

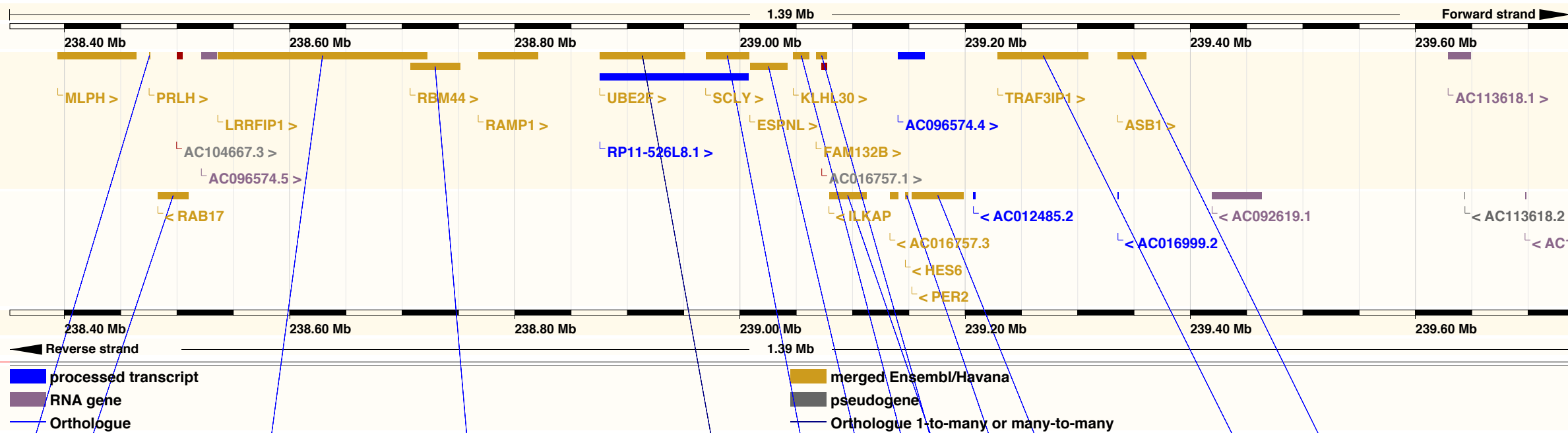
# “same” gene in different species

## Human chr2

Ensembl/Havana ...

Hsap Chr. 2

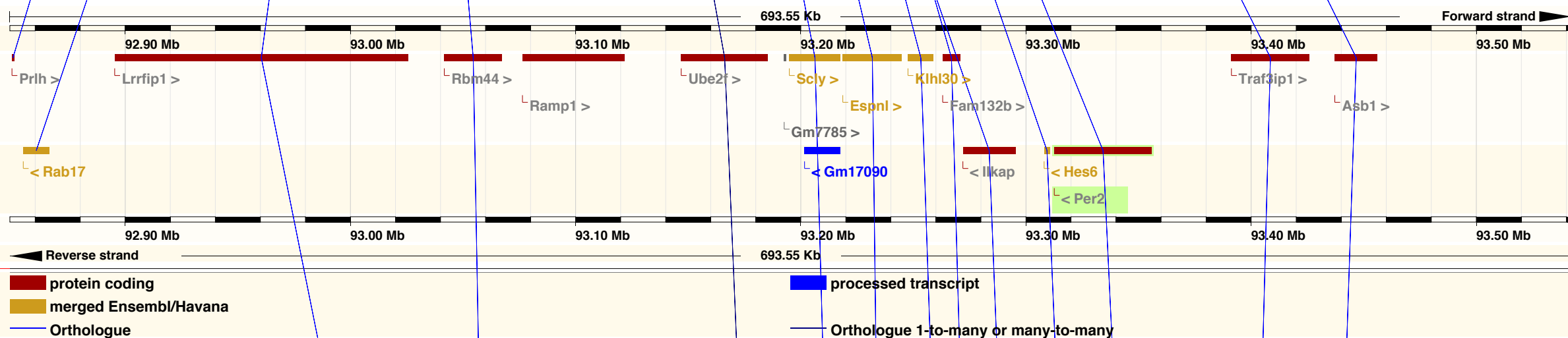
Gene Legend



## Mouse chr1

Ensembl/Havana ...

Gene Legend

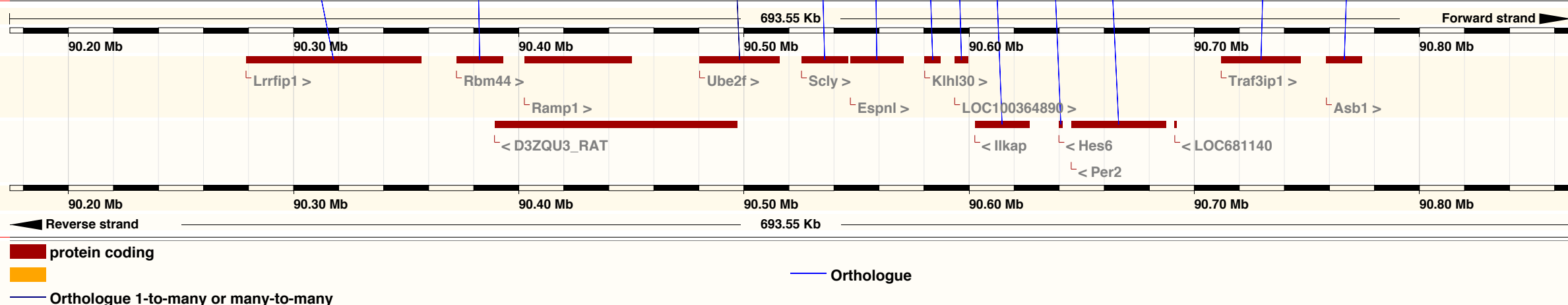


## Rat chr9

Ensembl

Rnor Chr. 9

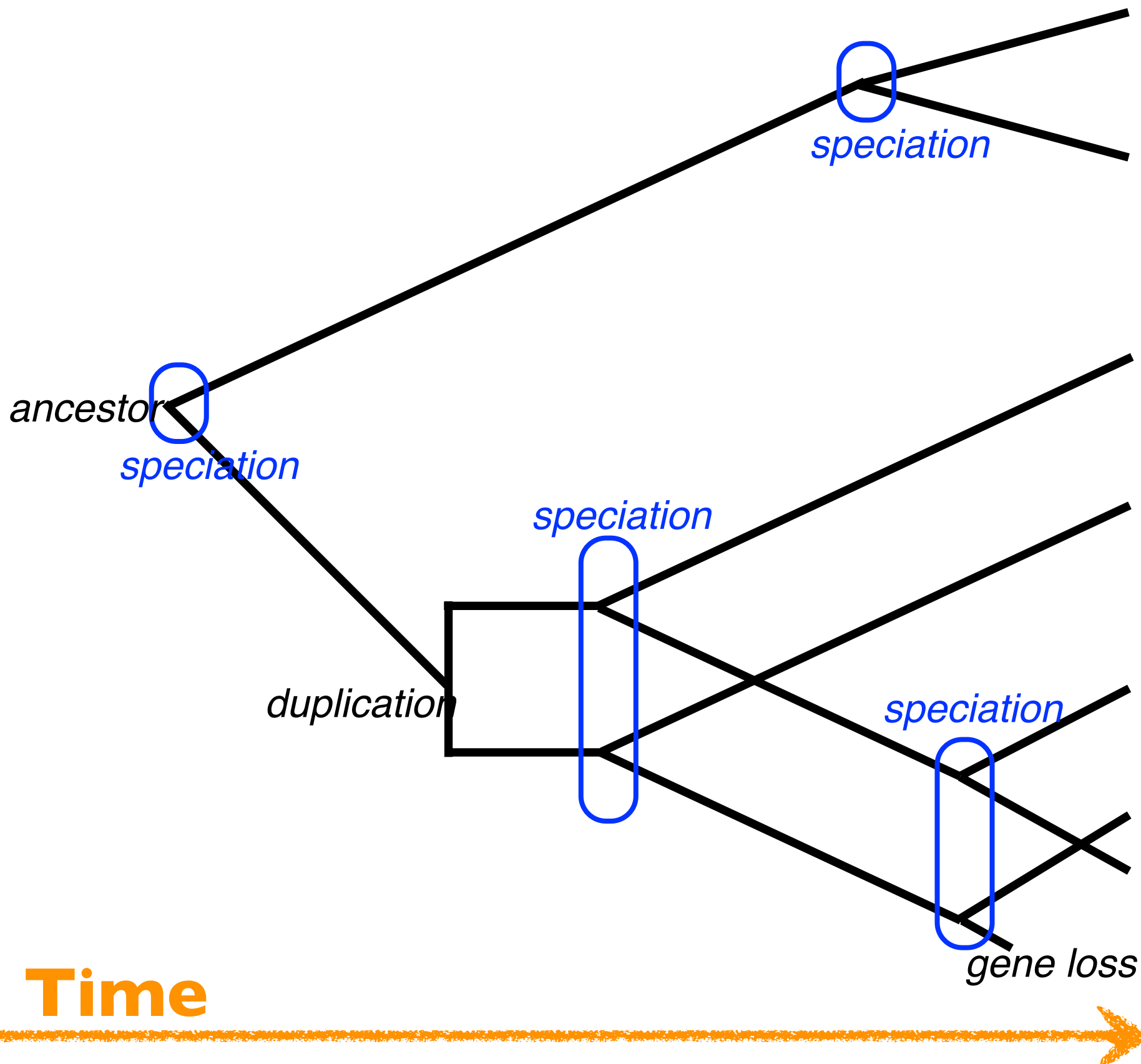
Gene Legend



# Homology

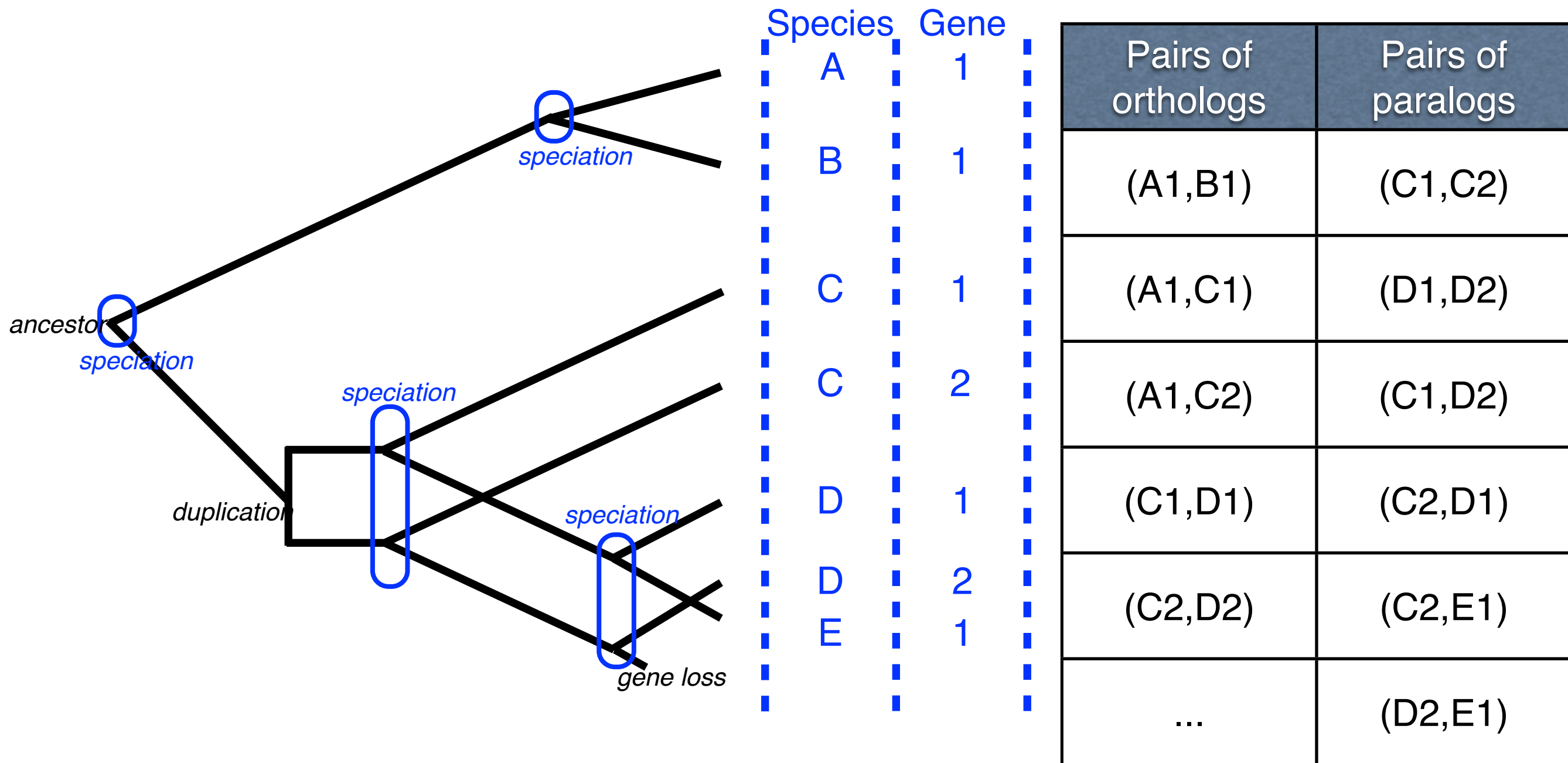
- Gene A in species S and gene B in species T are:
  - Homologous if they share a common ancestor
- Two homologous genes are:
  - Orthologous if they diverged at a speciation event
  - Paralogous if they diverged at a duplication event
  - Xenologous if one of them went through a lateral transfer

# Evolution of a gene



Species	Gene
A	1
B	1
C	1
C	2
D	1
D	2
E	1

# Evolution of a gene



**Time**



# Operational definitions

- Orthologs are more similar at nucleotide level than paralogs, because the paralog started diverging before speciation
- We ignore the existence of xenologs (lateral gene transfer) because otherwise evolution is not a tree anymore

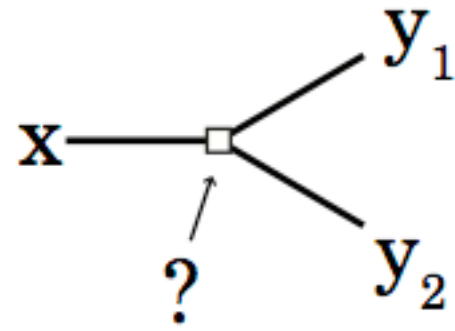
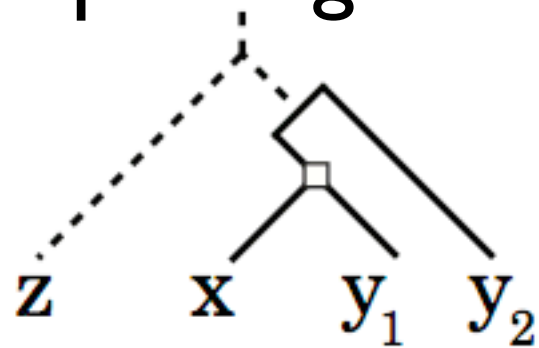
# Strategy

- Collect all protein sequences for each genome (translated from genes)
- Perform Smith-Waterman alignment of every pair of proteins from all genomes
- *Reciprocal best hits*: pairs (A,B) of genes from genomes (X,Y) such that A is the gene in X most similar to B, and B is the gene in Y most similar to A
- Look for a third genome as a proof of orthology
- Make clusters of triangular relations (orthologous groups)

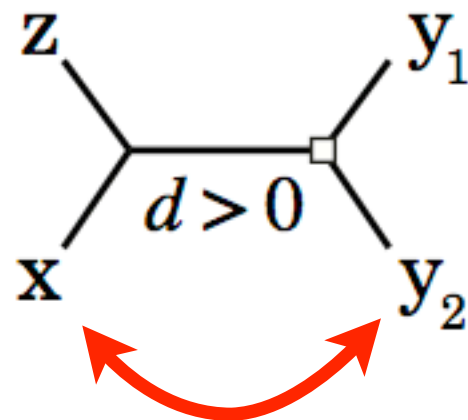
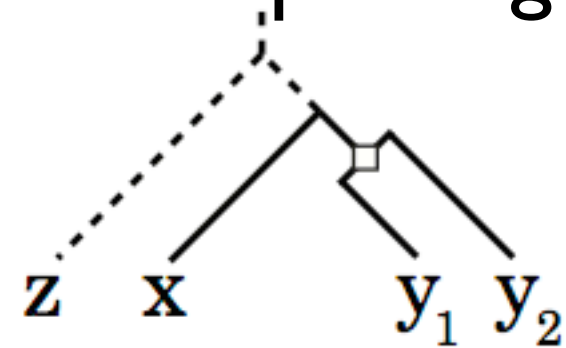
<http://omabrowser.org/>

# In/Out-paralogs

Out-paralogs

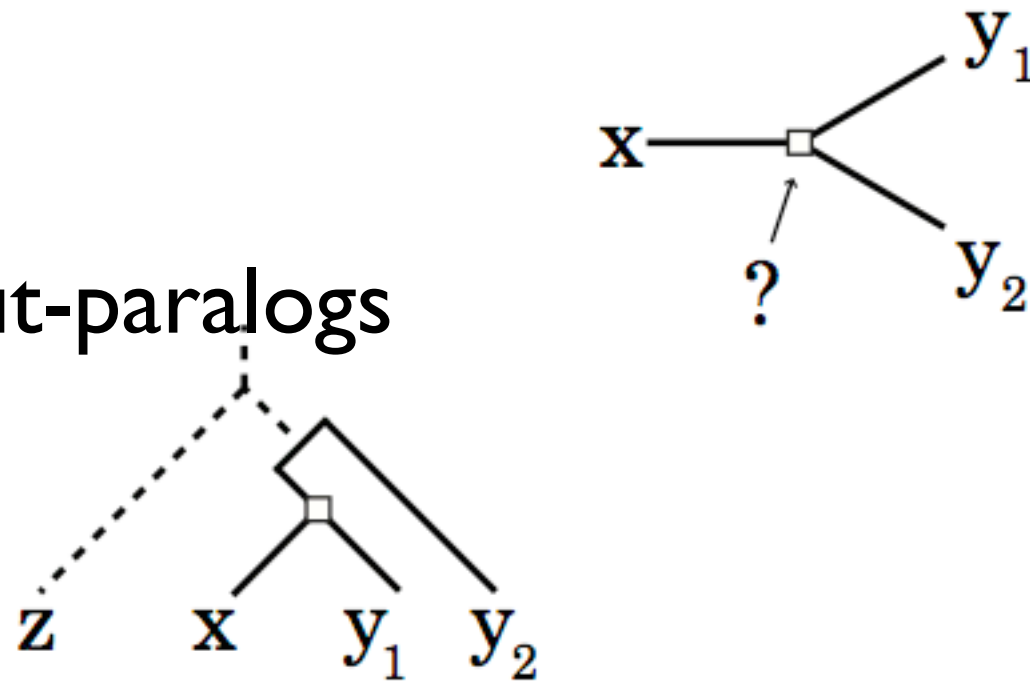


In-paralogs

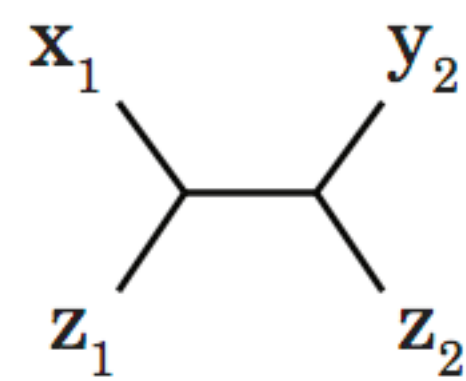
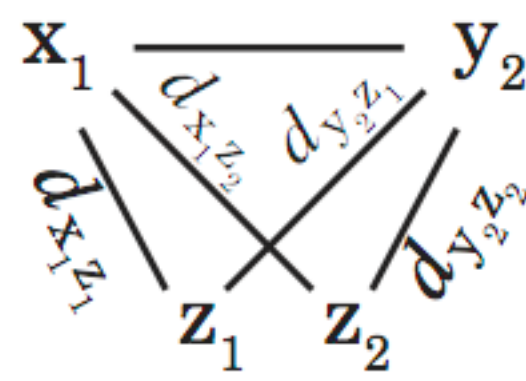
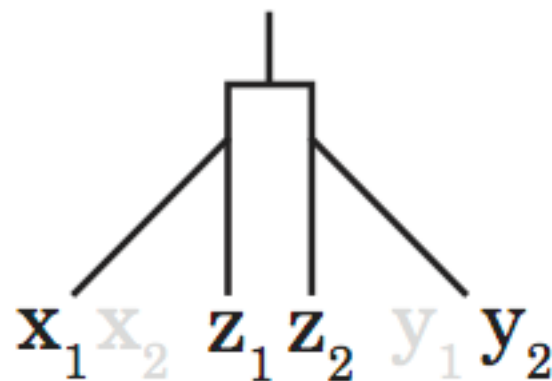
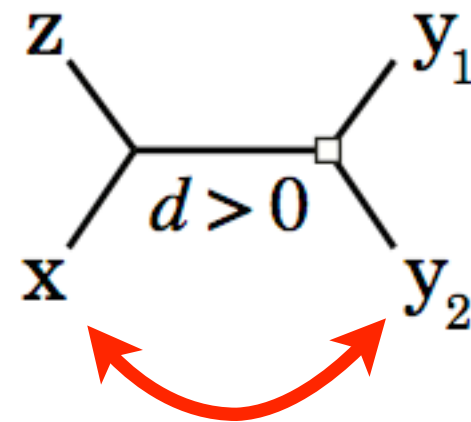
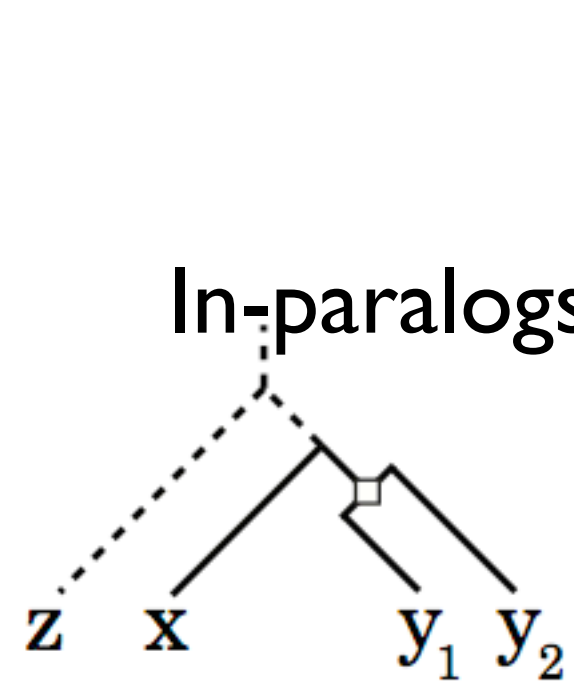


# In/Out-paralogs

Out-paralogs

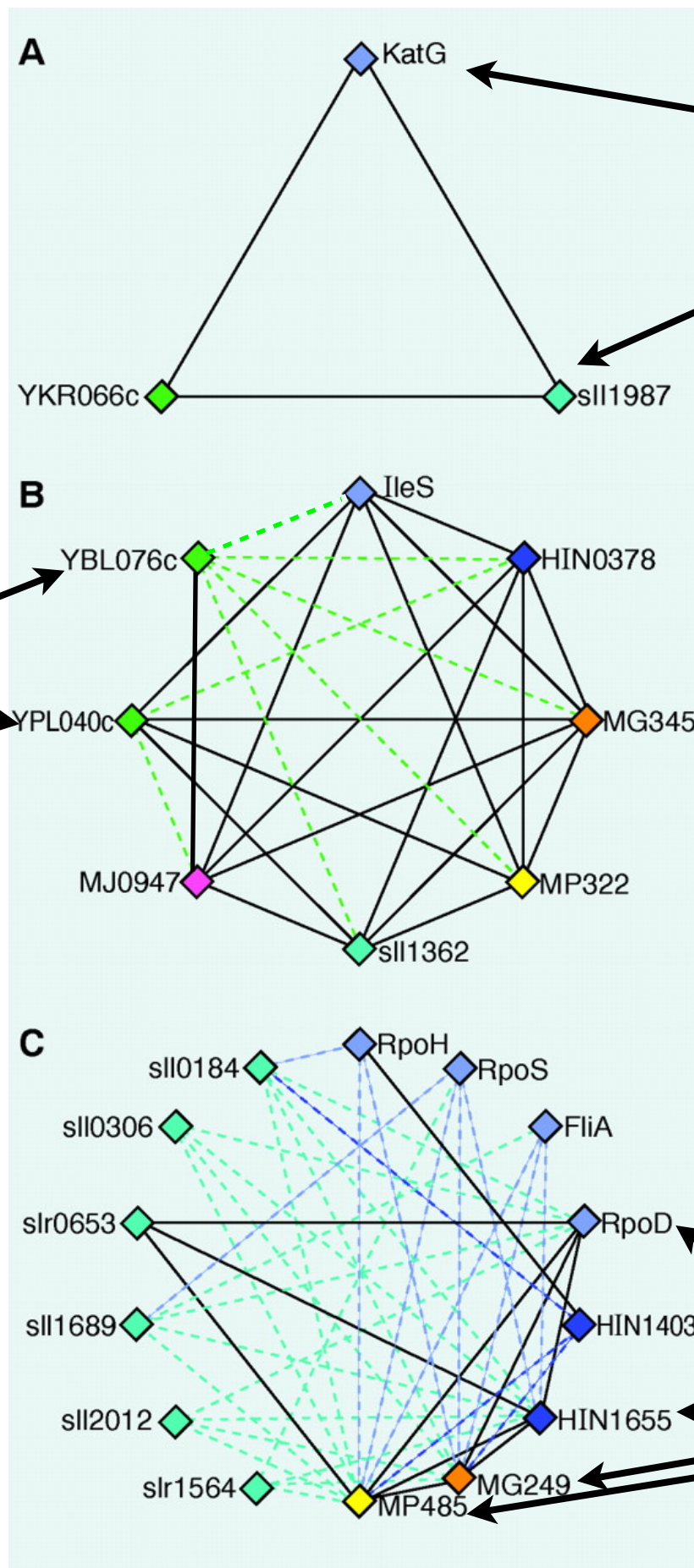


In-paralogs





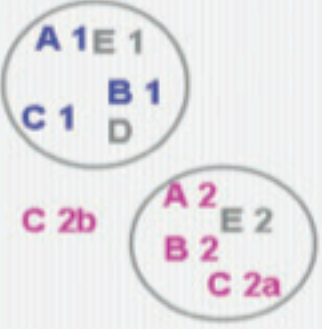
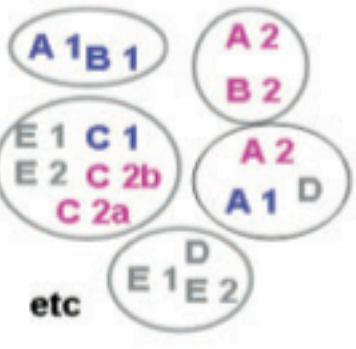
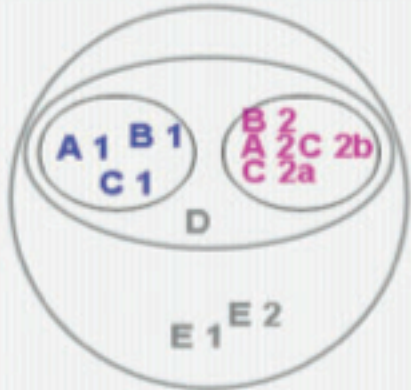
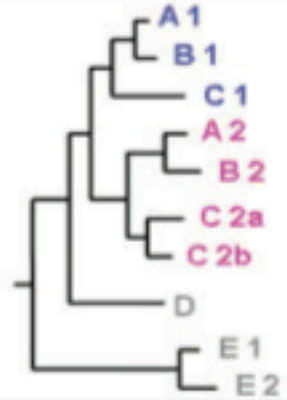
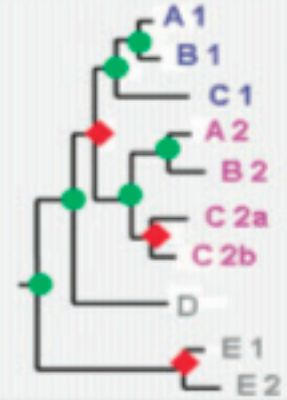
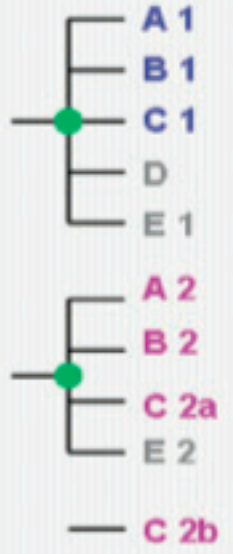
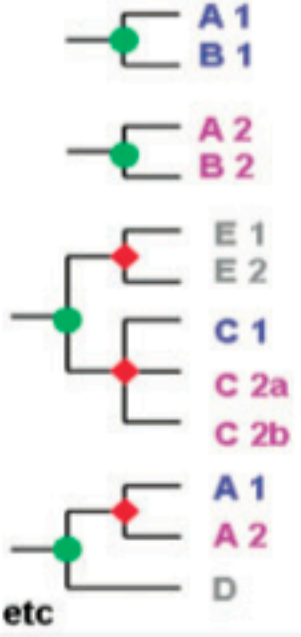
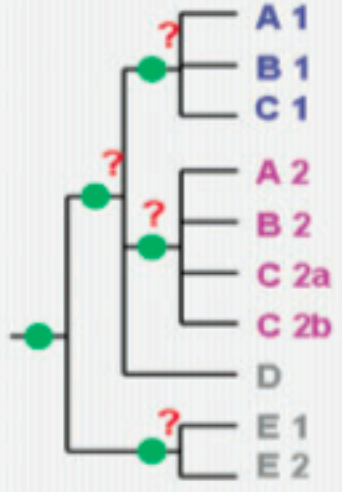
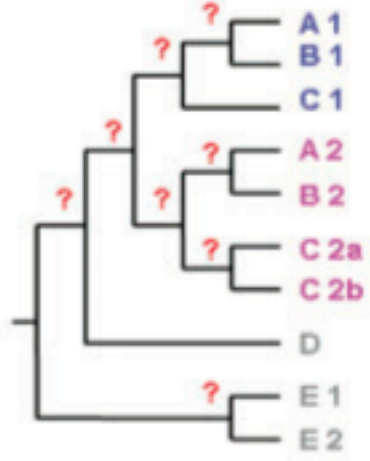
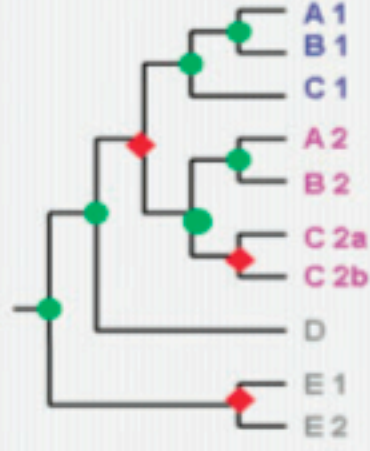
Yeast paralogs



RBH

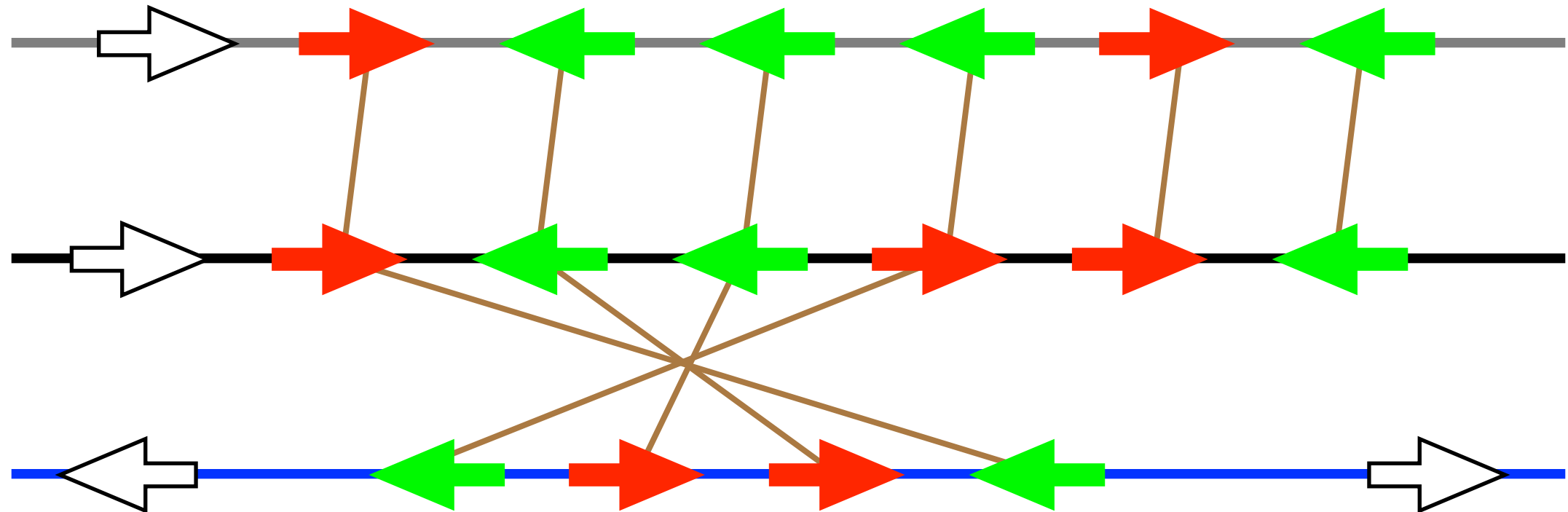
4-clique:

4 co-orthologs

	Pure ortholog groups	Pairwise groups	Hierarchical groups	Plain trees	Reconciled tree
Concept	All gene pairs of an ortholog group are orthologs.	All genes of a pairwise group are derived from a common gene that diverged subsequent to the given speciation event.	Groups of orthologs and inparalogs with respect to defined speciation events.	Unlabeled gene trees.	Gene trees with internal nodes labeled
Structure					
Prediction	Orthologs: not exhaustive Paralogs: none Hierarchy: none	Orthologs: exhaustive Paralogs: not exhaustive Hierarchy: 1-2 levels	Orthologs and paralogs: completeness depends on the grouping parameter Hierarchy: yes	Orthologs: none Paralogs: none Hierarchy: exhaustive	Orthologs: exhaustive Paralogs: exhaustive Hierarchy: exhaustive
Implied tree structure					
E.g	OMA groups	Compara pairwise, InParanoid, OMA pairwise	eggNOG, OrthoDB, OMA hierarchical groups	HOGENOM	Compara, Panther

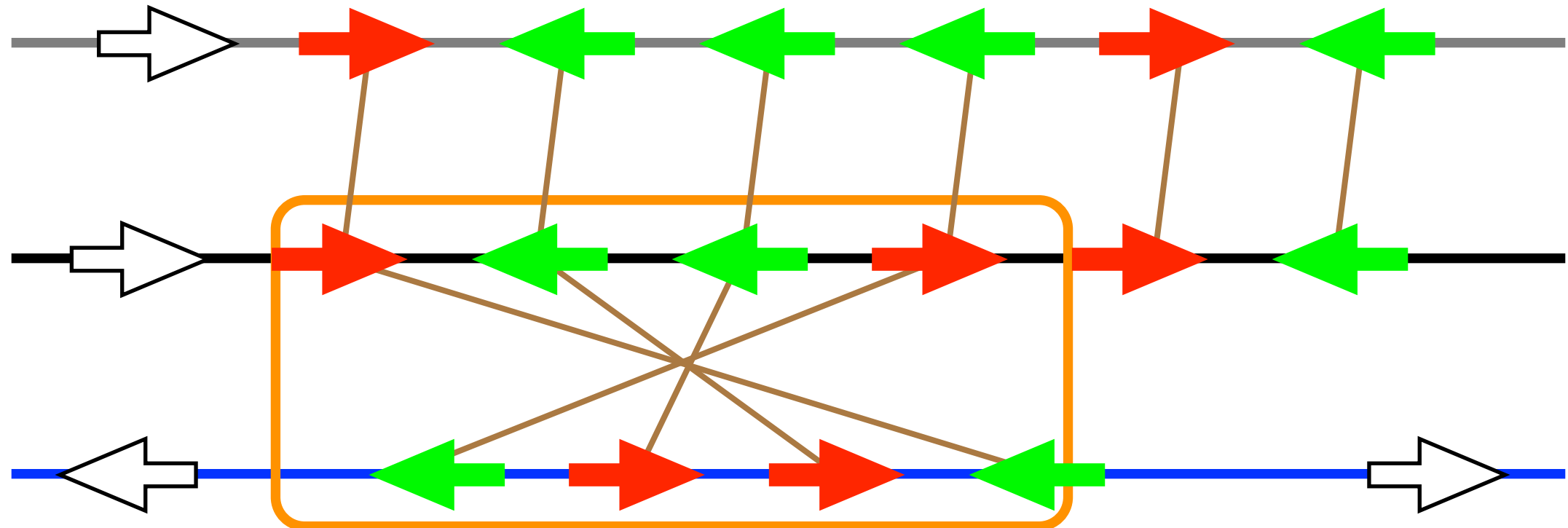
# Synteny

- Synteny is the conservation of gene order between species
- A *synteny block* is useful to make orthology inference when there is no other information



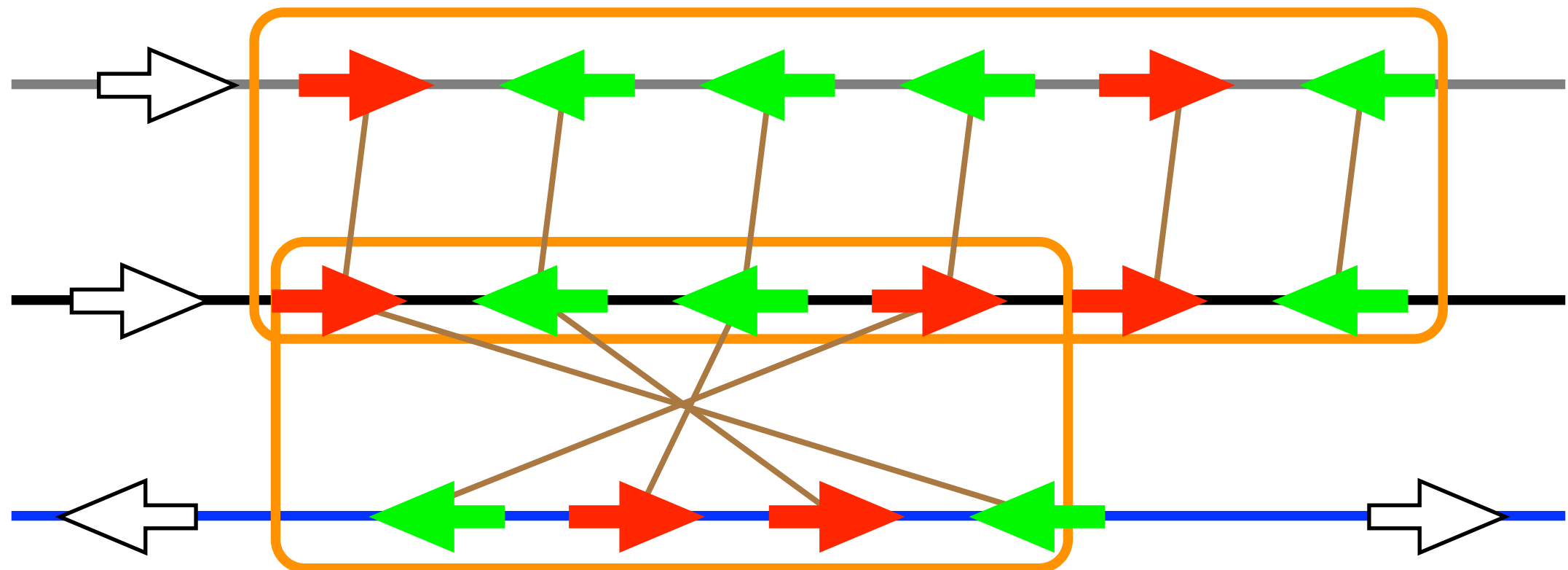
# Synteny

- Synteny is the conservation of gene order between species
- A *synteny block* is useful to make orthology inference when there is no other information



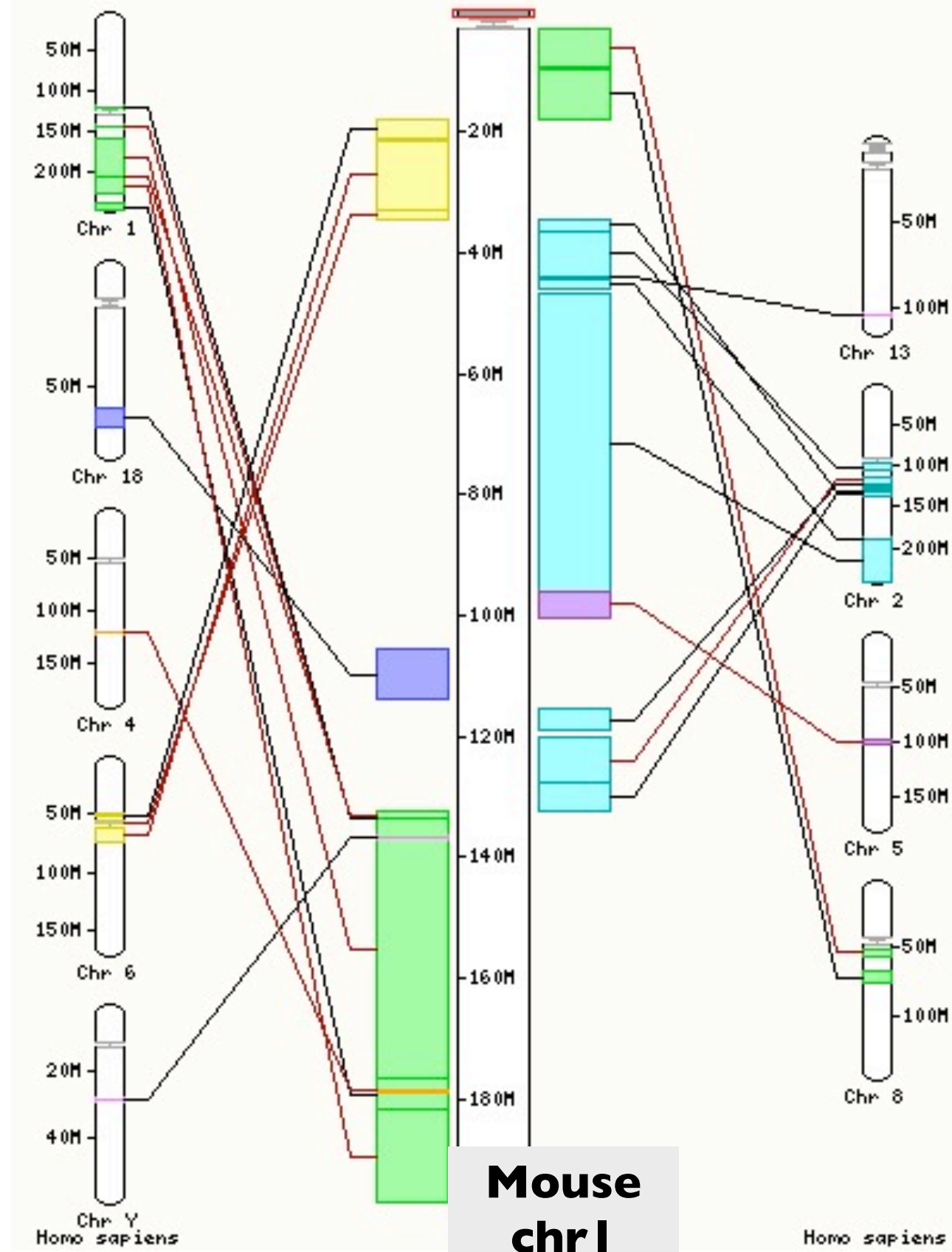
# Synteny

- Synteny is the conservation of gene order between species
- A *synteny block* is useful to make orthology inference when there is no other information





# Synteny



# References

- Fitch, W. M. (2000). Homology a personal view on some of the problems Trends in genetics, 16(5), 227–231.
- Sonnhammer, E. L. L., & Koonin, E. V. (2002). Orthology, paralogy and proposed classification for paralog subtypes Trends in genetics, 18(12), 619–620.
- Roth, A. C. J., Gonnet, G. H., & Dessimoz, C. (2008). Algorithm of OMA for large-scale orthology inference. BMC Bioinformatics, 9, 518.
- Boeckmann, B., Robinson-Rechavi, M., Xenarios, I., & Dessimoz, C. (2011). Conceptual framework and pilot study to benchmark phylogenomic databases based on reference gene trees Briefings in Bioinformatics, 12(5), 423–435.