

Figure 1: Overview of pseudogenes' life and implied mechanisms. The arrows' thickness reflects approximately the scale of each path.

Unitary pseudogenes have their locations unchanged compared to the genes they are originated from, while unprocessed pseudogenes can be located in the same locus or in a different locus than the gene they are originated from. As for the processed pseudogenes they are scattered in the genome.

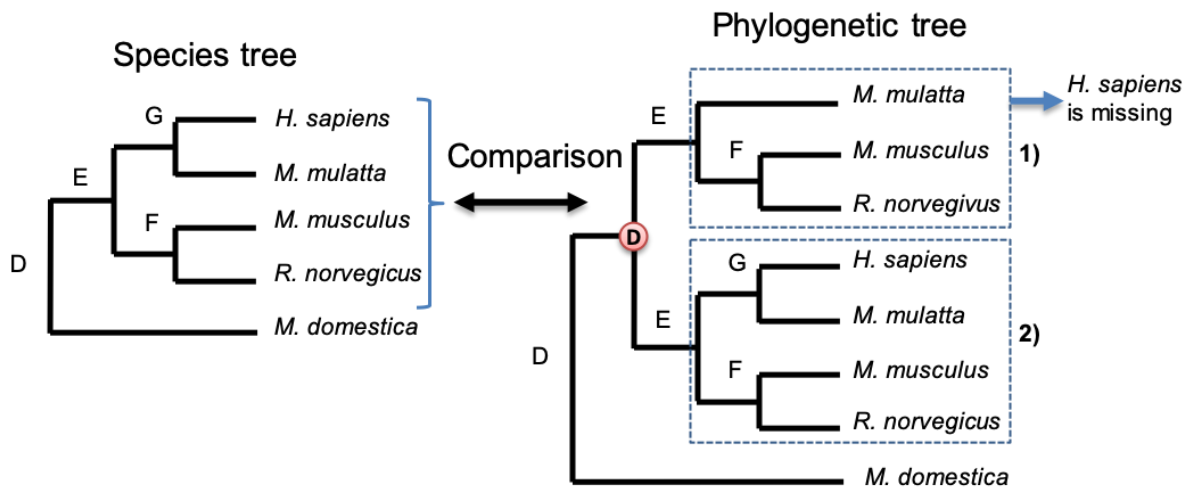


Figure 2: Comparison between the species tree of the investigated species and the phylogenetic tree of the gene of interest. Such a comparison allows determining putative gene losses, which should be further investigated to detect the presence of a pseudogene. Each species of the phylogenetic tree corresponds to a gene. D, E, F, G correspond to the ancestors, Theria, Eutheria, Murinae and Catarrhini, respectively. A duplication event occurred in the Eutheria ancestor and the orthologous groups within the subtrees 1 and 2 need to be studied separately. In the subtree 1 the gene of interest is lost in *H. sapiens*, whereas it is retained in the subtree 2.

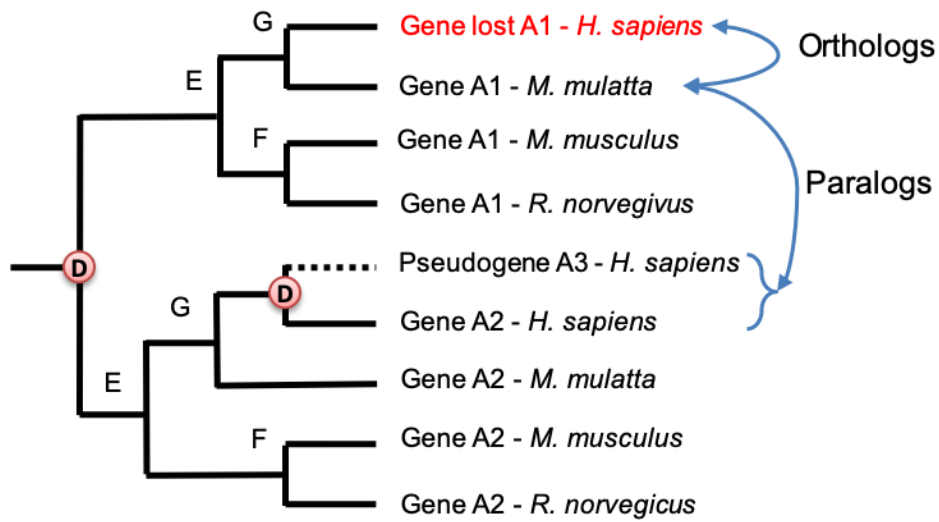


Figure 3: Use of the phylogenetic tree in order to confirm the orthology of a putative pseudogenic sequence. Gene A1 is lost in *H. sapiens* due to a pseudogenization event (red). In order to look for a putative A1 unitary pseudogene sequence, a Blast query is performed using *M. mulatta* A1 gene sequence. Such a query can lead to the detection of the *H. sapiens* A2 gene and A3 pseudogene. However, these sequences can be excluded because the phylogenetic tree indicates that they are not orthologs, but rather paralogs of the sequence used as query.

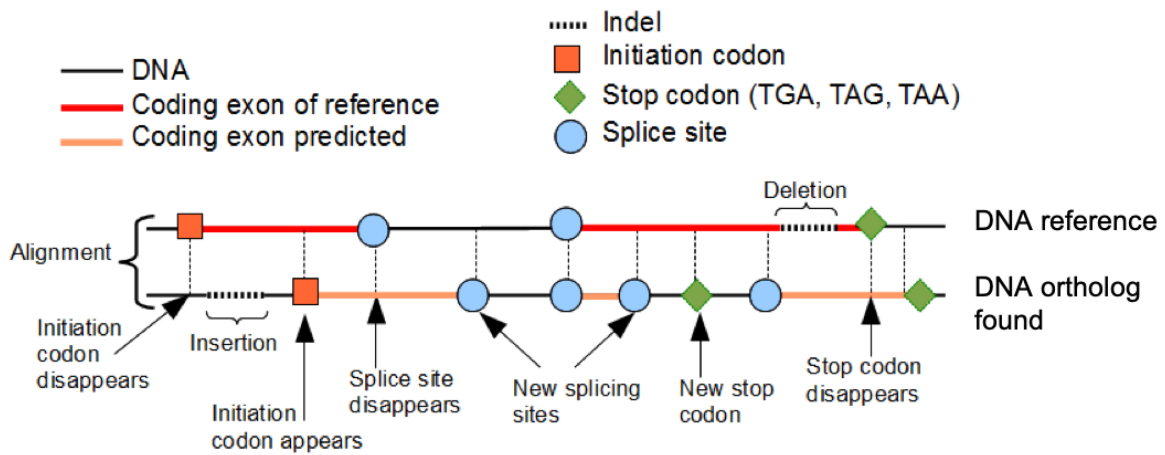


Figure 4: Illustration of the different types of micro-mutations. The micro-mutations occurred in the orthologous pseudogene retrieved by BLAST are found comparing its sequence with the reference sequence (functional gene). Some of these mutations are compatible with the appearance of a potential new ORF (orange).

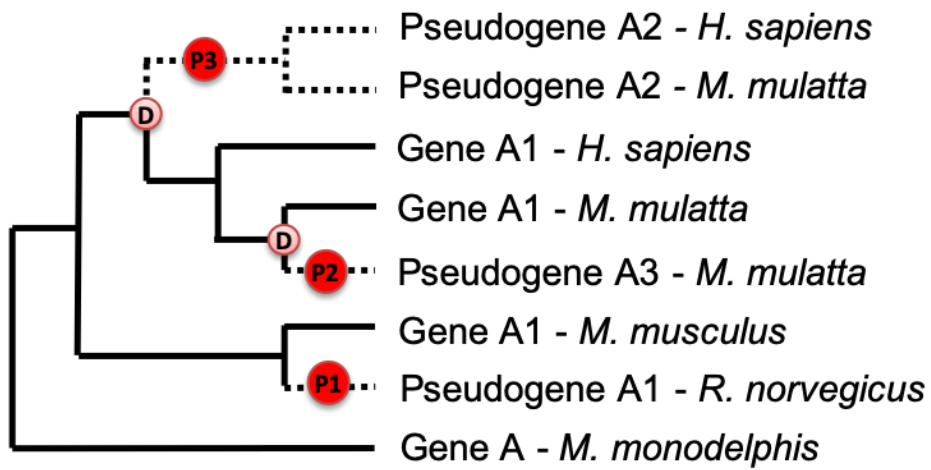


Figure 5: Phylogenetic tree of the Theria gene family in which 4 pseudogenes are present. P1, P2 and P3 refer to 3 distinct pseudogenization events.

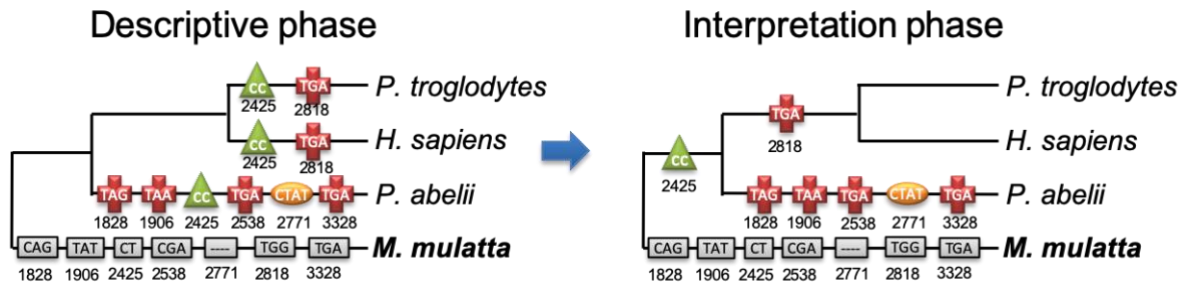


Figure 6: Study of the pseudogenization of Acyl3 gene in Catarrhini. The *M. mulatta* gene, which is functional, is taken as reference, while the other 3 species contain an orthologous unitary pseudogene. In the descriptive phase, all the known mutations are collected, while in the interpretation phase, the time frame of occurrence of the mutations is defined. The red crosses are stop codons; the green triangles are splice sites; the orange ellipses are insertions; the black rectangles include the reference nucleotide sequence. The numbers correspond to the positions in the alignment.

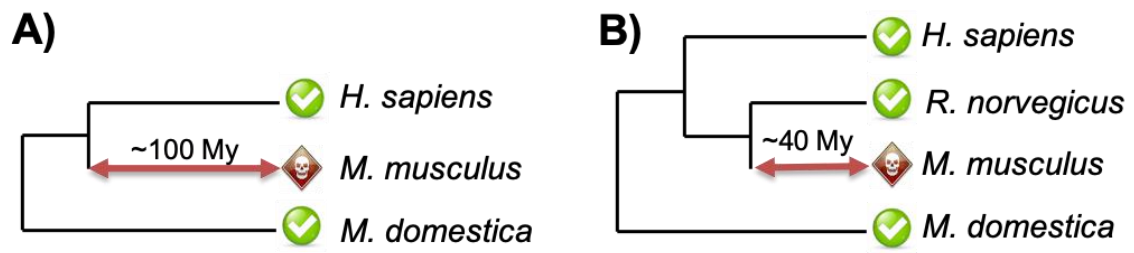


Figure 7: Species choice and impact on the detection of the date of gene loss.

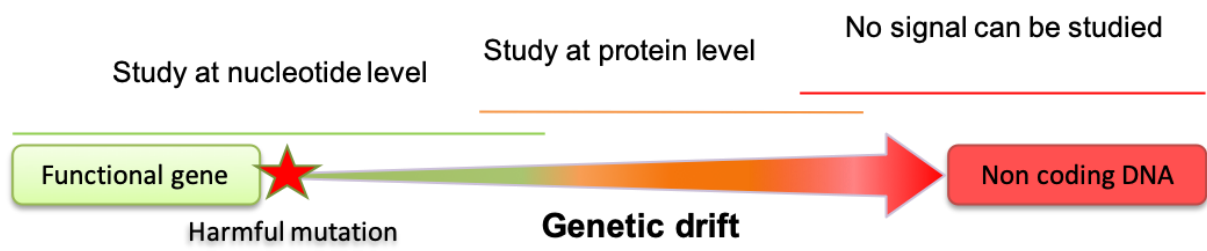


Figure 8: Description of pseudogene evolution and of the possibility to study them.