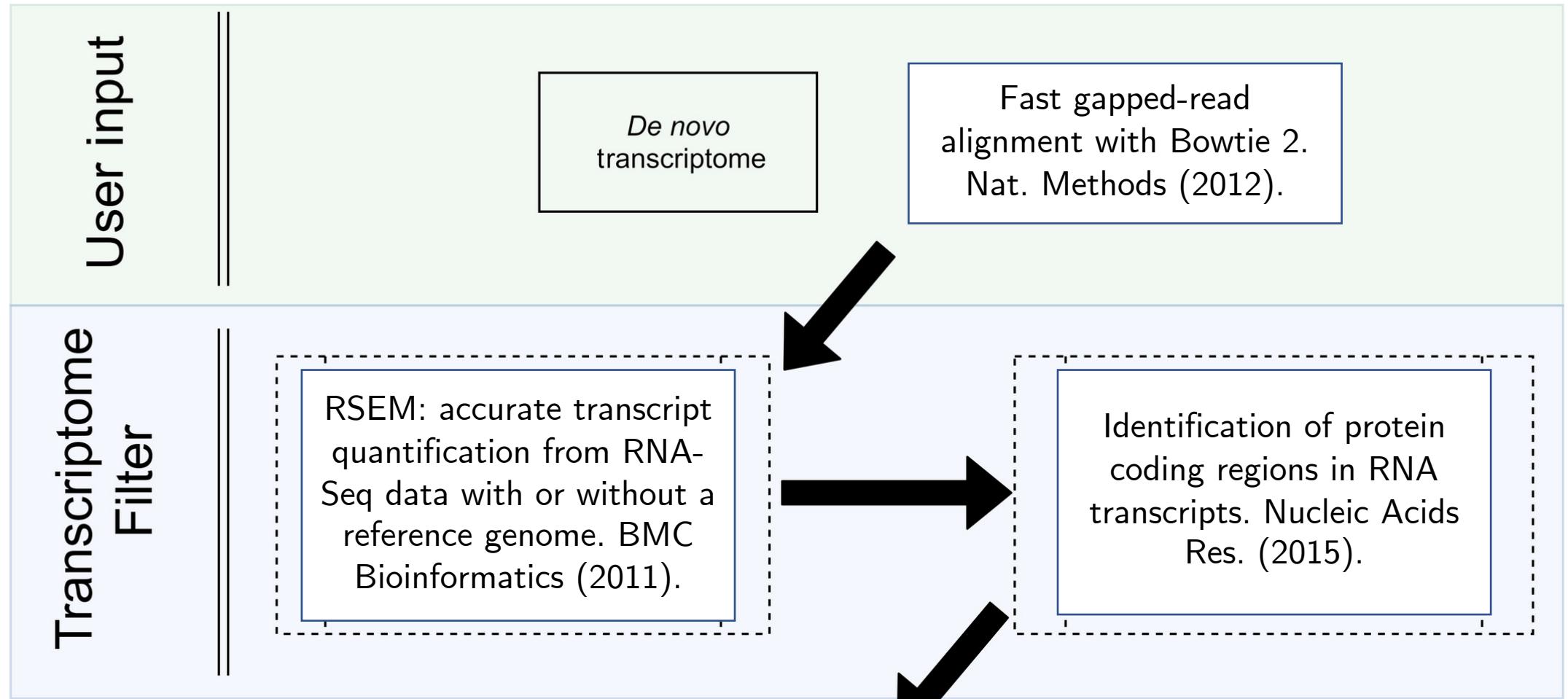
- E_NTAP: Eukaryotic Non-Model Transcriptome Annotation Pipeline
- Functional annotation with ‘frame selection’
- Alignment to orthology, domain, and protein databases

Annotation: E_NTAP



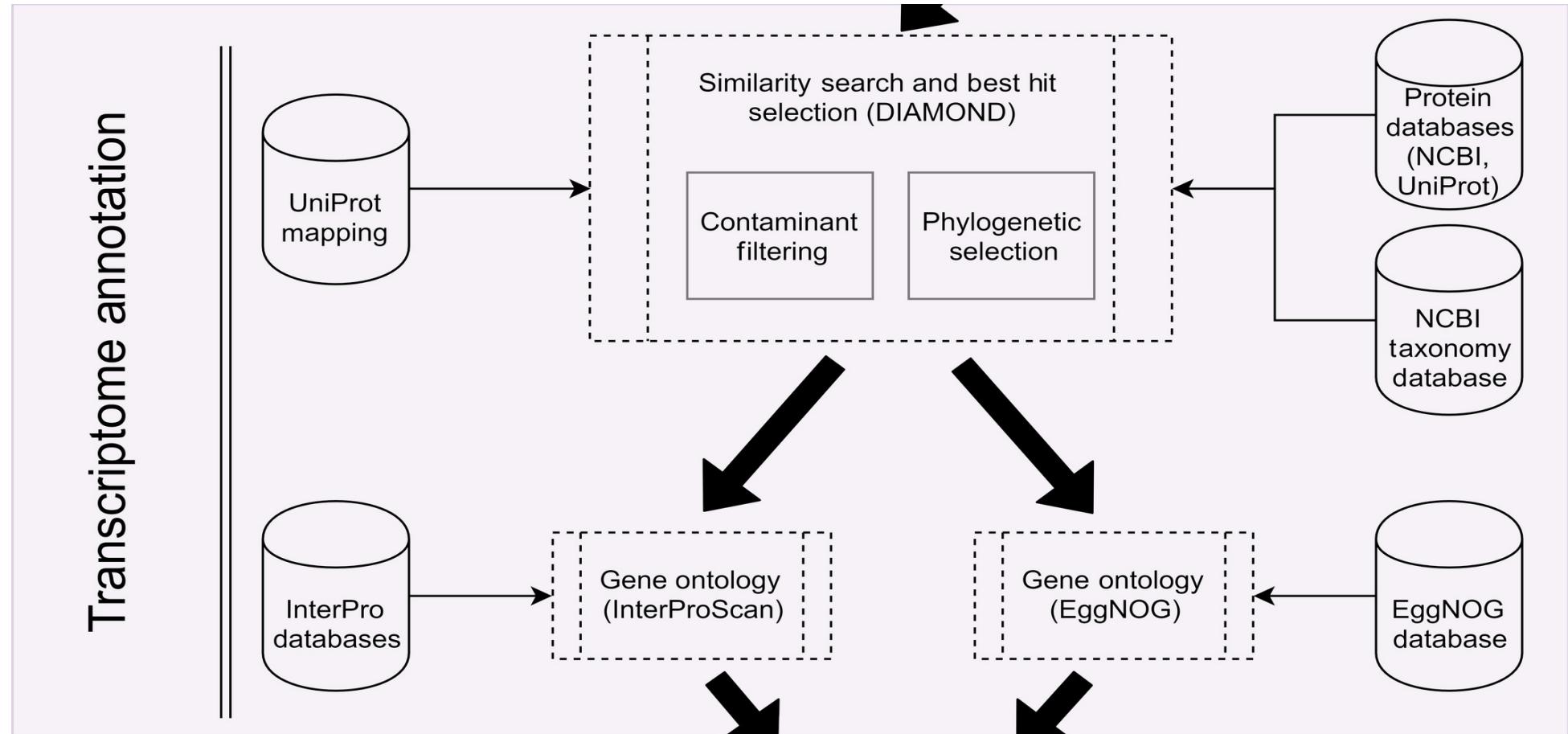
Source: Hart, A. J. et al. E_NTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20, 591–604 (2020).

Annotation: E_NTAP



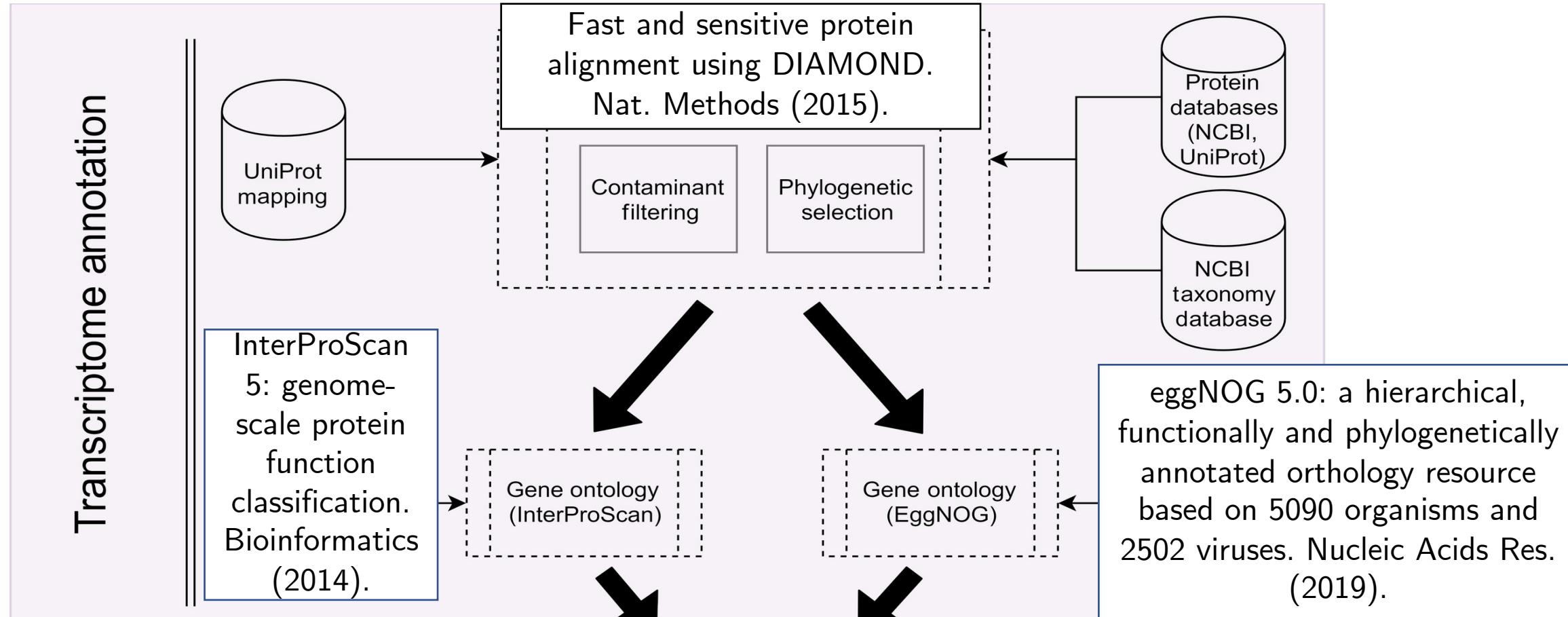
Source: Hart, A. J. et al. E_NTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20, 591–604 (2020).

Annotation: E_NTAP



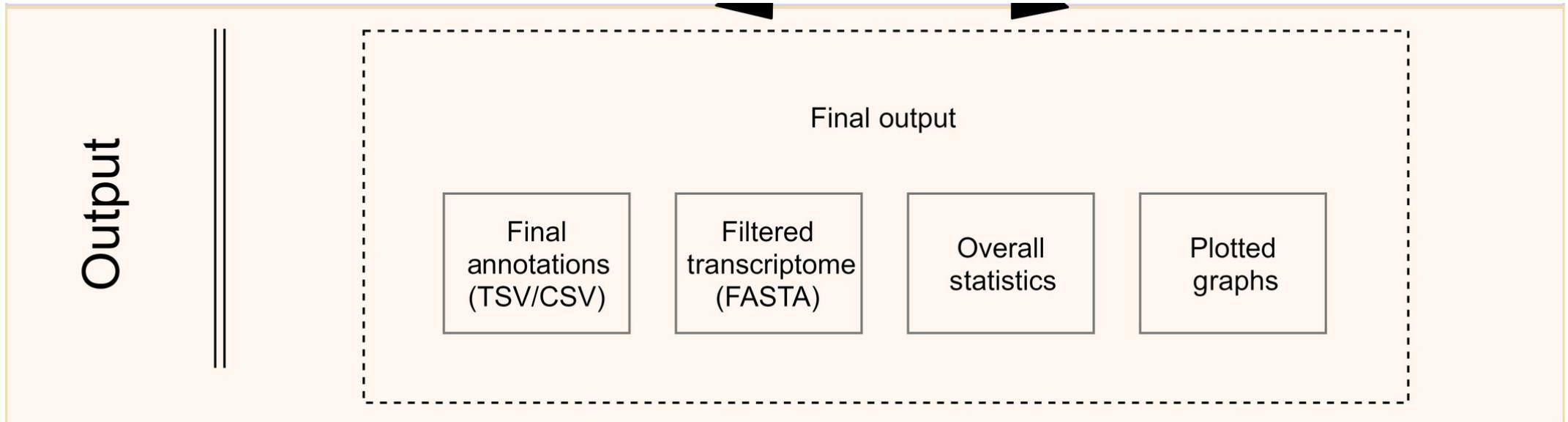
Source: Hart, A. J. et al. E_NTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20, 591–604 (2020).

Annotation: E_NTAP



Source: Hart, A. J. et al. E_NTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20, 591–604 (2020).

Annotation: E_NTAP



Source: Hart, A. J. et al. E_NTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20, 591–604 (2020).

Annotation: E_NTAP

Metric	BLAST2GO PRO	BLAST2GO Basic	TRINOTATE	EN-TAP	ANNOSCRIPT	DAMMIT
Open source/free software		‡	‡	‡	‡	‡
Command line integration	‡		‡	‡	‡	‡
Filtering assembly via short-read alignment (expression)	†				‡	
Frame selection	†		‡	‡	‡	‡
Custom database selection and indexing	‡		†	‡		†
Fast and sensitive NCBI BLAST alternative				‡	†	
Selection of optimal hit From several databases				‡	†	
Selection of optimal hit based on informativeness	†				‡	
Contaminant identification and filtering	‡				‡	
Orthologous gene family assignment	‡		†	‡		
Protein domain (CDD/InterProScan)	‡	‡	‡	‡	‡	‡
Gene ontology term and pathway assignment Sourced from protein alignments	‡	‡	‡		‡	
Gene ontology term and pathway assignment Sourced from orthologous genes	‡			‡		‡
Provides graphical user interface for annotation process	‡	‡				

Source: Hart, A. J. et al. E_NTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20, 591–604 (2020).



Homology Search: HMMER

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

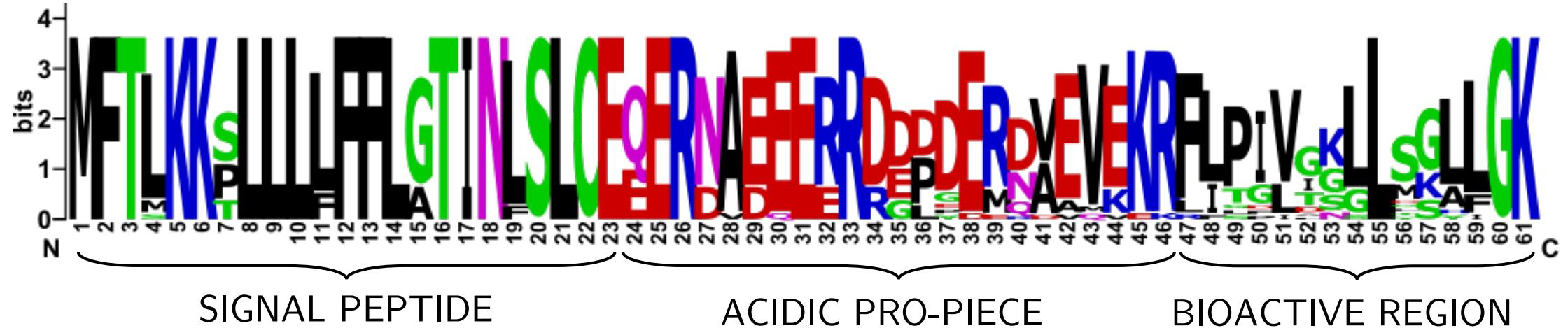
- Using antimicrobial precursor sequences from NCBI Protein database (including GenPept, UniProt, etc.)
- 2,792 amphibian precursors; 303 insect precursors



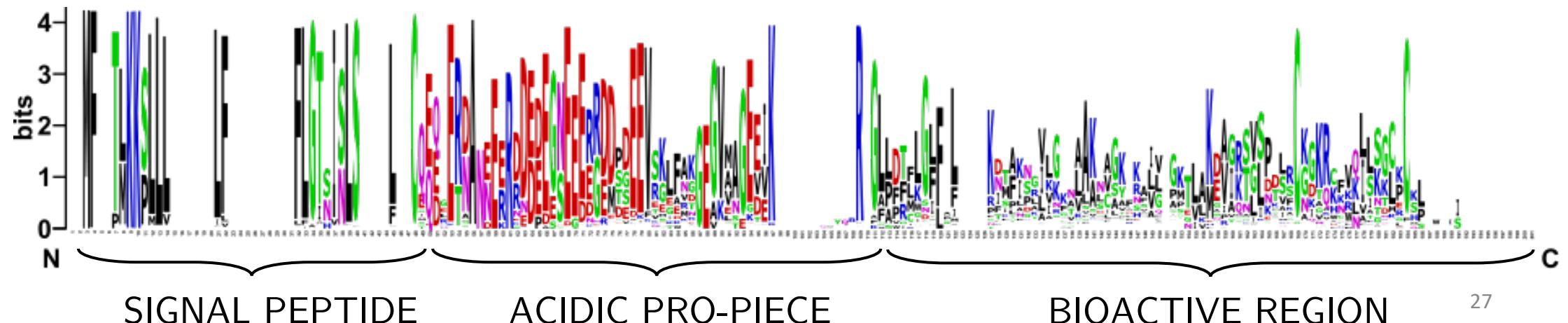


Homology Search: HMMER

Single AMP family:

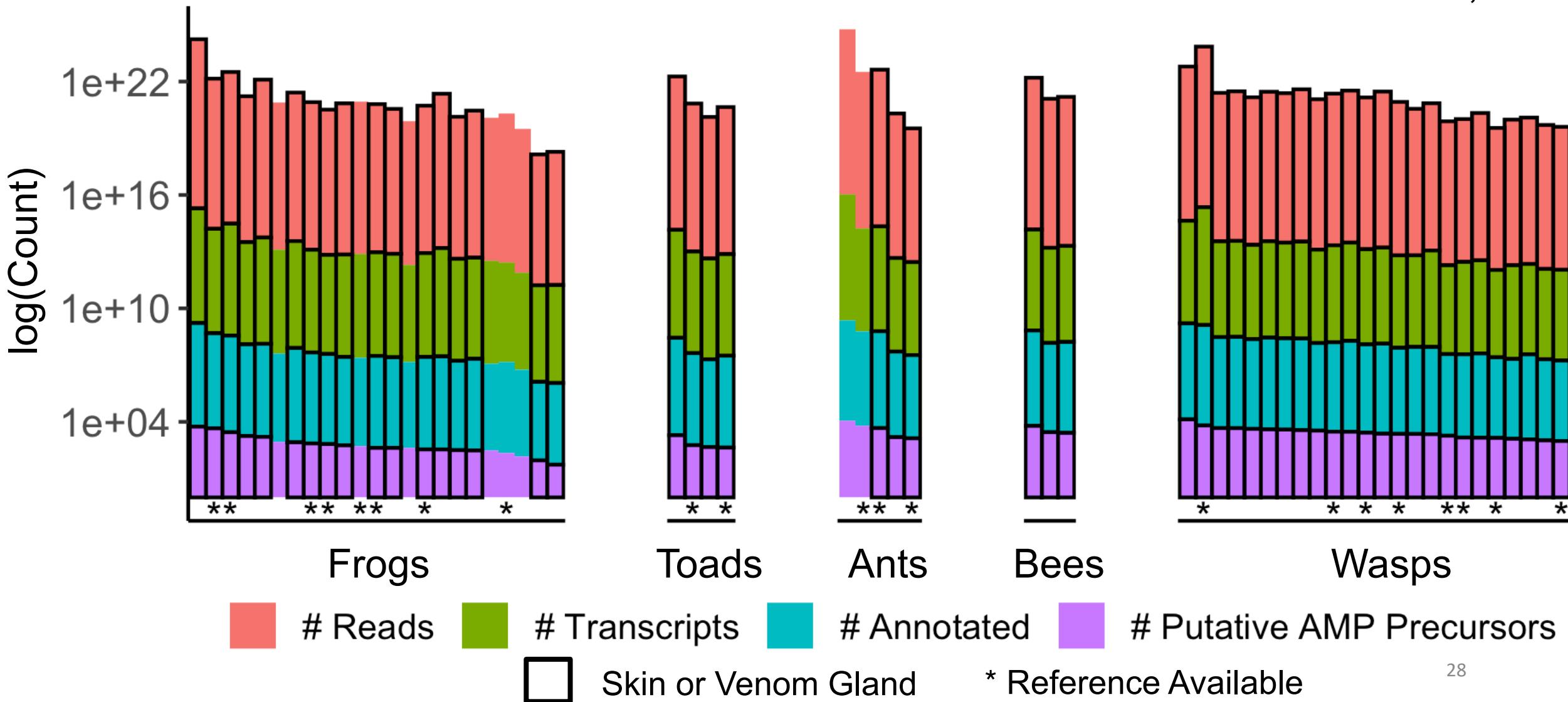


General AMP structure:

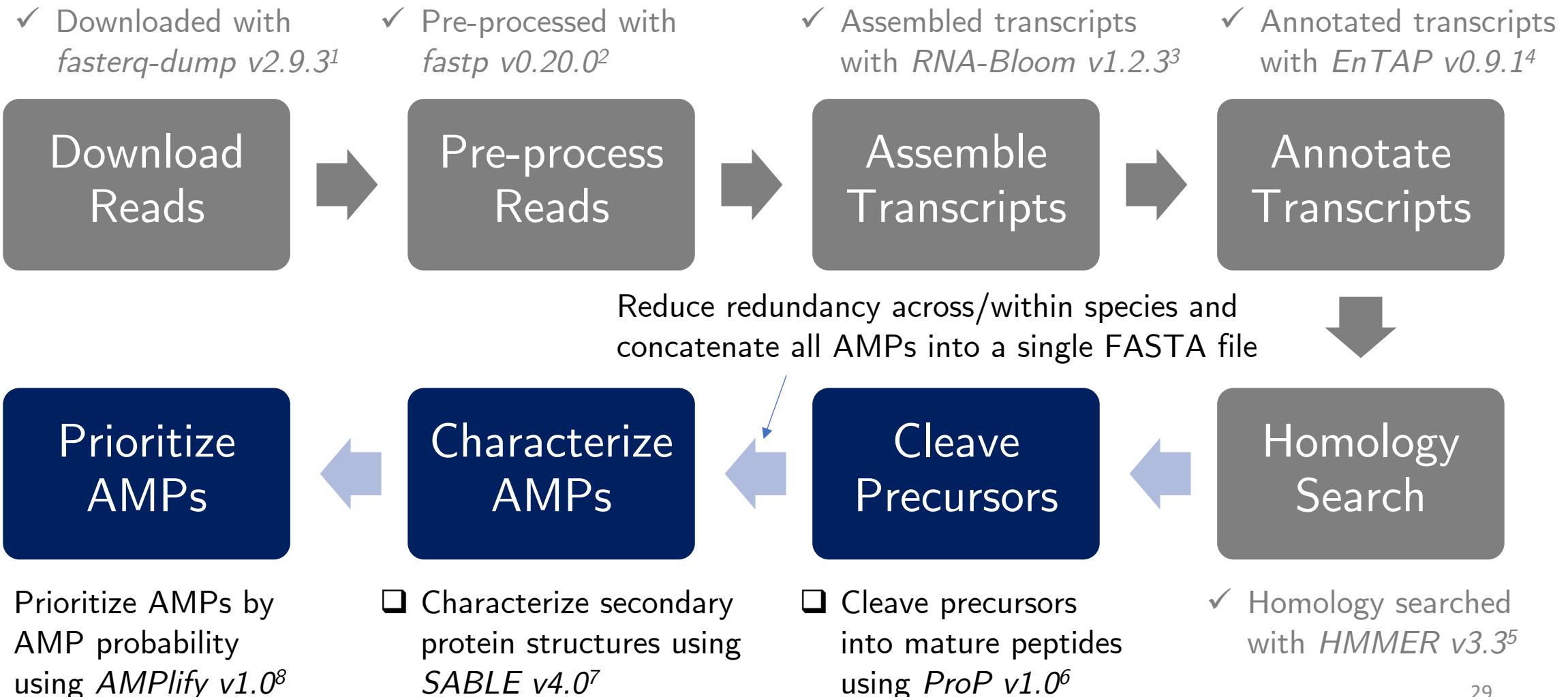


Preliminary Results

Total # Putative AMP Precursors: 140,054

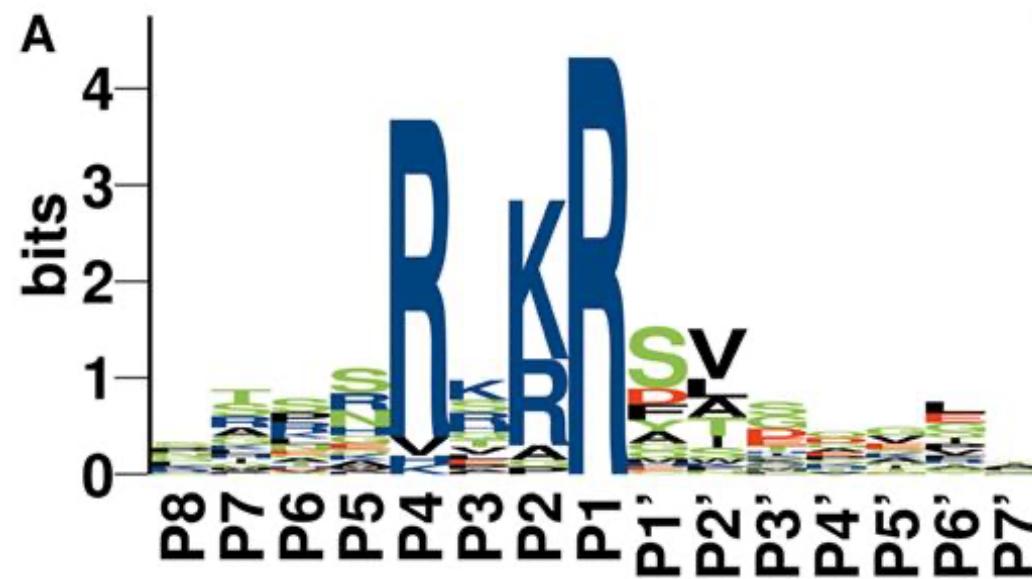


Next Steps



Precursor Cleavage: ProP

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



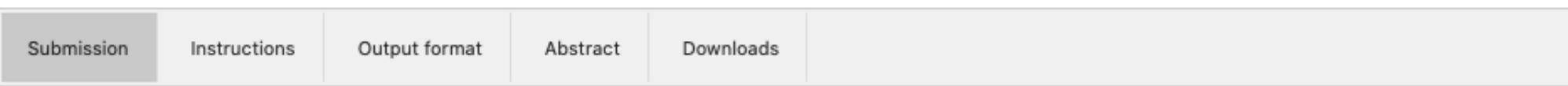
Precursor Cleavage: ProP

ProP - 1.0

Arginine and lysine propeptide cleavage sites in eukaryotic protein sequences

The ProP 1.0 server predicts arginine and lysine propeptide cleavage sites in eukaryotic protein sequences using an ensemble of neural networks. Furin-specific prediction is the default. It is also possible to perform a general proprotein convertase (PC) prediction.

For convenience, this server is integrated with the [SignalP-3.0](#) server predicting the presence and location of signal peptide cleavage sites.



>sp|P10891.2|DEFI_PROTE RecName: Full=Phormicin; AltName: Full=Insect defensin-A/B; Flags: Precursor

MKFFMFVVTFCLAVCFVSQSLAIPADAANDAHFVDGVQALKEIE
PELHGRYKRATCDLLSGTGINHSACAAHCLLRGNRGGYCNGKGVC
VCRN

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

Precursor Cleavage: ProP

Signal peptide cleavage site predicted: between pos. 23 and 24: SLA-IP

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

>sp|P10891.2|DEFI PROTE RecName: Full=Phormicin; AltName: Full=Insect defensin A/B; Flags: Precursor

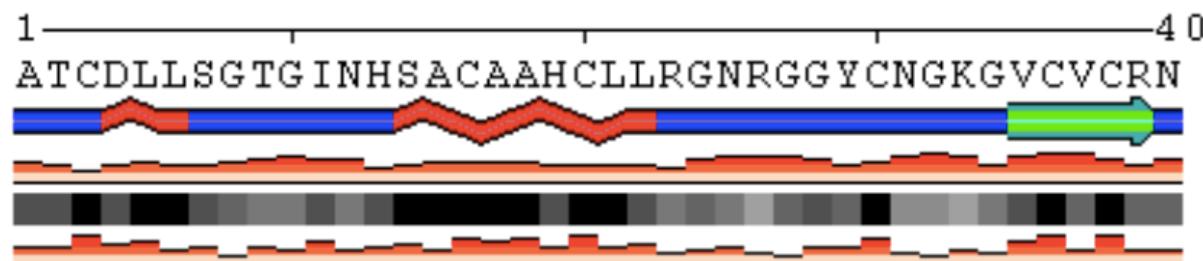
MKFFMVFVVTFCLAVCFVSQSLAIPADAANDAHFVDGVQALKEIEPELHGRYKR
ATCDLLSGTGINHSACAAHCLLRGNRGGYCNGKGVCVRN

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

Characterization: SABLE

Adamczak, R., Porollo, A. & Meller, J. Combining prediction of secondary structure and solvent accessibility in proteins. Proteins 59, 467–475 (2005).

>sp|P10891.2|DEFI_PROTE RecName: Full=Phormicin;
AltName: Full=Insect defensin A/B; Note: Full=Cleaved
ATCDLLSGTGINHSACAAHCLLRGNRGGYCNGKGVCVCRN

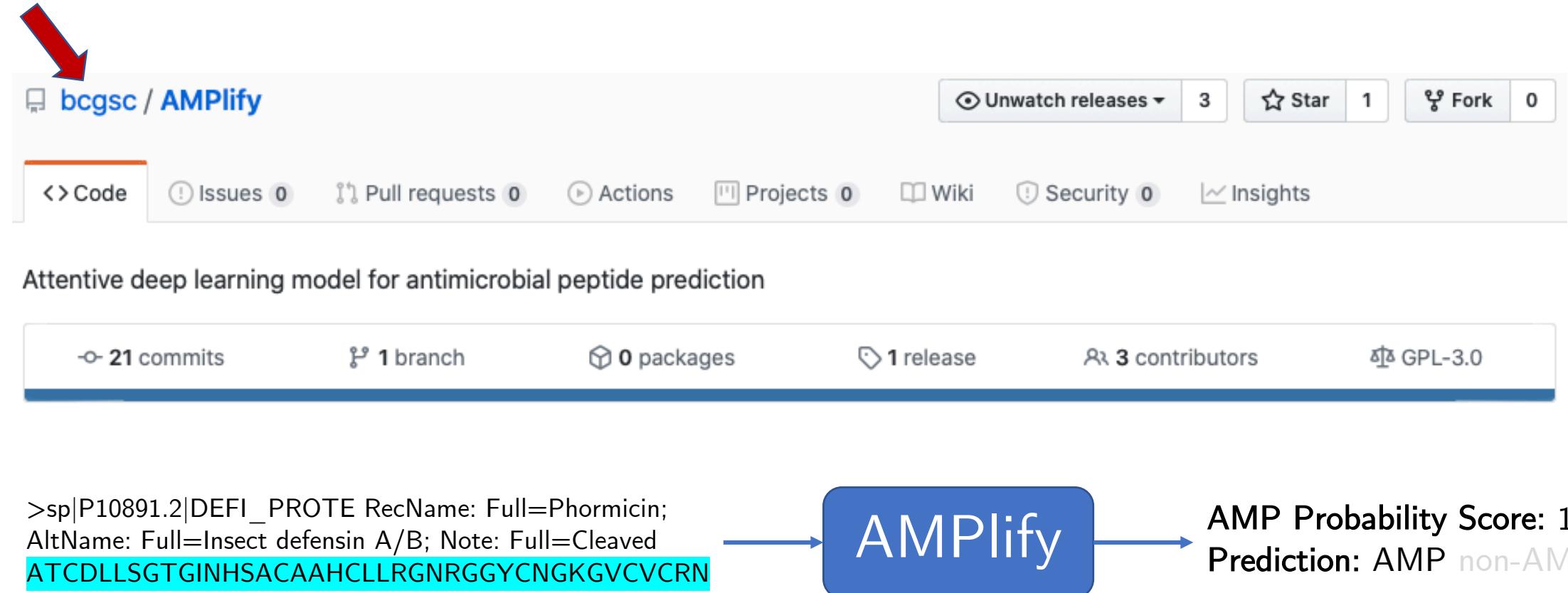


>sp|P10891.2|DEFI_PROTE RecName: Full=Phormicin;
AltName: Full=Insect defensin A/B; Note: Full=Cleaved

ATCDLLSGTGINHSACAAHCLLRGNRGGYCNGKGVCVCRN
CCC**H**HHCCCCCCC**H**HHHHHHHHCCCCCCCCCCCC**E**EEEEEC
110100123313100001001323521204453102022

Legend	Description
1 —————	Amino acid residue numeration
	Protein secondary structure H-alpha and other helices
	E-beta-strand or bridge
0 1 2 3 4 5 6 7 8 9	Relative solvent accessibility (RSA) 0-completely buried (0-9% RSA), 9-fully exposed (90-100% RSA)
0 1 2 3 4 5 6 7 8 9	Confidence level of prediction 0-the lowest level, 9-the highest level

Prioritization: AMPlify

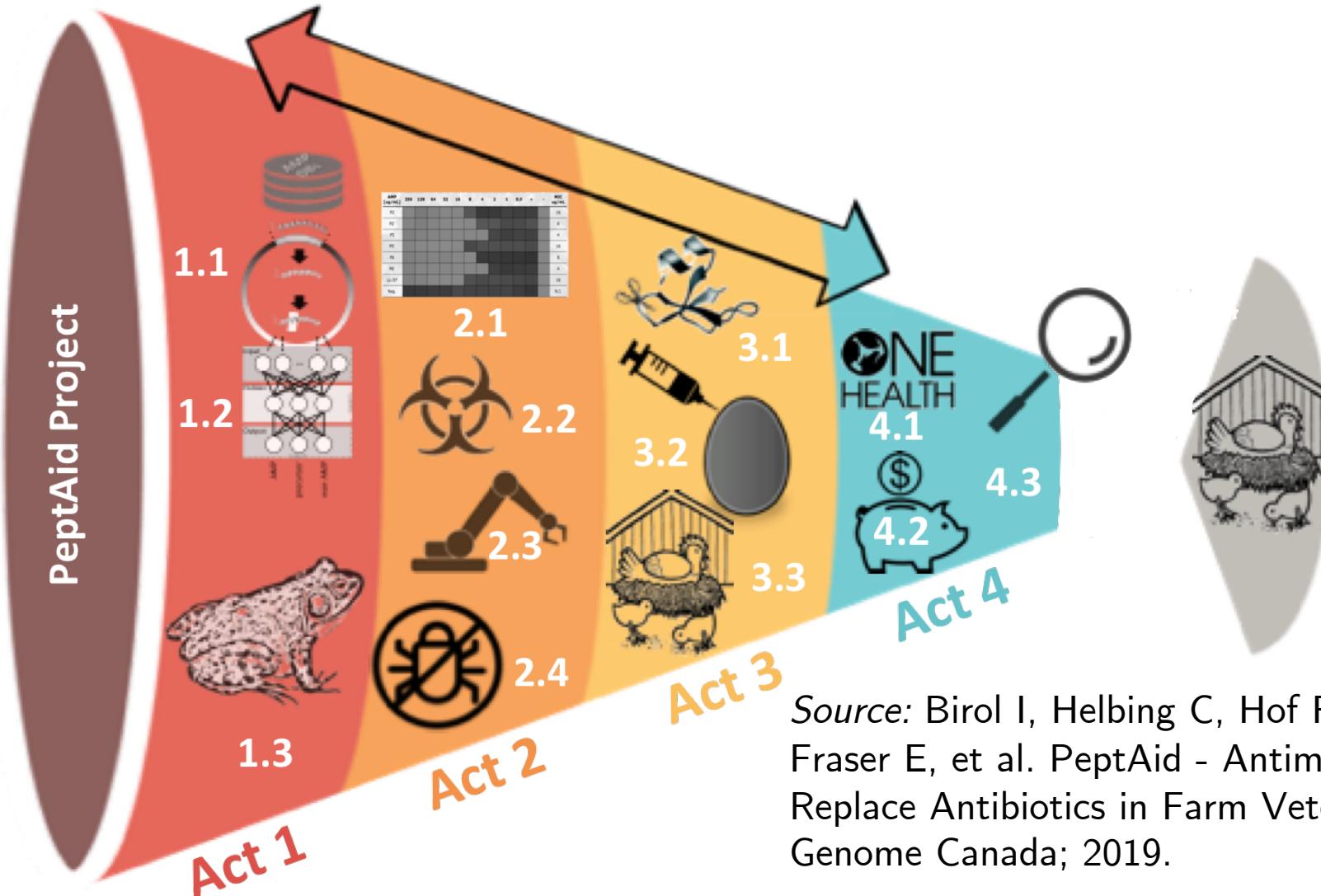


Source: Li, C. et al. AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens. bioRxiv 155706 (2020).

Putting It All Together...

- ✓ AMP Discovery Pipeline software package that runs all the tools from beginning to end
- ✓ Resulting from the pipeline, candidate AMPs for downstream analysis and *in vitro* bioactivity testing against various microbes

PeptAID: Antimicrobial Peptides to Replace Antibiotics in Farm Veterinary Practice



Source: Birol I, Helbing C, Hof F, Coope R, Hoang L, Fraser E, et al. PeptAID - Antimicrobial Peptides To Replace Antibiotics in Farm Veterinary Practice. Genome Canada; 2019.

Acknowledgements

Birol Lab



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

PeptAID Collaborators

- Caren Helbing

PeptAID Funding



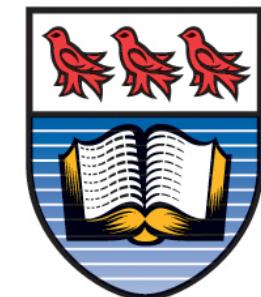
Genome
BritishColumbia



GenomeCanada



**Investment
Agriculture
Foundation
of British Columbia**



**University
of Victoria**

References

1. Leinonen, R., Sugawara, H., Shumway, M. & International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* **39**, D19–21 (2011).
2. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
3. Nip, K. M. et al. RNA-Bloom provides lightweight reference-free transcriptome assembly for single cells. *Genome Research* (under revision).
4. Hart, A. J. et al. E_NTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Mol. Ecol. Resour.* **20**, 591–604 (2020).
5. Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, 431 (2010).
6. Duckert, P., Brunak, S. & Blom, N. Prediction of proprotein convertase cleavage sites. *Protein Eng. Des. Sel.* **17**, 107–112 (2004).
7. Adamczak, R., Porollo, A. & Meller, J. Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins* **59**, 467–475 (2005).
8. Li, C. et al. AMPlify: Attentive Deep Learning Model for Discovery of Novel Antimicrobial Peptides Effective against WHO Priority Pathogens. *bioRxiv* 155705 (2020).