

M2 Mycologie

Approches phylogénétiques

- arbre phylogénétique : représentation graphique des relations (de proximité) entre différents organismes, dans un contexte d'évolution temporelle à partir d'un ancêtre commun à tous les descendants
- relations de similarité et phénomène de divergence extraites à partir de l'alignement des séquences
- phylogény moléculaire : analyse des données génétiques (ADN) et des différences moléculaires héréditaires
- objectif : caractériser le processus évolutif ayant permis de générer la diversité observée [1, 3]

- séquences non redondantes, séquences consensus
- aggrégation/classification (EST, OTU, famille de protéines)
- mesure de distance, blast, autre
- logiciels : cd-hit, vsearch, OrthoFinder

Etapes de construction d'un arbre

- sélection des organismes ou famille de gènes
- choix des marqueurs moléculaires
- amplification, séquençage et assemblage
- [*] alignement (ClustalW, MSA, MAFFT, T-Coffee)
- choix d'un modèle évolutif
- analyse phylogénétique (Paup, PAML, PHYLIP, MEGA, Raxml, MrBayes)
- construction de l'arbre phylogénétique
- évaluation de l'arbre phylogénétique (bootstrap, parsimonie/distance, vraisemblance)

- clustal, mafft, muscle¹
- algorithme progressif, temps de calcul, taille et nombre de séquences
- Mafft fournit une sortie Fasta par défaut, Mega dispose de son propre format d'exportation

¹<https://bioinformaticsreview.com/20151018/multiple-sequence-alignment/>

- plusieurs modèles de substitution : GTR +I/+G²
- GTR+I+ Γ = 10 paramètres
- compromis bias-variance, sur-ajustement, cadre de raisonnement hypothétique

²<https://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html> ; [2]

AIC-SELECTED MODEL: **GTR: General Time Reversible plus Gamma** (model 55)

LnL = -1315.630093

AIC = 2649.260186

[Parameter details](#)

The matrix on the right illustrates the different models considered by Findmodel.

Rate Parameters with the same color are assumed to have the same value.

Base Frequencies are the same when they are shown as f_N , and different otherwise.

General Time Reversible

	T	C	A	G
T	f_T	a	b	c
C	a	f_C	d	e
A	b	d	f_A	f
G	c	e	f	f_G

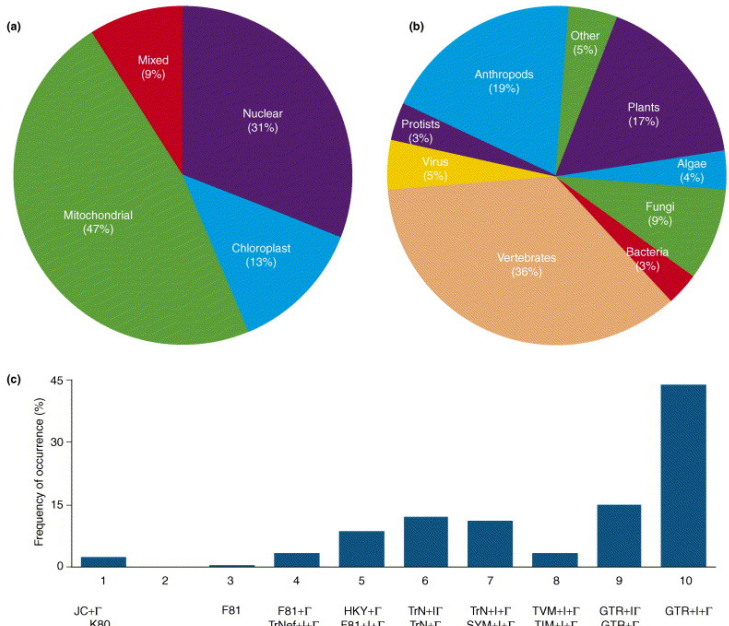
- utilisable directement depuis Seaview (sur séquences déjà alignées), ou en ligne
- utilisable également sous R (`plot` et `summary`)
- disponible en ligne sur NGPhylogeny

M5: Analysis Preferences

Options Summary

Option	Selection
Analysis	Phylogeny Reconstruction
Statistical Method	Maximum Likelihood
Phylogeny Test	
Test of Phylogeny	None
<i>No. of Bootstrap Replications</i>	<i>Not Applicable</i>
Substitution Model	
Substitutions Type	Amino acid
Model/Method	Jones-Taylor-Thornton (JTT) model
Rates and Patterns	
Rates among Sites	Uniform rates
<i>No of Discrete Gamma Categories</i>	<i>Not Applicable</i>
Data Subset to Use	
Gaps/Missing Data Treatment	Complete deletion
<i>Site Coverage Cutoff (%)</i>	<i>Not Applicable</i>
Tree Inference Options	
ML Heuristic Method	Nearest-Neighbor-Interchange (NNI)
Initial Tree for ML	Make initial tree automatically (Default - NJ/BioNJ)
<i>Initial Tree File</i>	<i>Not Applicable</i>

Modèles d'évolution [1]



- Evaluer la qualité du modèle
- Comparaison d'arbres
- Extraction de clades
- Combinaison d'arbres

- [1] Scot A. Kelchner and Michael A. Thomas. “Model Use in Phylogenetics: Nine Key Questions”. In: *TRENDS in Ecology and Evolution* 22.2 (2006), pp. 87–94.
- [2] Posada, D. and K. A. Crandall. “Modeltest: testing the model of DNA substitution”. In: *Bioinformatics* 14.9 (1998), pp. 817–818.
- [3] Ziheng Yang and Bruce Rannala. “Molecular phylogenetics: principles and practice”. In: *Nature Reviews Genetics* 13 (2012), pp. 303–314.