

**PHYLOGENY:** This tutorial explains how to use the online phylogeny website at <http://www.phylogeny.fr/> to do distance and parsimony analyses.

Méthodes  
et  
Algorithmes  
pour la  
Bio-informatique  
LIRMM

Information  
Genomique et  
Structurale

Home Phylogeny Analysis Blast Explorer Online Programs Your Workspace Documentation

\*One Click\*  
\*Advanced\*  
\*A la Carte\*

Phylogeny.fr  
Phylogenetic Analysis For The Non-Spec

Start by selecting "A la Carte".  
This all

## Workflow Settings

Name of the analysis (optional):

Choose processing steps to run and select software to use:

Multiple Alignment:

- MUSCLE
- ProbCons
- T-Coffee
- 3D-Coffee
- ClustalW

We are using a dataset that has already been aligned, so this is unchecked. **CAUTION:** Always check your alignment! Bad alignment means useless tree.

Alignment curation:

- Gblocks
- Remove positions with gaps

This feature is an automated check on your alignment.

Construction of phylogenetic tree:

Maximum Likelihood

- PhyML

Parsimony

- TNT

Distances

- ProtDist/FastDist + BioNJ
- ProtDist/FastDist + Neighbor

Bayesian inference

- MrBayes (limit: 30 sequences)

These 4 methods are the most commonly used to “build” trees.

ProtDist calculates distances for proteins, FastDist for DNA.

Visualisation of phylogenetic tree:

- TreeDyn
- Drawgram
- Drawtree

These programs draw the tree.

## Run workflow:

all at once

step by step

**Create workflow**

Click "Create workflow" to continue. This will take you to the Neighbor-Joining workflow where you upload or paste in your data.

## Input Data

Upload your set of sequences in FASTA, EMBL or 1

Choose File No file chosen

Or paste it here [\(load example of sequences\)](#)

Here, I have copy and pasted in a DNA sequence alignment in FASTA format.

```
>STLSOA
---ATGGTCAGGTAGGGTGGAGGGTCTGGCC-AGCCCTTATACCCACATGGCGCAACG---
TGGGCACCAGTAACTCCTATGCTATAATACC---TGCTCTTCG-AGATC-CCAGTCTAAC
TATGATCATCGCCCGACGGGGCGAGATAGTCGTGGGTTCCCTTTCTGGAGGGAGAGGGAA
TTCCACGTTGACCGGGGAACCGGCCAGGCCCGGAAGGGAGCAACCGTGCCCGGCTATCC
GCGTTCGTCGGTCTCCGATAGGAGGA---AGACTGGGGTAAATCTCGGGGAGTAAGGGT
TATGGCATAGGGGAGCTGACCA-----T-----
>STLSOB
---GGGGTCAGGGAGGGTGGGGGATCTCGCCAATCCCTATTACCCGCAAGGCCTAATG--
CGGGCACCAGTAACTCCTACCCTATGGTGTC---TCCTATCTGTAGGTC-CCAGTGGAGC
GATGAAGCCTGCCAGCGGGCTTGGCGGTCATGGGCTTTCTCTCCGGAGGGAGAGAAAAG
TACCATGATAGCTGGGGGAATCGGCGAGGCCCGGAAGGGAGCAGCCGTGCCTGGACGCCA
GCGTTCGCTGGTCAACAGCCAGAGTG---AAACTGGGGTAAACCTATAGATAGGTA-GGC
CATGGGGTAGGGGTCTGGCCC-----CAT--
>PYROCC
```

Maximum number of sequences is 200 for proteins and 200 for nucleic acids.  
Maximum length of sequences is 2000 for proteins and 6000 for nucleic acids.

## Phylogeny: Neighbor

### Settings

Number of **bootstraps**:

Substitution model:

Gamma distribution parameter (for amino-acid substitution models):

Transition / transversion ratio (for nucleic acids):

For this alignment, we will build a tree using the default distance calculations. But these can be easily changed.

---

## Tree Rendering: TreeDyn

### Settings

Tree can be customized using the [dynamic tree editing interface](#).

Email is optional, but is good for slower methods such as Maximum Likelihood and Bayesian.

To receive the results by e-mail, enter **your address(es)**:

And, of course, Submit!

### Tree Rendering results

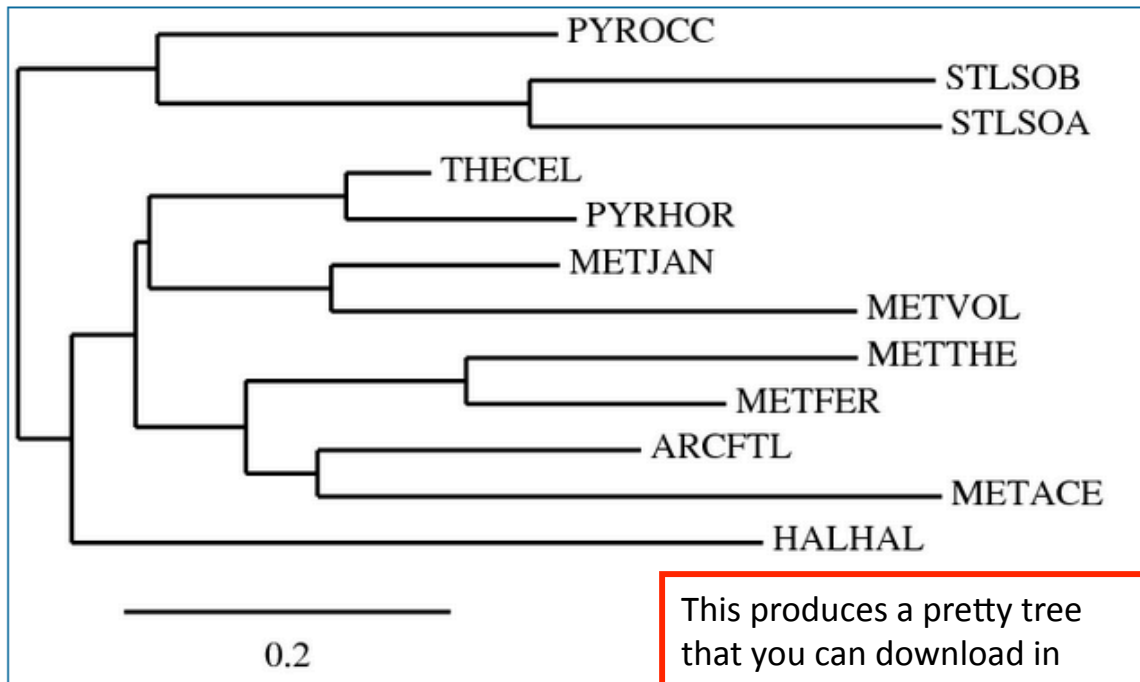


Figure 1: Phylogenetic tree.

⇒ Download the tree: [PNG](#) - [PDF](#) - [SVG](#) - [TGF \(Treedyn format\)](#) - [Newick](#) - [Text](#)

The TreeDyn Program has a lot of options to make your tree as you like it.

#### Select an action:

Reset (cancel all changes)

Mid-point rooting

Use Genbank information to automatically rename leaves by:  species and gi  species only  colorize

Collapse branches having branch support value smaller than  % or a number of bootstraps smaller than

#### Select an action and click leaf or internal branch:

Colorize  leaf  branch choose a color  and a legend label

Reroot (outgroup)

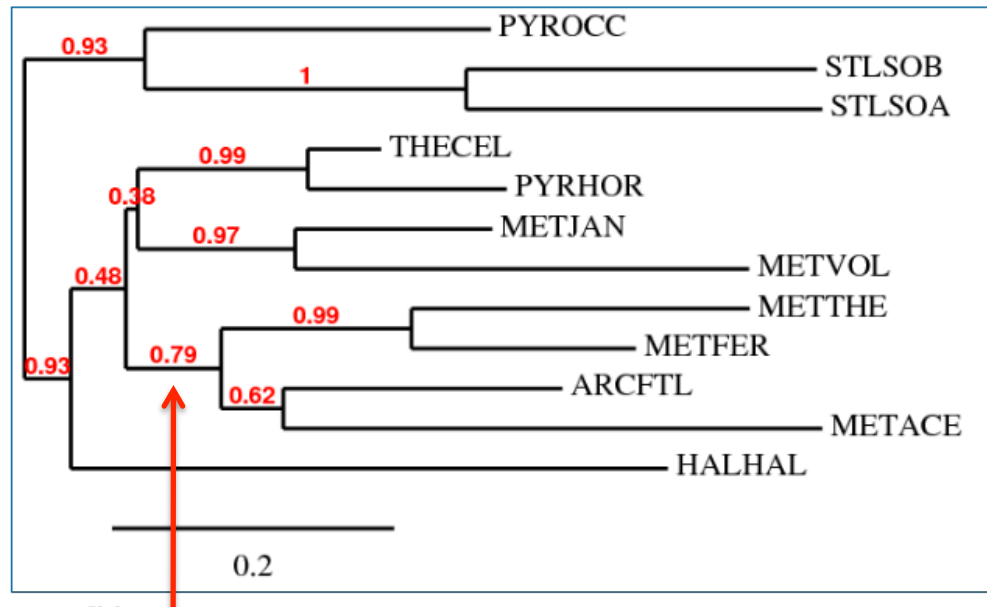
Flip (flip an entire tree at a node)

Swap (flip two branches at a node)

Change leaf name

Add leaf annotations color

Results of a bootstrap (100 replicates)  
using same dataset and Neighbor-Joining.



The numbers in RED indicate in what proportion of the 100 bootstrap replicates, this particular node was present. 0.79 = 79% of the time (79times out of the 100 bootstrap replicates)

To use the TNT program for Parsimony Analysis, you have to first agree to some “terms of service”. Then you can enter your data.

## Input Data

Upload your alignment (FASTA, Phylip, Clustal, EMBL or NEXUS format) from a file:

No file chosen

Or paste it here (load example of alignment)

```
#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=94 NCHAR=486;
FORMAT DATATYPE=PROTEIN INTERLEAVE MISSING=?

MATRIX
oOmpU_Van -----
MNKTLIALAVSAAAVVTGVN AG-----ELYNQDG TSLEMGGRAEARLSLKDGK-
oOmpU_Vch -----
MNKTLIALAVSAAAVATGAY ADGINQSGDKAGSTVYSAGK TSLEVGGRAEARLSLKDGK-
ompCTIGRPr1 -----MIINK
MRKSTIALSIFSALLVNTAN AA-----KVFNDGE SELNIHGRVQGMYYVSDDK
phoETIGRPr1 -----MIINK
MRKSTIALSIFSALLVNTAN AA-----KVFNDGE SELNIHGRVQGMYYVSDDK
```

In this example, we are using a protein sequence alignment in NEXUS format.

Maximum number of sequences: 200;  
Maximum alignment length: 6000.

## Phylogeny: TNT

### Settings

Search type:

- Traditional search
- New Technology search
  - Sectorial search (with RSS and CSS)
  - Ratchet
  - Drift
  - Tree fusing

Costs:

Amino-acids stepmatrix:

Please note that TNT is not very efficient (ie. slow) when amino-acids substitution stepmatrices are used.

Nucleic acids transversion cost:  (integer above 0)

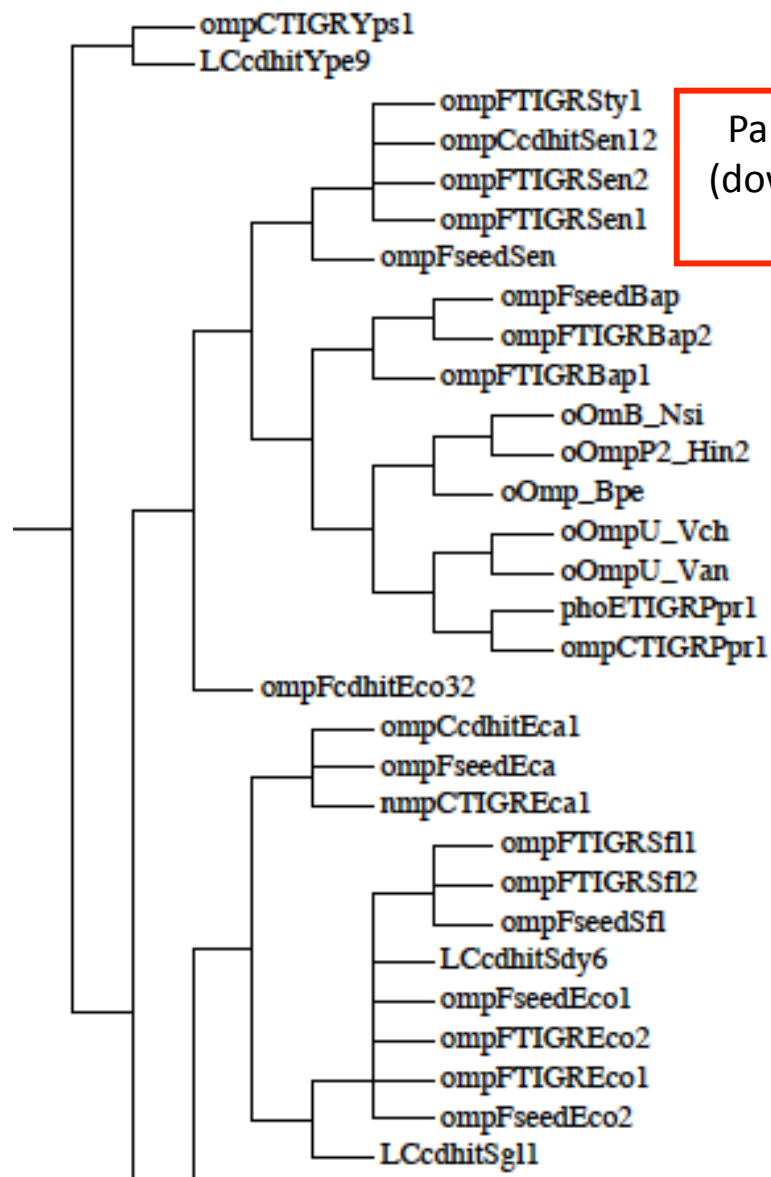
Resampling:

- No resampling
- Standard bootstrap
- Jackknife (with 36% removal probability)
- Symmetric resampling (with 33% change probability)

Number of replicats:

We left everything except the Search type as default. No bootstrap this time.

You know what to do now.



Part of the Tree  
(downloaded as a  
PDF file)