

Strong negative and positive selection can obscure ancestral signal in phylogenetic analysis

Alexander Tchourbanov

University of Wyoming, Laramie

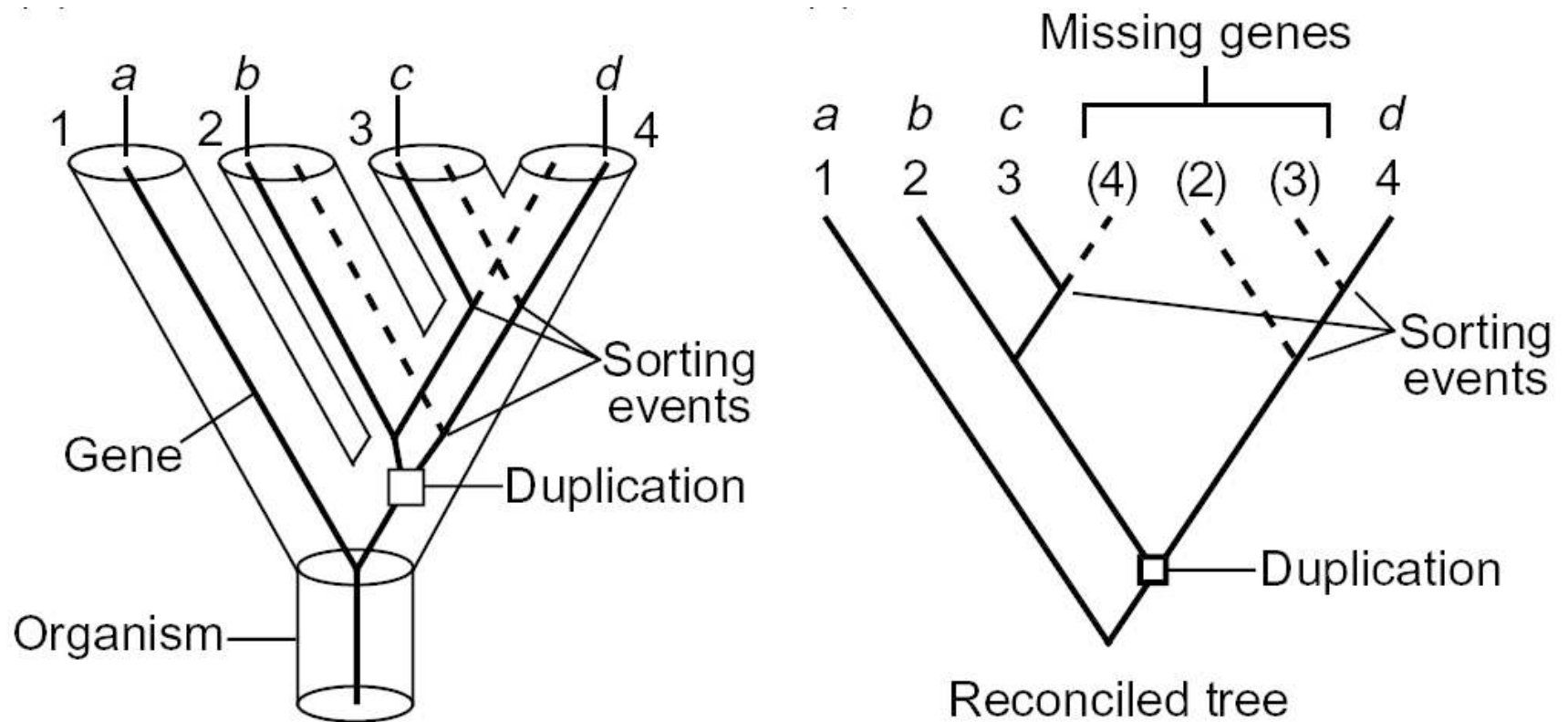
KaKsiTaxi project

- Genomes evolve through the consequence of duplication events followed by evolutionary adjustments
- Duplication events are classified as follows:
 - *Orthology*, i.e. duplication by speciation
 - *Paralogy*, i.e. duplication by replication within organism
- We are interested in effect of Ka/Ks ratio, characterizing the process of positive selection, on fidelity of phylogenetic reconstruction

Ka/Ks ratio

- In order to detect positive selection between gene sequences we compare, ratio of non-synonymous (Ka) to synonymous substitutions (Ks) is used:
 - $\frac{Ka}{Ks} < 1$ - purifying selection
 - $\frac{Ka}{Ks} = 1$ - neutral drift
 - $\frac{Ka}{Ks} > 1$ - positive selection

Reconciled trees

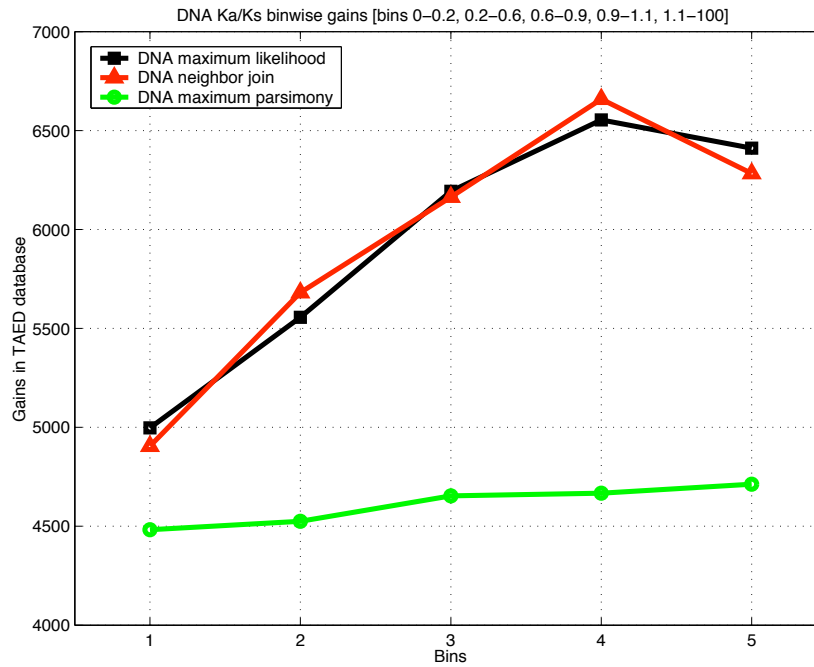


One explanation is to embed the gene tree in the species tree

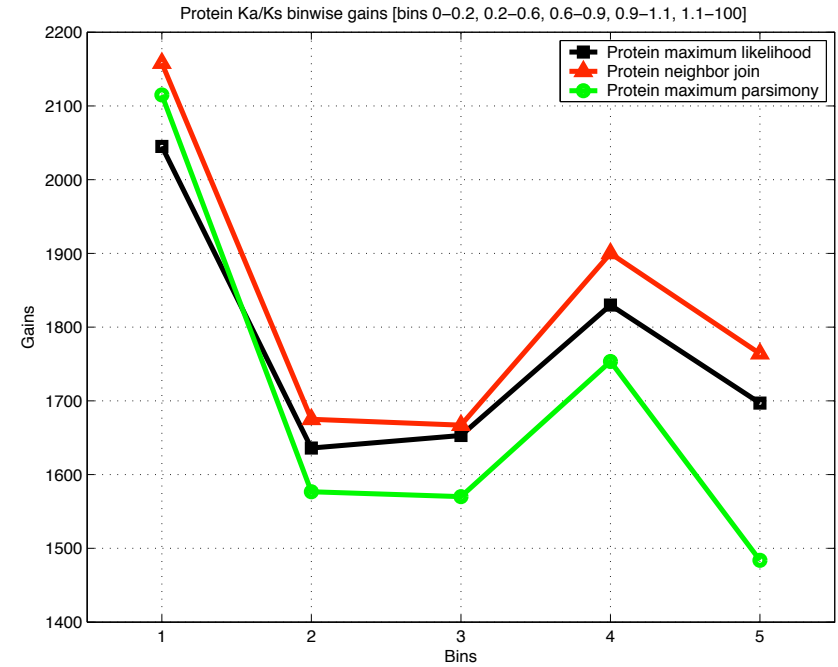
KaKsiTaxi framework

- Run MUSCLE to realign gene family sequences obtained from TAED database and then get rid of gaps;
- Using modified Nei-Gojobori technique we find position-specific $\frac{K_a}{K_s}$ ratios;
- Form the five bins according to $\frac{K_a}{K_s}$ ratio and check if they are have more than 3 sequences of more than 9 codons;
- For each bin we run `Phylip` tools:
 - Parsimony (both DNA/Protein)
 - Neighbor Join (both DNA/Protein)
 - Maximum likelihood (both DNA/Protein)
- We collect resulting trees. Special weighting scheme is used to process equiprobable trees from maximum Parsimony method.
- We transform the trees so that we get rid of branches and set posterior to 1.0;
- Run `SoftParseMap` on precompiled index file and collect gains and losses as reconciled with recent NCBI taxonomy file;
- Output cumulative statistics.

Results



DNA gains



Protein gains