



HARVARD
UNIVERSITY

Benasque 2026

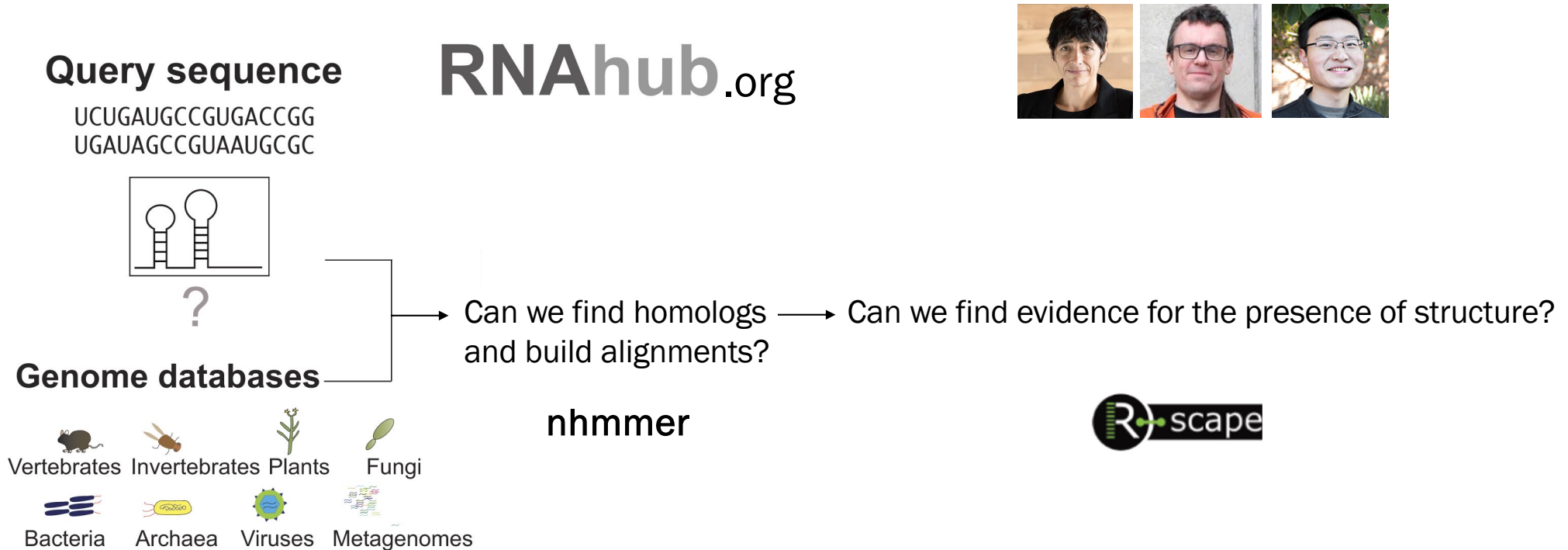
tbi



Computational identification of conserved structural RNAs in Nematoda

Agata M. Kilar, PhD

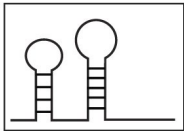
RNAhub – pipeline for structural RNAs discovery



RNAhub – pipeline for structural RNAs discovery

Query sequence

UCUGAUGCCGUGACCGG
UGAUAGCCGUAAUGCGC



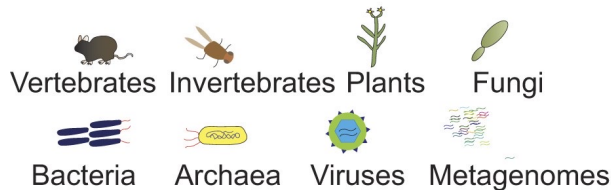
?

RNAhub.org

UPDATE is coming!



Genome databases



Can we find homologs and build alignments? → Can we find evidence for the presence of structure?

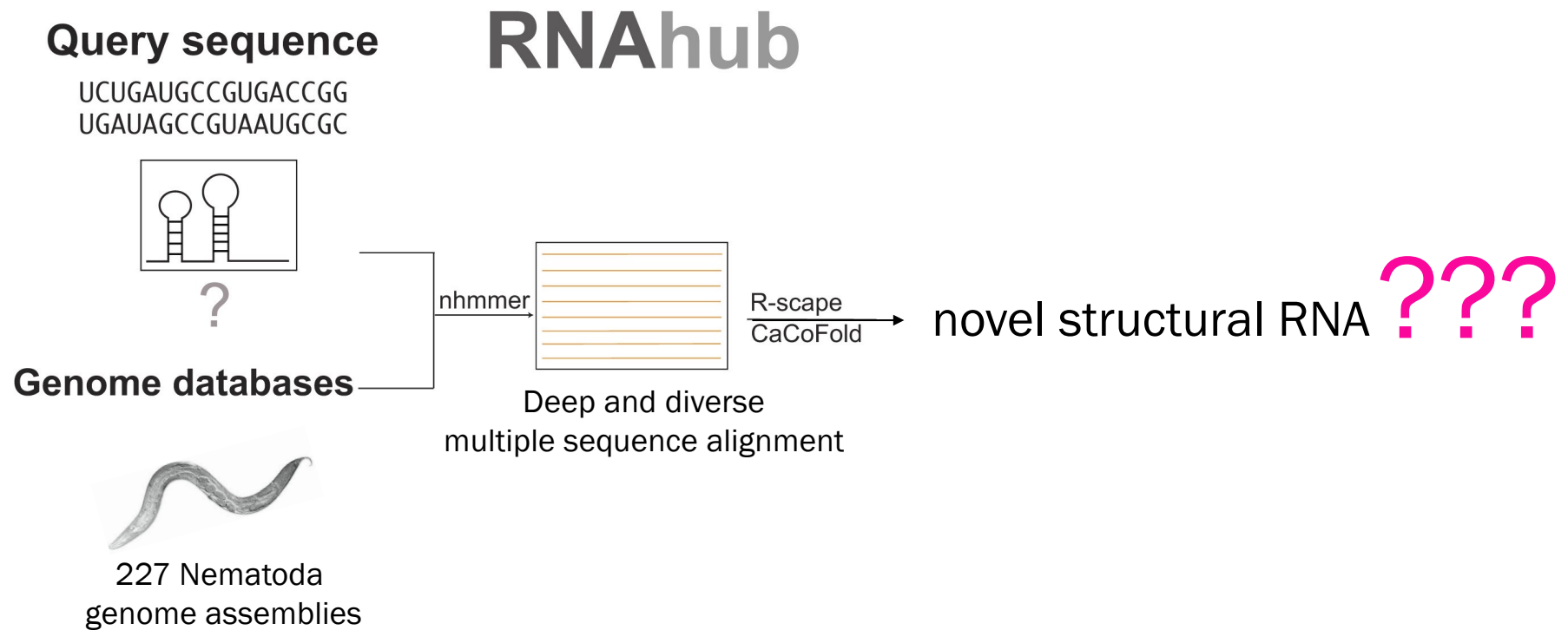
nhmmer



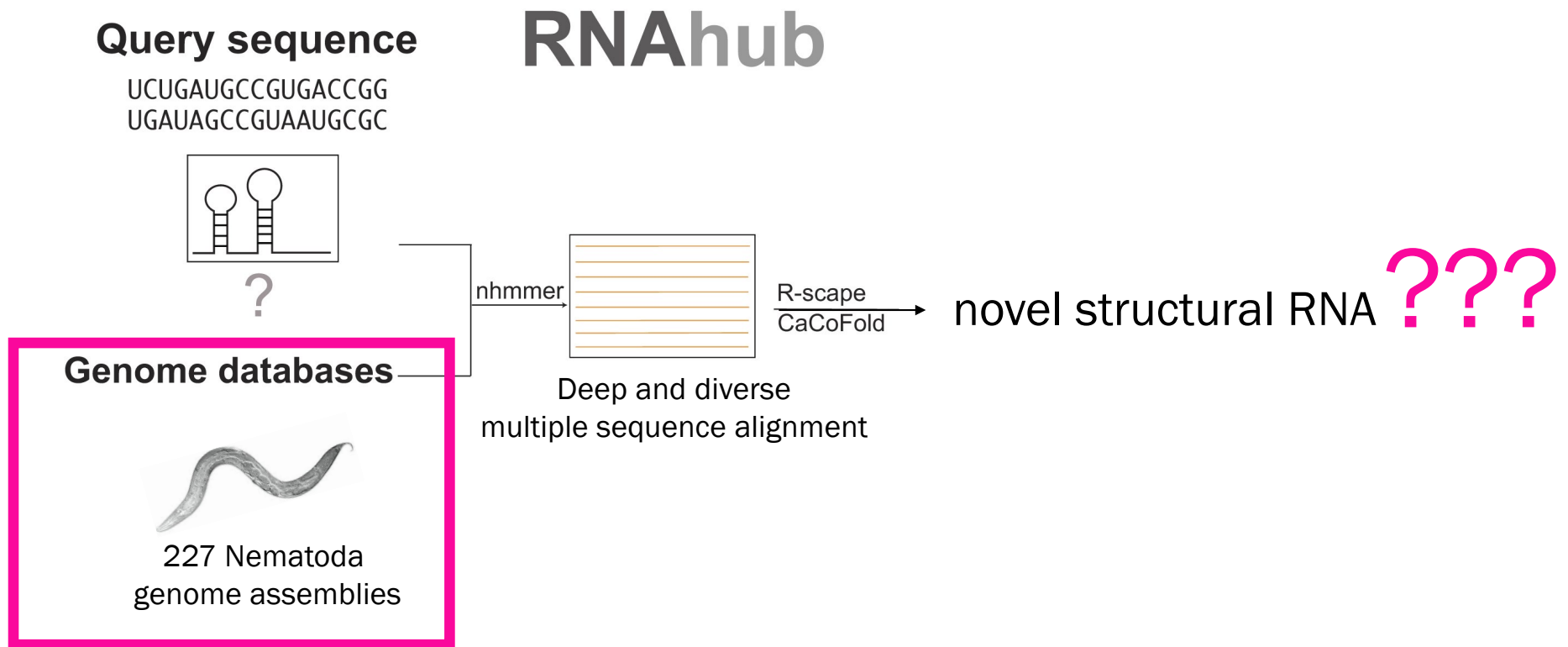
>>>MORE GENOMES>>>

>>>SPEED-UP>>>

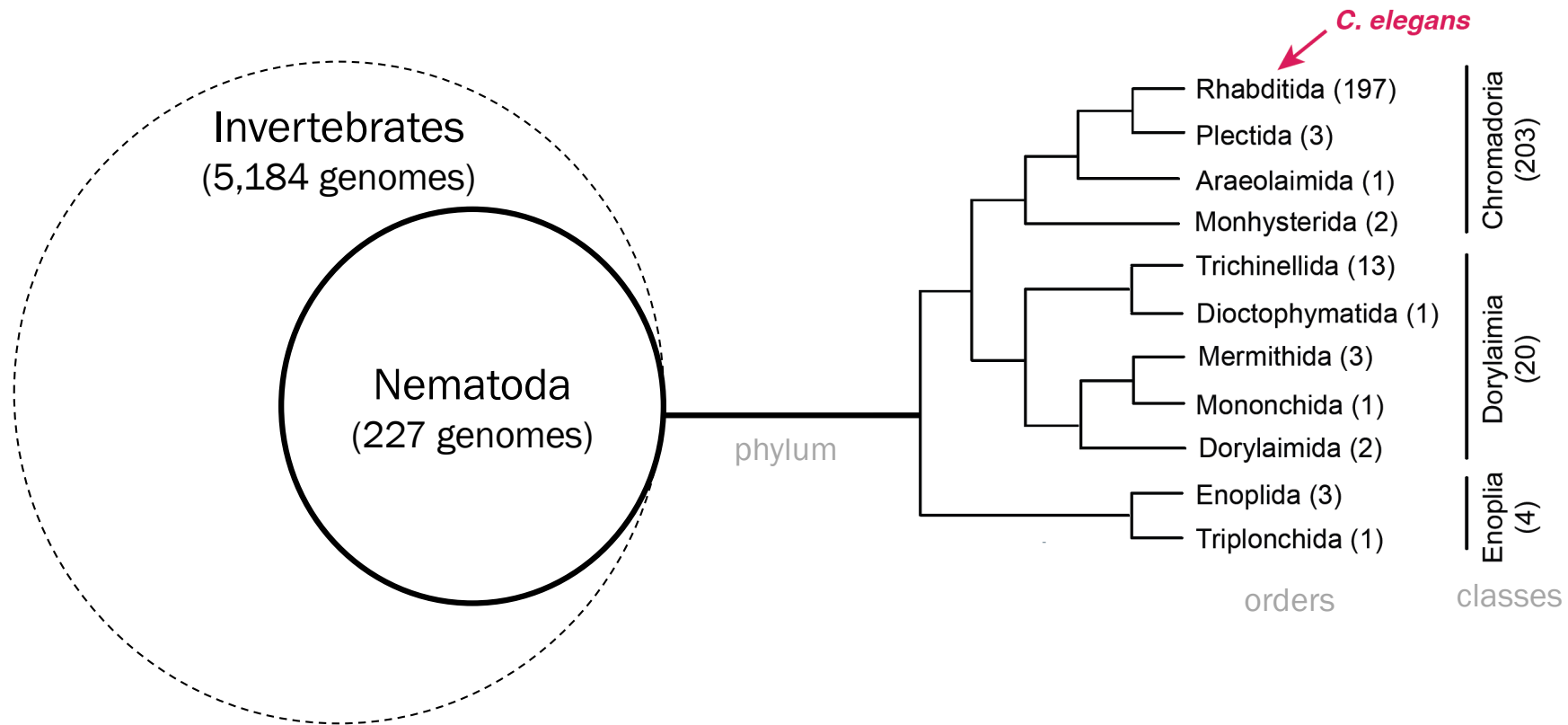
Pipeline for structural RNAs discovery in Nematoda



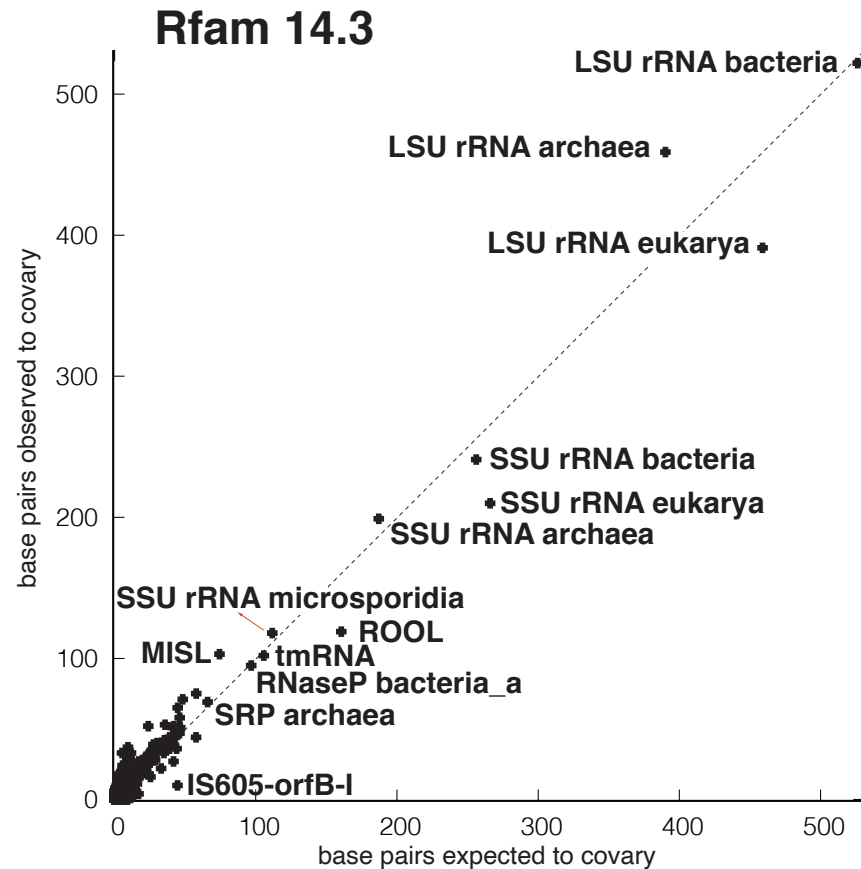
Pipeline for structural RNAs discovery in Nematoda



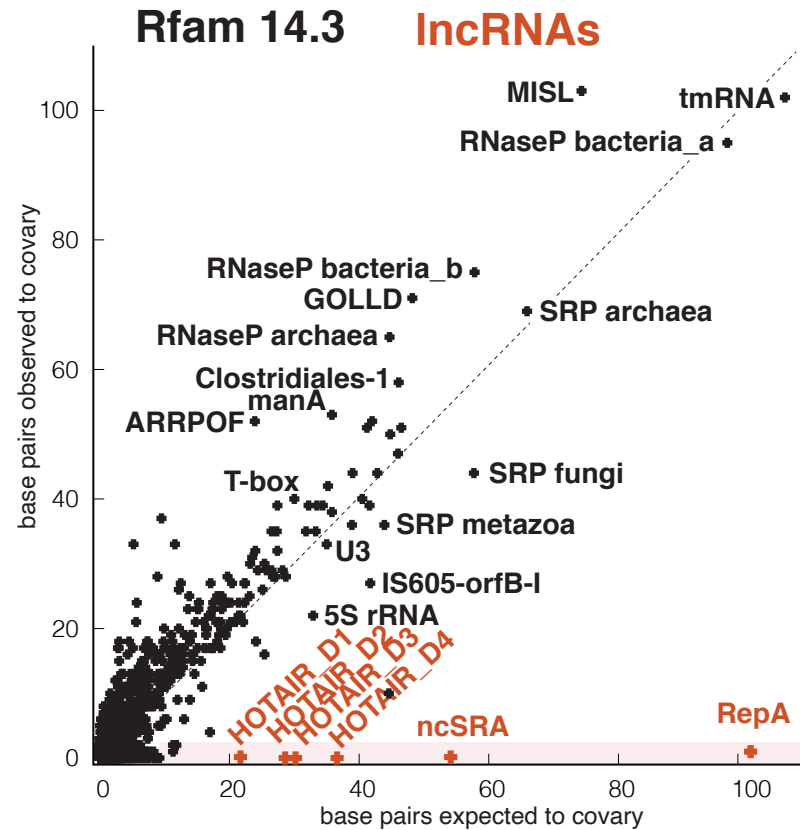
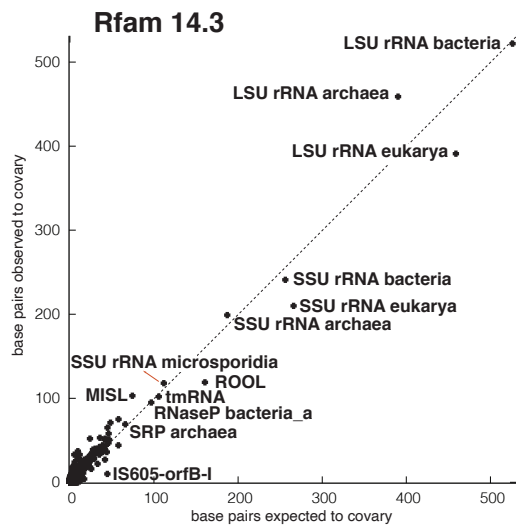
Taxonomy of the database (genomes pool)



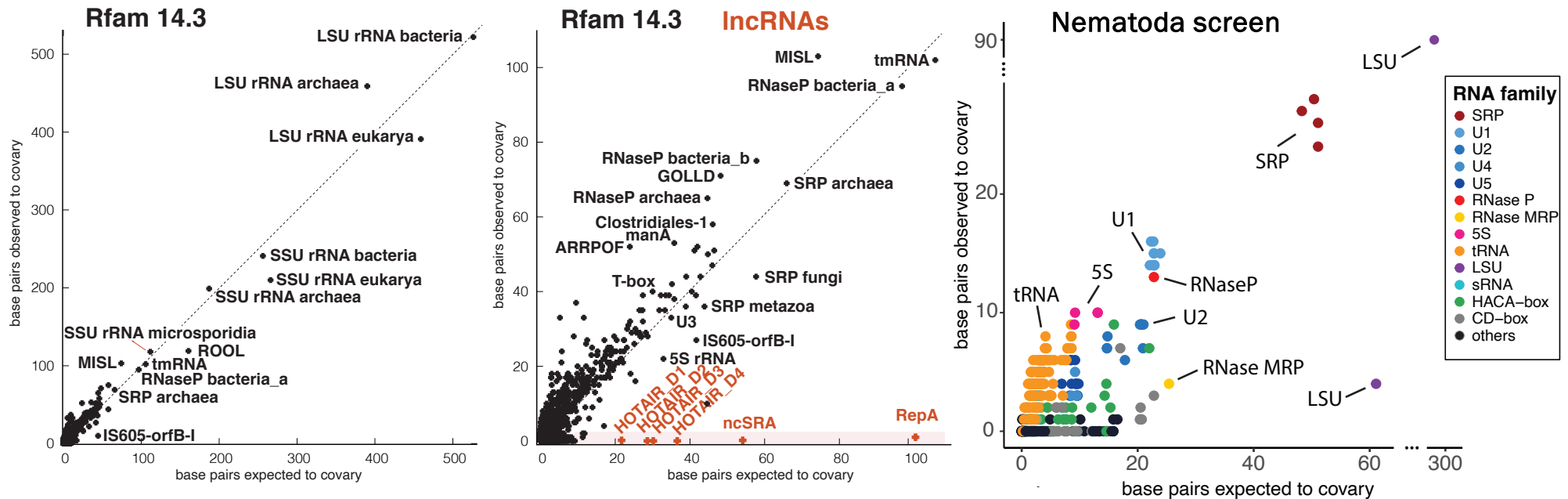
How many covarying base pairs do we find in Rfam?



How many covarying base pairs do we find in Rfam?

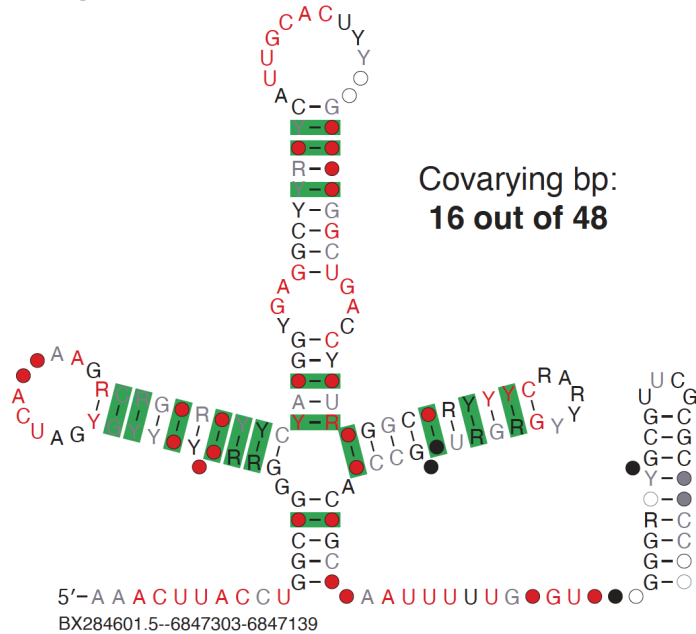


This kind of exercise can be performed in Nematoda

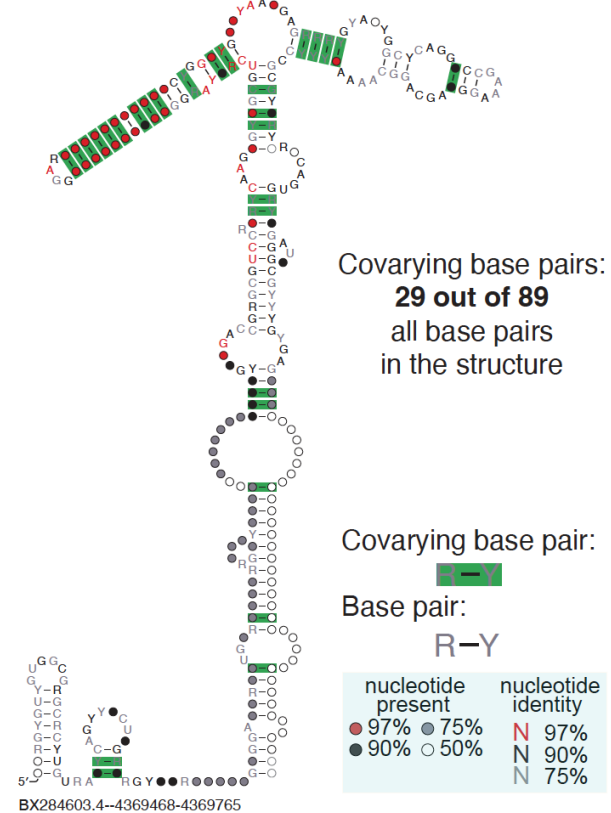


Example structures from positive control

U1 spliceosomal RNA



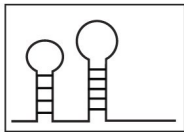
Signal Recognition Particle (SRP)



RNAhub

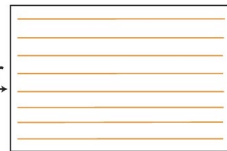
Query sequence

UCUGAUGCCGUGACCGG
UGAUAGCCGUAAUGCGC



?

nhmmer



R-scape
CaCoFold

novel structural RNA ???

Genome databases



227 Nematoda
genome assemblies

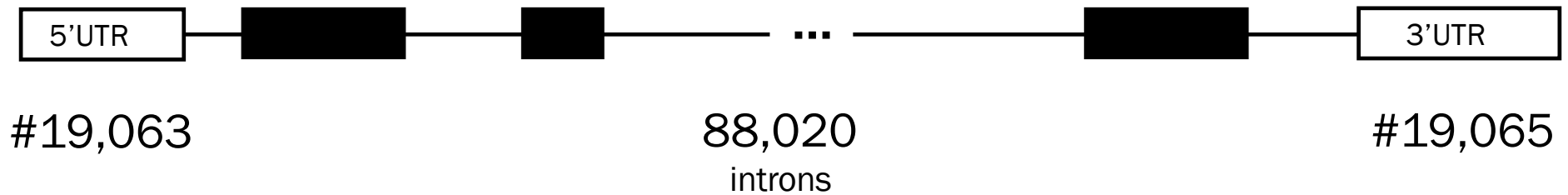
Deep and diverse
multiple sequence alignment

Screen for novel structural candidates - input

C. elegans

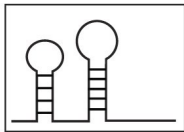


Annotations from: GCA_000002895.3_WBcel235_genomic.gff



Query sequence

UCUGAUGCCGUGACCGG
UGAUAGCCGUAAUGCGC



?

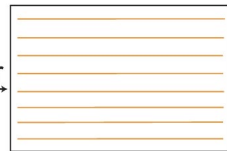
Genome databases



227 Nematoda
genome assemblies

RNAhub

nhmmer



Deep and diverse
multiple sequence alignment

R-scape
CaCoFold



novel structural RNA ???

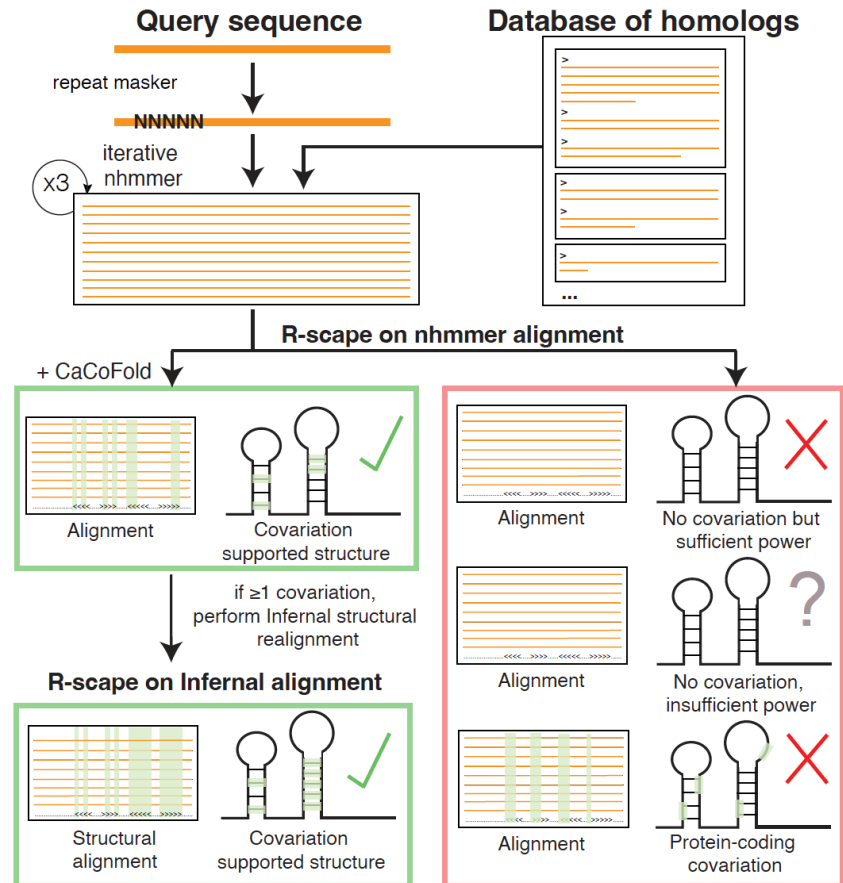
Pipeline for structural RNAs discovery in Nematoda

1. Creating database of homologs

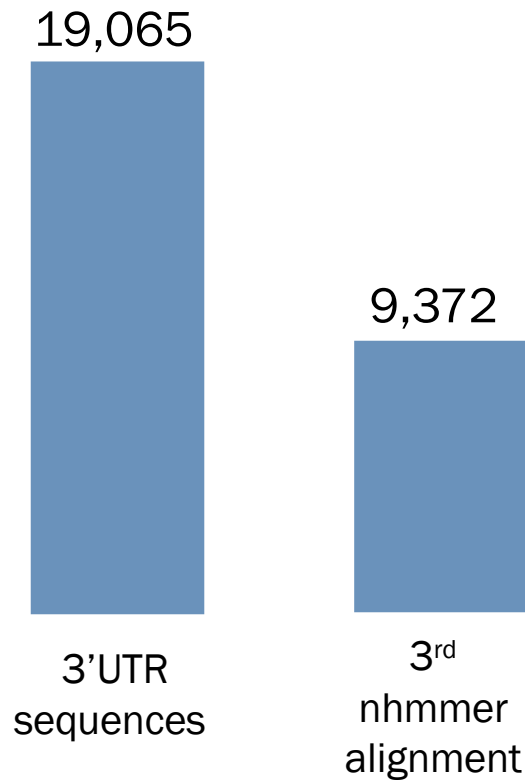
2. Creating multiple sequence alignment

3. Looking for covarying base pairs in the alignment

4. Looking for structural homologs with Infernal



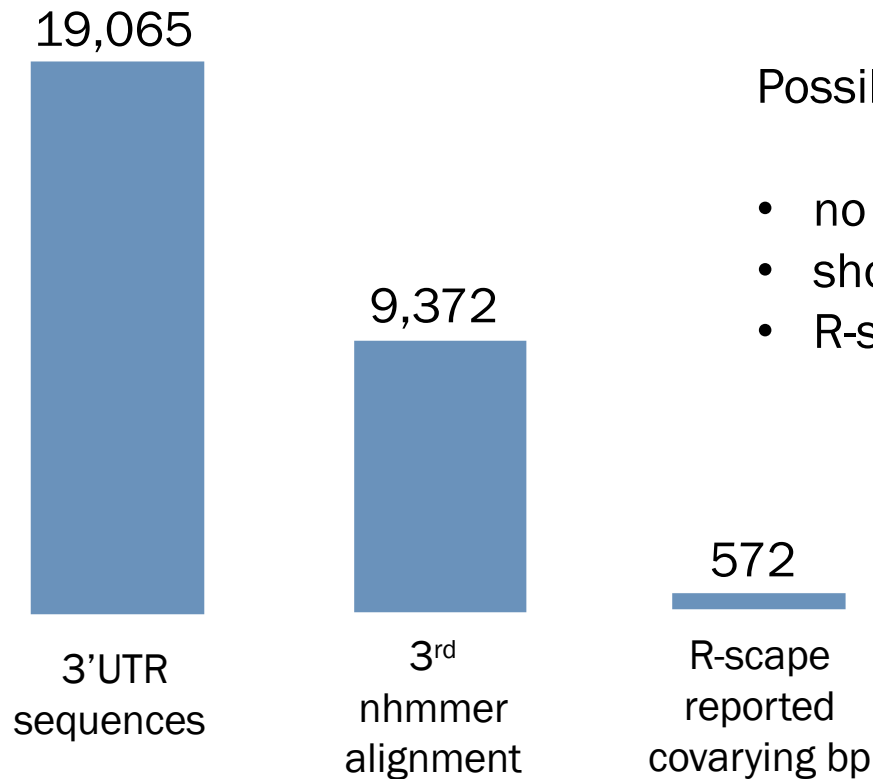
Summary of 3'UTRs screen - example



Possible reasons:

- no homologs
- short sequences

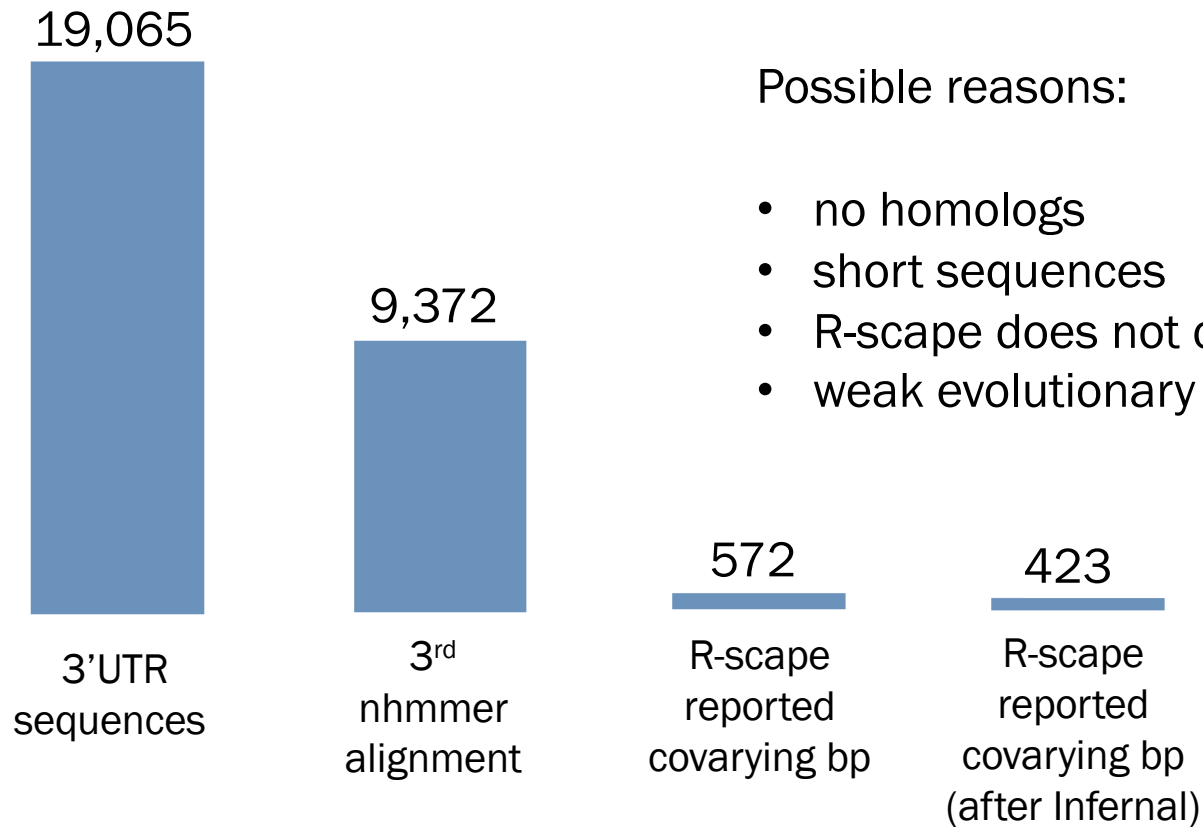
Summary of 3'UTRs screen - example



Possible reasons:

- no homologs
- short sequences
- R-scape does not detect an evolutionary signal

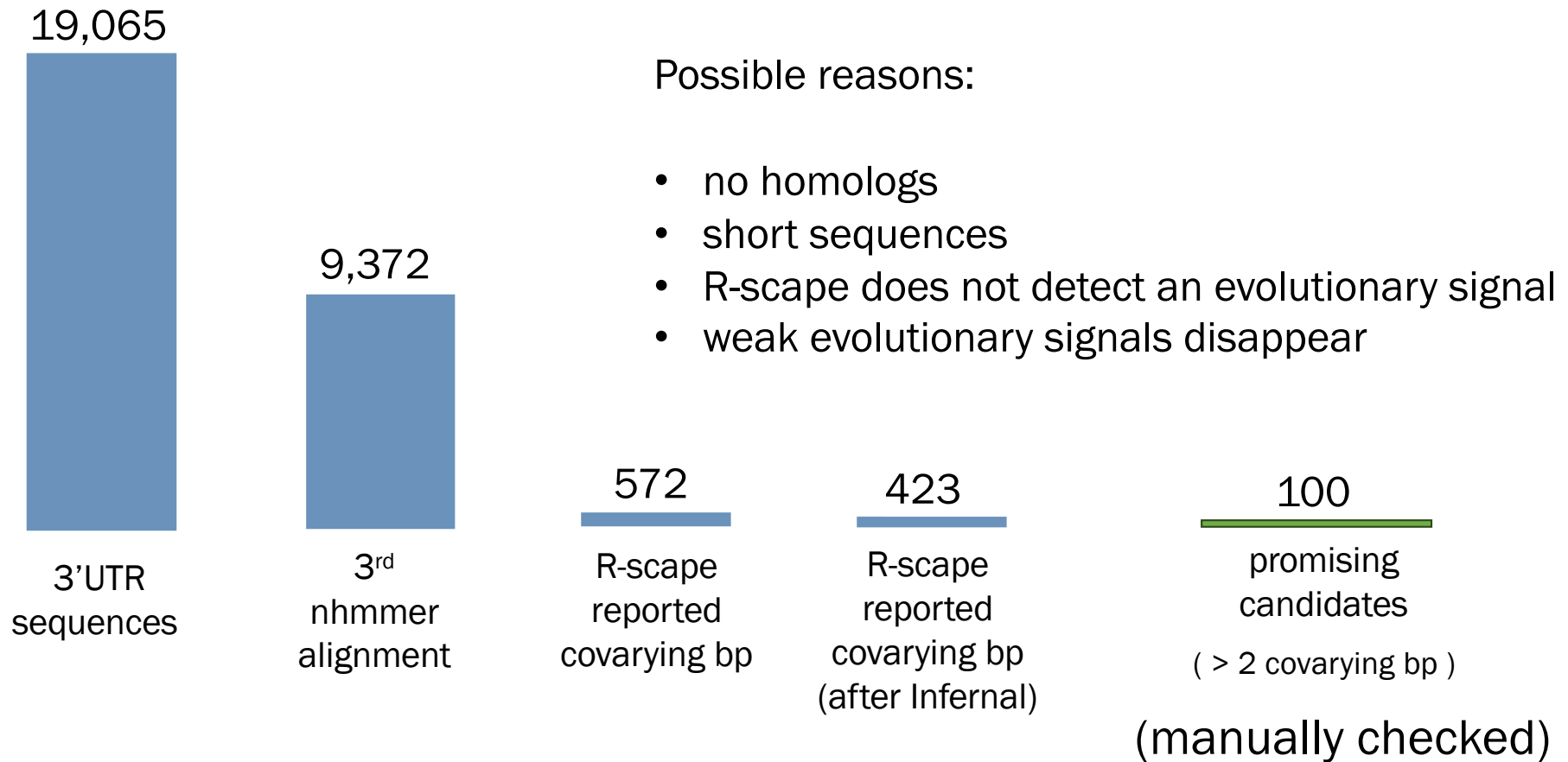
Summary of 3'UTRs screen - example



Possible reasons:

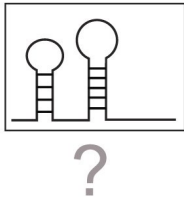
- no homologs
- short sequences
- R-scape does not detect an evolutionary signal
- weak evolutionary signals disappear

Summary of 3'UTRs screen - example



Query sequence

UCUGAUGCCGUGACCGG
UGAUAGCCGUAAUGCGC



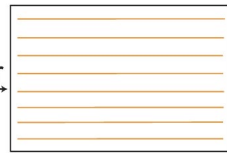
RNAhub

Genome databases



227 Nematoda
genome assemblies

nhmmer



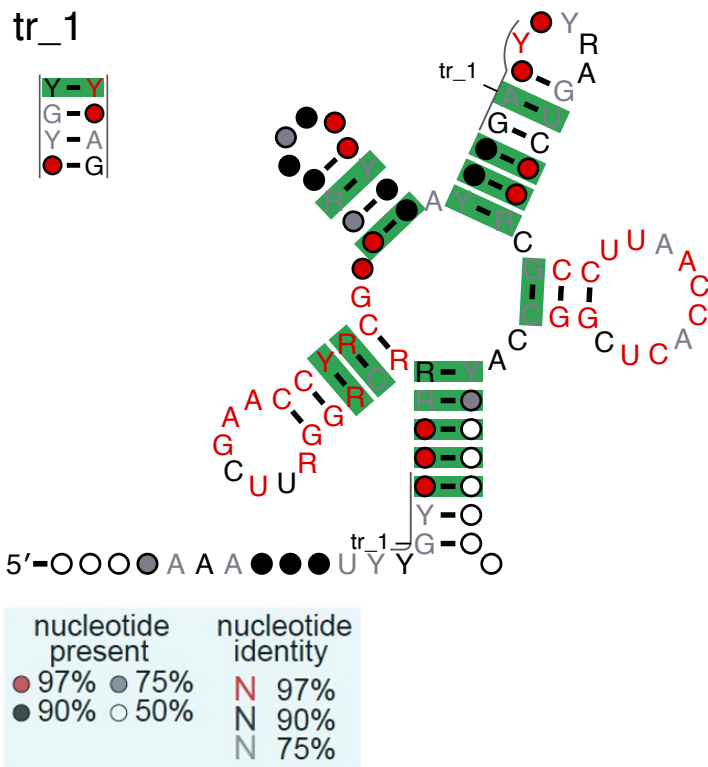
Deep and diverse
multiple sequence alignment

R-scape
CaCoFold

novel structural RNA ???

5 steps of candidates validation

Validation – step 1 – overall assessment



Last alignment statistics:

Number of sequences: 199
 Average length: 106.2
 Average identity: 67%

R-scape power outcome:

BPAIRS 29
 # avg substitutions per BP 26.1
 # BPAIRS expected to covary 17.5 +/- 2.1
 # BPAIRS observed to covary 15

Validation – step 2 – any clues about function?

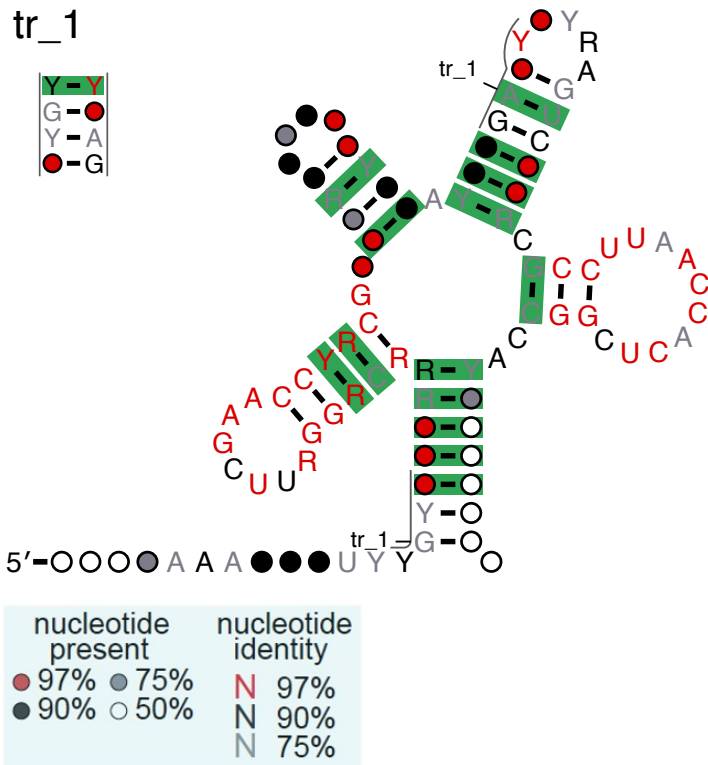
Checking in with Rfam:

#target name	accession	query name	seq from	score	E-value	inc
tRNA	RF00005	wdr-4_intron_521	136	55.2	2.9e-11	!
tRNA-Sec	RF01852	wdr-4_intron_521	136	29.6	0.0007	!

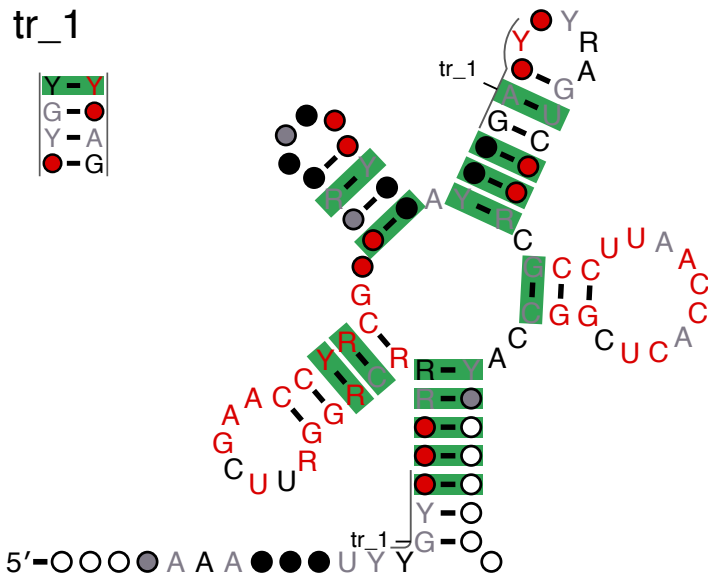
Program: cmscan

Can I find homologs in humans?
If yes, are those annotated?

- 138 hits to human
- 136 annotated as tRNAs



Validation – step 3 – additional, custom check



nucleotide present	nucleotide identity
● 97% ● 75%	N 97%
● 90% ○ 50%	N 90%
	N 75%

Checking in with Rfam:

```

#target name      accession  query name          seq from  score    E-value  inc
#-----
tRNA              RF00005    wdr-4_intron_521    136       55.2     2.9e-11  !
tRNA-Sec          RF01852    wdr-4_intron_521    136       29.6     0.0007   !
#
#  Program: cmscan

```

Can I find homologs in humans?
If yes, are those annotated?

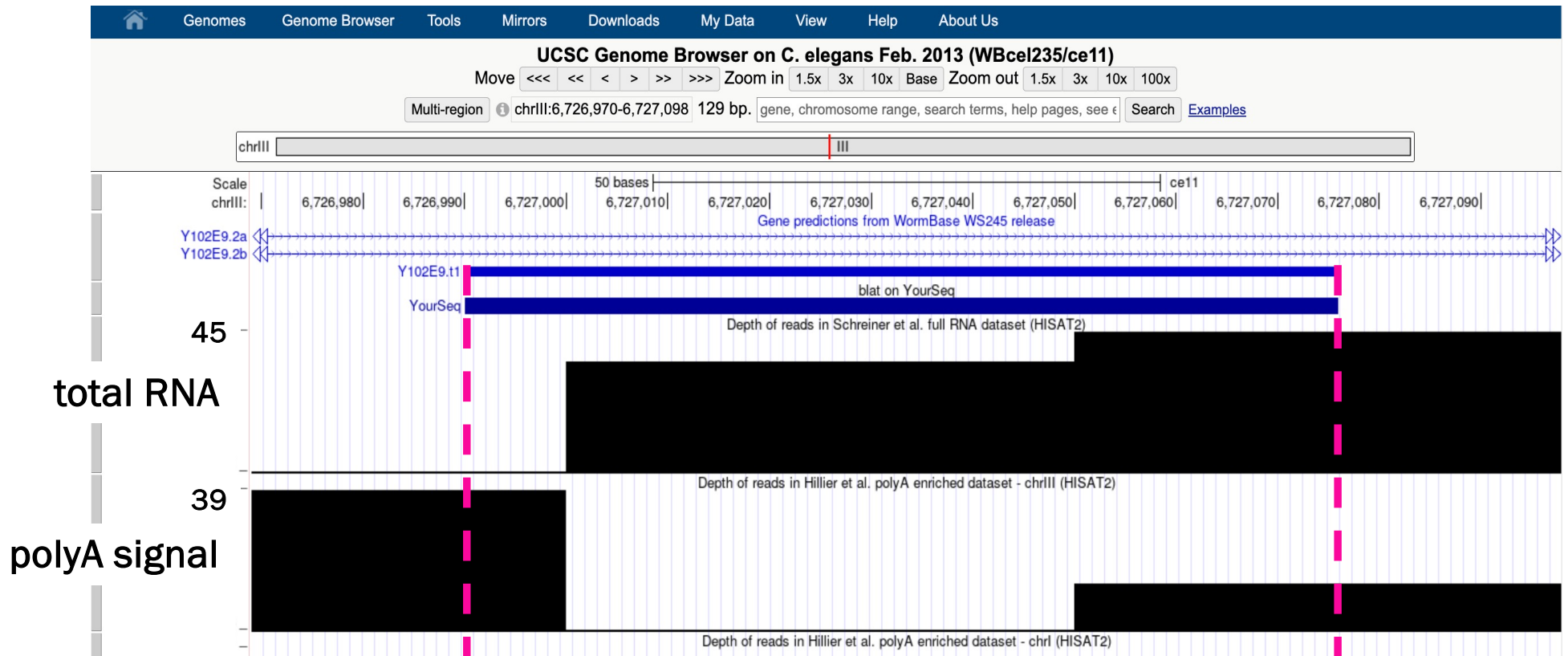
- 138 hits to human
- 136 annotated as tRNAs

tRNAscan-SE:

wdr-4_intron_521.trna1 (136-51) Length: 86 bp
Type: Arg Anticodon: CCT at 37-39 (100-98) Score: 68.7

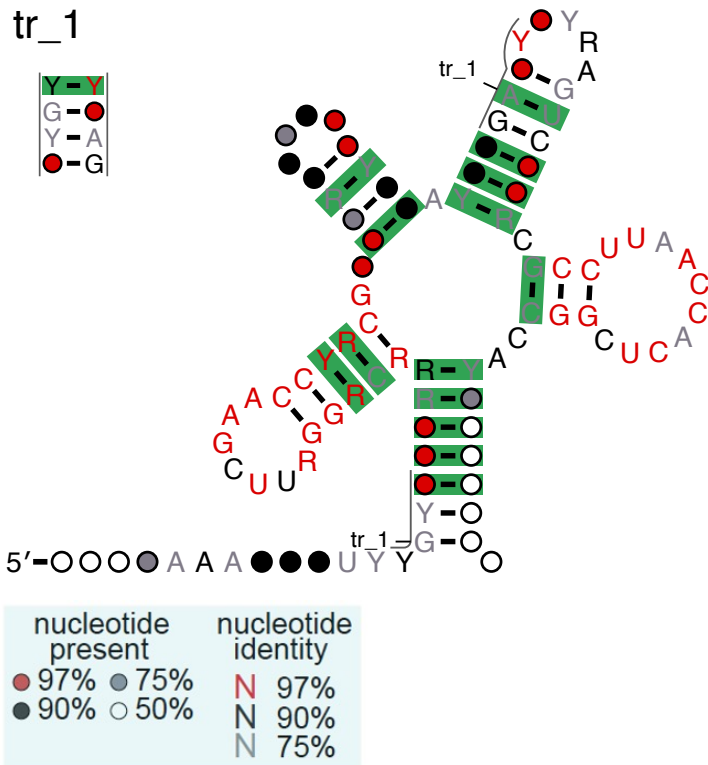
		*		*		*		*		*		*		*		*
Seq: GCCACGGTGGCCGAGTgTGGTcaAAGCGTGAGACTCCTGATCTCTTTTCGGCAACCGAtCGCAGGTTCGAATCTGCCCGTGGCA																
Str: >>>>>...>>.....<<<..>>>.....<<<<...>>>....<<<...>>>>.....<<<<<<<<<<<																

Validation – step 4 – is it expressed?

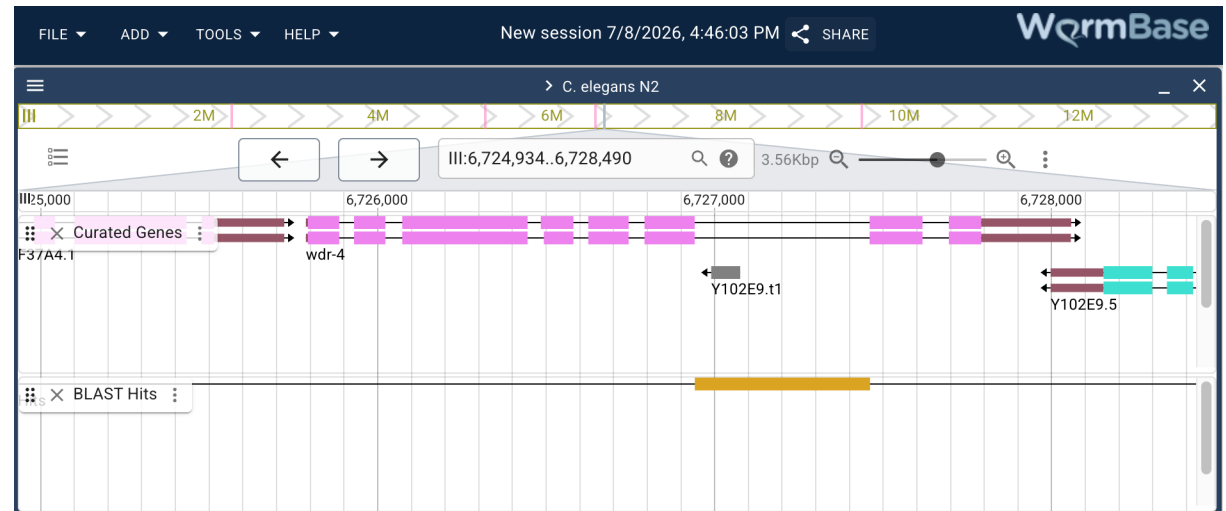


Klapproth, et al., (2026) <https://doi.org/10.3390/ncrna12010006>

Validation – step 5 – is it known?



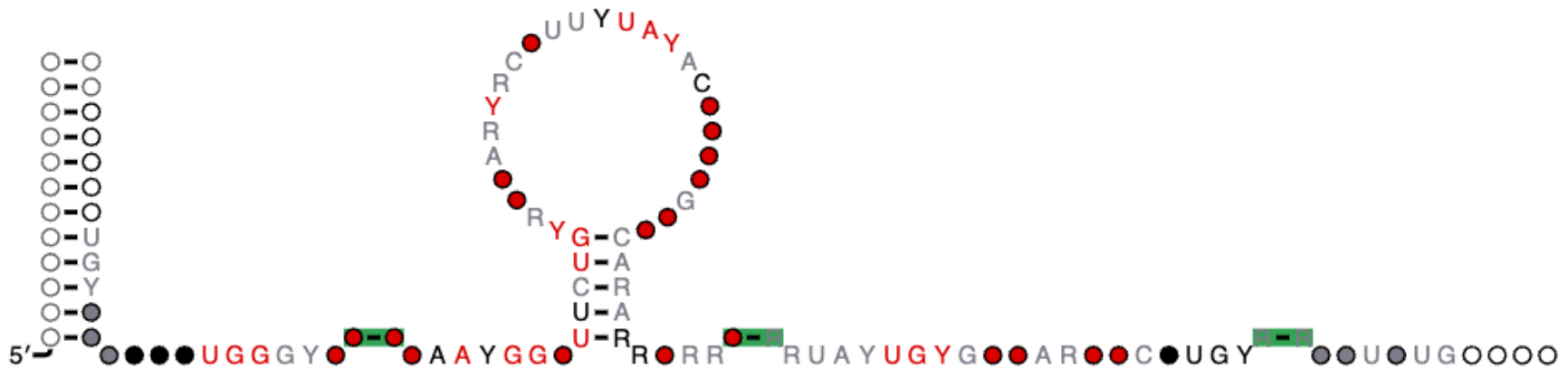
WormBase



RNAcentral

Not all covariation is due to the structure

F35E8.2_3_prime_UTR_397



nucleotide present		nucleotide identity	
● 97%	● 75%	N 97%	
● 90%	○ 50%	N 90%	
		N 75%	

Novel candidate (intron)

Nematoda

Last alignment statistics:

Number of sequences: 58

Average length: 108.6

Average identity: 63%

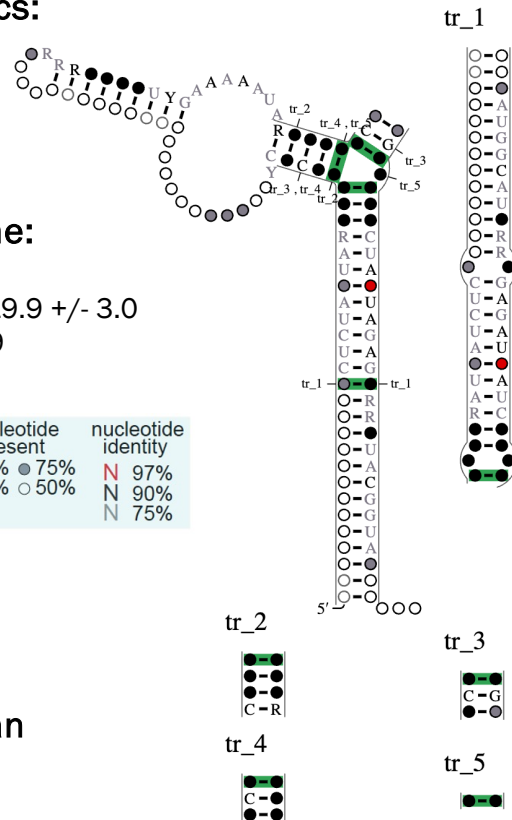
R-scape power outcome:

BPAIRS 55

BPAIRS expected to covary 19.9 +/- 3.0

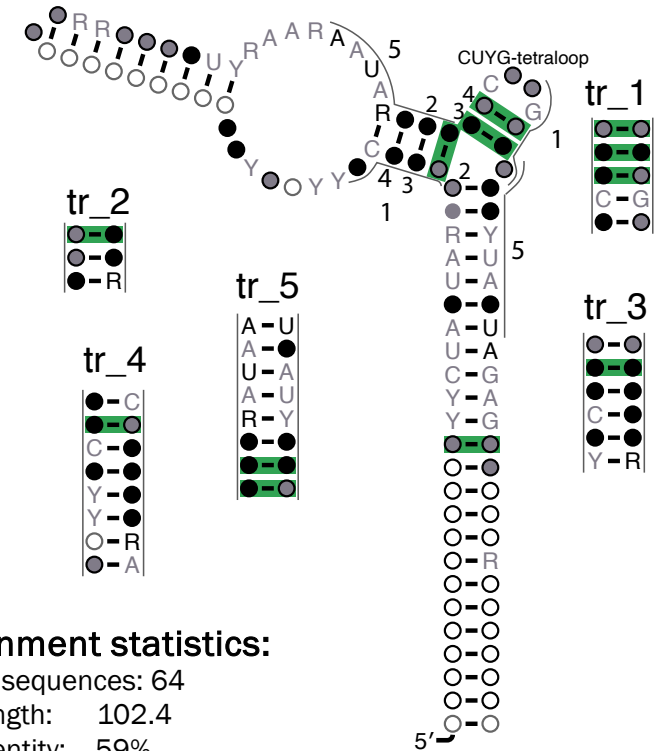
BPAIRS observed to covary 9

nucleotide present	nucleotide identity
● 97% ● 75%	N 97%
● 90% ○ 50%	N 90%
	N 75%



- no hit to Rfam
- one homolog to human
- expressed

Invertebrates



Last alignment statistics:

Number of sequences: 64

Average length: 102.4

Average identity: 59%

R-scape power outcome:

BPAIRS 71

BPAIRS expected to covary 32.7 +/- 3.6

BPAIRS observed to covary 12

Similar elements found in intergenic regions

- candidate 18245

Last alignment statistics:

Number of sequences: 54

Alignment length: 897

Average length: 107.2

Average identity: 61%

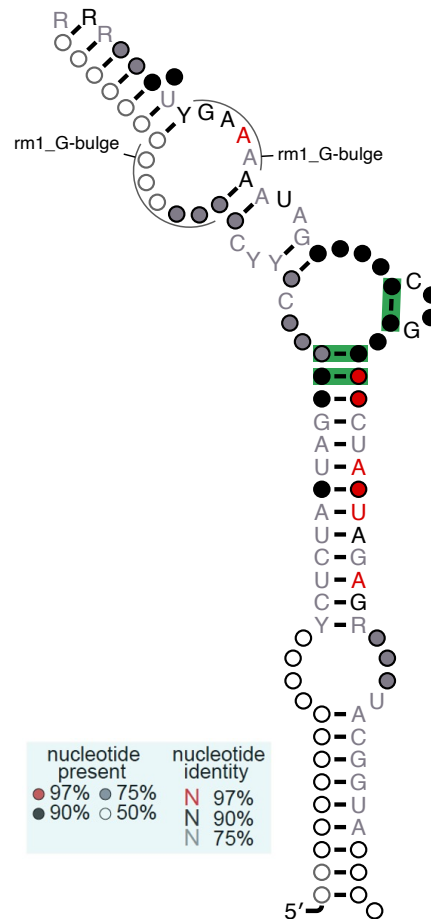
R-scape power outcome:

BPAIRS 59

BPAIRS expected to covary 23.0 +/- 3.2

BPAIRS observed to covary 12

- no hit to Rfam
- no homolog to human
- expressed



Analysis with all invertebrate genomes in progress...

It was a great trip to Benasque!
...and we look for recommendations for the way back

