

Calcul d'arbre. Phylogénie. – travail noté sur 20 à rendre pour dans 1 semaine (06/03/2017).

Merci de rédiger vos réponses – il est important d'arriver à se faire comprendre.

Calcul de la distance phylogénétique par analyse des tétramères

Dans le TD2 vous avez mis en œuvre les kmers.

Vous savez extraire une liste de kmers d'un fichier de nucléotides de type *.fna.

Si on considère les 4mers – kmers de taille 4 – vous pouvez extraire une liste de 4mers avec leur fréquences à partir du tableau donné.

Bacillus_subtilis	NC_000964
Bacillus_amyloliquefaciens_FZB42	NC_009725
Bacillus_pumilus_SAFR_032	NC_009848
Bacillus_thuringiensis_BMB171	NC_014171
Bacillus_cereus_03BB102	NC_012472
Bacillus_anthraxis_Ames	NC_003997
Bacillus_coagulans_2_6	NC_015634
Bacillus_atrophaeus_1942	NC_014639
Bacillus_licheniformis_ATCC_14580	NC_006322
Escherichia_coli_K_12_substr_MG1655	NC_000913
Pseudomonas_aeruginosa_LESB58	NC_011770
Rhodobacter_sphaeroides_ATCC_17025	NC_009428
Streptomyces_flavogriseus_ATCC_33331	NC_016114
Micrococcus_luteus_NCTC_2665_uid59033	NC_012803
Lactococcus_lactis_II1403	NC_002662

Vous calculerez une distance entre deux vecteurs et vous obtiendrez une matrice symétrique positive.

Cette matrice sera validée avec **R**. Vous savez utiliser R pour construire une clusterisation hiérarchique – et de ce fait pour la suite de ce TD, dans votre rapport vous afficherez les résultats obtenus avec R – en donnant les commandes que vous avez utilisées.

Vous comparerez votre résultat avec celui obtenu dans le TD3 (merci de remettre les figures si vous citez vos résultats) – et vous discuterez des différences et/ou des ressemblances.

Vous déposerez votre code sur la forge «eldarsoft » . Vous y inclurez votre rapport au format PDF.

Il est inutile de dupliquer les fichiers *.fna – merci de ne pas déposer dans votre espace de livraison les fichiers *.fna