

	<p align="center">ANNEE UNIVERSITAIRE 2018-2019 SESSION 1 D'AUTOMNE DECEMBRE 2018</p> <p>MENTION : BIOINFORMATIQUE Code UE : 4TBI703U Intitulé de l'épreuve : Omiques et Bioinformatiques : Date : 19/12/2018 Heure : 11H30 Durée : 1h30 Documents : autorisés / non autorisés Epreuve de M/Mme : R. Uricaru, P. Thébault et JC Taveau</p>	<p align="center">Collège Sciences et technologies</p> <p align="center">Masters</p>
--	--	--

Exercice 1 : Une étude a été menée par sur la diversité de 8 génomes de truffes par Murat C *et al.* (Nat Ecol Evol. 2018;2).

*Tuberaceae is one of the most diverse lineages of symbiotic truffle-forming fungi. To understand the molecular underpinning of the ectomycorrhizal truffle lifestyle, we compared the genomes of Piedmont white truffle (*Tuber magnatum*), Périgord black truffle (*Tuber melanosporum*), Burgundy truffle (*Tuber aestivum*), pig truffle (*Choiromyces venosus*) and desert truffle (*Terfezia boudieri*) to saprotrophic *Pezizomycetes*. Reconstructed gene **duplication/loss** histories along a time-calibrated phylogeny of *Ascomycetes* revealed that *Tuberaceae*-specific traits may be related to a higher gene diversification rate.*

Les deux figures suivantes sont extraites de cette publication :

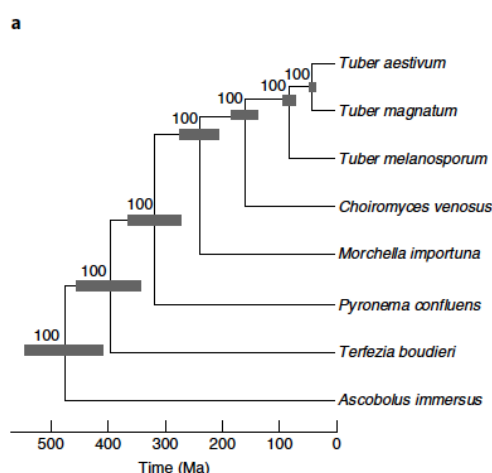


Fig1.1 : The maximum likelihood phylogeny is constructed using 2,093 conserved single copy orthologous *Pezyzomytes* genes.

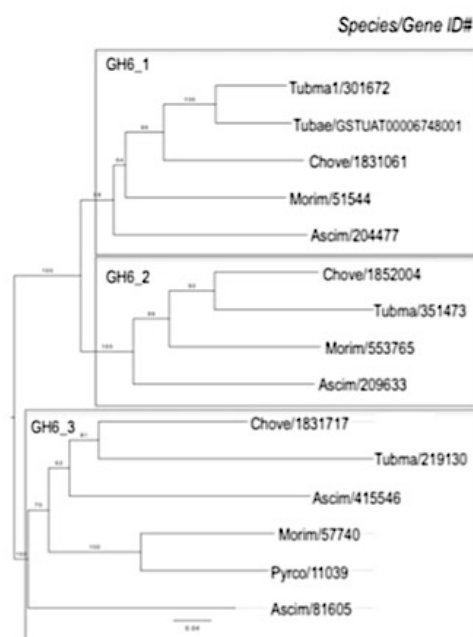


Fig 1.2 : Evolution of the GH6 cellobiohydrolase genes in *Tuberaceae*. Neighbour-joining tree of GH6_1, GH6_2, and GH6_3 cellobiohydrolases in the sequenced *Pezizomycetes*. The name of each species is given with the id of the sequence (*species/ID*). Tubma: *Tuber magnatum*; Tubae: *Tuber aestivum*; Chove: *Choiromyces venosus*; Morim: *Mochella importuna*; Ascim: *Ascobolus immersus*. Pyrco: *Pyronema confluens*.

1- Quel type d'histoire évolutive est décrit dans chacune des deux figures.

2- Décrivez le fonctionnement de la méthode utilisée pour construire la phylogénie de la famille des gènes GH6.

3- En vous appuyant sur les deux figures:

- Pour chacune des 8 espèces, identifiez le nombre de gènes GH6.
- Interprétez l'évolution de la famille de gènes.

Exercice 2 : Des séquences et du Python

Note 1 : Les fonctions demandées dans cet exercice seront écrites en langage Python.

Note 2 : Pensez à réutiliser vos fonctions.

1.1. Étant donnée la matrice de substitution suivante (à utiliser dans la suite de l'exercice) :

	A	C	G	T
A	4	0	2	0
C	0	4	0	2
G	2	0	4	0
T	0	2	0	4

```
nucIMat = [  
    #A C G T  
    [4,0,2,0], # A  
    [0,4,0,2], # C  
    [2,0,4,0], # G  
    [0,2,0,4] # T  
]
```

Fig. 2.1 :Matrice de substitution. Pour le code Python, on définira la matrice comme une **List** de **List** (à droite)

Écrire une fonction **computeAlignmentScore(seq1, seq2,matrix)** qui calcule et renvoie le score d'alignement pour 2 séquences nucléiques, **seq1** et **seq2**, de même longueur. L'alignement entre les 2 séquences se fera uniquement sur le brin sens et ne comportera que des identités et des substitutions (pas d'insertions/délétions).

Exemple : **computeAlignmentScore('AGGCT','AGGAT',nucIMat)** renverra 16.

1.2. Écrire une fonction **reverseComplement(seq,matrix)** qui calcule et renvoie la séquence complémentaire d'une séquence seq passée en argument.

Exemple : **reverseComplement ('AGGCT',nucIMat)** renverra 'AGCCT'.

1.3. Écrire une fonction **bestAlignmentScore(seq1, seq2,matrix)** qui renvoie le meilleur score d'alignement parmi toutes les combinaisons possibles entre les brins sens et anti-sens, entre les 2 séquences **seq1** et **seq2**.

Exemple : **bestAlignmentScore ('ATGGC', 'GCCAC',nucIMat)** renverra 18.

1.4. Écrire une fonction **bestPair(seq, sequences,matrix)** qui prend en argument une séquence nucléique seq et un ensemble de N séquences nucléiques (ayant toutes la même longueur que seq) et qui renvoie la séquence parmi les N qui donne le meilleur score d'alignement avec seq (en prenant en compte les brins sens et anti-sens).

Exemple : **bestPair('ATGGC',['GCCAC','GGCAA','GAGAA'],nucIMat)** renverra 'GCCAC'.

1.5. Modifier la fonction précédente (1.4) pour écrire une nouvelle fonction **alignAllPairs(sequences,matrix)** qui prend en argument un ensemble de N séquences nucléiques de même longueur et qui renvoie le meilleur score d'alignement parmi les brins sens et anti-sens pour toutes les paires de séquences de cet ensemble.

Exemple : **alignAllPairs(['ATGGC','GCCAC','GGCAA','GAGAA'],nucIMat)** renverra 18 (le meilleur score d'alignement obtenu pour le couple de séquences 'ATGGC' et 'GCCAC').