

Tutoriel pour la délimitation des MOTUs d'Aselloidea

1) Récupérer dernier Alignement COI

2) Retirer les séquences qui doivent être enlevée (NUMT, SHORT, POLYM, CONTAM, SALE, primers, ...)
nom du fichier : *COI_Asellidae_clean.fas*

3) Remplacer les X dans l'alignement par des N
ambiguity2N.pl COI_Asellidae_clean.fas COI_Asellidae.fas

4) Renommer les séquences (sinon noms trop long, non acceptés dans certains logiciels ensuite)

Récupérer les noms de l'alignement :
seq2id.pl COI_Asellidae.fas fasta sequence_name.txt

Créer des nouveaux noms :
recode_e3s_seq.pl sequence_name.txt recod_sequence_name.tab

Renommer l'alignement :
rename_aln.pl COI_Asellidae.fas recod_sequence_name.tab -format fasta

Retirer annotations après noms alignement :
sed 's/.//g' COI_Asellidae.fas.recod > COI_Asellidae.fas.recod2*

5) Aligner les séquences

muscle
→ Attention remettre dans le cadre de lecture

6) Gblocks
-b2 → mettre la moitié des séquences
Gblocks COI_Asellidae.fas.recod2 -t=c -b5=h -p=t -b2=866

7) Collapser les séquences indentiques (https://github.com/TristanLefebure/collapse_to_uniq_seq)
collapse_to_uniq_seq.pl COI_Asellidae.fas.recod2-gb COI_Asellidae.fas.recod2-gb_haplo COI_seq_haplo.tab

8) Convertir au format Phylip
aln2aln.pl COI_Asellidae.fas.recod2-gb_haplo

9) Construire l'arbre d'haplotype
phyml -i COI_Asellidae.fas.recod2-gb_haplo.phylip -b -4 -m GTR -f m -v e -a e -s BEST

10) Enraciner l'arbre
sed s/\^//g COI_Asellidae.fas.recod2-gb_haplo.phylip_phyml_tree.txt > COI_Asellidae.fas.recod2-gb_haplo.phylip_phyml_tree_r.txt

RootedTree.R COI_Asellidae.fas.recod2-gb_haplo.phylip_phyml_tree_r.txt outgroup.txt

11) Définir les MOTUs

- En threshold
MotuTh.R COI_Asellidae.fas.recod2-gb_haplo.phylip_phyml_tree_r.txt_rooted

- En ptp
mptp --mcmc 400000 --single --mcmc_sample 400 --mcmc_burnin 40000 --tree_file COI_Asellidae.fas.recod2-gb_haplo.phylip_phyml_tree_r.txt_rooted --seed 123 --output_file MOTU_COI_Asellidae.fas.ptp_single.txt

12) Transforme sortie en tableau utilisable dans R
ptp2tab.py MOTU_COI_Asellidae.fas.ptp_single.txt.123.txt MOTU_COI_Asellidae.fas.ptp_single.txt.123.tab