"same" gene in different species
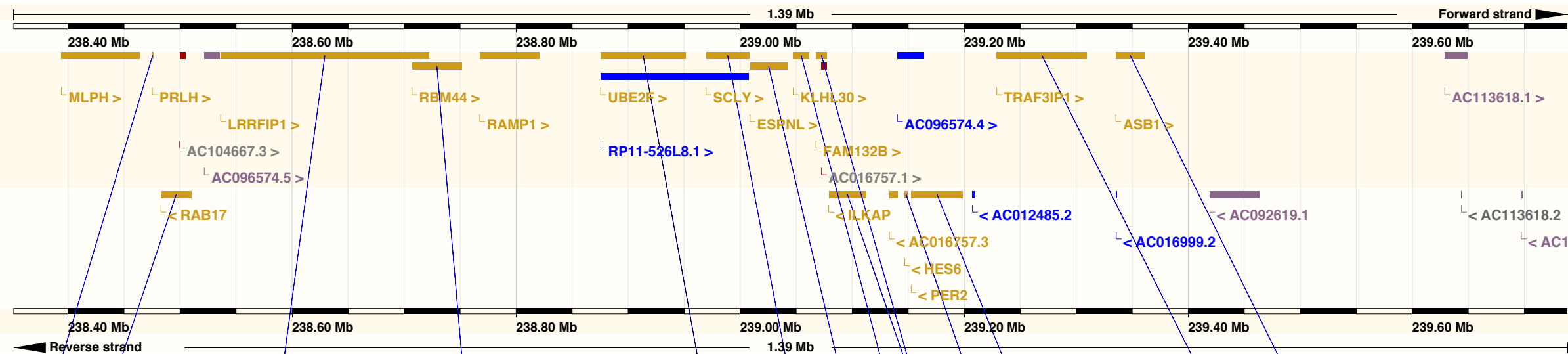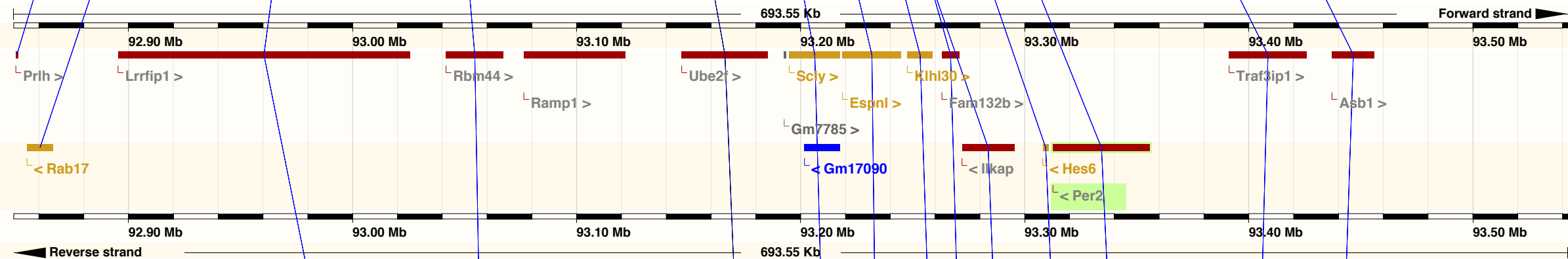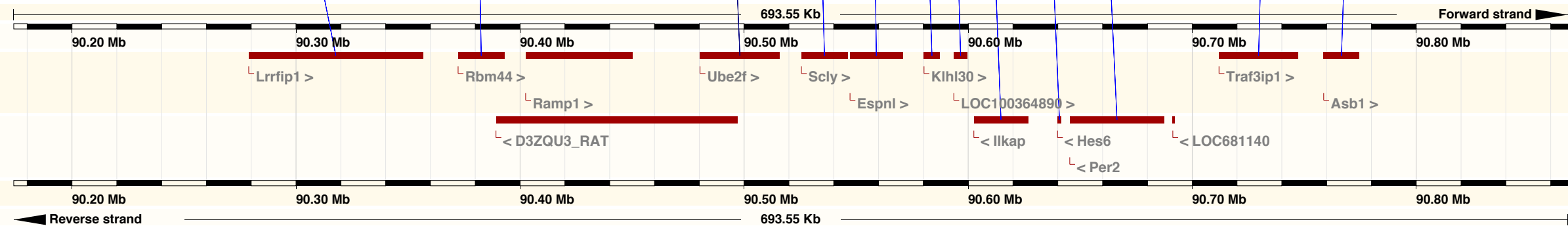
Homology

- Gene A in species S and gene B in species T are:
    - <u>Homologous</u> if they share a common ancestor

- Two homologous genes are:
    - <u>Orthologous</u> if they diverged at a speciation event
    - <u>Paralogous</u> if they diverged at a duplication event
    - <u>Xenologous</u> if one of them went through a lateral transfer

Evolution of a gene

speciation

ancestor

speciation

speciation

duplication

speciation

gene loss
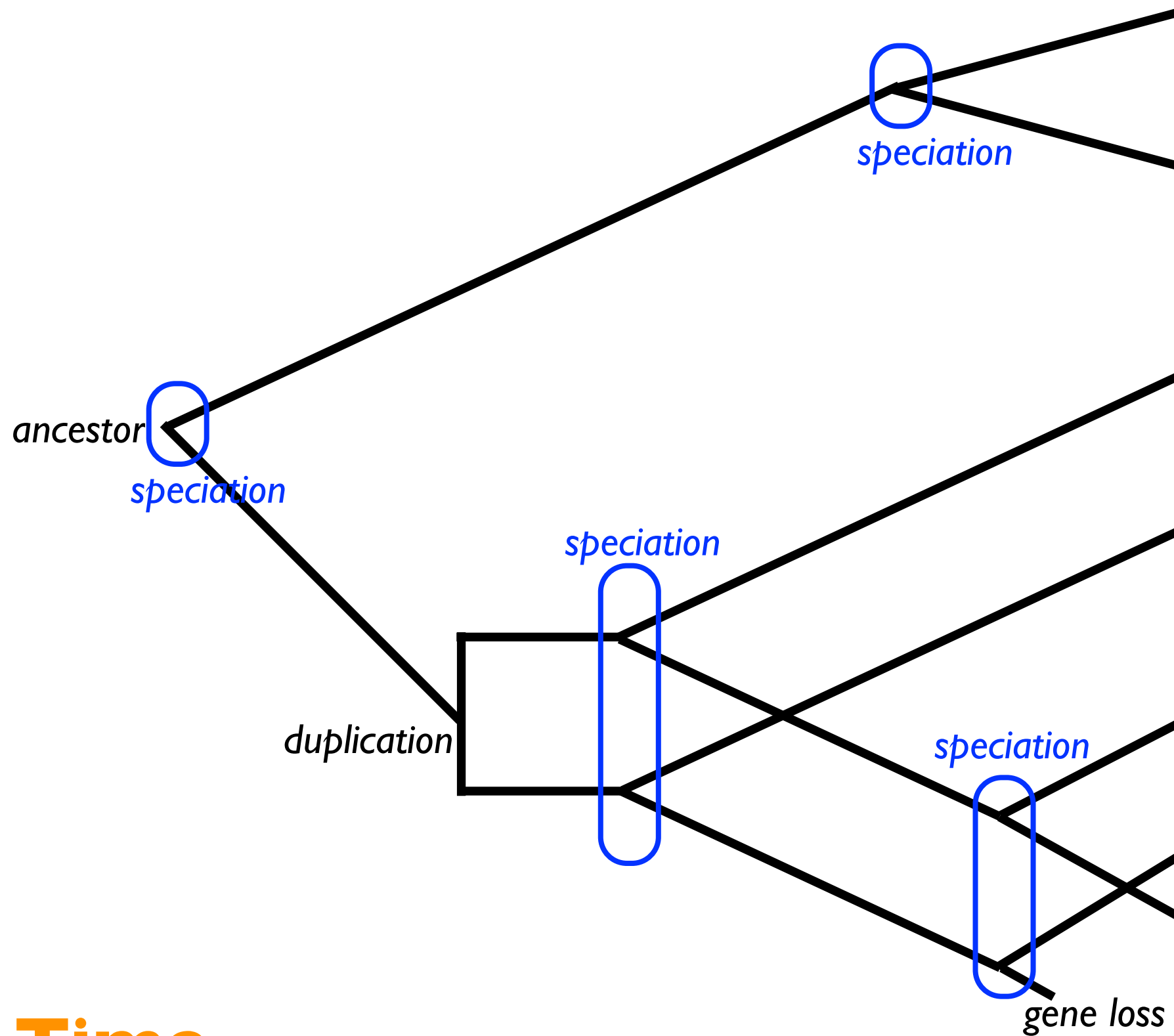
Time

Species A	Gene I
Pairs of orthologs	Pairs of paralogs
(A1,B1)	(C1,C2)
(A1,C1)	(D1,D2)
(A1,C2)	(C1,D2)
(C1,D1)	(C2,D1)
(C2,D2)	(C2,E1)
...	(D2,E1)

C

C2

D

D2

E

I

Evolution of a gene

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

speciation

ancestor

speciation

speciation

duplication

speciation

gene loss

Time
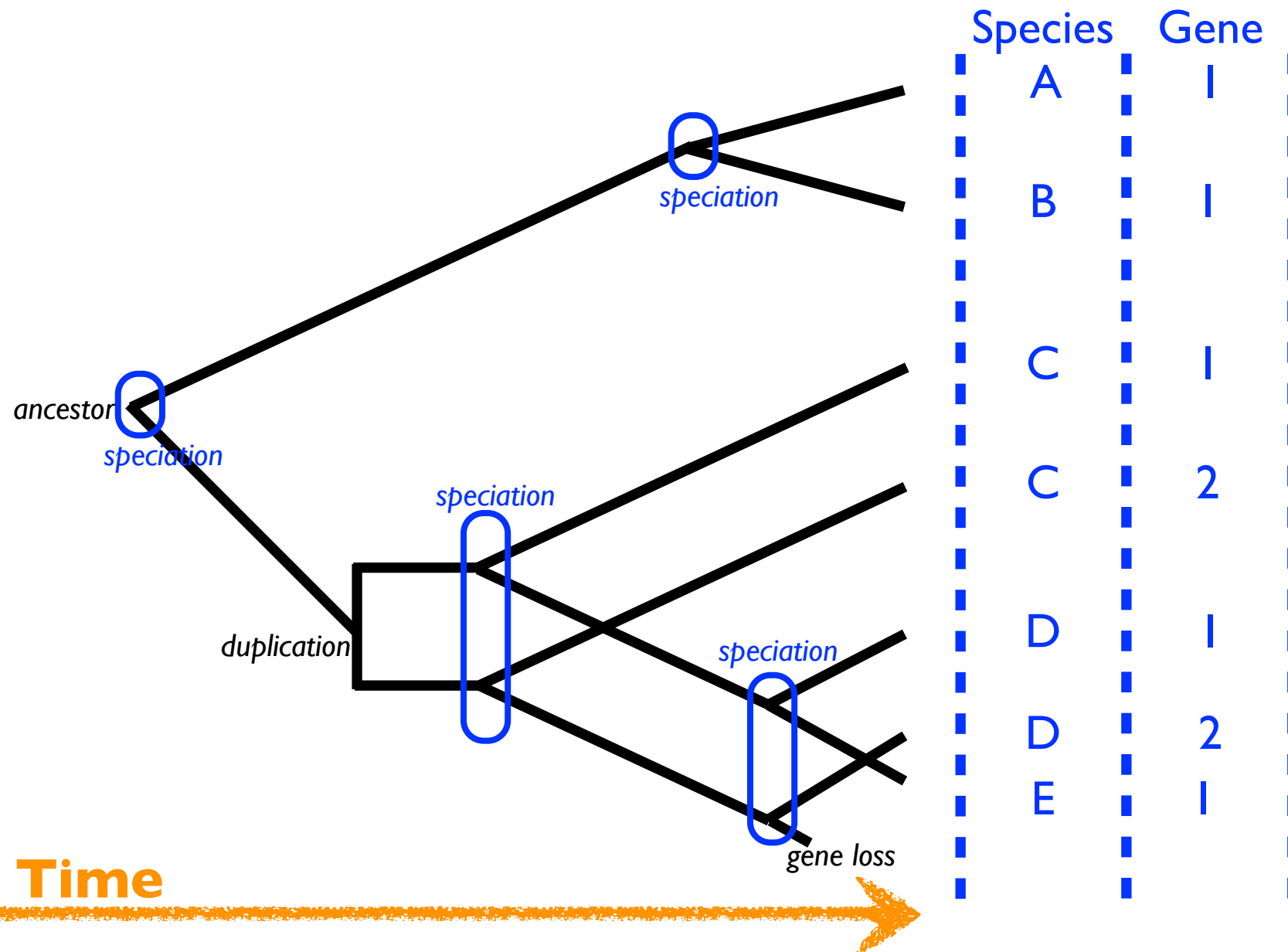
Pairs of orthologs	Pairs of paralogs
(A1,B1)	(C1,C2)
(A1,C1)	(D1,D2)
(A1,C2)	(C1,D2)
(C1,D1)	(C2,D1)
(C2,D2)	(C2,E1)
...	(D2,E1)

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
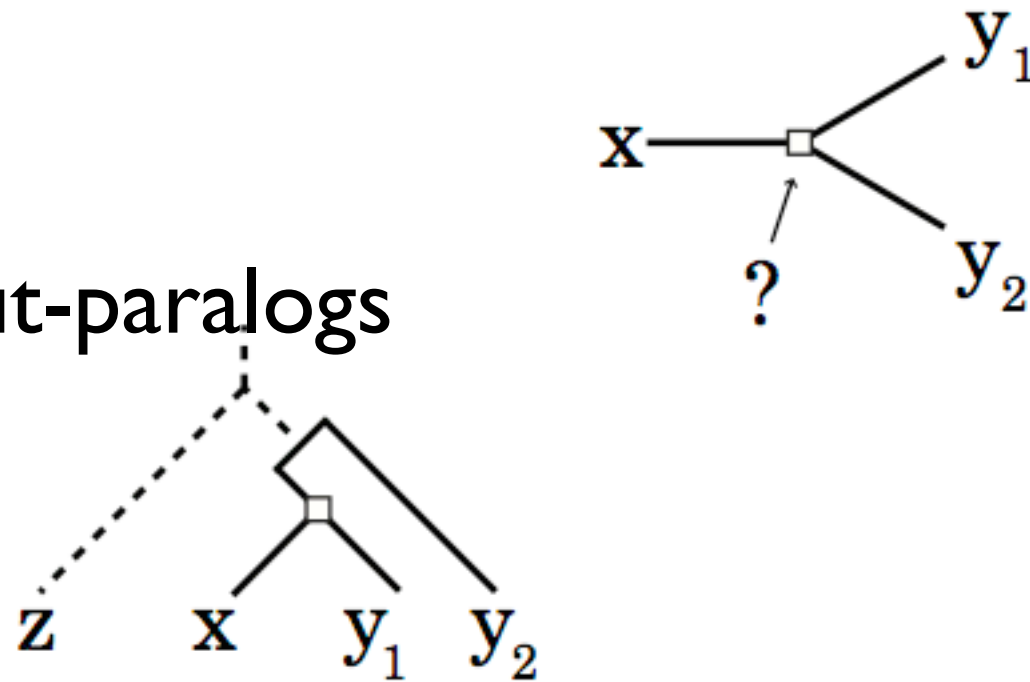
Strategy

- 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 of all genes in X, A is the most similar to B, and from all genes in Y, B is the most similar to A

- Look for a third genome as a proof of orthology

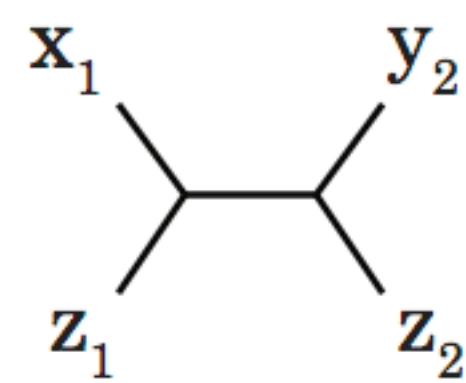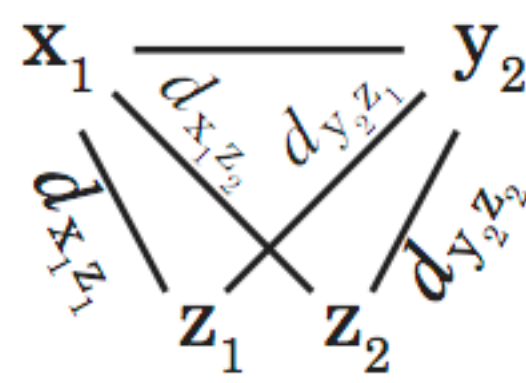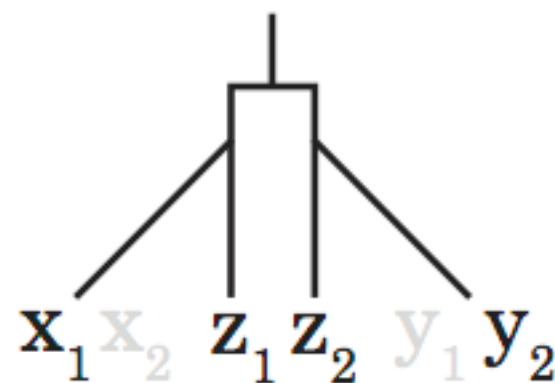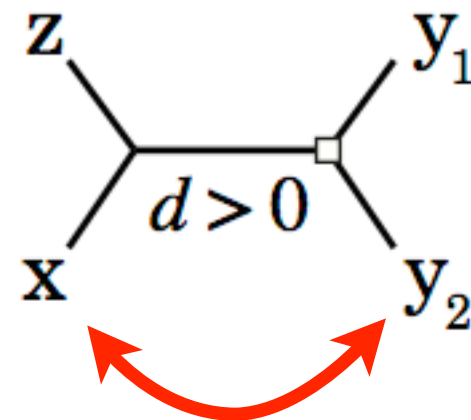- Make clusters of triangular relations (orthologous groups)

http://omabrowser.org/

In/Out-paralogs

A KatG, YKR066c, sII1987 — RBH

B IleS, YBL076c, HIN0378, YPL040c, MG345, MJ0947, MP322, sII1362 — Yeast paralogs

C sII0184, RpoH, RpoS, sII0306, FliA, slr0653, RpoD, sII1689, HIN1403, sII2012, HIN1655, slr1564, MP485, MG249 — 4-clique: 4 co-orthologs

R L Tatusov et al. Science 1997;278:631-637
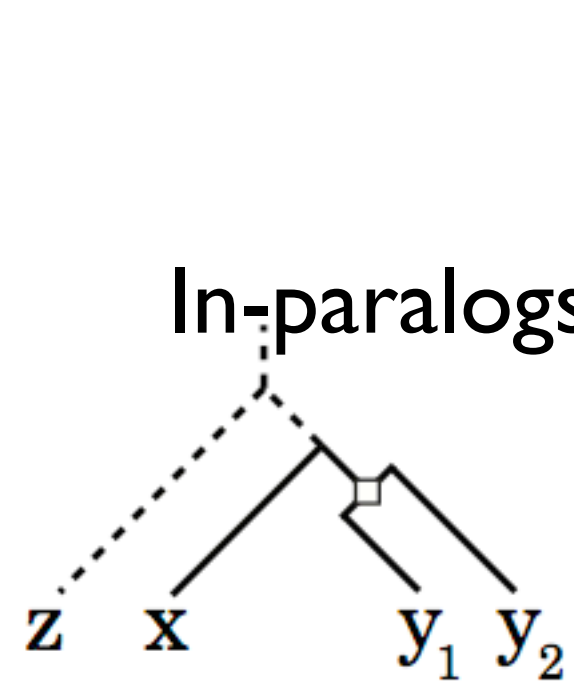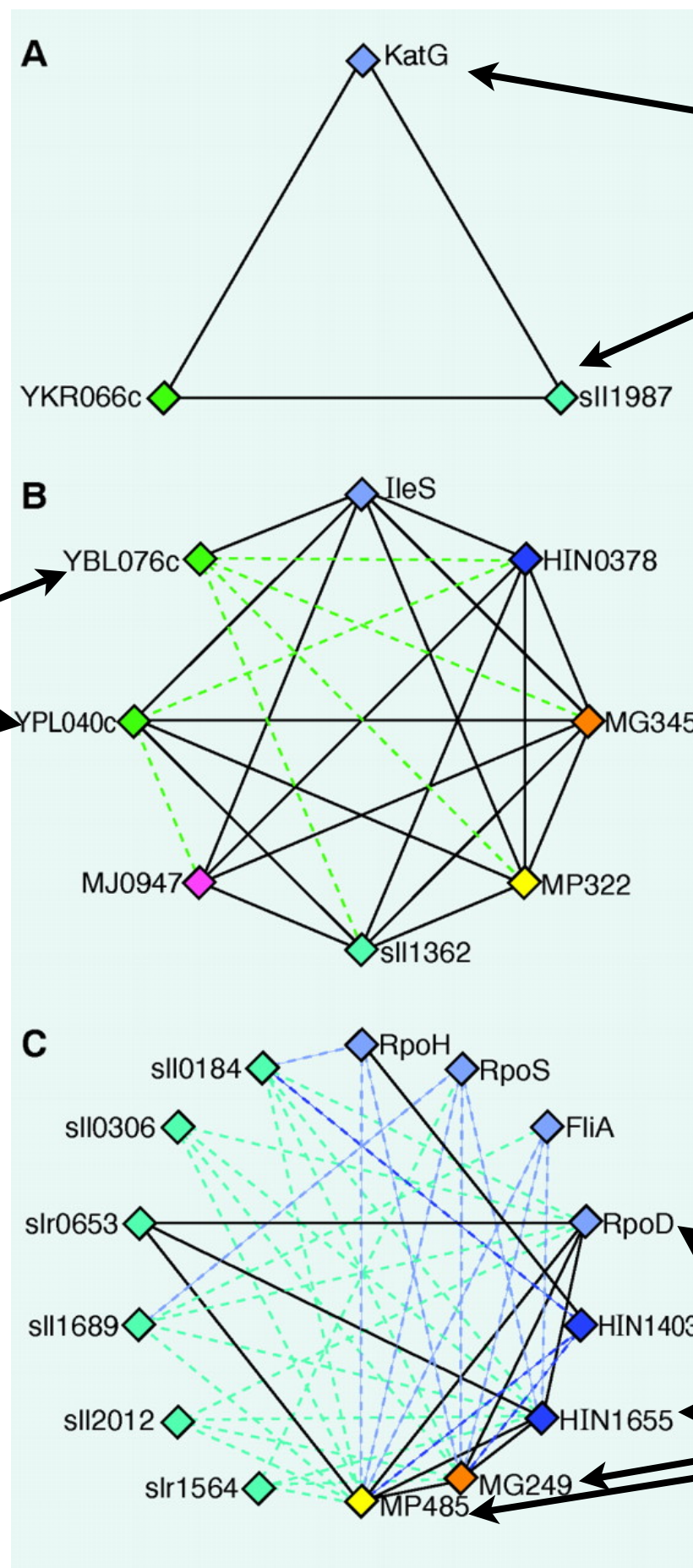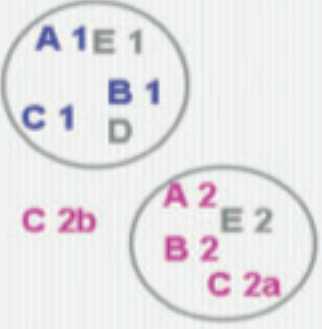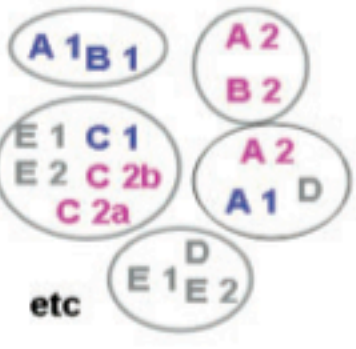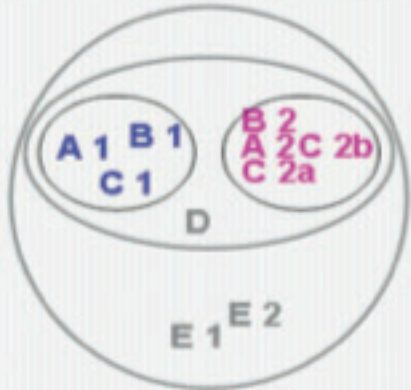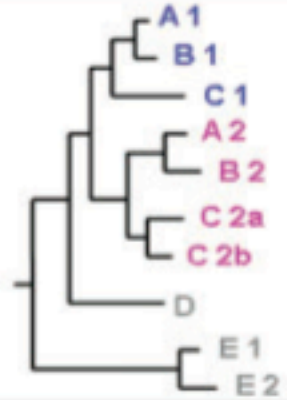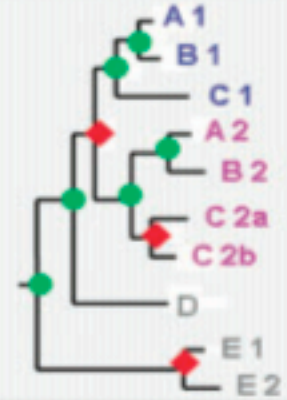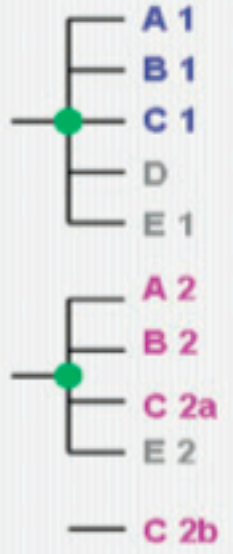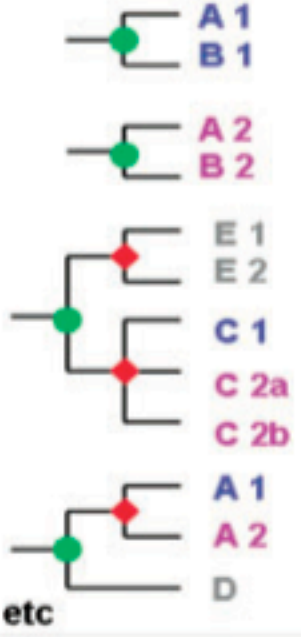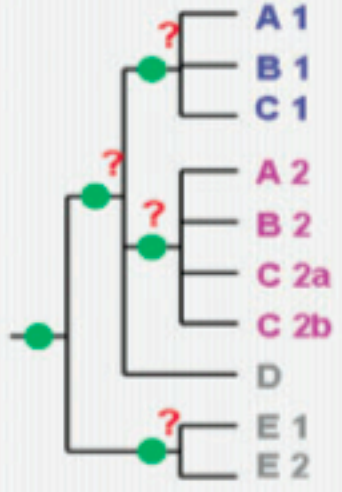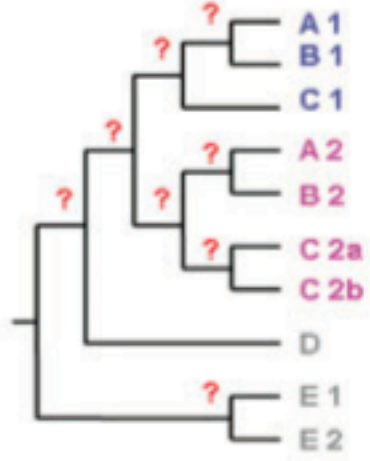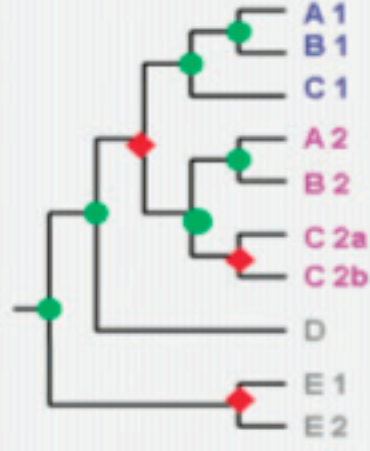
	Pure ortholog groups	Pairwise groups	Hierarchical groups	Plain trees	Reconciled tree
Concept	All gene pairs of an ortholog group are orthologs.	All genes of a pairwise group are derived from a common gene that diverged subsequent to the given speciation event.	Groups of orthologs and inparalogs with respect to defined speciation events.	Unlabeled gene trees.	Gene trees with internal nodes labeled
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
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
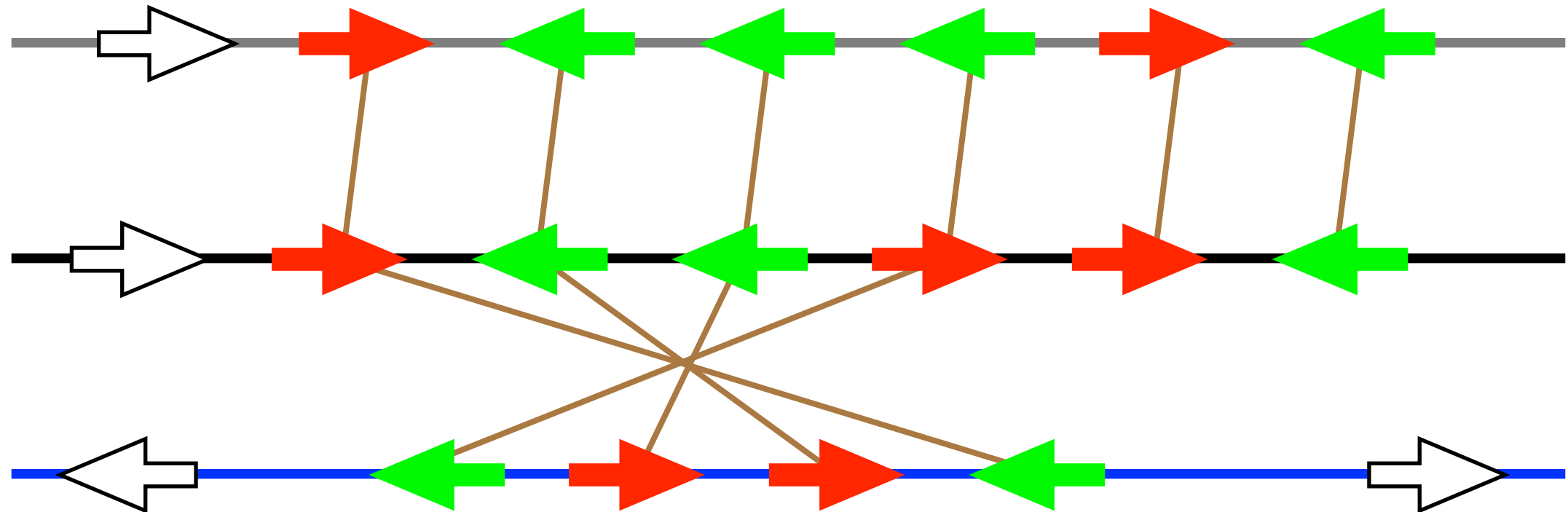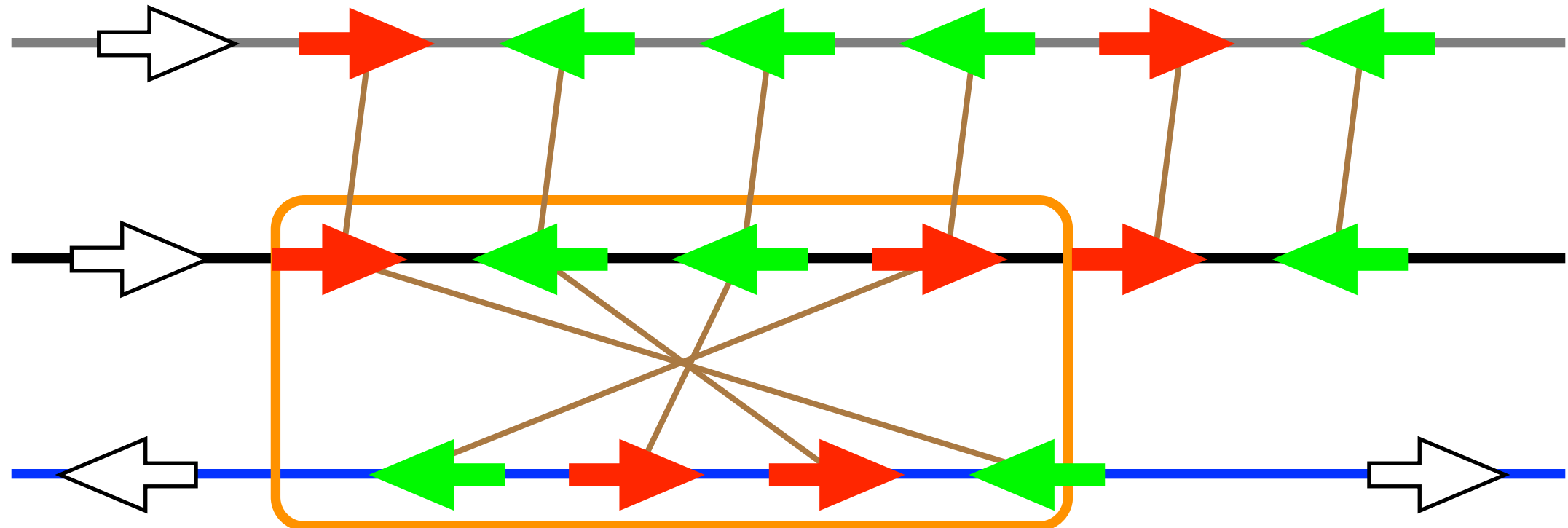
Synteny

- Synteny is the conservation of gene order between species

- A *synteny block* is useful to make orthology inference when there is no other information
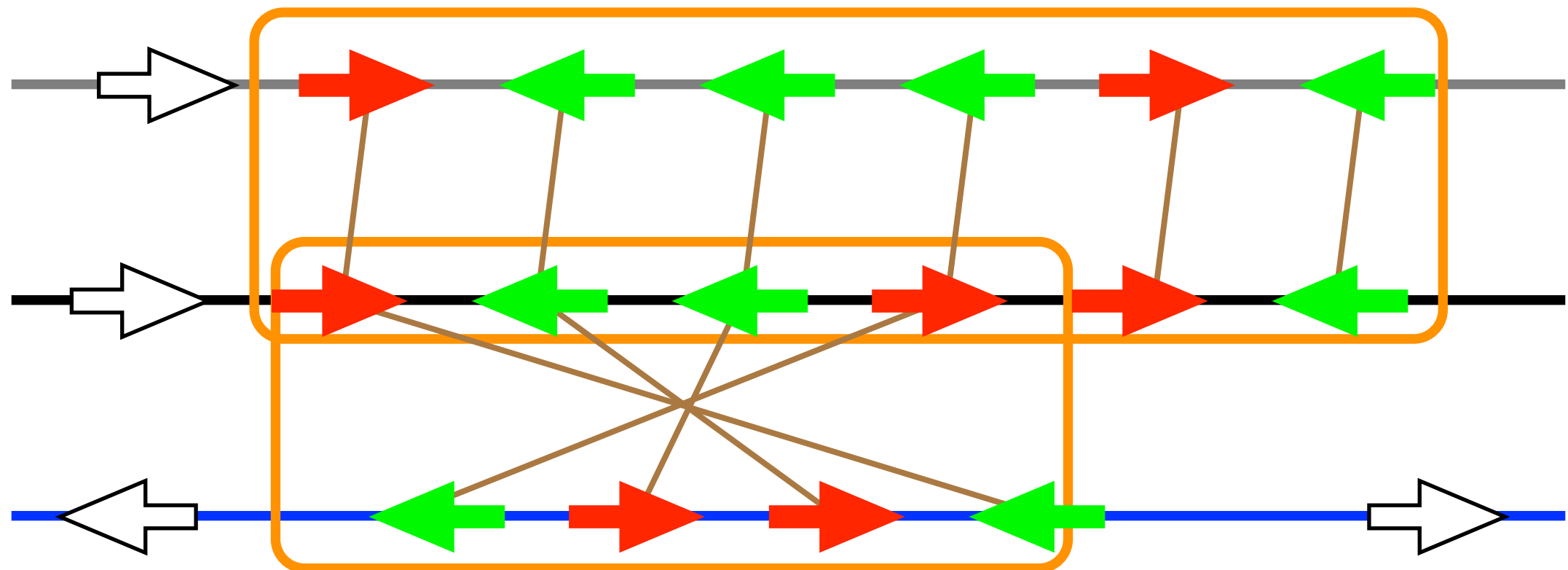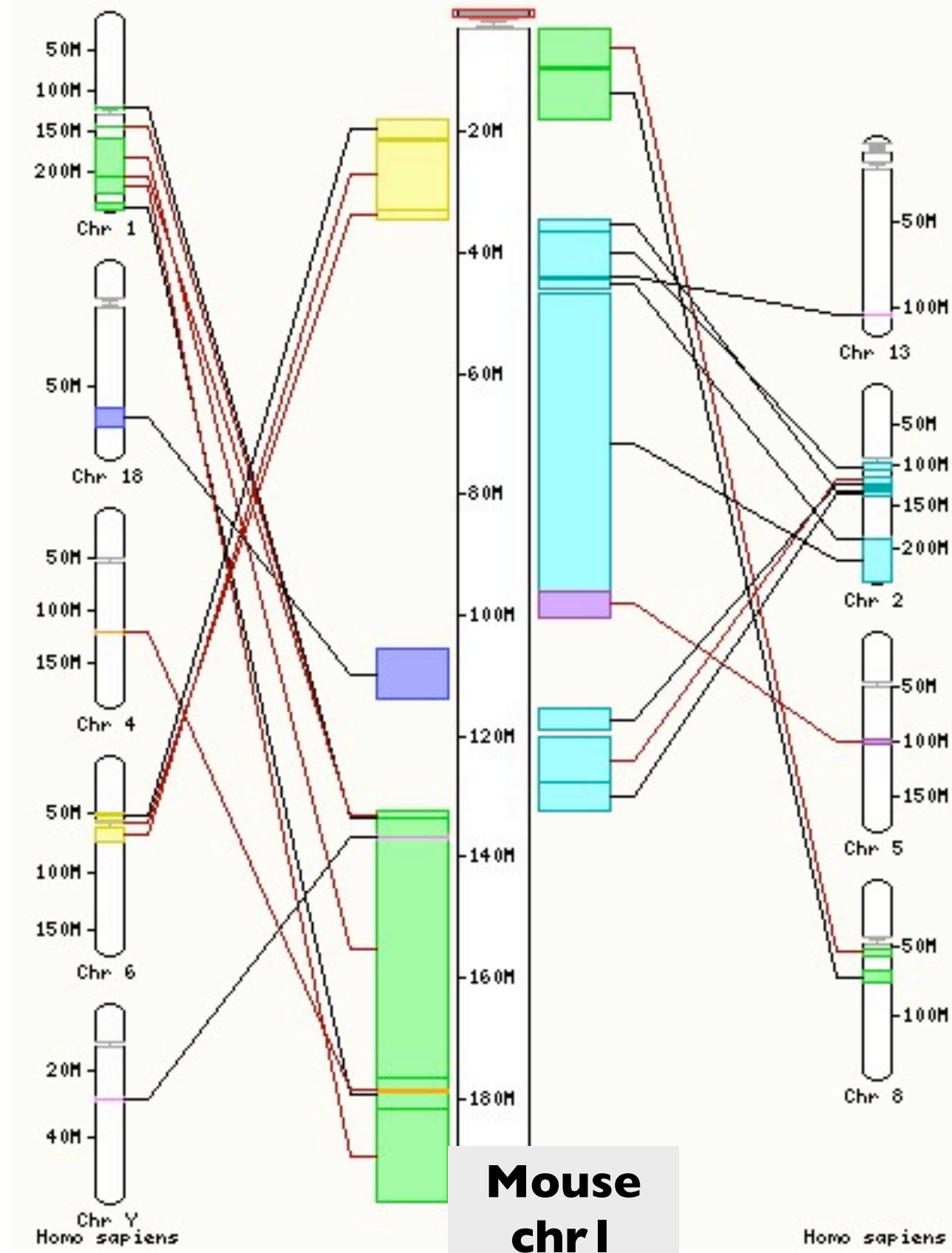
Synteny

- 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.