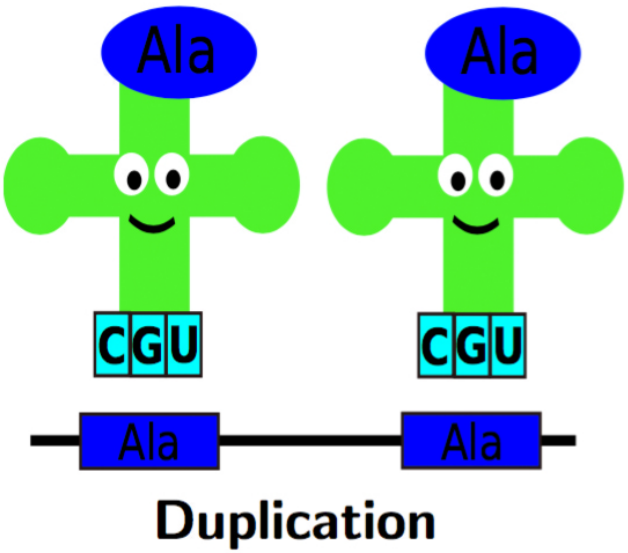
Orthologs, turn-over, and remolding of tRNAs in primates

Anne Hoffmann

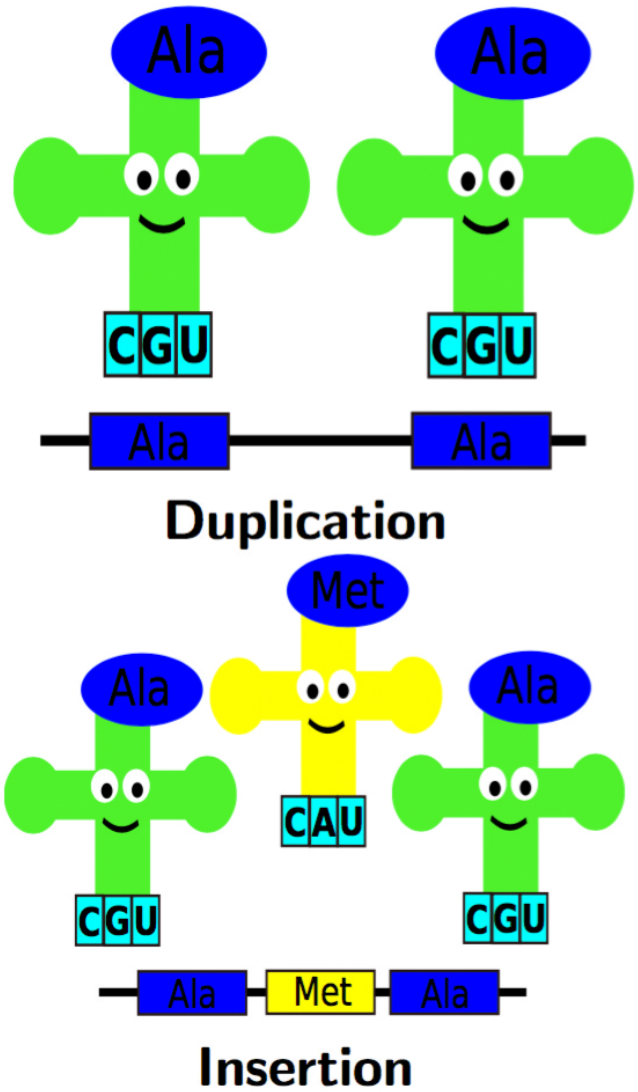
Department of Computer Science, and Interdisciplinary Center for Bioinformatics
Universität Leipzig

19th February 2016

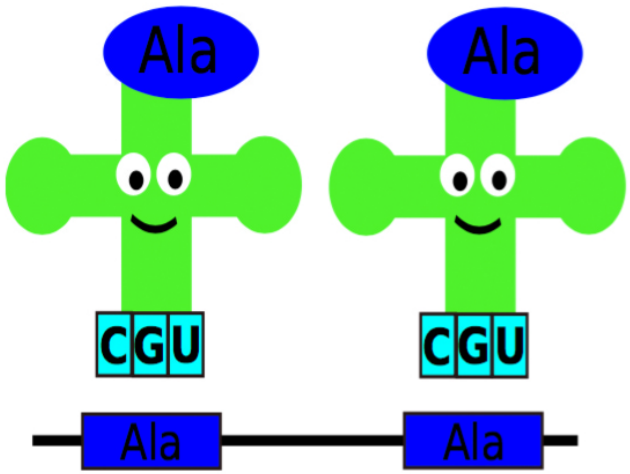
Evolution of eukaryotic tRNAs I



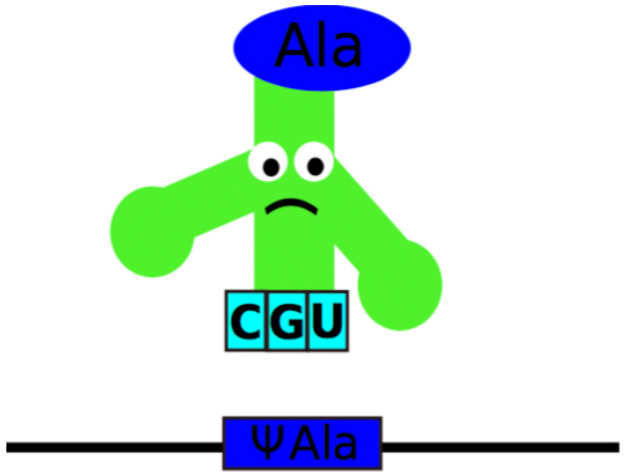
Evolution of eukaryotic tRNAs I



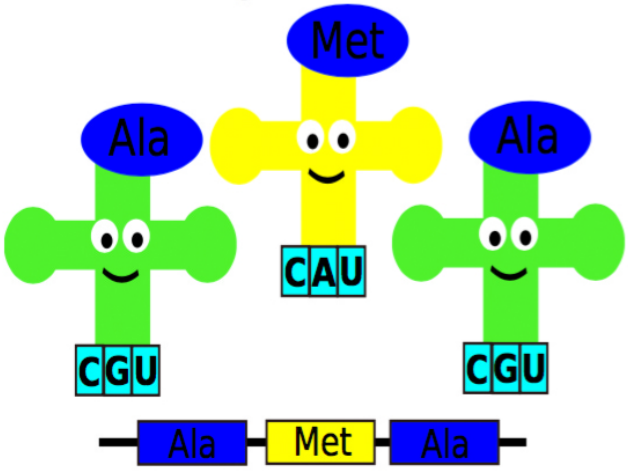
Evolution of eukaryotic tRNAs I



Duplication

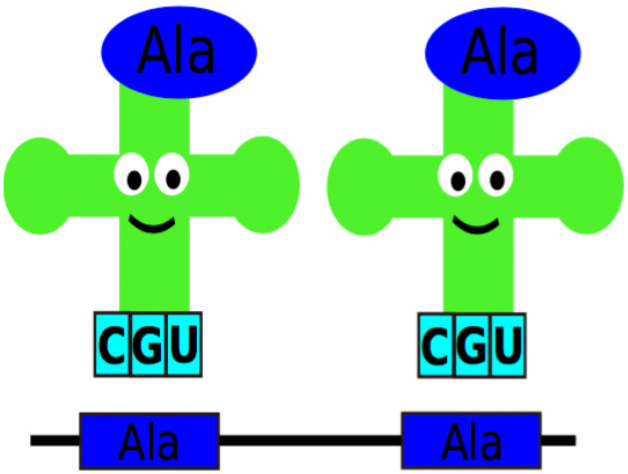


Pseudogenization

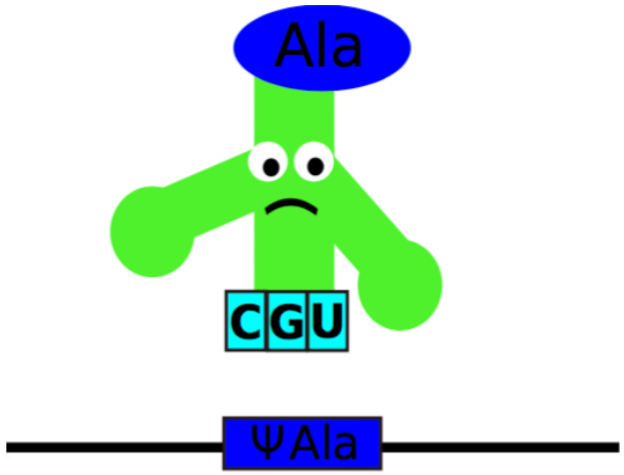


Insertion

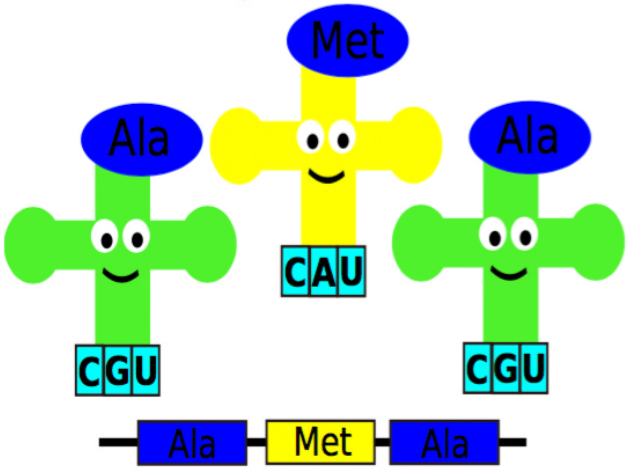
Evolution of eukaryotic tRNAs I



Duplication



Pseudogenization



Insertion

Loss

Evolution of eukaryotic tRNAs II

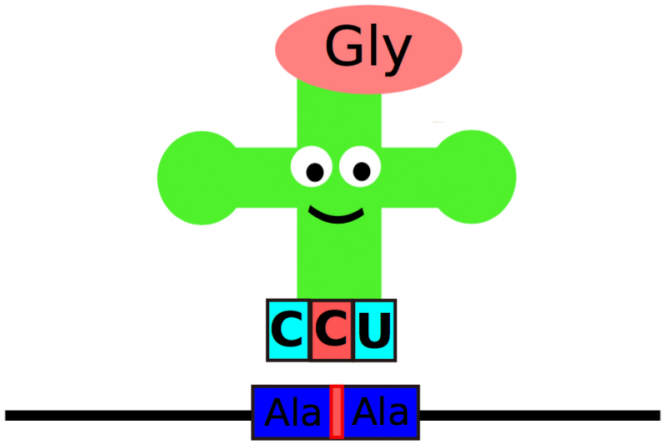
Remolding

Mutation inside the anticodon such that the tRNA identity changes

Evolution of eukaryotic tRNAs II

Remolding

Mutation inside the anticodon such that the tRNA identity changes

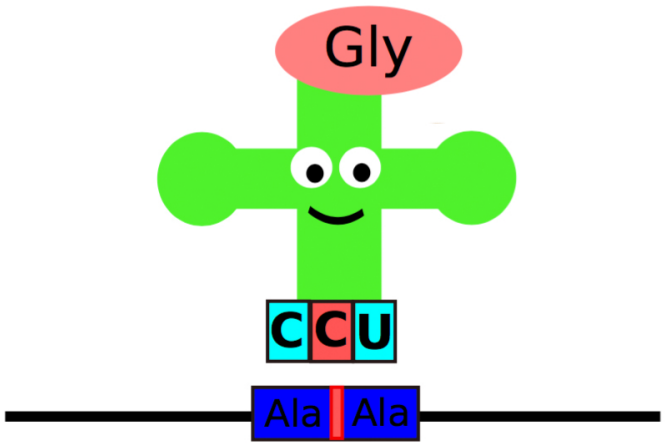


Alloacceptor
anticodon change

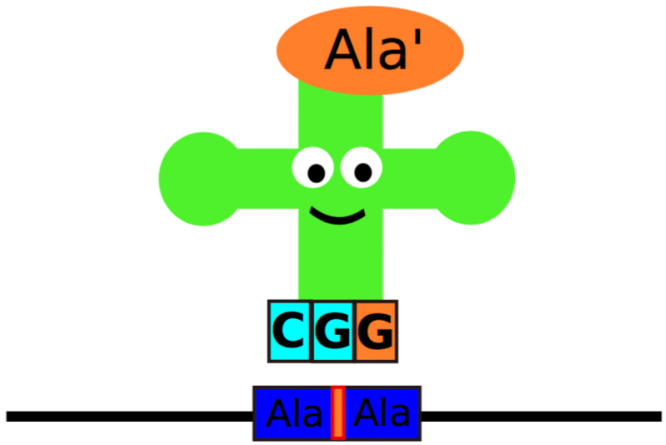
Evolution of eukaryotic tRNAs II

Remolding

Mutation inside the anticodon such that the tRNA identity changes



Alloacceptor
anticodon change



Isoacceptor
anticodon change

Motivation

- tRNA genes act as repetitive elements
- loss of tRNAs leads to large differences in the tRNA complements
- reconstruction of the history of a gene familiy is not possible with common methods
  - reciprocal best match
  - phylogenetic tree-based

→ **Aim:** create a framework to distinguish paralogs and orthologs in gene families of related species

Step-wise orthology identification

1. Annotation of tRNAs

$\rightarrow$ tRNA gene detection in genome sequences applying `tRNAscan-SE`

- It is able to detect ~99% of eukaryotic nuclear tRNA genes
- 5 primates and human
- 12 drosophilids

Step-wise orthology identification

2. Determine anchors of tRNAs

I) MSA block based approach

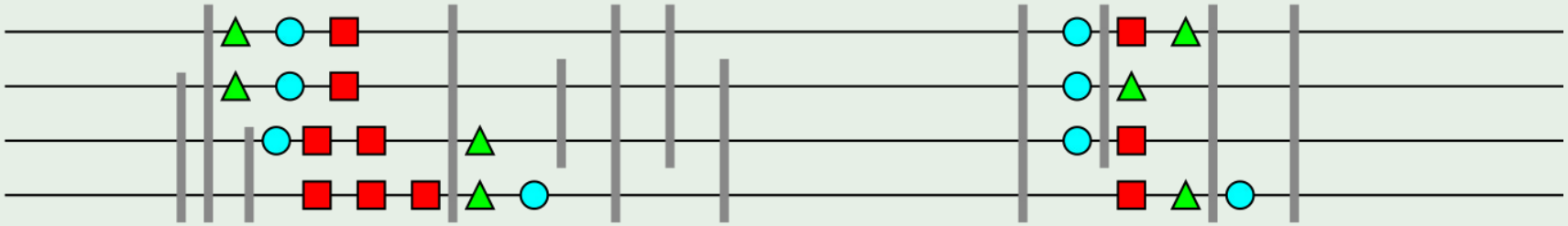
→ Sequence-unique MAF blocks of `multiz` alignments

Step-wise orthology identification

2. Determine anchors of tRNAs

I) MSA block based approach

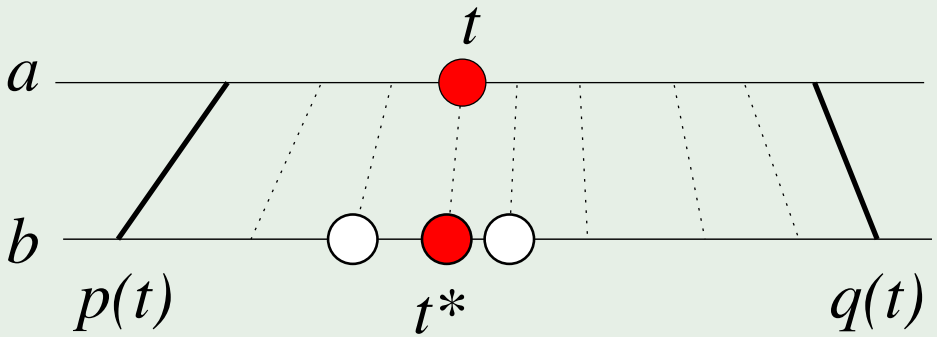
→ Sequence-unique MAF blocks of `multiz` alignments



II) Orthologous proteins approach

→ 1 : 1 set of orthologs

→ Linear coordinate interpolation



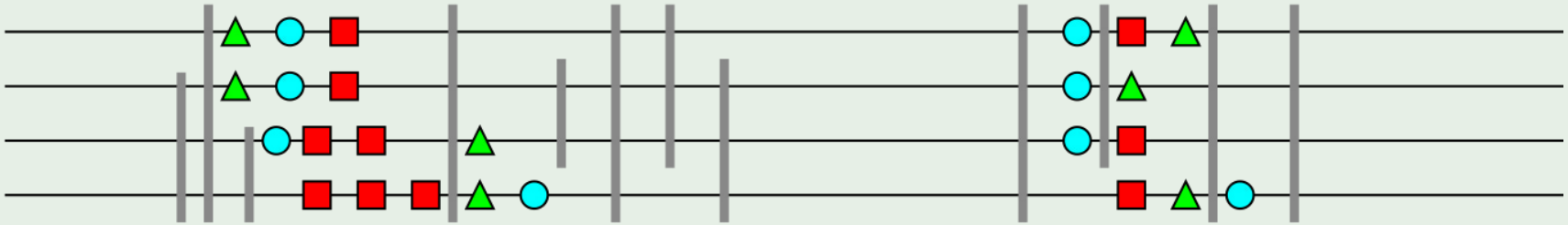
$$t^* = p^b + \frac{q^b - p^b + 1}{q^a - p^a + 1} t^a$$

Step-wise orthology identification

2. Determine anchors of tRNAs

I) MSA block based approach

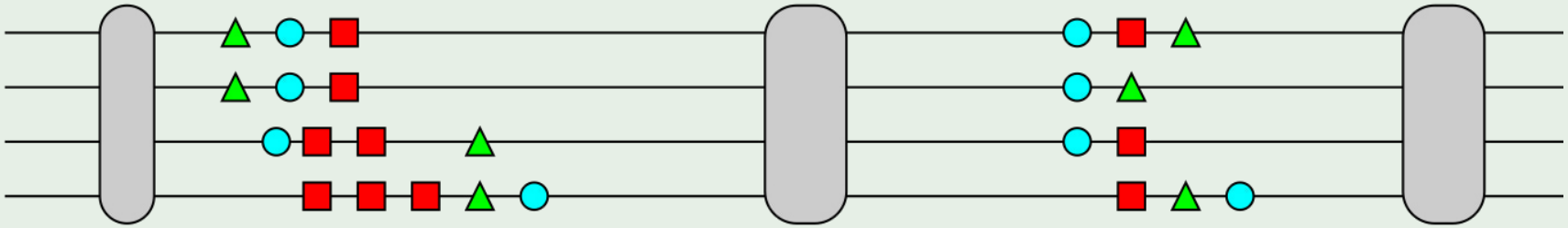
→ Sequence-unique MAF blocks of `multiz` alignments



II) Orthologous proteins approach

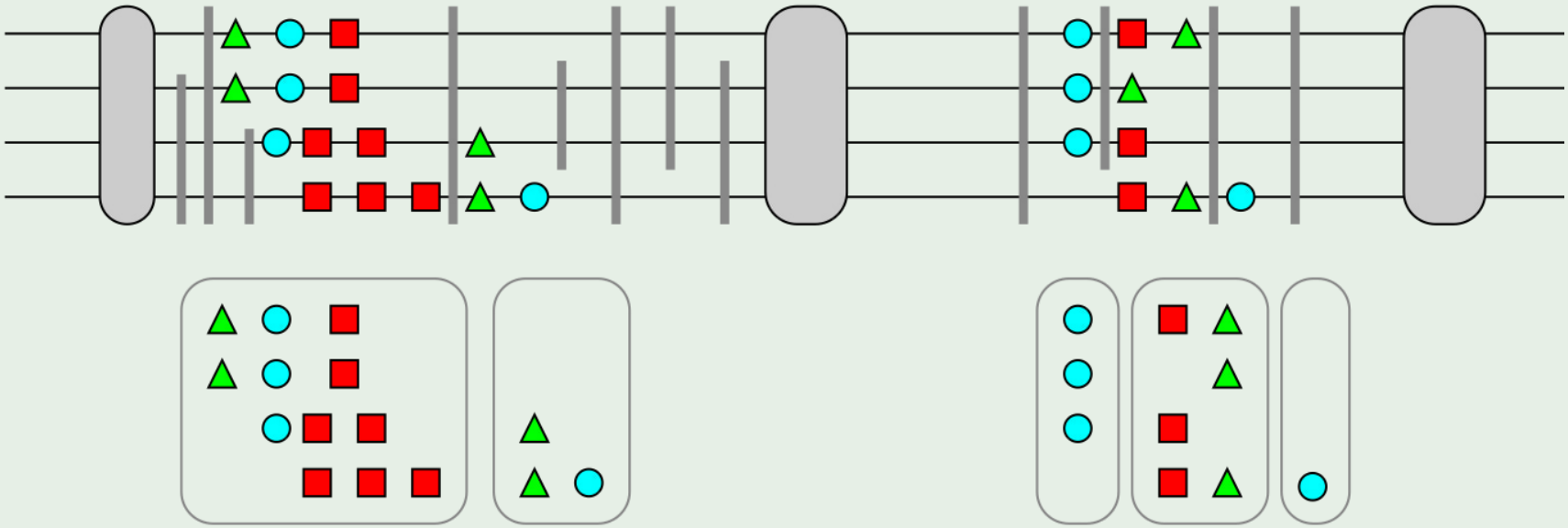
→ 1 : 1 set of orthologs

→ Linear coordinate interpolation



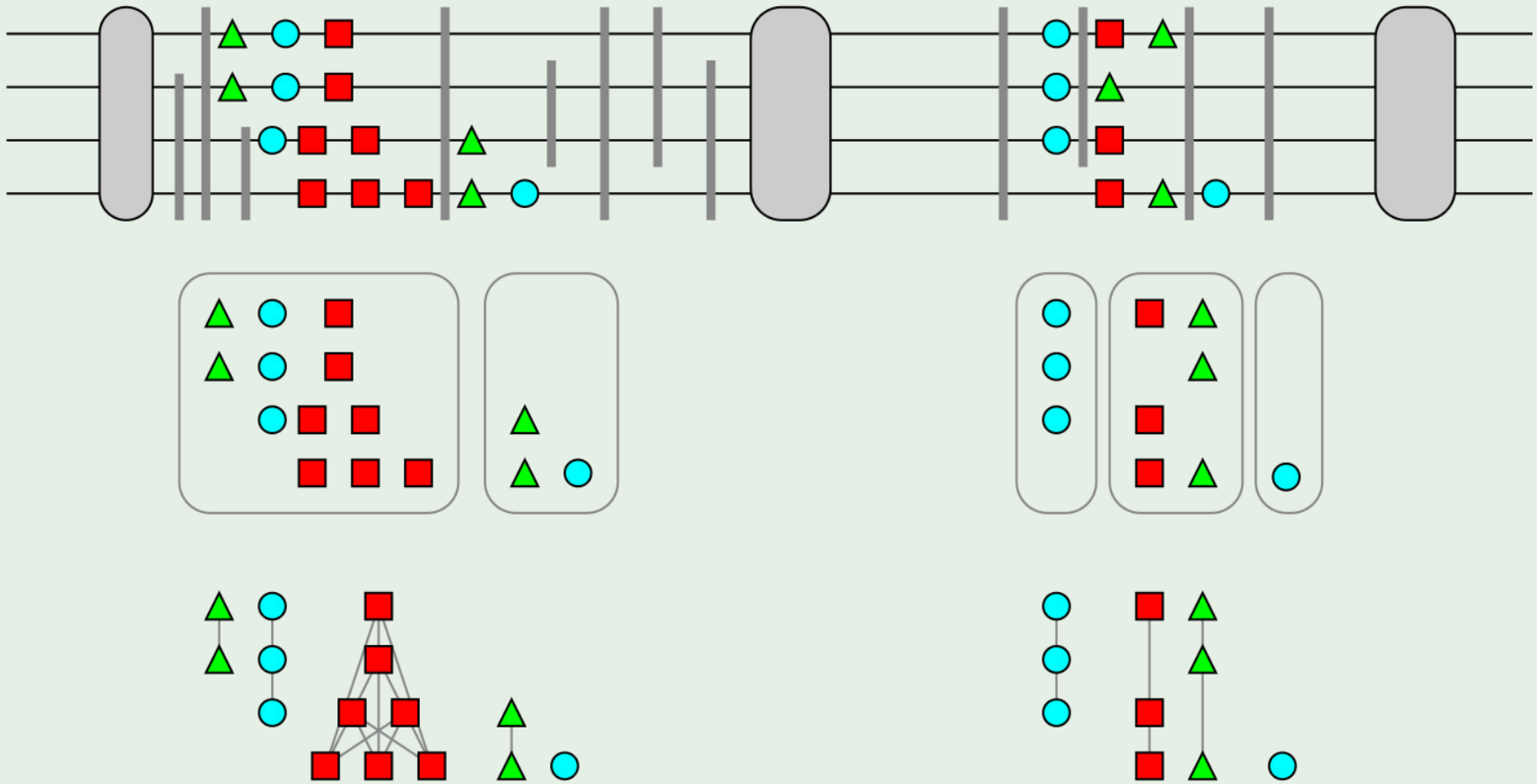
Step-wise orthology identification

3. Clustering of tRNAs based on anchors



Step-wise orthology identification

4. Orthology reconstruction



Step-wise orthology identification

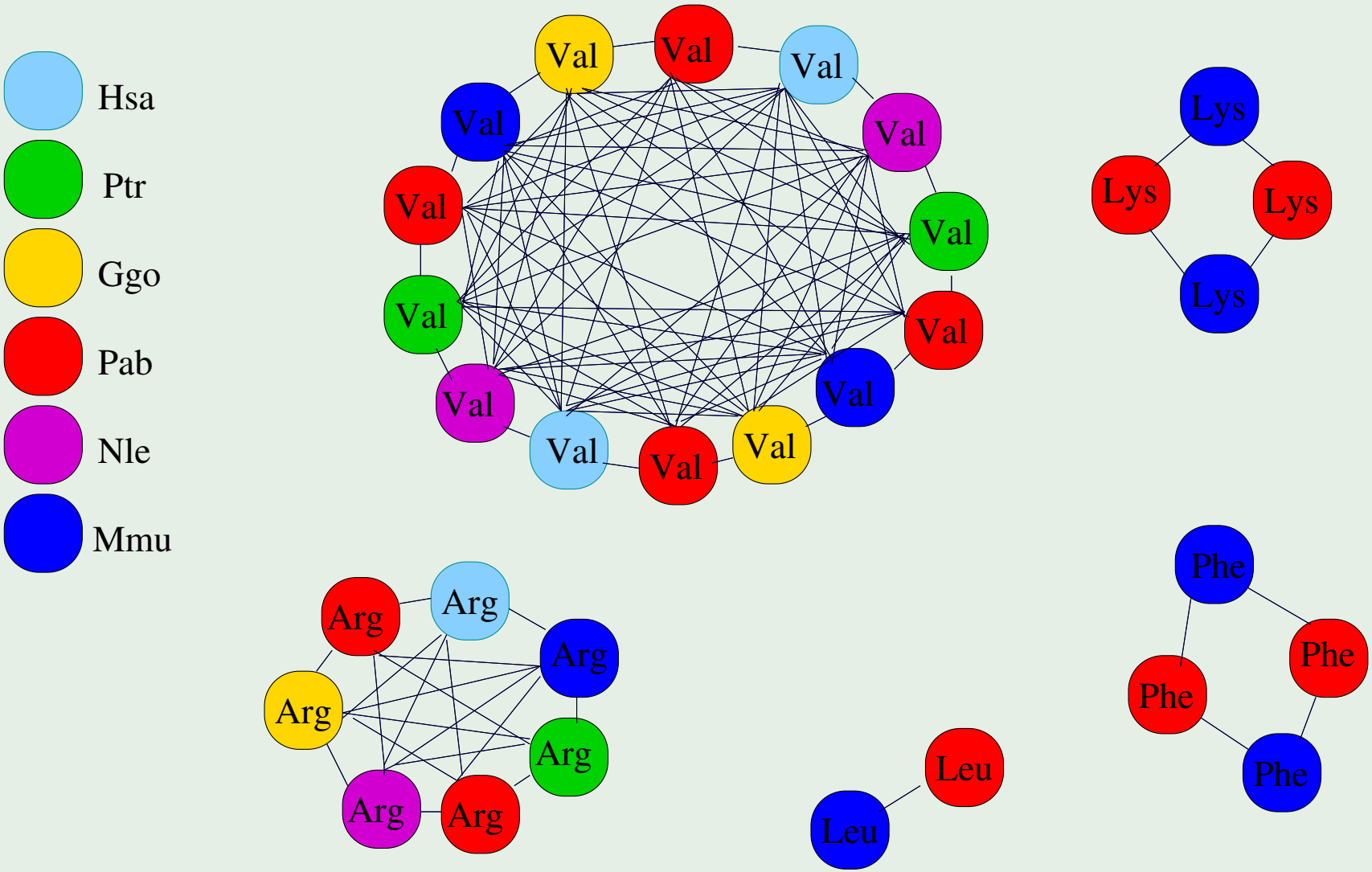
4. Orthology reconstruction

→ Generalized list alignments

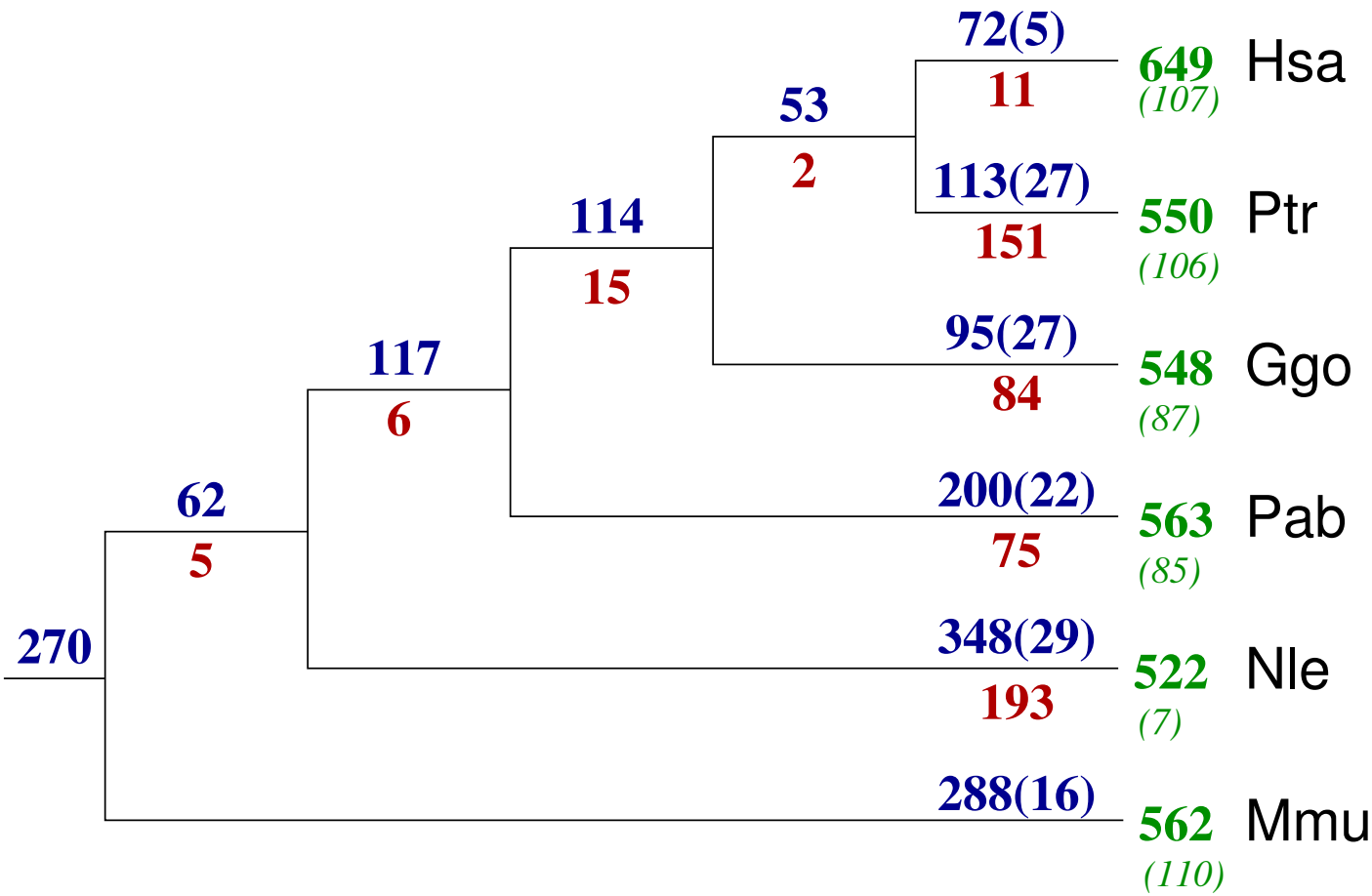
```
Ptr  - - - - - V ~ V ~ R ~
Hsa  - - - - - V ~ V ~ R ~
Ggo  - - - - - V ~ V ~ R ~
Pab  F F K K L V V V V R R
Nle  - - - - - V ~ V ~ R ~
Mmu  F F K K L V ~ V ~ R ~
```

Step-wise orthology identification

4. Orthology reconstruction



Gain, loss and duplications of tRNAs in primates (MSA)

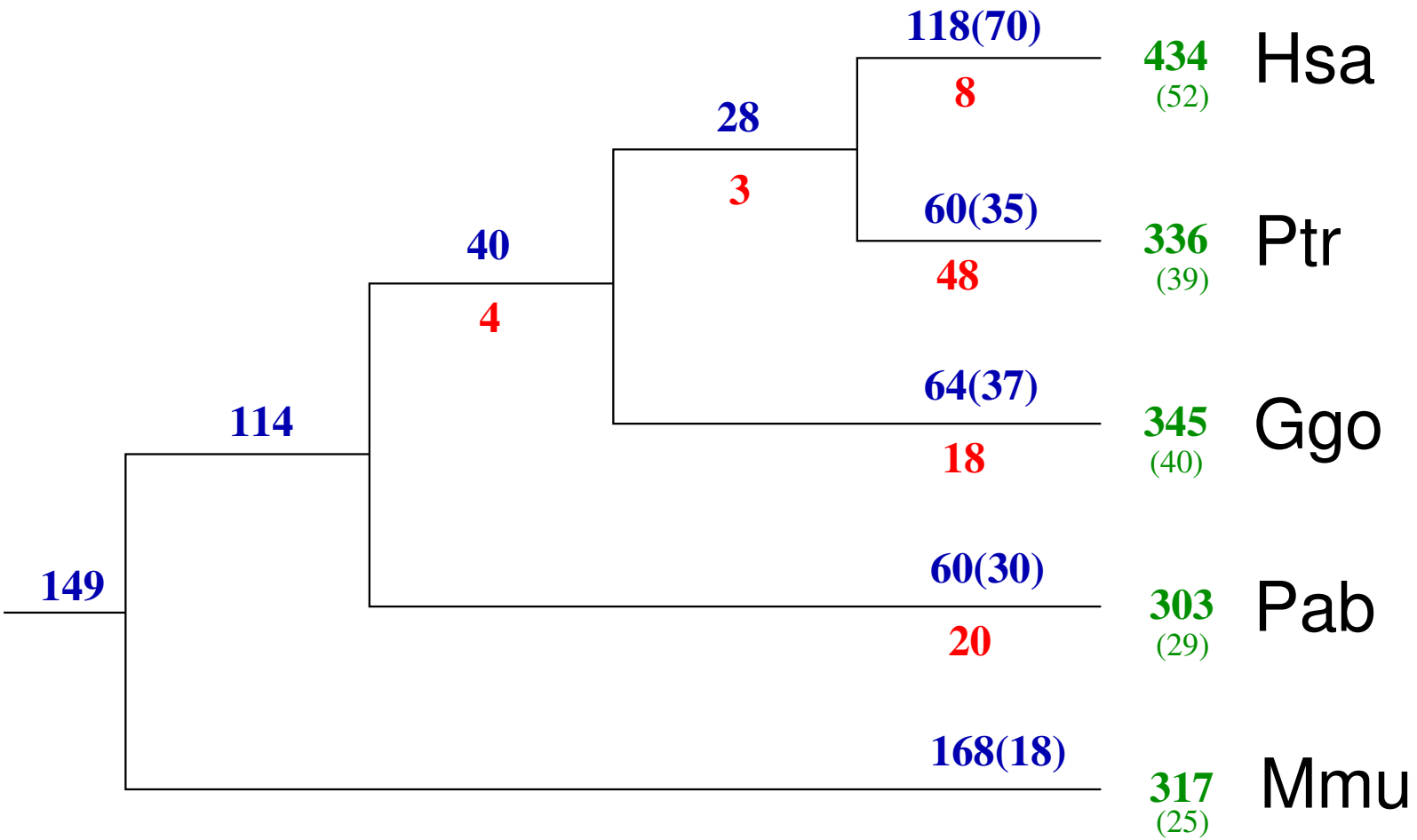


- 482 singeltons
- 64 tRNAs of a single species

$\rightarrow \frac{2}{3}$ syntenically conserved tRNAs

- 327 cliques, without duplications
- 206 including duplications

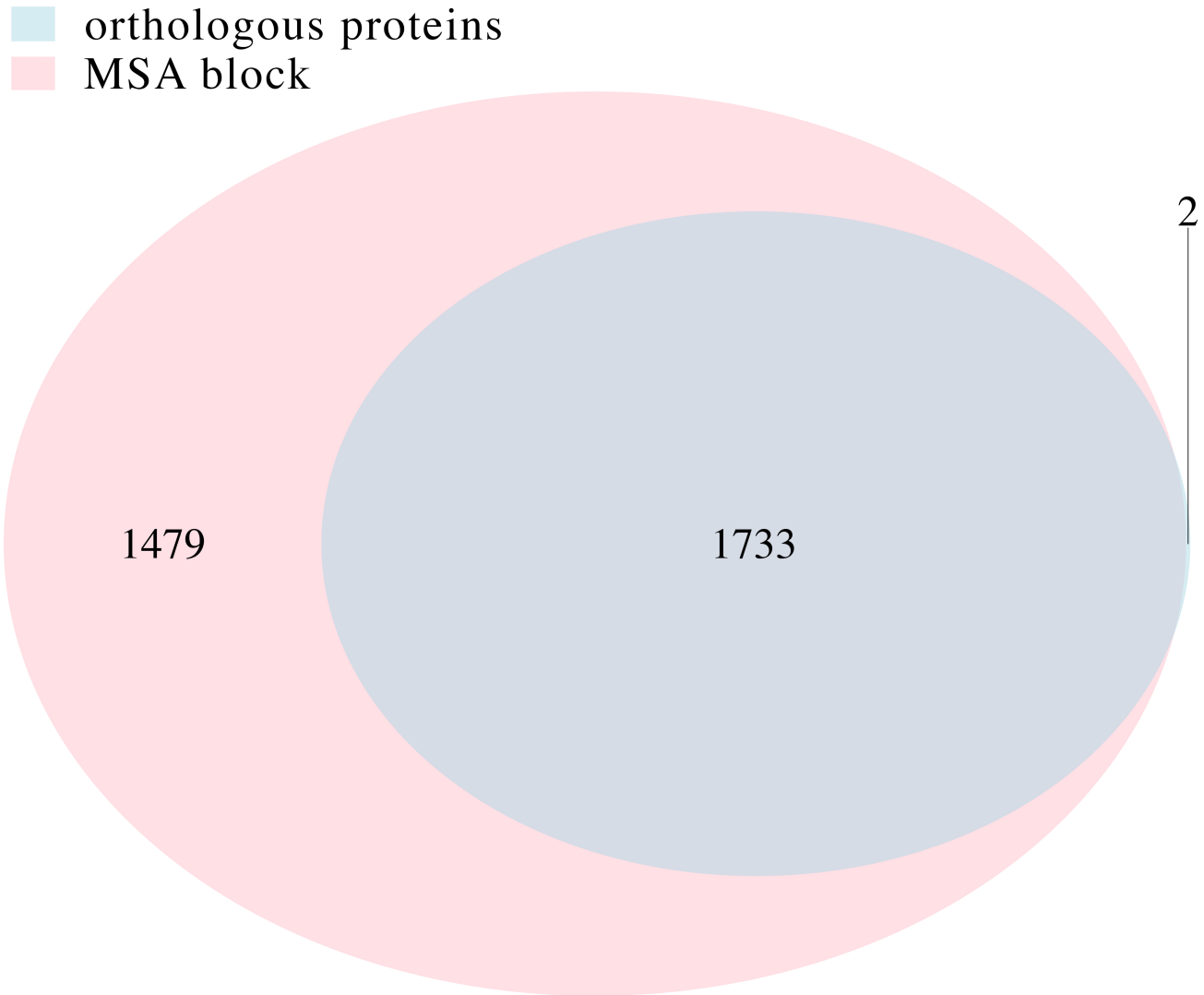
Gain, loss and duplications of tRNAs in primates (OP)



- 231 genomic clusters of tRNAs, of which 166 are nontrivial

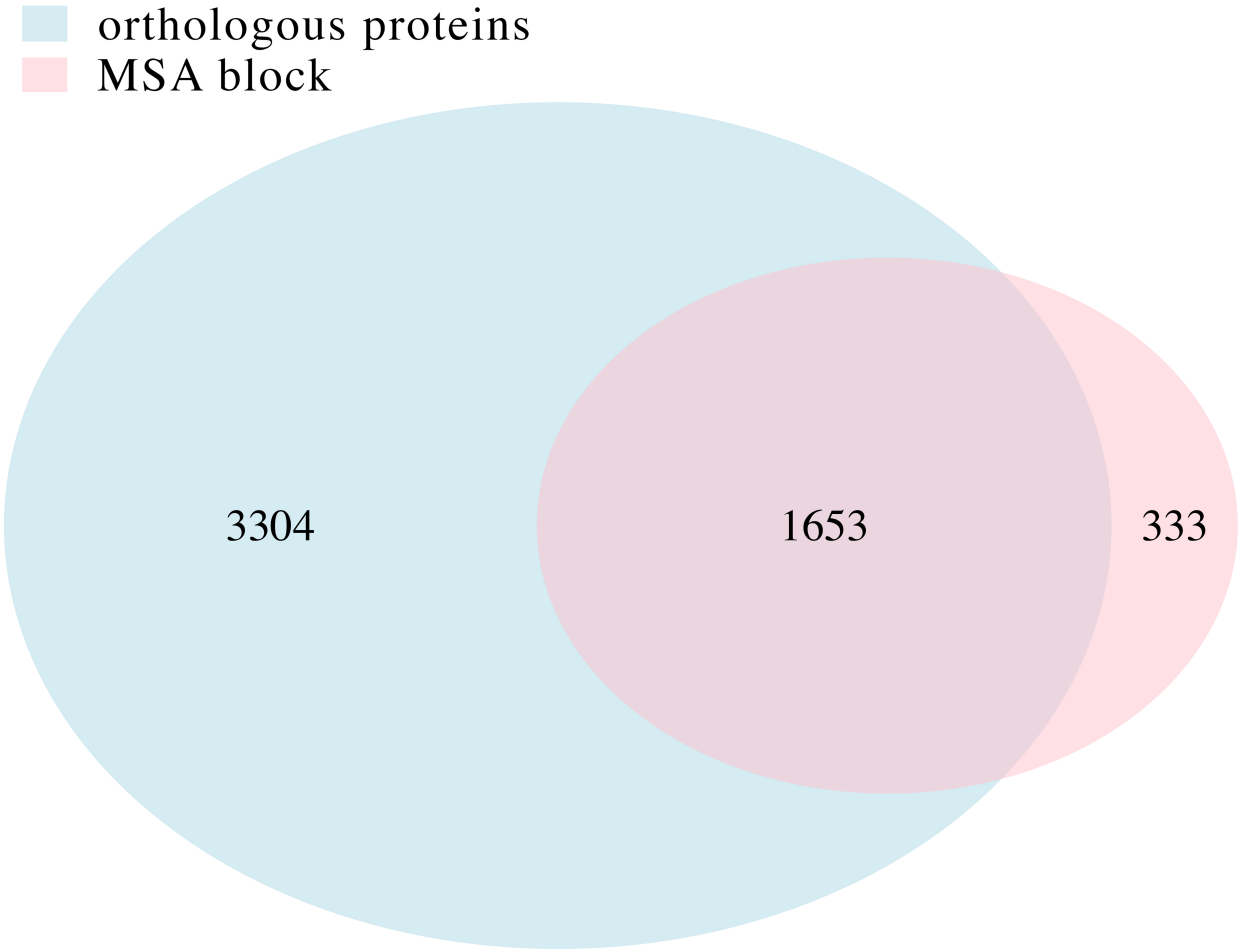
→ Less numbers of evolutionary events detectable

Comparison of the different methods



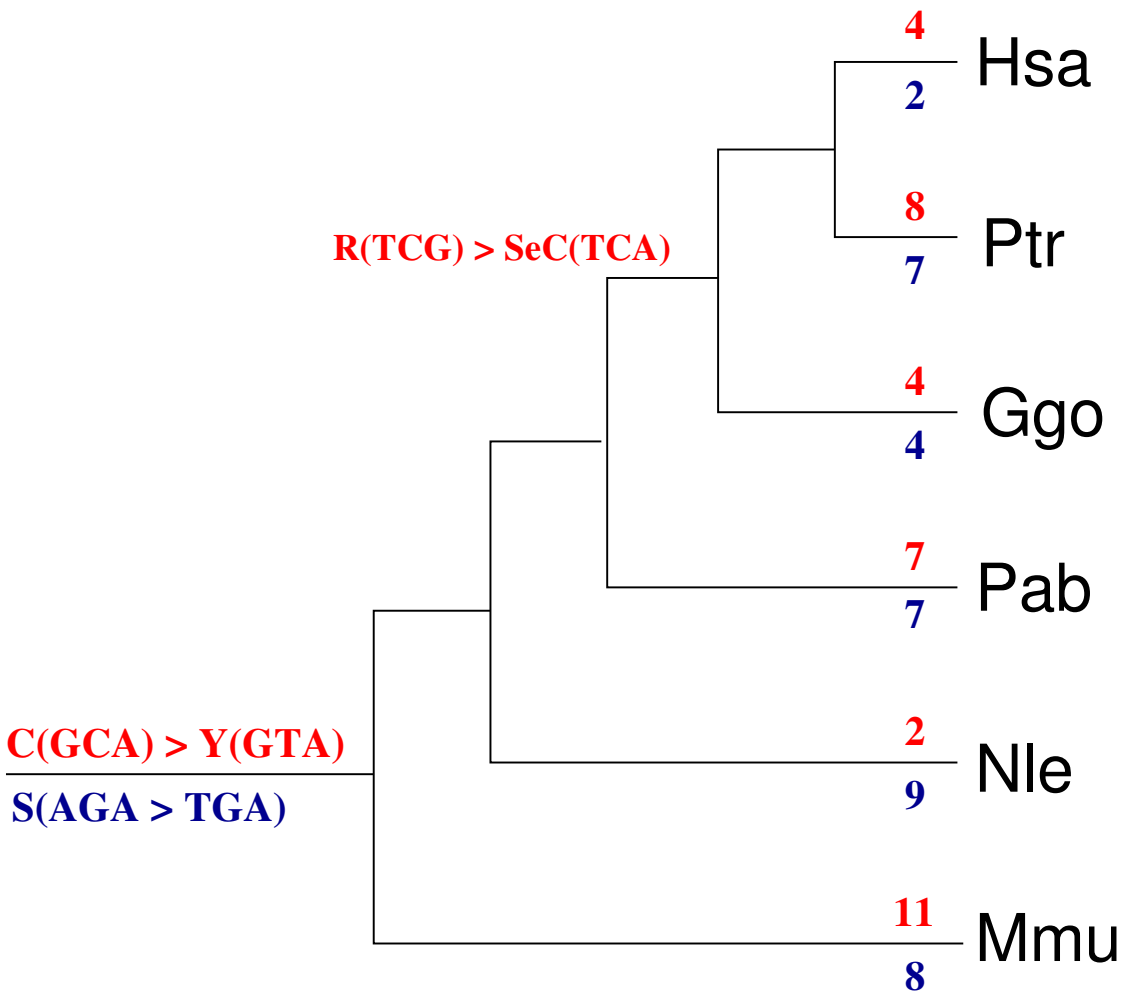
- Number of clustered tRNAs

Comparison of the different methods



- Number of ortholog edges
- 83% of the edges of MSA are recovered with PO

Remolding events in primates (MSA)



- Alloacceptor (red), isoacceptor (blue)
- Most remolding events constitute pseudogenes

Summary

- Created significant methods to detect phylogenetic events on tRNAs
- MSA is sensitive for predictions of remolding events
- Remolding events mostly occur in pseudogenes
- A large number of gain and loss of tRNAs are species specific

Acknowledgements

Many thanks to:

- Peter F. Stadler
- Clara I. Bermúdez Santana
- Maribel Hernández Rosales
- Sarah J. Berkemer
- Cristian A. Velandia-Huerto
- Nancy Retzlaff
- Liliana Romero Marroquin

Thank you for your attention!