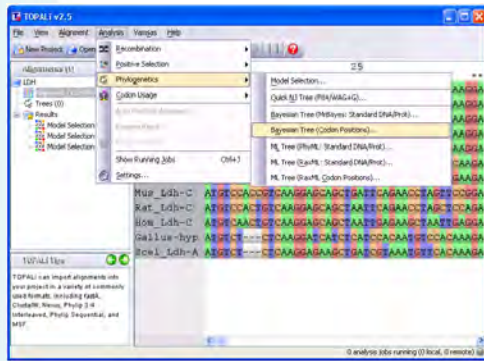


TOPALi v2

TOPALi software for statistical and evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops

INTRODUCTION

There are a growing number of biological questions that can be answered by analysing multiple sequence alignment data. We have extended the original TOPALi Java application, beyond recombination detection, to launch a range of statistical and evolutionary analyses of multiple sequence alignments as web services. TOPALi version 2 provides phylogenetic model selection, Bayesian analysis and Maximum Likelihood phylogenetic tree estimation, and also detection of sites under positive selection.



MULTIPLE SEQUENCE ALIGNMENT DATA

TOPALi inputs a wide range of alignment formats for nucleic acid and protein sequence data. TOPALi focuses particularly on protein alignments and cDNA alignments and allows the user to carry out protein-guided cDNA alignment to facilitate the use of a codon position model and positive selection analysis.

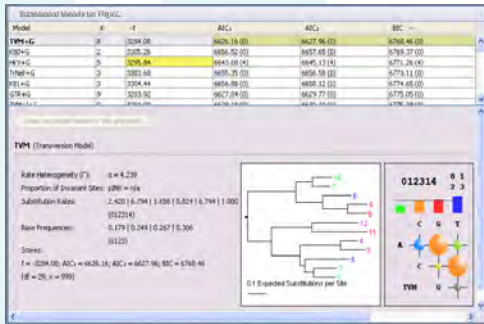
Alignment Overviews



Using the Alignment Overview feature you can view the entire alignment in a floating window while still working with a close-up area in detail. Scrolling around the overview automatically moves the main display too.

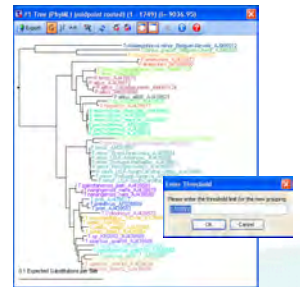
PHYLOGENETIC MODEL SELECTION

Model-based phylogenetic tree methods require the evolutionary model to be optimised prior to tree estimation. TOPALi has its own model selection web service to rank 88 amino acid (for protein) or 56 nucleotide substitution models (for DNA/RNA) according to four statistical criteria, including codon position models for protein-coding DNA.



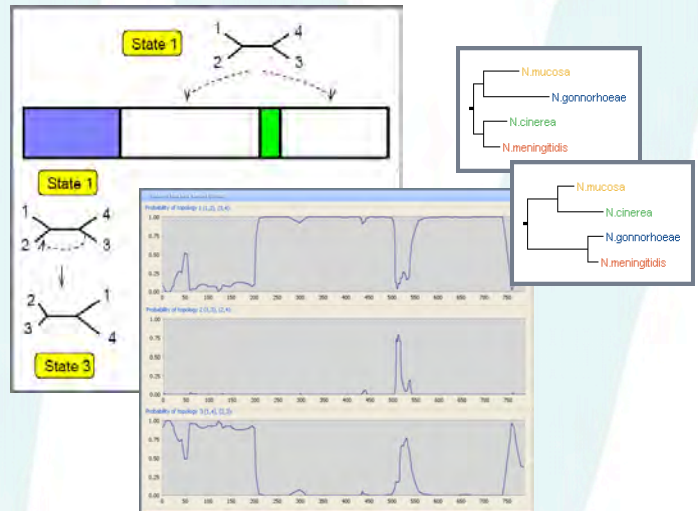
Tree-based approach to sample sequences

You can use sequence similarity to group sequences that lie close on the tree, and to distinguish the groups using colour. One sequence from each group can then be automatically selected for further analysis thus reducing computational time.

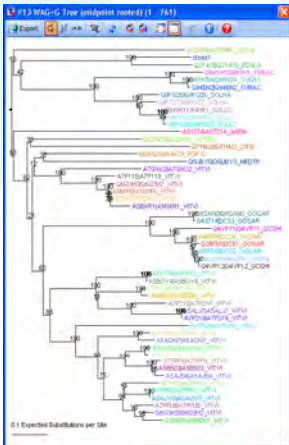


RECOMBINATION BREAKPOINT PREDICTION

Checking for recombination is an important step. If recombination breakpoints are detected then tree estimation and selection analyses should be undertaken on phylogenetically homogeneous partitions.



PHYLOGENETIC TREE ESTIMATION



Once the model is chosen, there are 3 tree estimation methods using Maximum Likelihood (PhyML and RaxML software) and Bayesian Analysis (MrBayes software) approaches for protein and nucleic acids (general DNA/RNA plus cDNA codon position model). There is also a fast approximate Neighbor Joining method with fixed model for very large numbers of sequences.

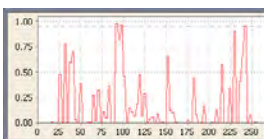
Trees are initially arbitrarily rooted but can be midpoint rooted or saved in standard tree format for rerooting outwith TOPALi.



ANCESTRAL SEQUENCES

TOPALi uses a Maximum Likelihood method (FASTML program) for ancestral sequence prediction. FASTML produces trees and alignments containing labelled ancestral nodes (N1, N2, and so on).

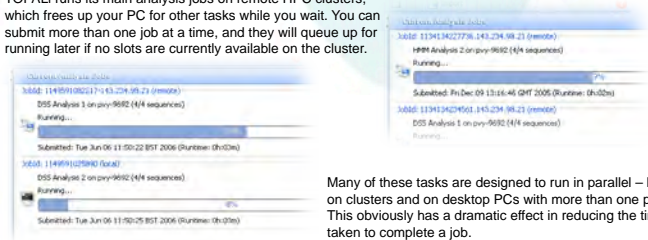
POSITIVE SELECTION ANALYSIS



TOPALi provides web service access to the PAML codeml program to carry out diversifying selection tests: (1) selection differences among branches within the tree, and (2) selection differences among sites (shown on left).

RUNNING JOBS REMOTELY AND IN PARALLEL

TOPALi runs its main analysis jobs on remote HPC clusters, which frees up your PC for other tasks while you wait. You can submit more than one job at a time, and they will queue up for running later if no slots are currently available on the cluster.



Many of these tasks are designed to run in parallel – both on clusters and on desktop PCs with more than one processor. This obviously has a dramatic effect in reducing the time taken to complete a job.

External PROGRAMS setup as WEB SERVICES:

Many thanks to the authors of the external programs setup as web services: PhyML, RaxML, FASTML, MrBayes, PAML/codeml, CODONW