Evolution of mitogenome-encoded RNAs: everything goes

B. Franz Lang, Matt Sarrasin, Samuel Prince, Jacob Beaulac, Gertraud Burger

Robert Cedergren Centre for Bioinformatics and Genomics, Département de Biochimie et Médecine Moléculaire, Université de Montréal, Canada

.

***Email: b.franz.lang@gmail.com**

Nothing in biology makes sense except in the light of Evolution.

Theodosius Dobzhansky

Mitochondria are Nature's evolutionary playground *par excellence*.

The case for our mission - automated genome annotation

GenBank records provide basic information on genes, proteins, RNAs etc, but **cannot be trusted**.

Genome annotations unpredictably have :

- **missing genes**
- **incorrect gene structures**: missing introns and exons, shifted intron positions, extra invalid exons and introns ...
- NCBI **curation is limited**
- virtually **no update of records**

The case for our mission - automated genome annotation

In addition, some gene structures are excessively difficult to resolve with current tools.

Example: *Orbilia brochopaga* fungal mitogenome encodes a *cox1* gene (protein-coding) of 66.7 kbp with 24 introns ...with numerous small exons (mini exons) that are too small to be identified by profile HMM search.

The case for our mission - automated genome annotation

The **most reliable value of GenBank records :** nucleotide **sequence** plus related species **taxonomy**.

Solution: Develop tools and annotators that infer state-of-the-art annotations that do not require substantial expert validation.

MFannot: mitogenomes other than bilaterian animals
Eukan: nuclear genomes (submitted)

This talk: two mito ncRNAs – tmRNAs, RNase P RNAs

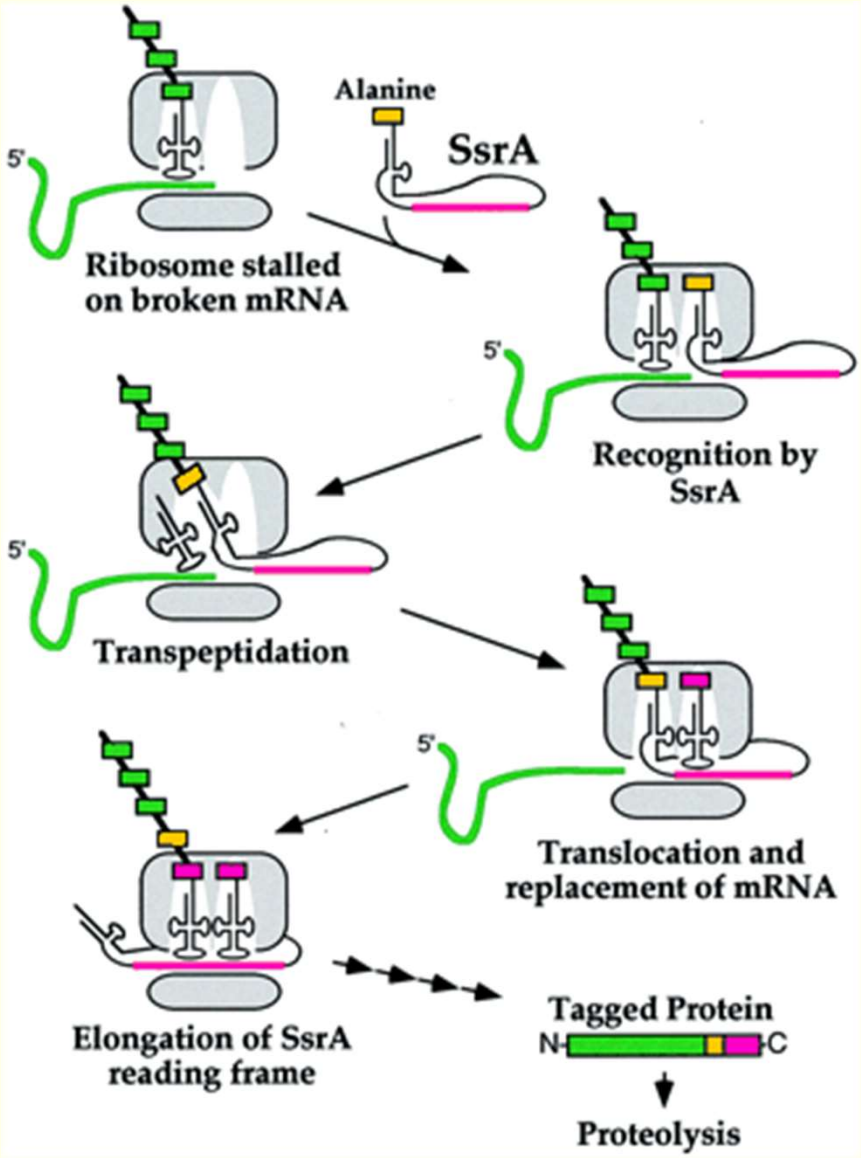
ncRNAs encoded by mitogenomes

Variable presence across eukaryotes, sequence and structures tend to be **highly derived**, with variable-size insertions and deletions, some in pieces, intron-containing, edited … and therefore **difficult to identify**.

Currently known :

- **tRNAs, small and large subunit rRNAs** (almost universal, RNA editing … PMID 22708551, 15247432, 15037771, 11861890, 9016605, 23223758, 12626702)
- **5S rRNA** (plants, protists; lacks in animals, fungi … 25429974)
- **Group I and II intron RNAs** (widespread, frequent in fungi; 37469769, 23823571, 15037770)
- **tmRNA** (jakobids, oomycetes, some other protists … 35369492, 37469769, 17280737, 22411852)

  Release of ribosomes stalled on 'non-stop' mRNAs
- **RNase P RNA** (fungi and protists; lacks in animals, plants … 12923256, 15689432, 9168110)

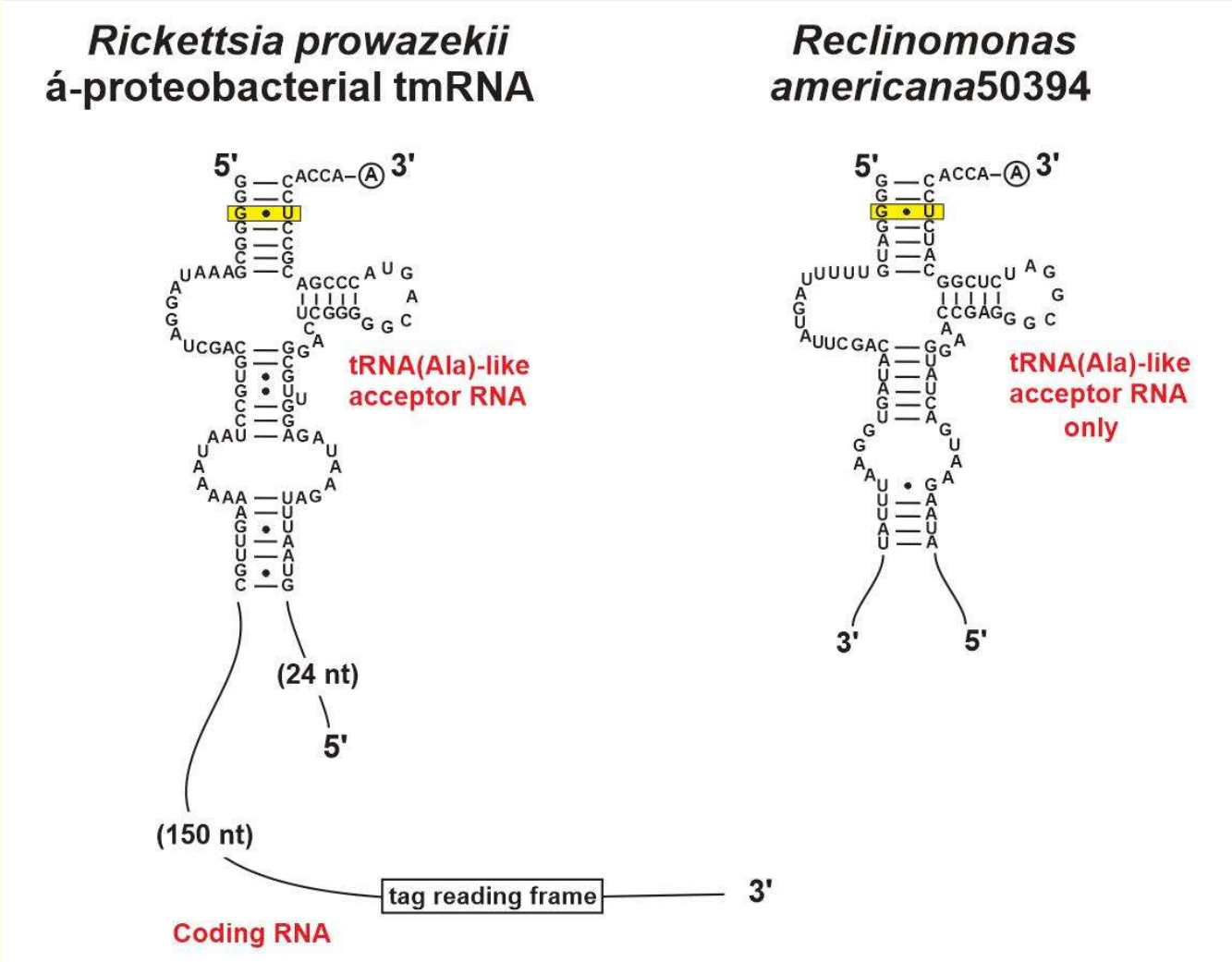


What are tmRNAs?

In bacteria, tmRNAs permit the liberation of stalled ribosomes at the end of broken mRNAs. The addition of a non-encoded alanine by a tRNA-like domain (without anticodon) releases the ribosome, and translation continues in an mRNA domain of the tmRNA (red tag sequence). The addition of the peptide tag to the truncated protein earmarks it for degradation.

In other words, this is programmed ribosomal hopping from one transcript to another (trans-translation).

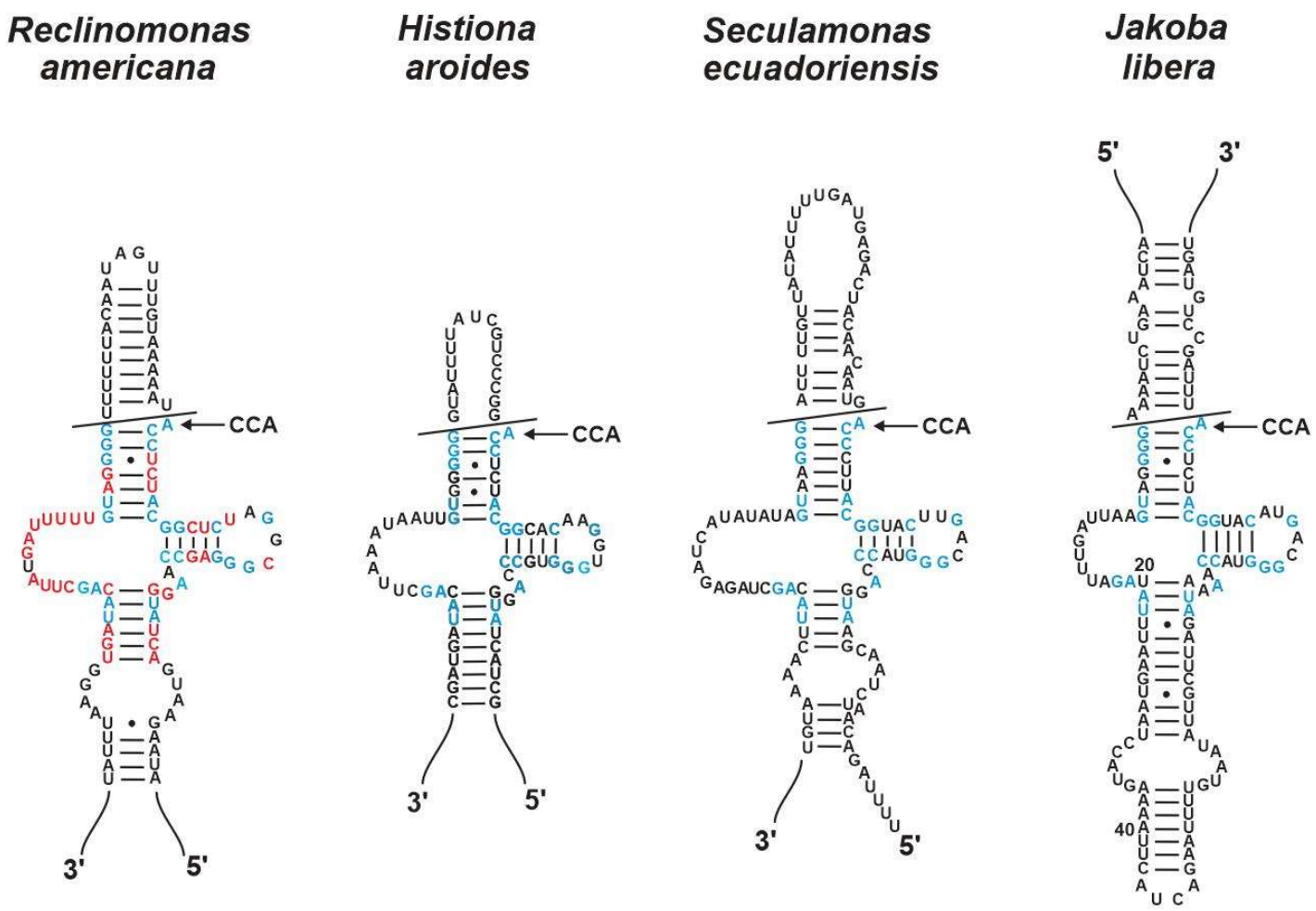
Reduced tmRNA structure in mitochondria



Typical structure of a bacterial *versus* mitochondrial tmRNA (loss of tag reading frame region).

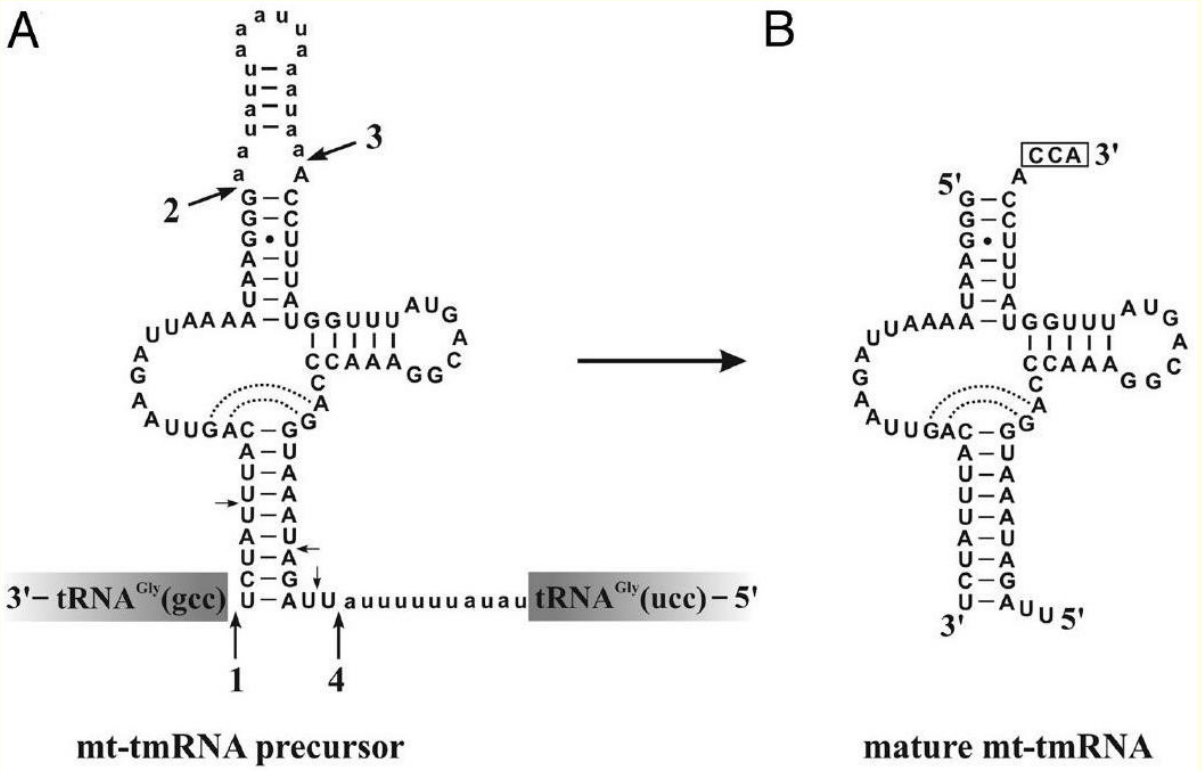
Note the G-U pair in position three of the acceptor stem specifying Alanine.

Comparison of jakobid mt tmRNA genes



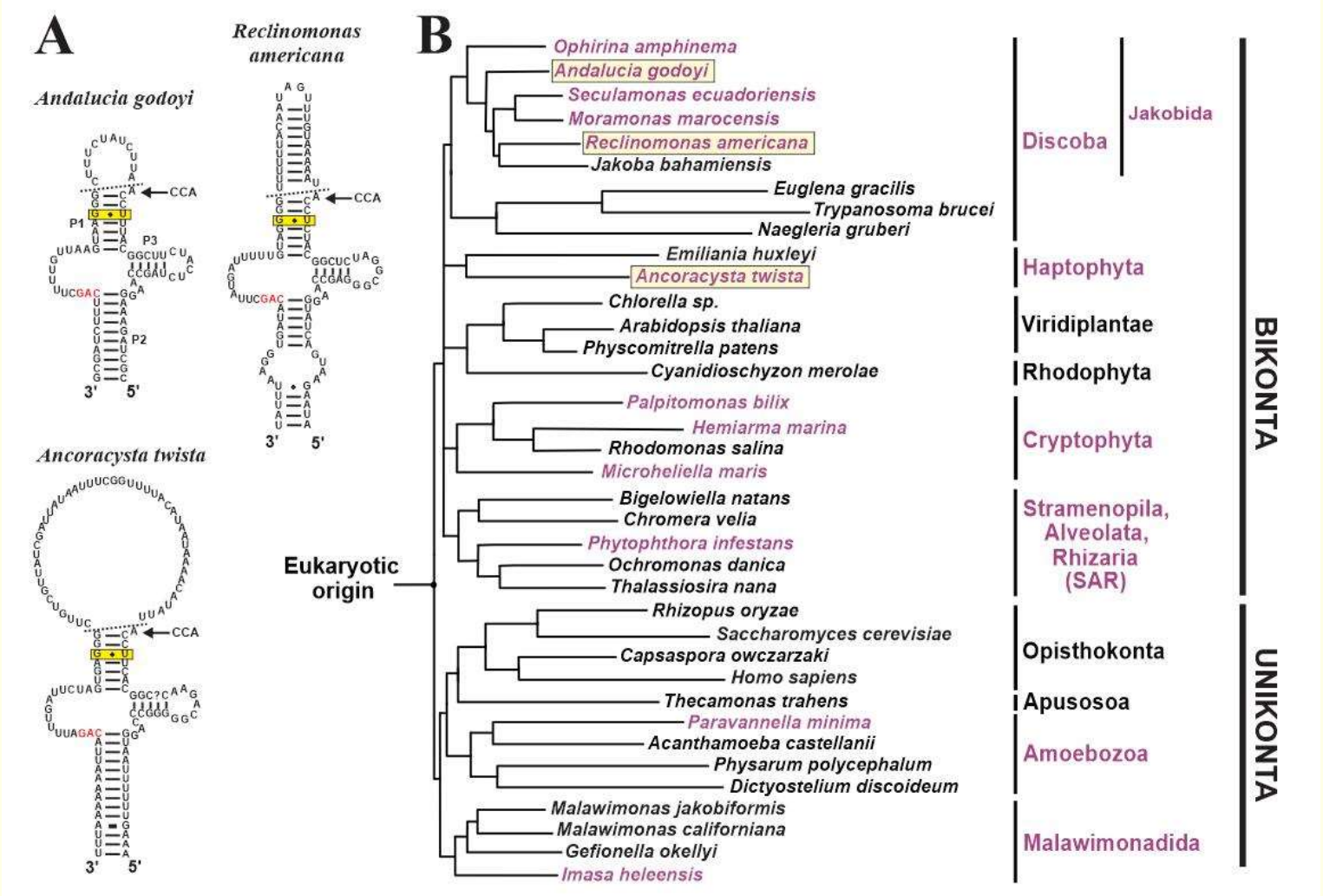
In *J. libera*, the structure is encoded in one piece, similar to a tRNA. In all others, a circular permutation leaves two RNA pieces after processing.

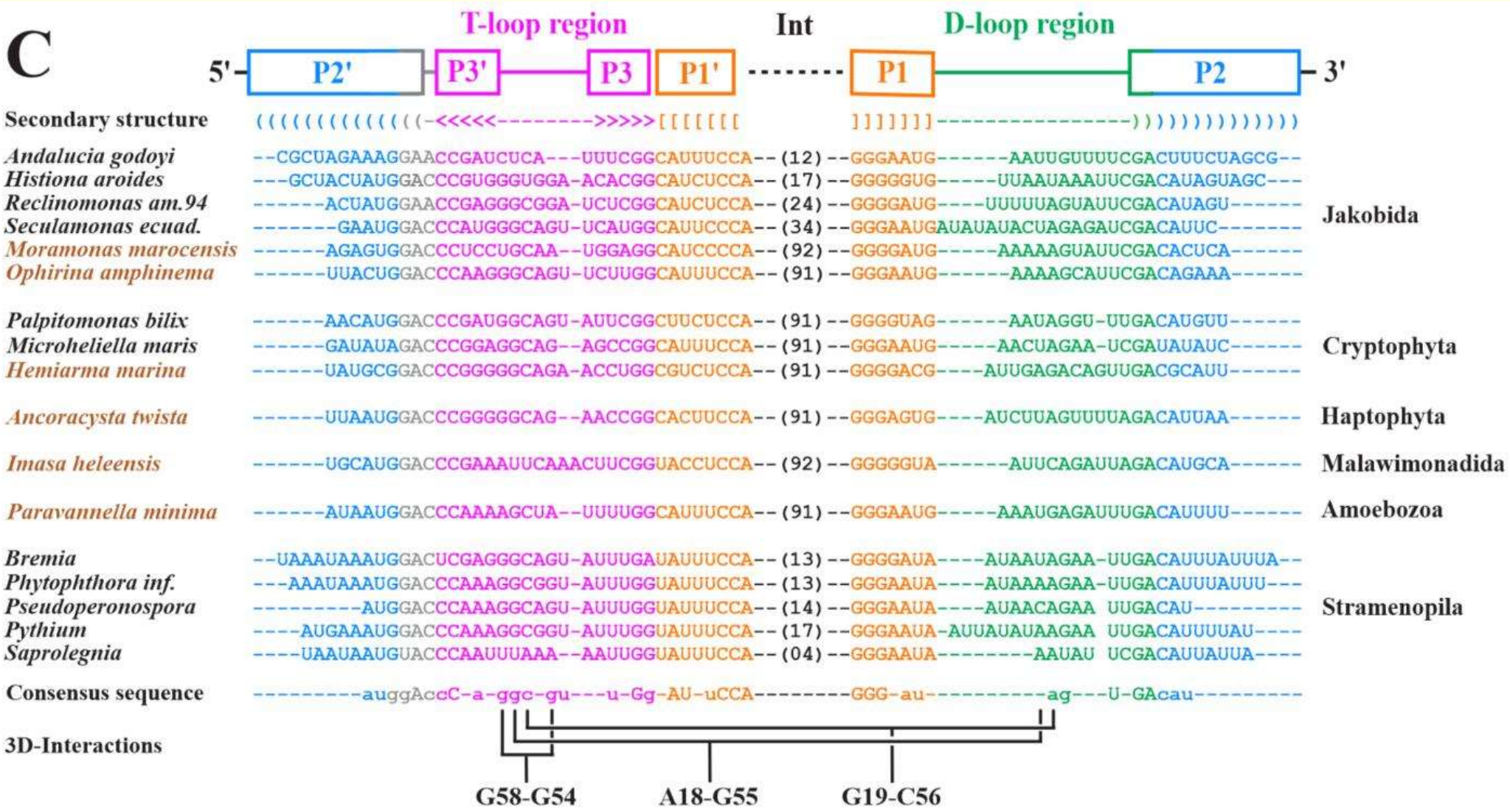
RNA processing of mt-tmRNA in the oomycete *Phytophthora*



Step 1 and 4 process the anticodon-stem region, via 5' processing of tRNA-Gly and a specific endonucleolytic cut. Acceptor stem processing follows regular tRNA processing biochemistry (5': RNase P), followed by addition of the non-encoded CCA. Note the G-U pair at position three of the acceptor stem, which is the landmark recognition site of Ala tRNA synthetase.

Identified mt tmRNAs across eukaryotes

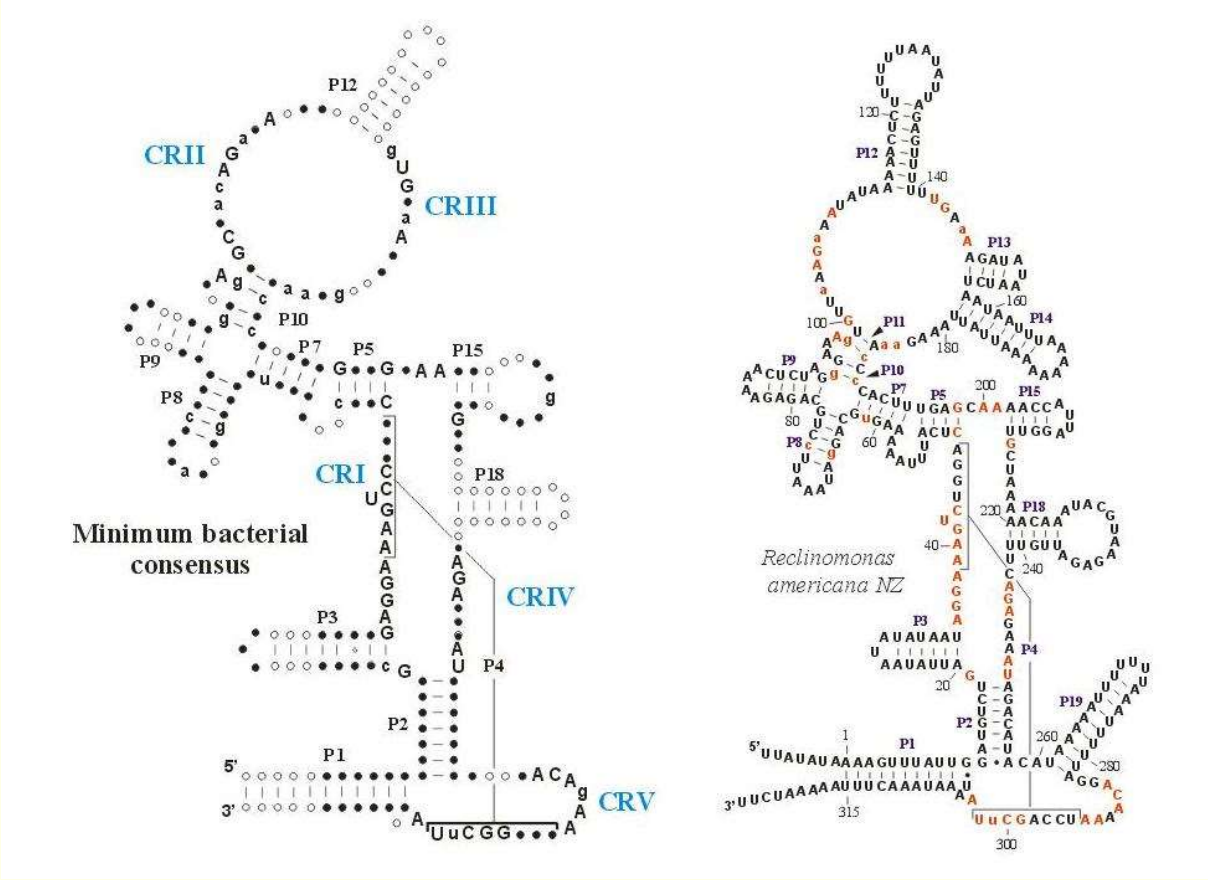



C
T-loop region
Int
D-loop region
5'
P2'
P3'
P3
P1'
P1
P2
3'
Secondary structure
(((((((((((((-<<<<<--------->>>>>[[[[[[[ ]]]]]]]----------------)))))))))))))
Andalucia godoyi --CGCUAGAAAGGAACCGAUCUCA---UUUCGGCAUUUCCA--(12)--GGGAAUG------AAUUGUUUUCGACUUUCUAGCG--
Histiona aroides ---GCUACUAUGGACCCGUGGGUGGA-ACACGGCAUCUCCA--(17)--GGGGGUG-----UUAAUAAAUUCGACAUAGUAGC---
Reclinomonas am.94 ------ACUAUGGAACCGAGGGCGGA-UCUCGGCAUCUCCA--(24)--GGGGAUG----UUUUUAGUAUUCGACAUAGU------
Seculamonas ecuad. -------GAAUGGACCCAUGGGCAGU-UCAUGGCAUUCCCA--(34)--GGGAAUGAUAUAUACUAGAGAUCGACAUUC-------
Moramonas marocensis ------AGAGUGGACCCUCCUGCAA--UGGAGGCAUCCCCA--(92)--GGGGAUG-----AAAAAGUAUUCGACACUCA------
Ophirina amphinema ------UUACUGGACCCAAGGGCAGU-UCUUGGCAUUUCCA--(91)--GGGAAUG------AAAAGCAUUCGACAGAAA------
Palpitomonas bilix ------AACAUGGACCCGAUGGCAGU-AUUCGGCUUCUCCA--(91)--GGGGUAG------AAUAGGU-UUGACAUGUU------
Microheliella maris ------GAUAUAGACCCGGAGGCAG--AGCCGGCAUUUCCA--(91)--GGGAAUG-----AACUAGAA-UCGAUAUAUC------
Hemiarma marina ------UAUGCGGACCCGGGGGCAGA-ACCUGGCGUCUCCA--(91)--GGGGACG----AUUGAGACAGUUGACGCAUU------
Ancoracysta twista ------UUAAUGGACCCGGGGGCAG--AACCGGCACUUCCA--(91)--GGGAGUG----AUCUUAGUUUUAGACAUUAA------
Imasa heleensis ------UGCAUGGACCCGAAAUUCAAACUUCGGUACCUCCA--(92)--GGGGGUA------AUUCAGAUUAGACAUGCA------
Paravannella minima ------AUAAUGGACCCAAAAGCUA--UUUUGGCAUUUCCA--(91)--GGGAAUG-----AAAUGAGAUUUGACAUUUU------
Bremia --UAAAUAAAUGGACUCGAGGGCAGU-AUUUGAUAUUUCCA--(13)--GGGGAUA----AUAAUAGAA-UUGACAUUUAUUUA--
Phytophthora inf. ---AAAUAAAUGGACCCAAAGGCGGU-AUUUGGUAUUUCCA--(13)--GGGAAUA----AUAAAAGAA-UUGACAUUUAUUU---
Pseudoperonospora ---------AUGGACCCAAAGGCAGU-AUUUGGUAUUUCCA--(14)--GGGAAUA----AUAACAGAA UUGACAU---------
Pythium ----AUGAAAUGGACCCAAAGGCGGU-AUUUGGUAUUUCCA--(17)--GGGAAUA-AUUAUAUAAGAA UUGACAUUUUAU----
Saprolegnia ----UAAUAAUGUACCCAAUUUAAA--AAUUGGUAUUUCCA--(04)--GGGAAUA--------AAUAU UCGACAUUAUUA----
Jakobida
Cryptophyta
Haptophyta
Malawimonadida
Amoebozoa
Stramenopila
Consensus sequence ---------auggAccC-a-ggc-gu---u-Gg-AU-uCCA--------GGG-au----------ag---U-GAcau---------
3D-Interactions
G58-G54
A18-G55
G19-C56

What are RNase P RNAs?

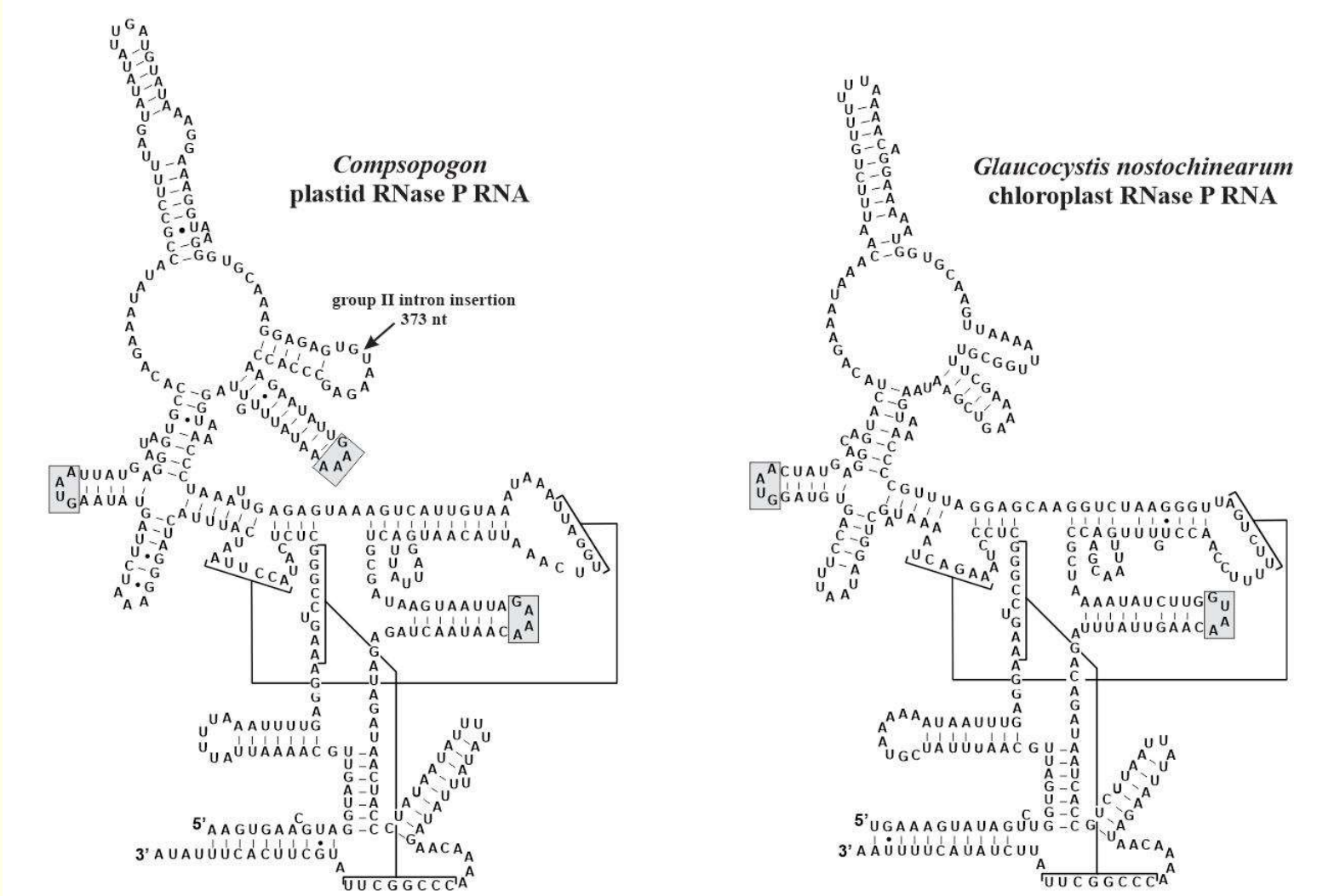
RNase P RNA is a catalytic RNA molecule that is part of the ribonuclease P (RNase P) enzyme complex. RNase P is an essential ribozyme found in almost all living organisms (exception plants, metazoans), including plastids and mitochondria. Its primary function is to process precursor tRNAs (and a few other RNAs such as tmRNAs) by cleaving their 5' leader sequences to generate mature 5’ termini.

Mitochondrial RNase P RNA structures – bacteria-like to highly derived



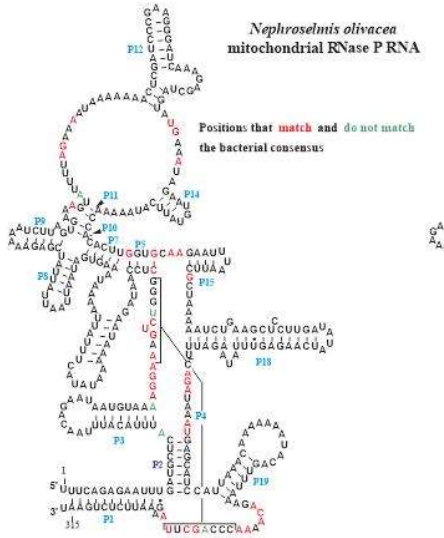
Jakobid mitochondrial RNase P RNA (right, *Reclinomonas*) has an essentially bacteria-like structure (left, bacterial consensus structure).

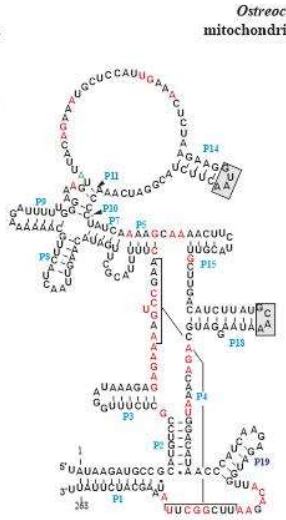
Plastid RNase P RNA structures are as bacteria-like

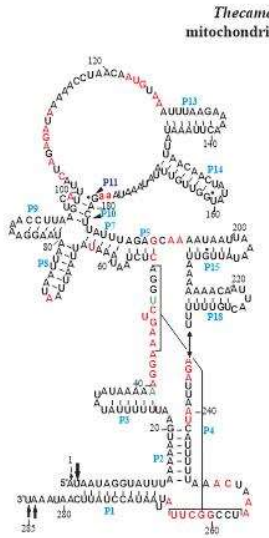


Mt RNase P RNA structures come in various degrees of structural reduction

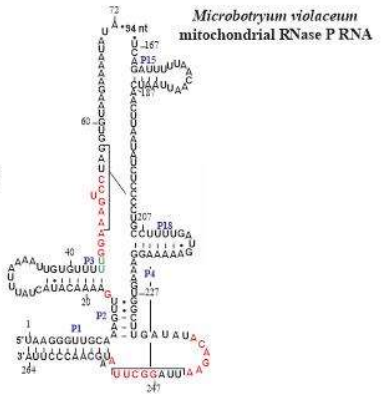
Protists

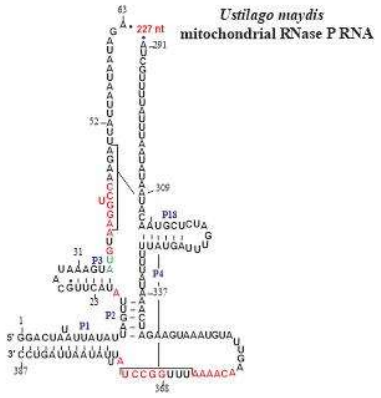


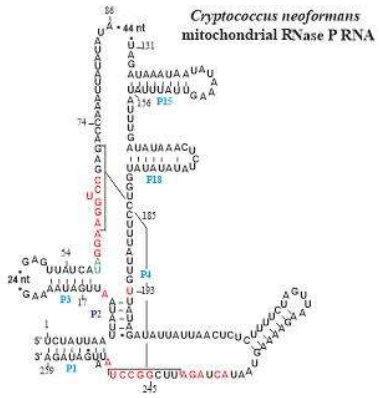


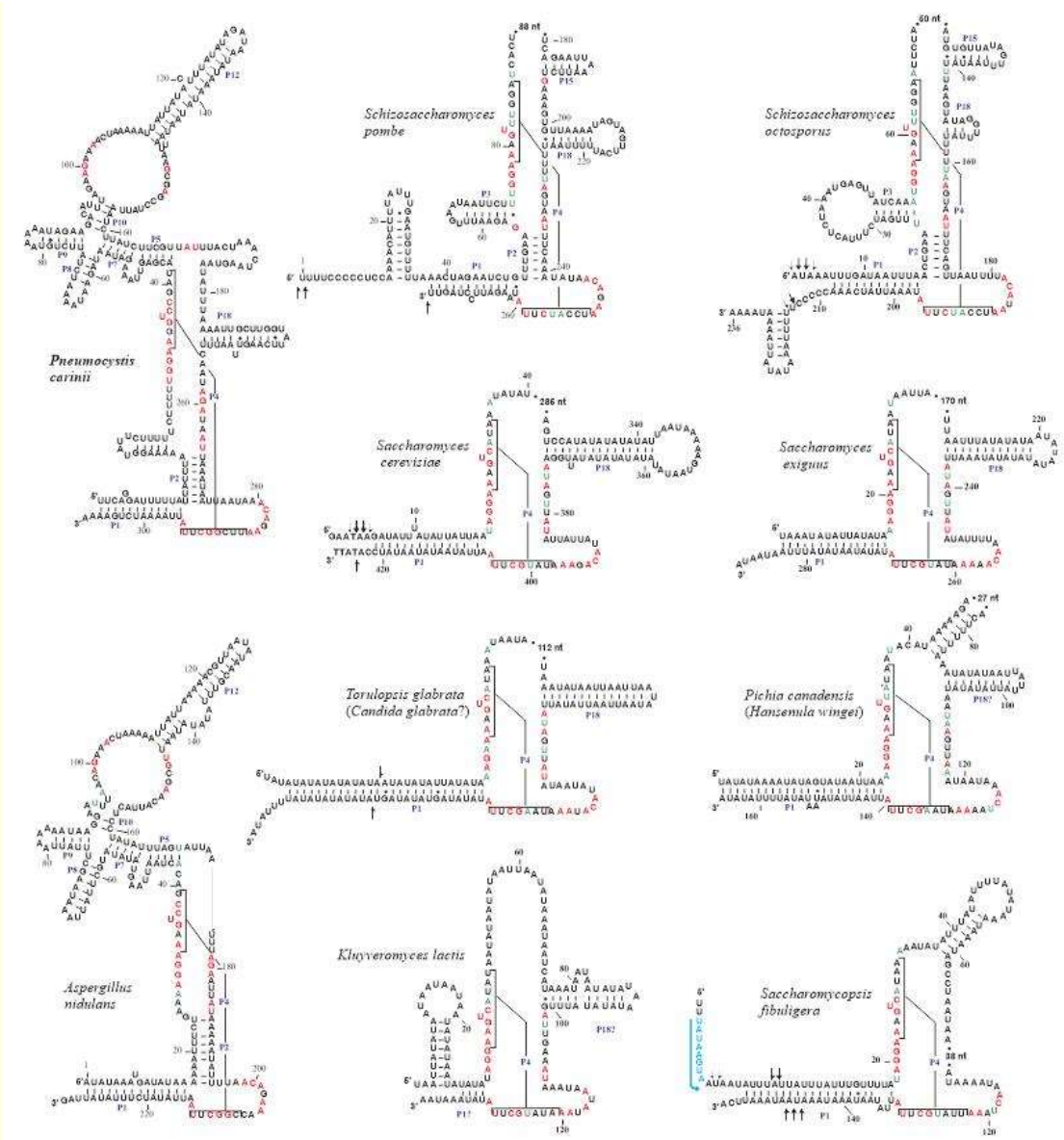


Basidiomycete Fungi



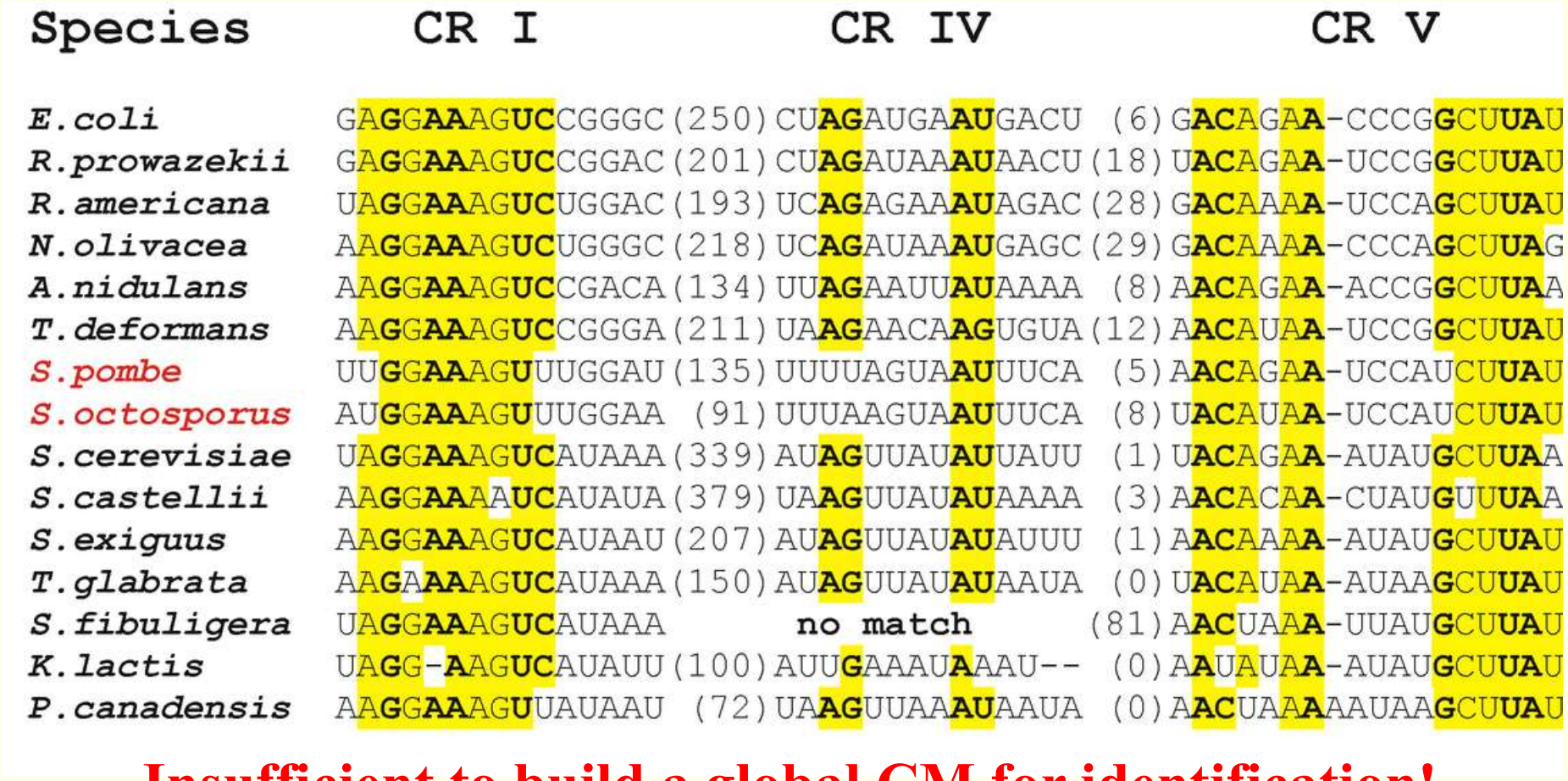






Extreme range of structural mutilations in ascomycetes

Minimum consensus structure of mt *rnpB* (P4)



Insufficient to build a global CM for identification!

How then can we identify *rnpB* genes?

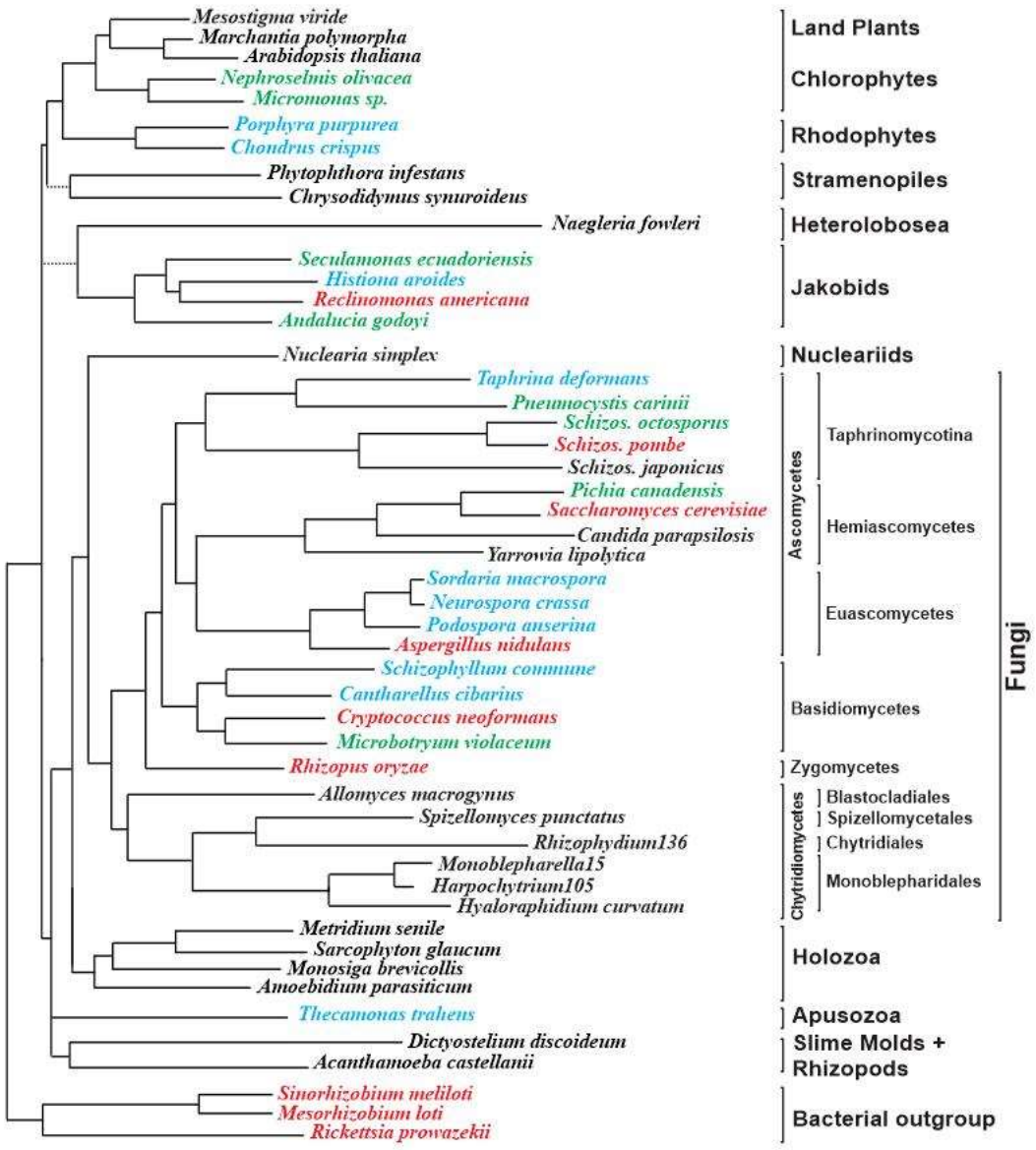
Alignment of basic known set, with P4 (no phylogenetic grouping, ~12)

CMs from large phylogenetic groups → iterations (~ 30)

CMs from increasingly smaller phylogenetic groups, → iterations (~80)

CMs from orphan groups, based on curator's seed → iterations (~100)
(using gene synteny as a starter and lots of hours)

Still inconsistent filling of hits within taxonomic groups



Evolution of mt *rnpB* identification

Red → Green → Blue

Continues to be spotty, similar to tmRNA identification

Current status of *rnpB* identification:

- 12 CMs across eukaryotes
- incomplete filling of hits within taxonomic groups

Open questions:

- Multiple independent losses of RNase P RNAs?
- Hidden genes with (i) highly derived structures, (ii) large sequence insertions (iii) genes in pieces, (iv) encoded in nucleus?
- The developer does not know what he is doing?

Directions:

Merge available CMs into one and add all 2D interactions?

Acknowledgements

Special thanks to Eric Nawrocki and Sean Eddy for continuing advice on how to get the best out of Infernal and hmmsearch tools; and to the advice of participants at the Benasque meetings

M.W.Gray
Jozef Nosek
Michaela Jakubkova
Marek Eliáš
Dieter Söll
Yiqiang Ling
Broňa Brejová
Tomáš Vinař
Ľubomír Tomáška
Peter Griač

Lise Forget
Matt Sarrasin
Elias Seif
Matus Valach
Nadia El-Mabrouk
Emmanuel Noutahi
Yannick Jacob
Rachid Daoud
Marie-Josee Laforest
Dennis Lavrov
Charles Bullerwell

Financial support

NSERC CRSNG

Fonds de recherche Québec