

## TRAVAUX PRATIQUES BCH 4037-Bioinformatique

### Énoncé du problème

Plusieurs patients hospitalisés dans une région forestière du Cameroun présentent des symptômes sévères : fièvre, fatigue intense, et éruptions cutanées. Les médecins suspectent une infection liée à un pathogène émergent. Rapidement ils sont isolés et des échantillons de sang sont prélevés sur chacun d'eux pour identifier l'agent pathogène afin de mettre des stratégies en place pour contrôler l'infection.

Pour ce faire, le Ministre de la Santé fait appel à deux étudiants de l'Université de Yaoundé 1, spécialisés dans les Biotechnologies et la Bio-informatique pour identifier l'agent pathogène. Après séquençage d'un extrait du matériel génétique de l'agent pathogène, le premier étudiant obtient une **séquence 1** de 2238 bp tandis que le deuxième obtient une **séquence 2** de 2237 bp.

Votre mission est de :

1. Identifier le pathogène responsable.
2. Étudier les variations génétiques éventuelles.
3. Ressortir l'arbre phylogénétique et Identifier des régions conservées avec des espèces apparentées pour envisager des applications en santé (diagnostics, thérapies).

### Activité 1 : Identification de l'agent pathogène

Question :

- 1- Quels outils bioinformatiques utiliserez-vous pour identifier les séquences obtenues ?
- 2- Quelle est l'espèce à laquelle appartiennent les deux séquences ?
- 3- Quel est le meilleur pourcentage d'identité obtenu pour chaque séquence par rapport à la base de données ?
- 4- Comment interpréter ces résultats pour confirmer l'agent pathogène responsable ?

Étapes :

- 1- Ouvrir NCBI <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

- 2- Sélectionner le BLAST qui convient
- 3- Copier la première séquence 1 et coller dans la boîte en dessous de « Enter Query Sequence »
- 4- Se rendre à la fin de la page ouverte et faire un clic sur « BLAST » exécuter le programme. Patienter jusqu'à l'obtention du résultat du Blast.
- 5- Refaire l'étape 1 à 4 avec la séquence 2.

### **Activité 2 : Étude des variations génétiques éventuelles.**

Après exécution d'un programme bioinformatique, l'un des étudiants trouve une séquence appartenant à un organisme avec un pourcentage d'identité de 100% tandis que l'autre trouve un pourcentage d'identité de 99,96% pour la même espèce. Voyant ces résultats, votre superviseur vous dit qu'il s'agirait probablement d'une mutation qui a eu lieu dans le gène et vous demande donc d'identifier la mutation en question ainsi que sa position.

### **Questions**

- 1- Quels outils utiliseriez-vous pour localiser cette mutation et pourquoi ?
- 2- De quel type de mutation s'agit-il, et quelle est sa position exacte ?
- 3- Quelles pourraient être les conséquences biologiques de cette mutation (pathogénicité, résistance, etc.) ?

### **Étapes :**

- 1- Ouvrez EMBOSS [https://www.ebi.ac.uk/jdispatcher/psa/emboss\\_needle](https://www.ebi.ac.uk/jdispatcher/psa/emboss_needle)
- 2- Au niveau de « **Sequence type** » choisir DNA
- 3- Coller la première séquence 1 dans le premier box
- 4- Coller la deuxième séquence 2 dans le deuxième box
- 5- Aller à la fin de la page et cliquer sur « submit »
- 6- Patienter jusqu'à l'obtention du résultat pour voir les alignements entre les séquences.

### **Activité 3 : Identification d'un motif conservé entre plusieurs espèces**

Des recherches supplémentaires révèlent que plusieurs espèces proches du pathogène identifié sont disponibles dans les bases de données publiques. Votre superviseur souhaite comprendre les relations évolutives entre ces espèces afin d'éclairer la dynamique de transmission et de proposer des approches ciblées pour le contrôle de la maladie.

### Questions

1. Pourquoi est-il essentiel d'étudier les relations phylogénétiques entre des espèces apparentées ?
2. Utilisez l'outil Clustal Omega pour ressortir l'arbre phylogénétique, entre les séquences des espèces suivantes : *Homo sapiens*; *Plasmodium vivax* ; *Plasmodium knowlesi* ; *Plasmodium falciparum* ; *Plasmodium malariae*?
3. À partir de votre rapport et des résultats obtenus, le MINSANTE vous informe disposer d'un médicament capable d'inhiber la fonction du gène que vous avez identifié chez l'agent pathogène. En analysant les alignements réalisés entre les séquences humaines et celles des espèces apparentées à l'agent pathogène identifié, la question qui vous est posée est la suivante : ce médicament peut-il être administré à l'humain pour traiter l'infection causée par ce pathogène ?

### Étapes :

- 1- Ouvrir <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>
- 2- Au niveau de « **Sequence Type** » choisir DNA
- 3- Copier et coller dans le boxe les séquences contenues dans les fichiers fournis, y compris les entêtes de chaque fichier. (NB : pour ouvrir ces fichiers, faites un clique droit et choisissez « ouvrir avec » ensuite sélectionner « TextEdit »
- 4- Aller à la fin de la page et cliquer sur « submit »
- 5- Patienter jusqu'à obtention du résultat.
- 6- Naviguer dans les différentes entrées pour voir les alignements et l'arbre phylogénétique.

Séquence 1 :

ATGTCAACGGAAACATTTCGCATTTAACGCCGACATCAGGCAGTTGATGAGTTTGAT  
TATCAACACTTTTTACAGTAACAAAGAAATATTTTTAAGAGAATTGATTAGTAATGC  
TAGTGATGCCTTAGATAAAAATAAGATATGAATCAATTACAGATACTCAAAAATTATCT  
GCTGAGCCTGAATTTTTTATTCGTATCATTCTGACAAAACCAACAATACATTAACT  
ATTGAAGATTCAGGTATTGGTATGACAAAAAATGATTTAATTAATAACCTTGGTACT  
ATTGCAAGATCAGGAACCAAAGCTTTTATGGAAGCCATACAAGCCAGTGGAGATAT  
ATCTATGATTGGTCAATTTGGTGTGGTTTTTATTCAGCCTATTTAGTTGCTGATCAT  
GTTGTTGTTATCTCCAAAATAATGATGATGAACAATATGTTTGGGAATCTGCTGCA  
GGAGGTTCCCTCACAGTTACTAAGGATGAAACCAATGAAAACTTGGAAAGAGGTA  
CGAAAATTATTCTTCATTTAAAAGAAGATCAATTAGAATATCTTGAAGAAAAACGTA  
TCAAAGATTTAGTTAAGAAACACTCTGAATTTATCTCTTTCCCAATCAAGTTATACT  
GTGAAAGGCCAAAATGAAAAAGAAATCACCGCATCTGAAGAAGAAGAAGGAGAAG  
GAGAAGGAGAAAGAGAAGGAGAAGAAGAAGAAGAAAAAAAAAAAAAAAAAACAGG  
CGAAGATAAAAATGCTGATGAAAGTAAAGAAGAAAATGAAGATGAAGAAAAAAA  
AGAAGATAACGAAGAAGATGATAACAAAACACTGATCATCCAAAAGTTGAAGATGTT  
ACCGAAGAATTAGAAAATGCTGAAAAGAAAAAAAAAAGAAAAAGAAAAAAAAAA  
AATACACACAGTTGAACATGAATGGGAAGAATTAATAAACAAAACCATTATGGA  
TGAGAAAACCAGAAGAAGTTACAAATGAAGAATATGCAAGCTTCTATAAATCATT  
ACAAATGATTGGGAAGACCATTTAGCTGTAAACATTTCTCTGTTGAAGGACAATT  
AGAATTTAAAGCCTTATTATTTATACCAAAAAGAGCACCTTTTGATATGTTTCGAAAA  
TAGAAAAAAAAAGAAATAATATCAAATTATATGTAAGAAGAGTTTTTATTATGGATGA  
TTGTGAAGAAATTATTCCAGAATGGTTAAATTTTGTAAAGGGTGTGTCGATTCAGA  
AGATTTACCACTTAATATTTCAAGAGAATCATTACAACAAAATAAAATACTTAAGGT  
TATCAAAAAAAAAACCTTATCAAAAATGTTTAGACATGTTCTCAGAATTAGCTGAAA  
ATAAGGAAAACACTACAAAAGTTTTATGAACAATTCAGCAAAAACCTTAAAGTTGGG  
TATCCACGAGGATAACGCAAATCGTACAAAGATCACCGAATTACTCCGATTCCAAA  
CCTCAAAATCAGGAGACGAAATGATCGGATTAAGAATACGTAGACAGAATGAA  
GGAAAACCAAAGGATATTTACTATATCACCGGTGAATCCATCAATGCTGTTTCTAA  
TTCTCCATTCTTAGAAGCTTTGACCAAAAAGGATTCGAAGTTATTTATATGGTTGA  
TCCTATTGATGAATATGCAGTACAACAATTAAGAATTTTGATGGTAAGAAATTGAA  
ATGTTGTACCAAAGAAGGTTTAGATATTGATGATTCAGAAGAAGCCAAAAAAGATT  
TCGAAACCTTGAAAGCTGAATATGAAGGATTATGCAAAGTTATTAAGACGTATTA  
CACGAGAAAGTTGAAAAAGTTGTTGTAGGACAAAGAATTACAGATTCTCCATGTG  
TATTAGTCACATCAGAATTTGGATGGTCCGCAAACATGGAAAGAATTATGAAAGCT

CAAGCATTAAGAGATAATTCCATGACTAGCTATATGTTATCCAAAAAATTATGGAA  
ATCAATGCTCGTCACCCAATTATATCAGCATTAAAACAAAAAGCTGATGCAGATAA  
ATCAGATAAAACCGTTAAAGATTTAATCTGGTTATTATTTGATACCTCTTTATTAACA  
TCTGGTTTTGCTCTTGAAGAACCAACTACCTTTTCTAAAAGAATCCACAGAATGAT  
TAAATTAGGTTTATCAATAGATGAAGAAGAAAACAATGATATCGATTTACCACCTCT  
TGAAGAACTGTAGATGCAACCGATTCTAAAATGGAAGAAGTTGACTAA

Séquence 2 :

ATGTCAACGGAAACATTTCGCATTTAACGCCGACATCAGGCAGTTGATGAGTTTGAT  
TATCAACACTTTTTACAGTAACAAAGAAATATTTTTAAGAGAATTGATTAGTAATGC  
TAGTGATGCCTTAGATAAAAATAAGATATGAATCAATTACAGATACTCAAAAATTATCT  
GCTGAGCCTGAATTTTTTATTCGTATCATTCTGACAAAACCAACAATACATTA  
ACTATTGAAGATTCAGGTATTGGTATGACAAAAAATGATTTAATTAATAACCTTGGTATAT  
TGCAAGATCAGGAACCAAAGCTTTTATGGAAGCCATACAAGCCAGTGGAGATATAT  
CTATGATTGGTCAATTTGGTGTGGTTTTTATTCAGCCTATTTAGTTGCTGATCATGT  
TGTTGTTATCTCCAAAATAATGATGATGAACAATATGTTTGGGAATCTGCTGCAGG  
AGGTTCTTCACAGTTACTAAGGATGAAACCAATGAAAACTTGGAAGAGGTACG  
AAAATTATTCCTCATTAAAAGAAGATCAATTAGAATATCTTGAAGAAAAACGTATC  
AAAGATTTAGTTAAGAAACACTCTGAATTTATCTCTTTCCCAATCAAGTTATACTGT  
GAAAGGCAAATGAAAAAGAAATCACCGCATCTGAAGAAGAAGAAGGAGAAGG  
AGAAGGAGAAAGAGAAGGAGAAGAAGAAGAAGAAAAAAGAAAAAAGAAAAAAGGC  
GAAGATAAAAATGCTGATGAAAGTAAAGAAGAAAATGAAGATGAAGAAAAA  
GAAGATAACGAAGAAGATGATAACAAAACCTGATCATCCAAAAGTTGAAGATGTTA  
CCGAAGAATTAGAAAATGCTGAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA  
ATACACACAGTTGAACATGAATGGGAAGAATTAATAAACAACCAATATGGAT  
GAGAAAACCAGAAGAAGTTACAAATGAAGAATATGCAAGCTTCTATAAATCATTAA  
CAAATGATTGGGAAGACCATTTAGCTGTAAACATTTCTCTGTTGAAGGACAATTA  
GAATTTAAAGCCTTATTATTTATACCAAAAAGAGCACCTTTTGATATGTTTCGAAAAT  
AGAAAAAAGAAATAATATCAAATTATATGTAAGAAGAGTTTTTATTATGGATGAT  
TGTGAAGAAATTATCCAGAATGGTTAAATTTTGTTAAGGGTGTGTCGATTCAGA

AGATTTACCACTTAATATTTCAAGAGAATCATTACAACAAAATAAAATACTTAAGGT  
TATCAAAAAAACCTTATCAAAAAATGTTTAGACATGTTCTCAGAATTAGCTGAAA  
ATAAGGAAAACCTACAAAAAGTTTTATGAACAATTCAGCAAAAACTTAAAGTTGGG  
TATCCACGAGGATAACGCAAATCGTACAAAGATCACCGAATTACTCCGATTCCAAA  
CCTCAAAATCAGGAGACGAAATGATCGGATTAAGAATAACGTAGACAGAATGAA  
GGAAAACCAAAAGGATATTTACTATATCACCGGTGAATCCATCAATGCTGTTTCTAA  
TTCTCCATTCTTAGAAGCTTTGACCAAAAAAGGATTCGAAGTTATTTATATGGTTGA  
TCCTATTGATGAATATGCAGTACAACAATTAAGAATTTTGATGGTAAGAAATTGAA  
ATGTTGTACCAAAGAAGGTTTAGATATTGATGATTCAGAAGAAGCCAAAAAAGATT  
TCGAAACCTTGAAAGCTGAATATGAAGGATTATGCAAAGTTATTAAGACGTATTA  
CACGAGAAAGTTGAAAAAGTTGTTGTAGGACAAAGAATTACAGATTCTCCATGTG  
TATTAGTCACATCAGAATTTGGATGGTCCGCAAACATGGAAAGAATTATGAAAGCT  
CAAGCATTAAAGAGATAATTCCATGACTAGCTATATGTTATCCAAAAAATTATGGAA  
ATCAATGCTCGTCACCCAATTATATCAGCATTAAAACAAAAAGCTGATGCAGATAA  
ATCAGATAAAACCGTTAAAGATTTAATCTGGTTATTATTTGATACCTCTTTATTAACA  
TCTGGTTTTGCTCTTGAAGAACCAACTACCTTTTCTAAAAGAATCCACAGAATGAT  
TAAATTAGGTTTATCAATAGATGAAGAAGAAAACAATGATATCGATTTACCACCTCT  
TGAAGAACTGTAGATGCAACCGATTCTAAAATGGAAGAAGTTGACTAA