# Supplementary Material: Unsorted Homology within Locus and Species Trees

Diego Mallo[1], Leonardo de Oliveira Martins[2] and David Posada[3]

*Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo, 36310, Spain*

**E-mails:** dmallo@uvigo.es[1], leomrtns@uvigo.es[2], dposada@uvigo.es[3]

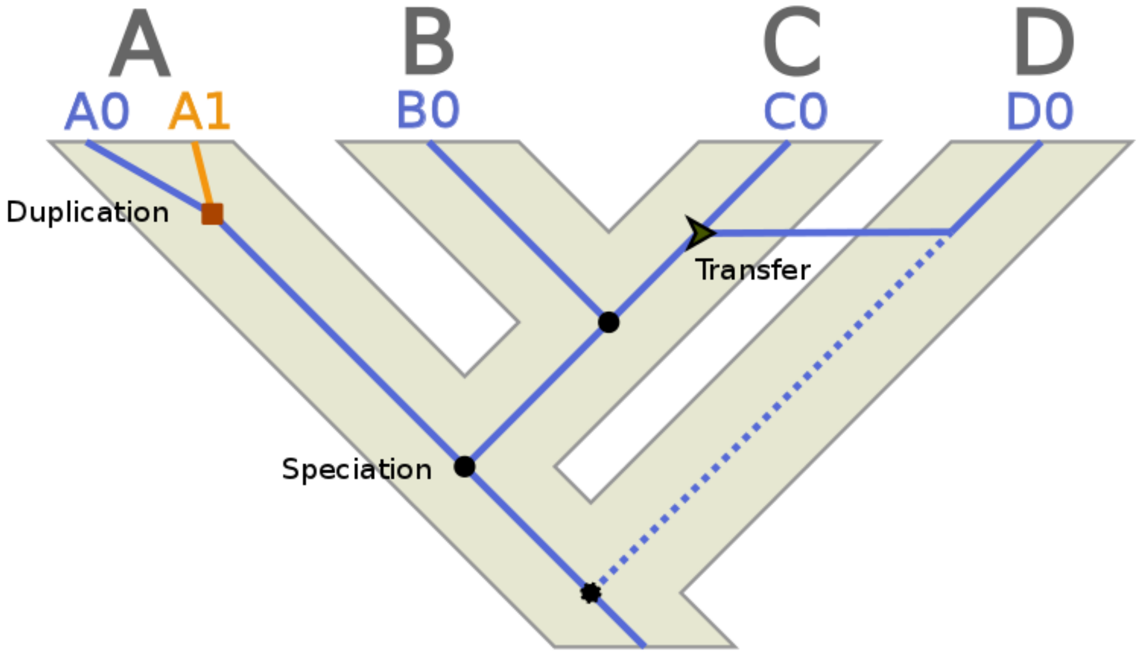


Figure S1. Gene tree (thin blue and orange lines) inside a species tree (light tree in the background) showing the traditional orthology, paralogy and xenology relationships. There are four species (A, B, C and D) and 5 gene copies (A0, A1, B0, C0 and D0). Gene copies B0 and C0 are orthologs, because their MRCA corresponds to a coalescence/speciation event. On the other hand, A0 and A1 are paralogs, since their MRCA corresponds to a duplication event. D0 and C0 are xenologs, as D0 was transferred from the C lineage to the D lineage. In this case the transfer implies a replacement of the original gene copy present in the D lineage, and the branch leading to the latter is represented using dashed lines.

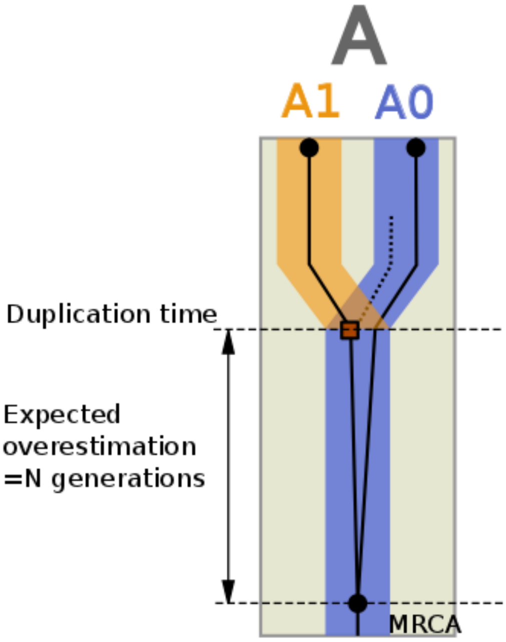


Figure S2. Expected overestimation on the duplication time for two paralogous copies (A1, A0) in one individual (A). The gene tree (thin black lines) evolves inside the locus tree (medium-thick blue and orange lines) along one species tree branch (light gray shadow in the background). The duplication takes place in the gene copy represented by a red square, whose lineage in the blue locus does not get to the present (dashed line). The coalescence process starts (backwards in time) at the duplication time, since previously the lineages pertain to different loci. For diploid populations, the expected overestimation equals to 2N generations.

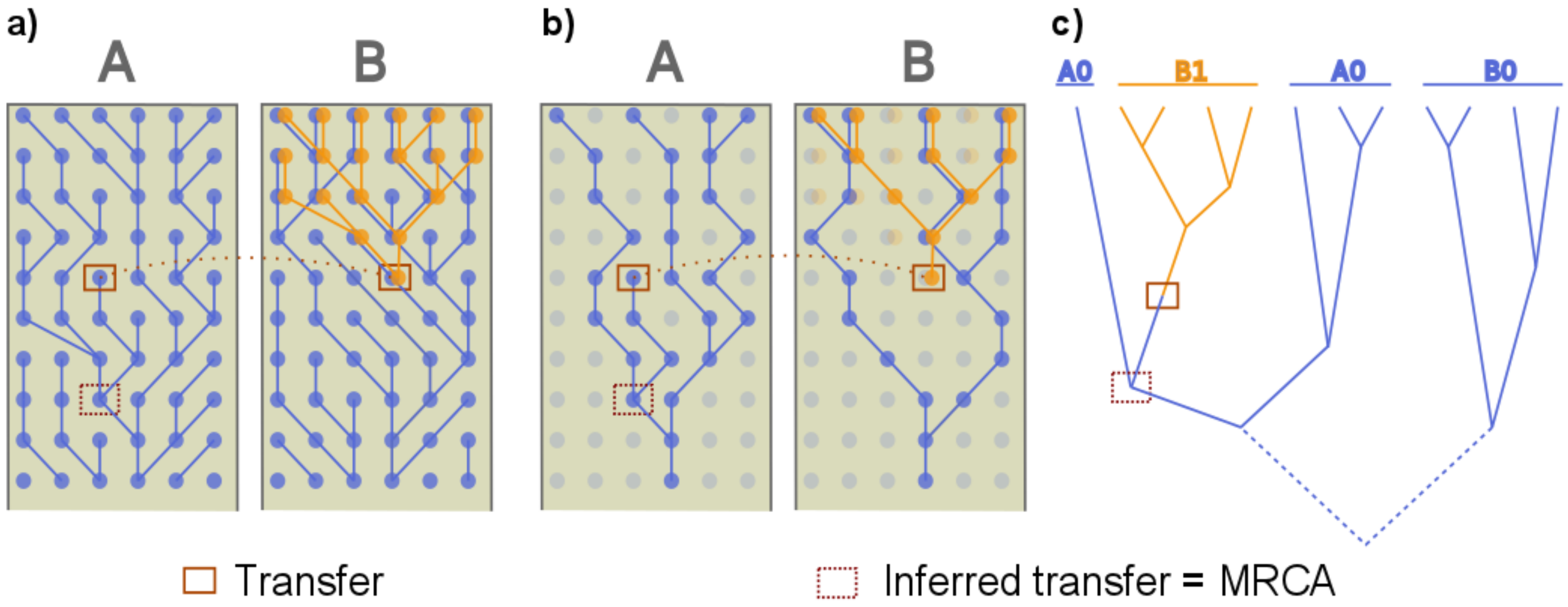
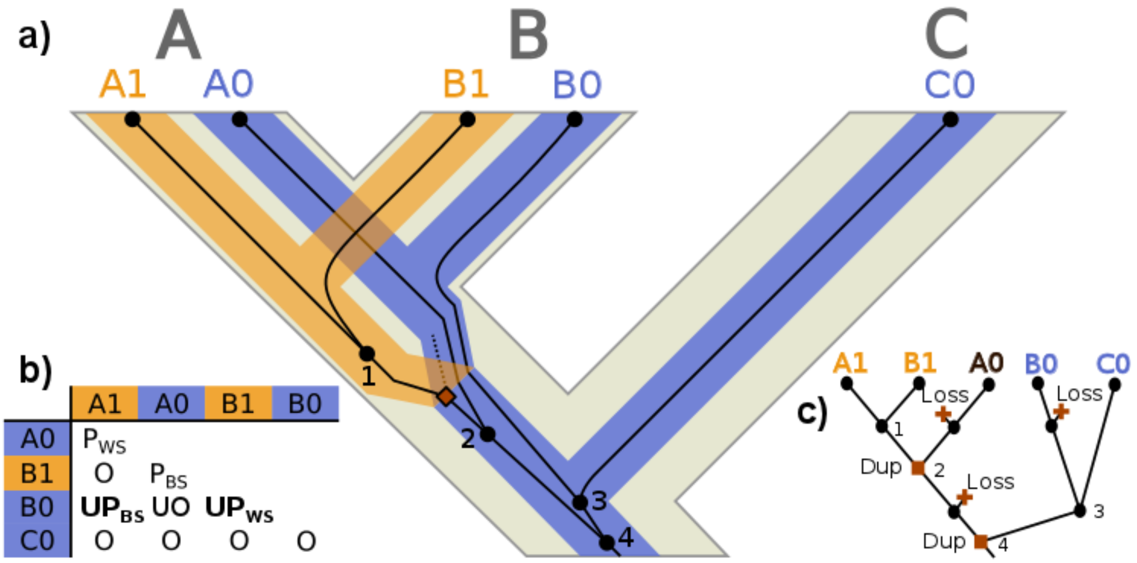


Figure S3. Genealogy of two xenologous loci in two different species (A and B). Filled circles represent segregating gene copies, with colors indicating the loci. The new locus (orange) was originated by horizontal transfer of a gene copy (surrounded by a solid square) from species A to species B. It is assumed that both loci (orange and blue) are unlinked and evolve independently. Recombination events occur when the two loci are inherited from different individuals, resulting in non-parallel branches in the figure. a) Complete genealogy. b) Genealogy of the sampled gene copies. c) Reconstructed (true) gene tree for the sample.



| | A1 | A0 | B1 | B0 |
|---|---|---|---|---|
| A0 | $P_{WS}$ | | | |
| B1 | O | $P_{BS}$ | | |
| B0 | **$UP_{BS}$** | UO | **$UP_{WS}$** | |
| C0 | O | O | O | O |

Figure S4. Unsorted paralog evolution with ILS. a) Evolution of a gene tree (black lines) inside a locus tree (medium-thick blue and orange lines) embedded in a species tree (light gray tree in the background) with ILS at the locus tree level. A, B and C represent species/populations, while A1, A0, B1, B0 and C0 identify gene copies. Black circles represent nodes in the gene tree (only internal nodes are numbered), where the dashed line represents an extinct/unsampled lineage. Squares signal duplication events. b) Homology relationships between the gene copies (O: orthologs; UO: unsorted orthologs; $P_{BS}$: between-species paralogs; $P_{WS}$: within-species paralogs; $UP_{BS}$: between-species unsorted paralogs; $UP_{WS}$: within-species unsorted paralogs). c) Most parsimonious duplication/loss reconciliation of the gene and species trees. Label colors indicate different estimated loci, while text refers to the real loci.

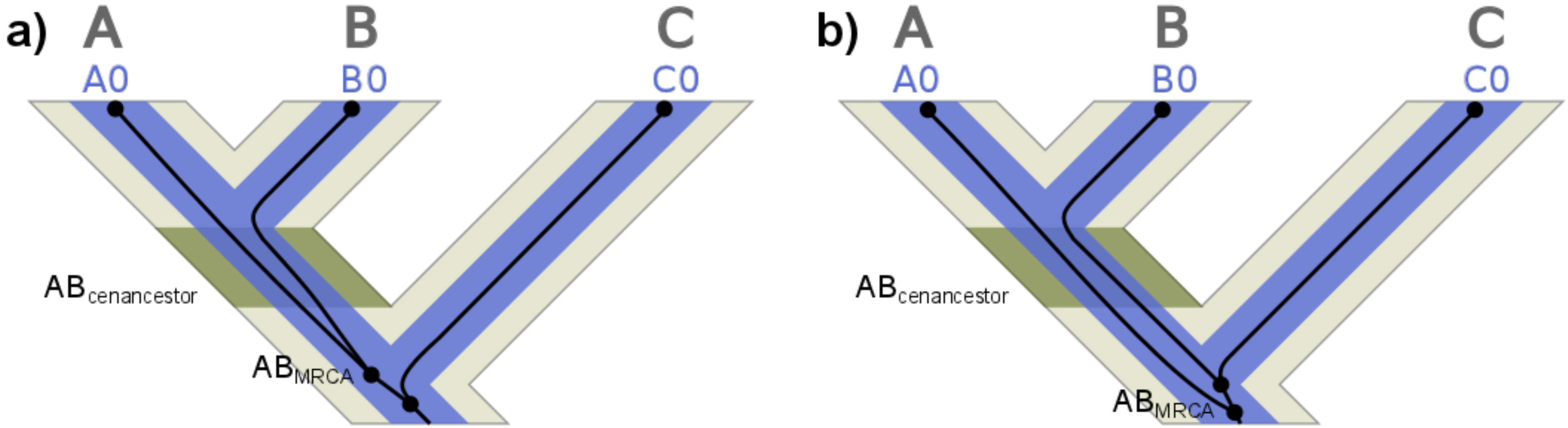


Figure S5. Unsorted orthologs. A, B and C represent species/populations, while A0, B0 and C0 identify gene copies of the same locus. Both panels represent the evolution of a gene tree (thin black lines) inside a locus tree (medium-thick blue branches) embedded in a species tree (light gray tree in the background). Black circles represent gene tree nodes. The MRCA of A0 and B0 does not occur in the AB cenancestor (green species tree branch), but in the ABC ancestral species. Therefore, A0 and B0 are unsorted orthologs. a) The species, locus and gene trees are congruent. b) The gene tree is incongruent with the locus and species trees.