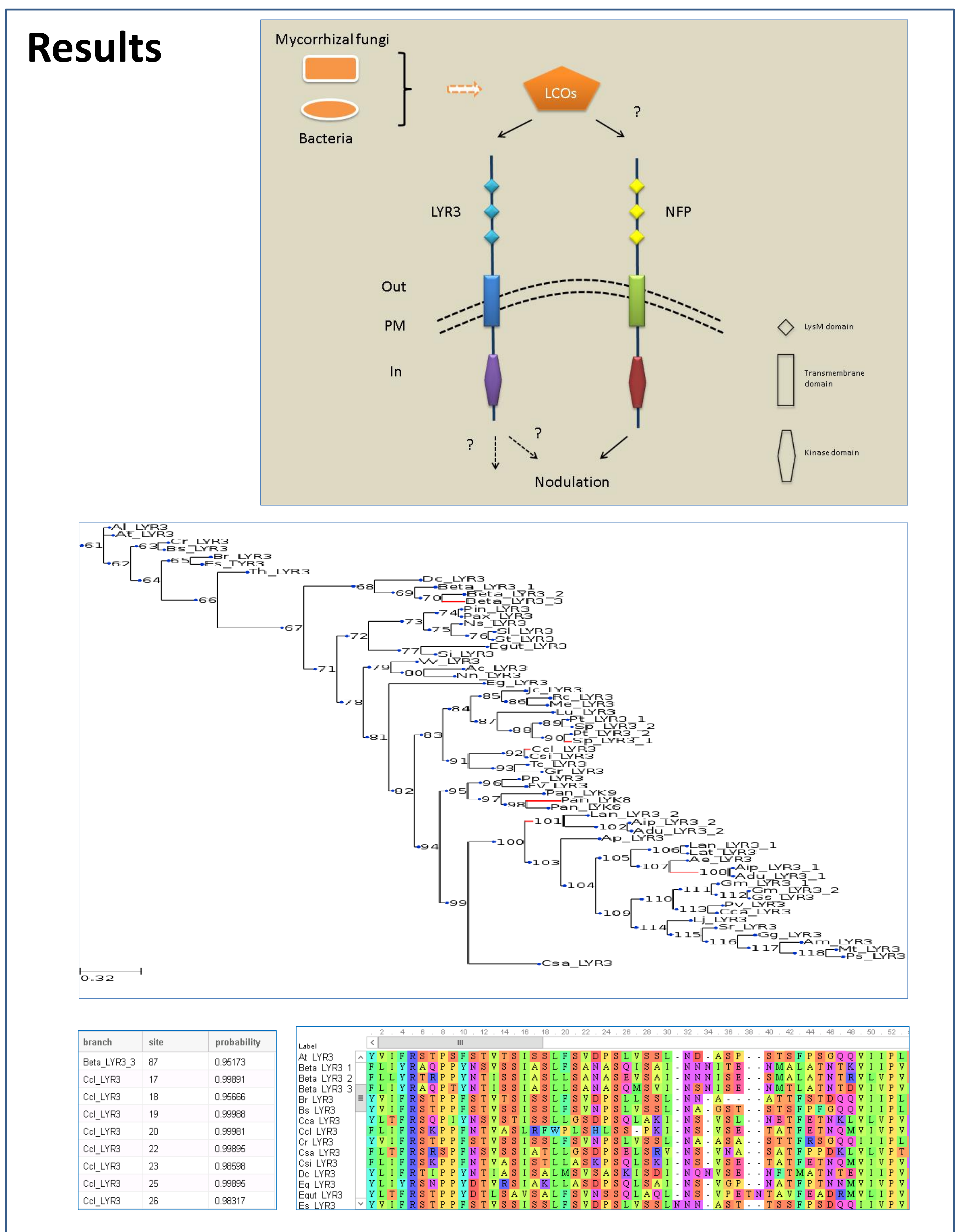
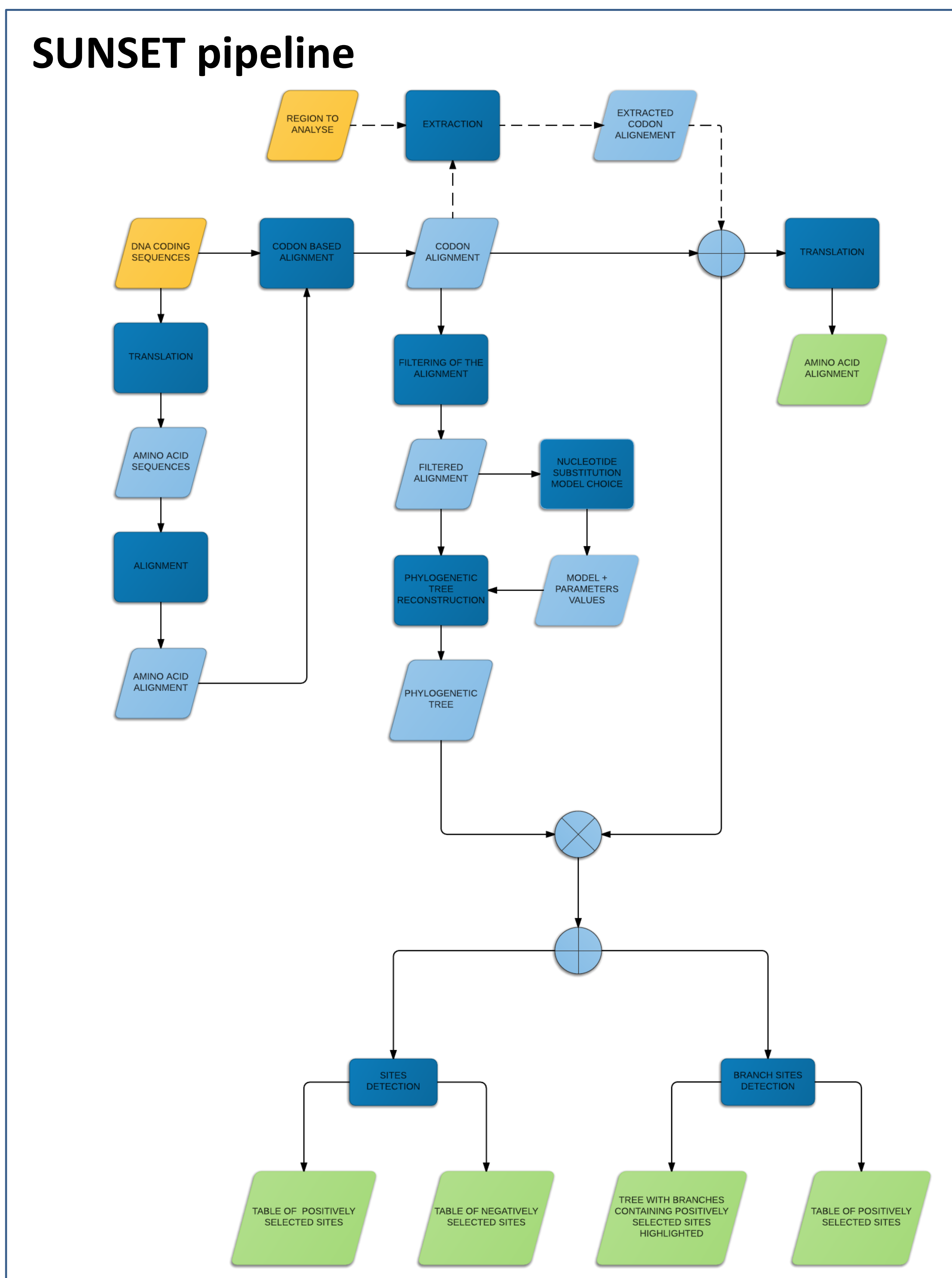
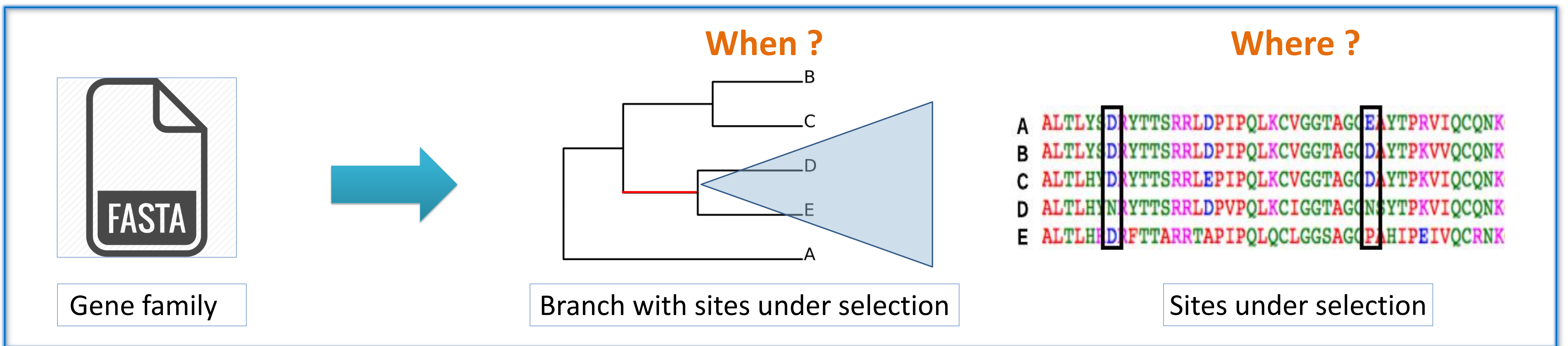


Maude Maréchaux, Jean-Jacques Bono, Clare Gough, Benoit Lefebvre, Ludovic Cottret
LIPM, Université de Toulouse, INRA, CNRS, Castanet-Tolosan, France

Abstract

Inferring patterns of natural selection in a family of genes can provide interesting insights about the evolutionary history of these genes and the divergence of their functions. **Phylogenetic maximum likelihood methods** are particularly appropriate and classically used for identification of natural selection in gene families. However, tools to apply this type of method are quite complex to use. First, a deep knowledge of the models involved in these methods is required. Furthermore, they necessitate most of the time a local installation and many file manipulations, what restricts their use to people with programming skills. The set of pipelines, **Sites Under SElection (SUNSET)** was designed to facilitate the use of these methods by uninitiated users. This set includes currently a pipeline allowing the **detection of sites under selection** and a pipeline allowing the **detection of sites under selection on various branches of a phylogenetic tree**. They were implemented in Python and use the ETE3 library for the calculations involved in detection of the selection. These pipelines require only one file and the choice of very few parameters. Besides, the calculations involved in detection of the selection by phylogenetic maximum likelihood methods, can be time-consuming, that is why SUNSET allows the realization of these **calculations in parallel**. Moreover, SUNSET was integrated into Galaxy to facilitate its use by biologists. Finally, SUNSET allowed the identification of sites under selection within two families of genes of legumes involved in the formation of mycorrhizal/rhizobial symbiotic interactions.



Implementation

References

Yang, Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586-1591

ETE 3: *Reconstruction, analysis and visualization of phylogenomic data*. Jaime Huerta-Cepas, Francois Serra and Peer Bork. *Mol Biol Evol* 2016

De Mita S. and M. Siol. 2012. EggLib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genet.* 13:27