

M2 Mycologie

Approches phylogénétiques



Phylogénie moléculaire

- 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]

1

Classification de séquences

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

2

Etapas 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)

3

Alignement multiple

- clustal, mafft, muscle¹
- algorithmes progressifs, temps de calcul, taille et nombre de séquences
- Mafft fournit une sortie Fasta par défaut, Mega dispose de son propre format d'exportation
- Matrices de substitution (PAM, BLOSSUM)

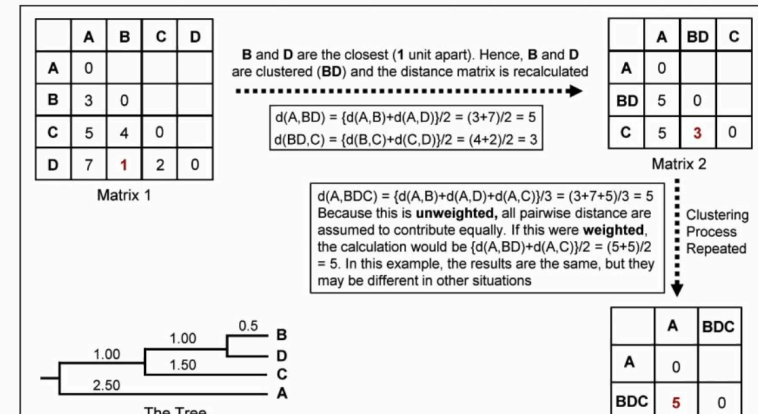
¹<https://bioinformaticsreview.com/20151018/multiple-sequence-alignment/>

4

Modèles basés sur les distances ou la parsimonie

(A)

Sequences being compared	# of substitutions out of 10 nt	A simple distance matrix computed			
		1	2	3	4
1-2: 2 (2/10 = 0.2)		1	0.2	0.3	0.1
1-3: 3 (3/10 = 0.3)			2	0.1	0.3
1-4: 1 (1/10 = 0.1)				3	0.4
2-3: 1 (1/10 = 0.1)					4
2-4: 3 (3/10 = 0.3)					
3-4: 4 (4/10 = 0.4)					



5

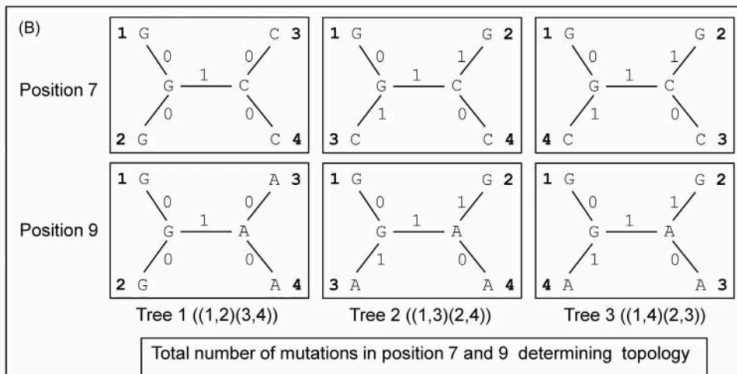
Modèles basés sur les distances ou la parsimonie (2)

(A)

Sequence	Position
1	A G G T G T G A G
2	A A C T G T G A G
3	A G A T G T C A A
4	A G T C G T C A A

Non-informative sites because no two bases occur in at least two sequences

Informative sites because each of the two alternative bases occurs in at least two sequences



6

Modèle de substitution

- 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]

7

Modèle GTR

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

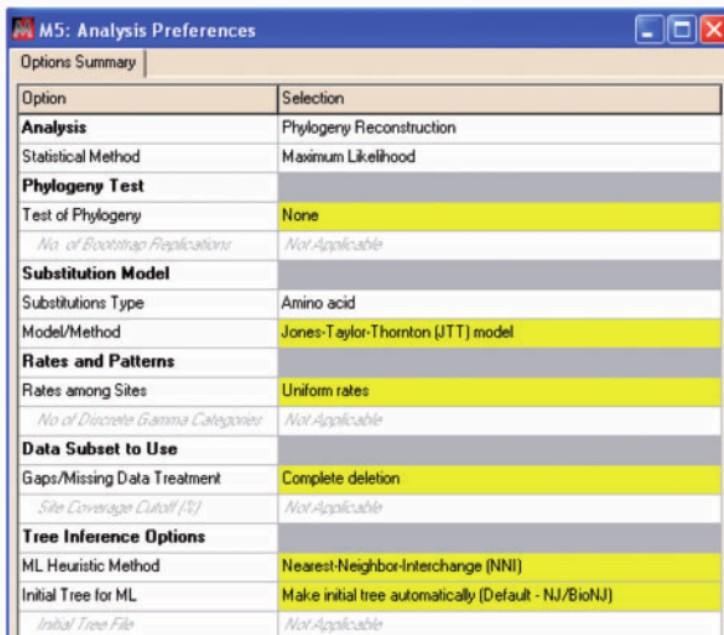
8

Phyml

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

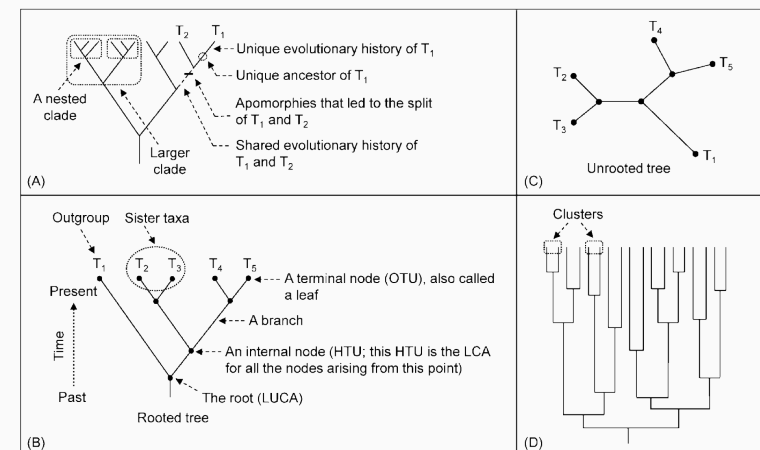
9

MEGA



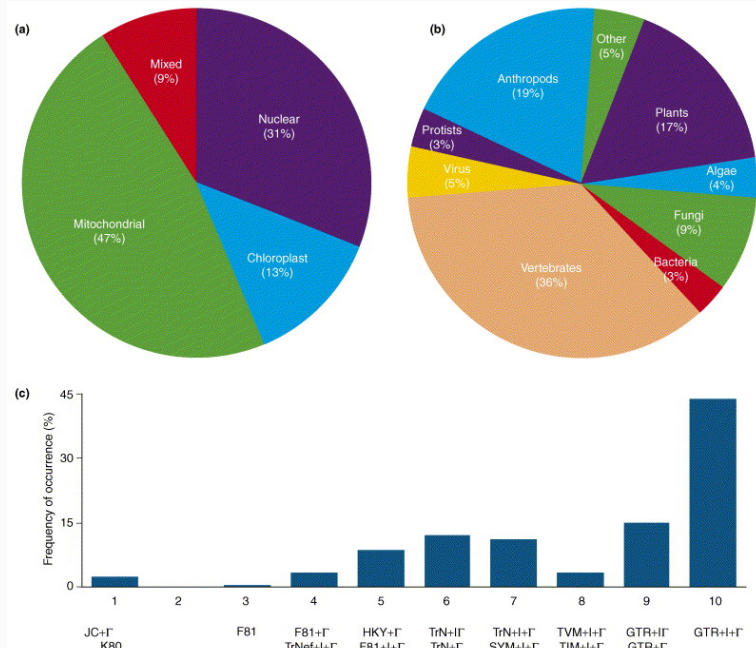
10

Nomenclature



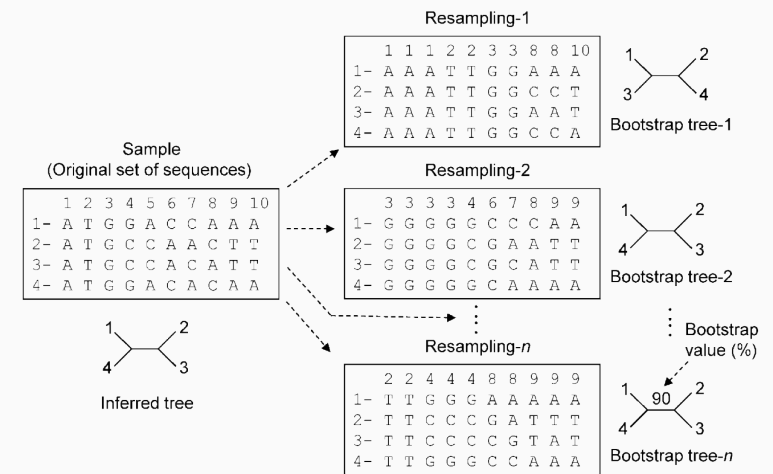
11

Modèles d'évolution [1]



12

Stabilité de l'arbre (bootstrap)



13

Applications

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

14

Références i

- [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.

15