

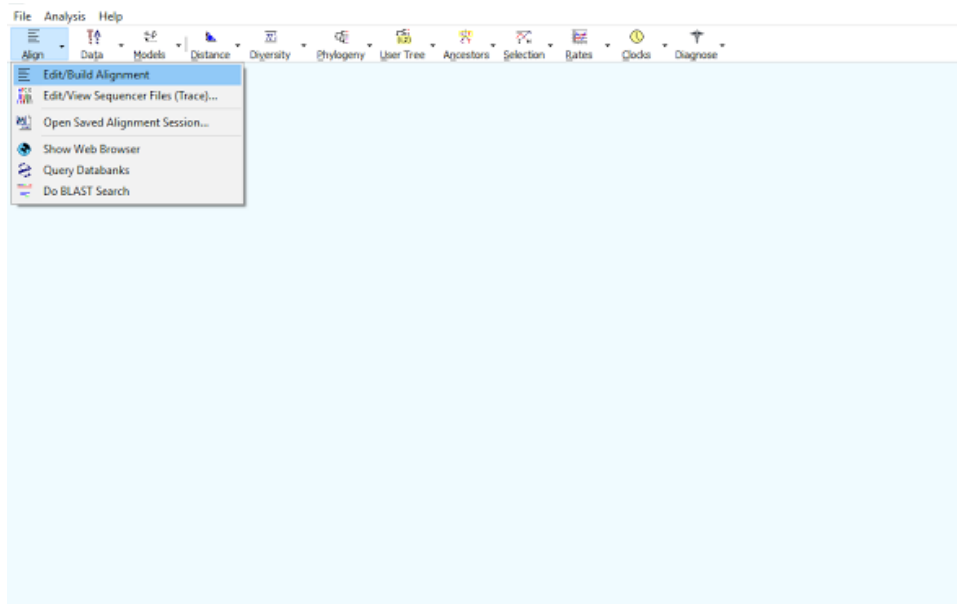
# Construire un arbre phylogénétique avec MEGA7 (Molecular Evolutionary Genetics Analysis Version7)

## Avant de commencer:

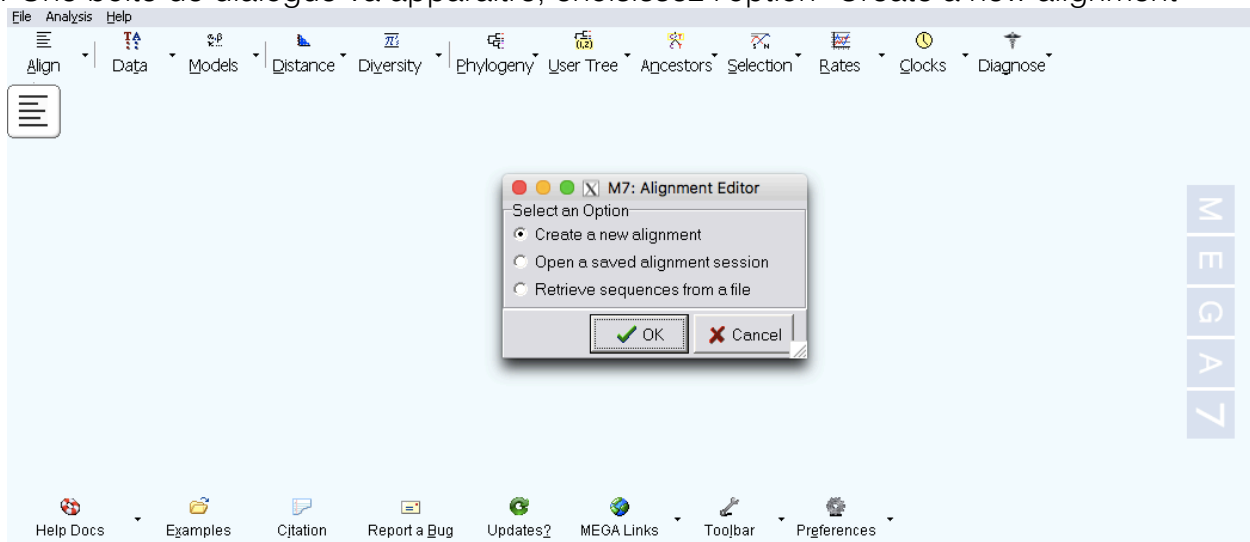
1. Si vous n'avez toujours pas MEGA7 sur votre ordinateur, téléchargez la version graphique du programme sur <http://www.megasoftware.net/> et installez-le.
2. Assurez-vous d'avoir téléchargé les fichiers de séquence dans une extension fasta. Vous aurez besoin d'au moins 3 fichiers pour chaque scénario: les séquences de référence, l'outgroup et le (s) fichier (s) avec les séquences à étudier.

## Première étape : importer les séquences dans MEGA7

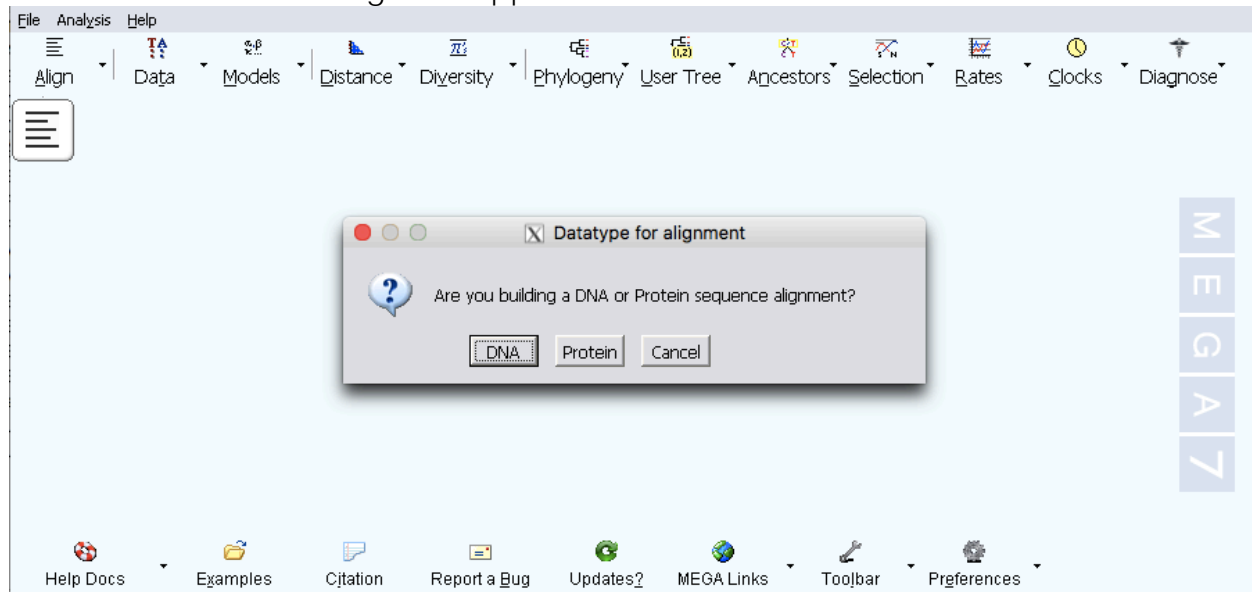
1.0 Sélectionnez l'option "Align" et choisissez "Edit/build Alignment" en haut à gauche dans la barre d'outils de MEGA7



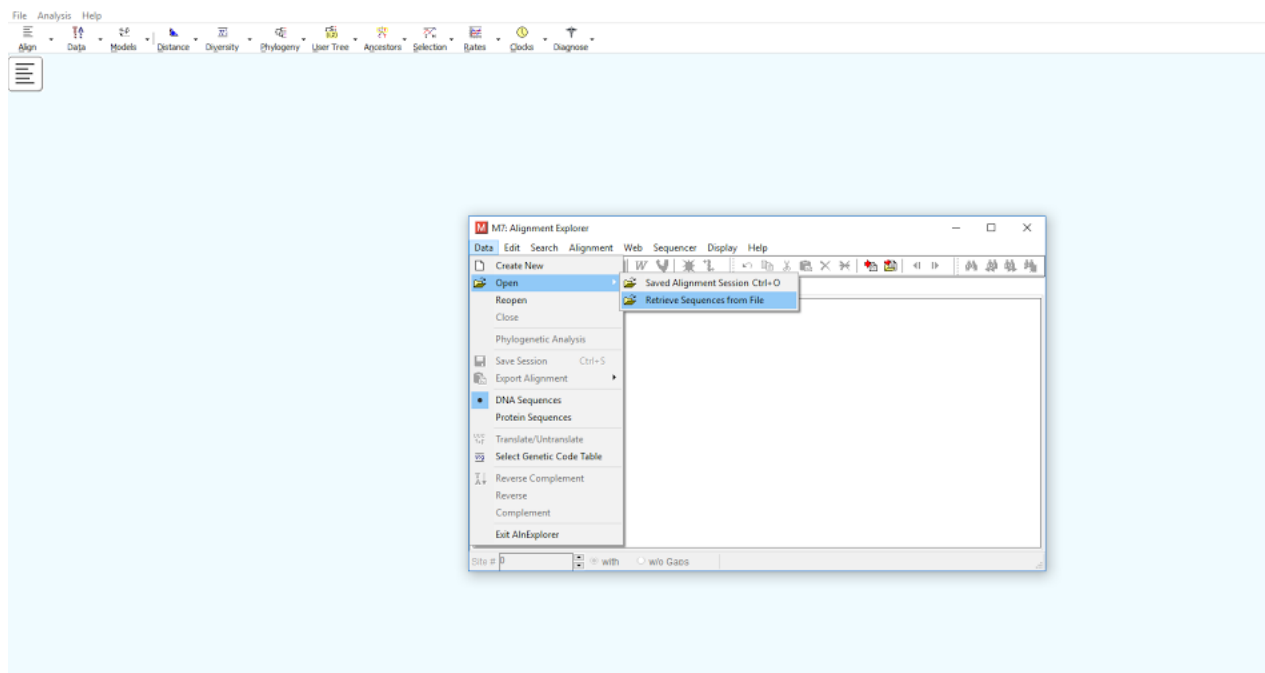
1.1 Une boîte de dialogue va apparaître, choisissez l'option "Create a new alignment"



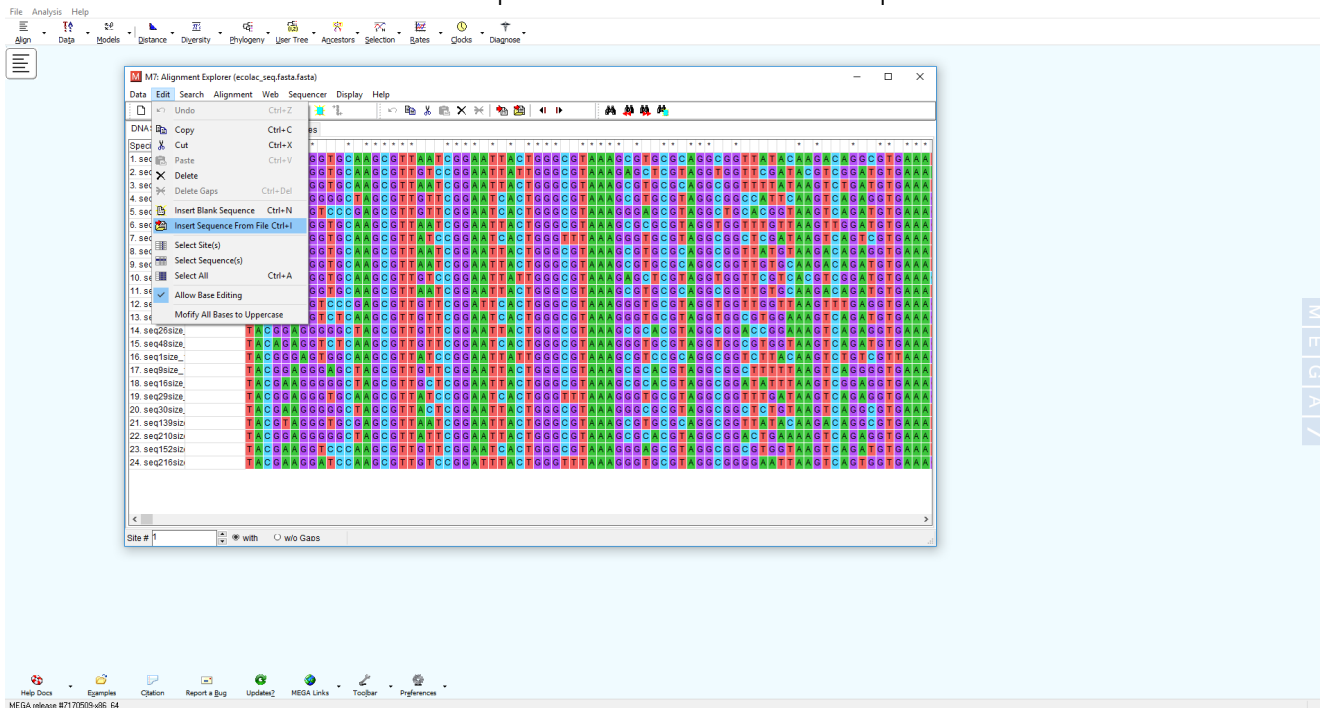
## 1.2 Une autre boîte de dialogue va apparaître : sélectionnez “DNA”



## 1.3 Maintenant, allez dans “Data → Open → Retrieve sequences from file” et choisissez le fichier fasta avec les séquences que vous voulez étudier.

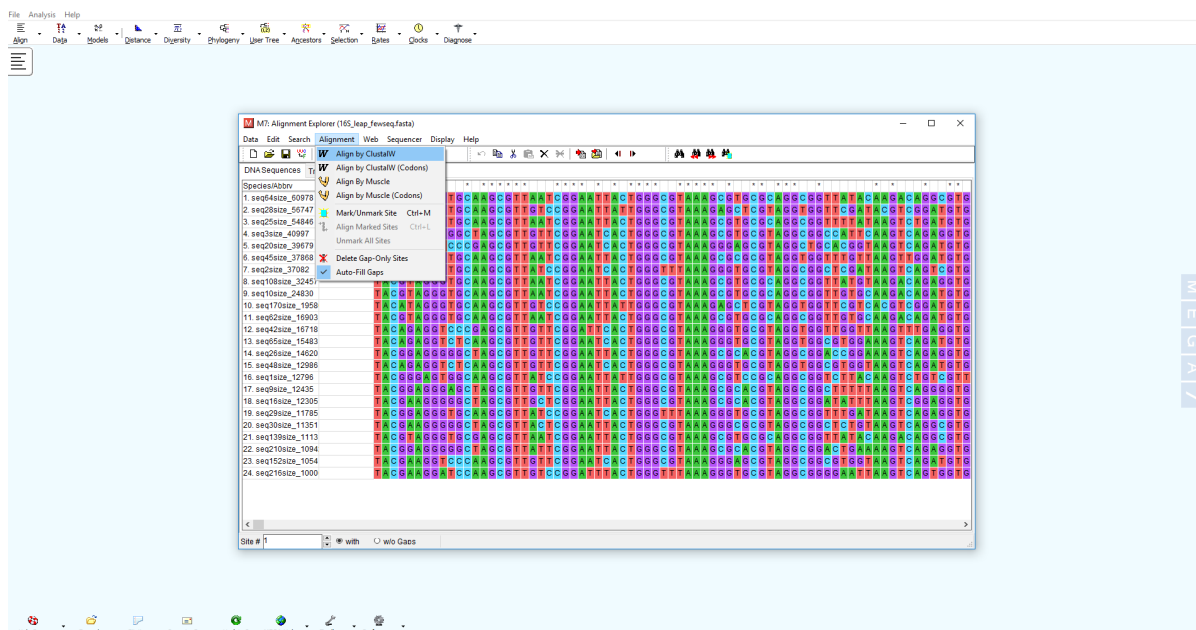


1.4 Pour ajouter les autres séquences que vous souhaitez aligner, allez dans “Edit → Insert Sequence From File”. N'oubliez pas d'ajouter le groupe extérieur (outgroup), au moment de placer la racine de l'arbre phylogénétique, et les séquences de référence qui aident à avoir une idée de l'identification des séquences non cultivées lorsque l'arbre est fait.

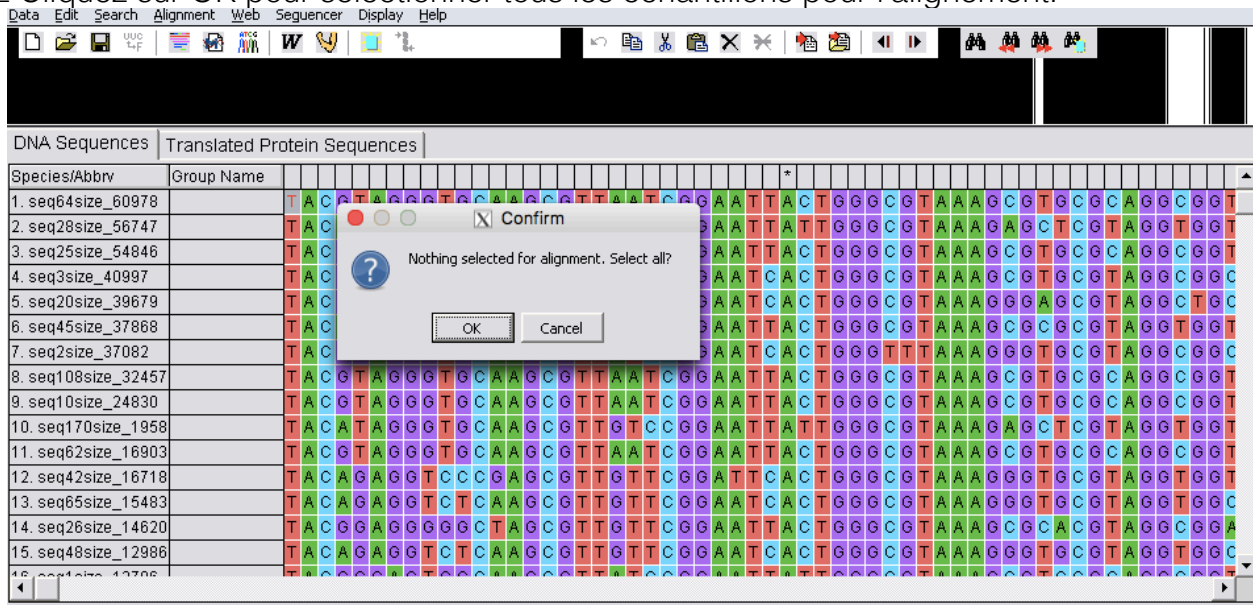


## Deuxième étape : réaliser l'alignement

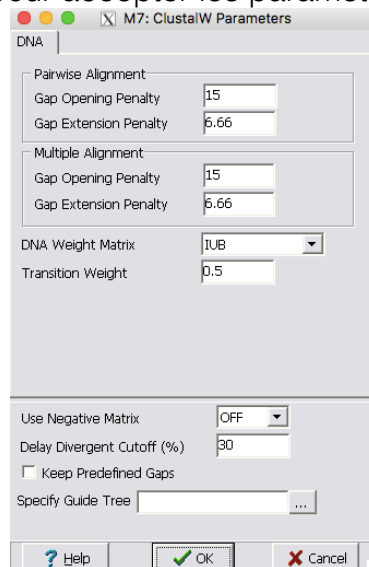
2.1 Cliquez n'importe où en dehors de la première colonne pour éviter de sélectionner seulement une rangée de séquences à aligner. Sélectionnez l'option “Alignment” et choisissez une méthode d'alignement des séquences, soit ClustalW ou Muscle. Dans ce cas, nous avons utilisé ClustalW.



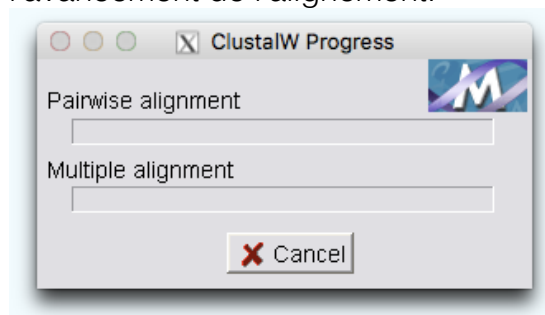
2.2 Cliquez sur OK pour sélectionner tous les échantillons pour l'alignement.



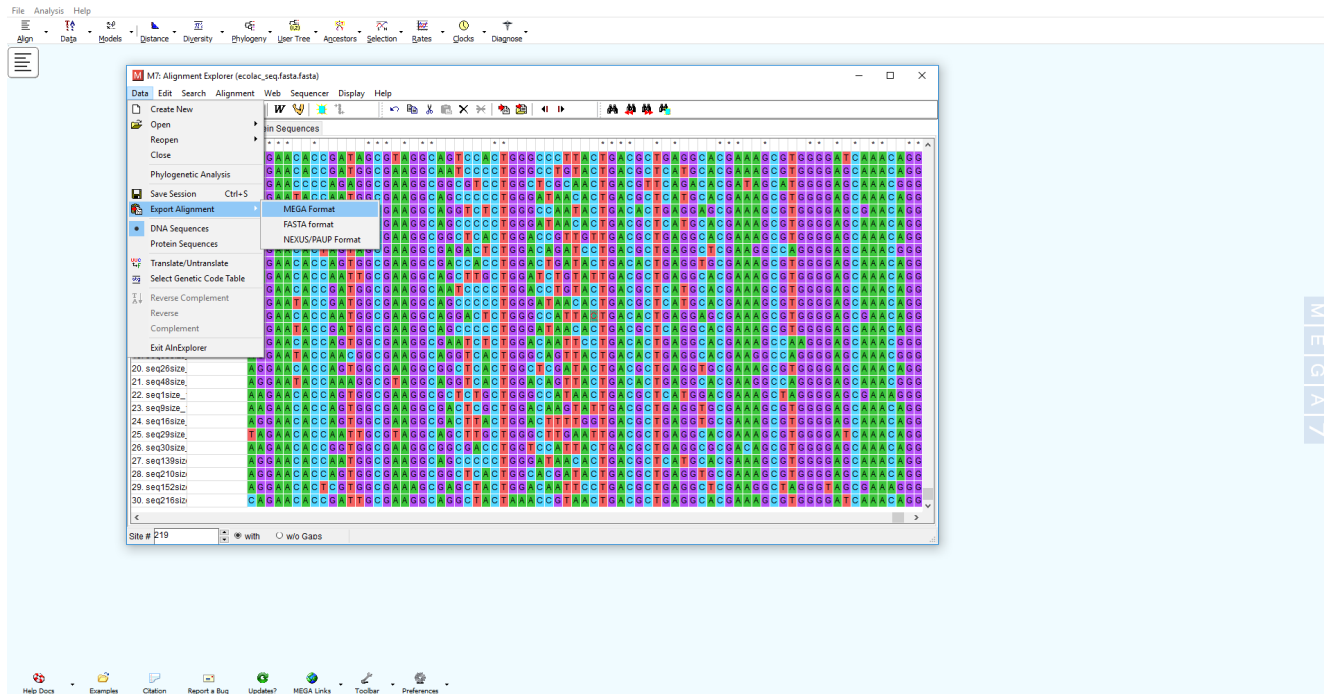
2.3 Cliquez sur OK de nouveau pour accepter les paramètres par défaut pour l'alignement



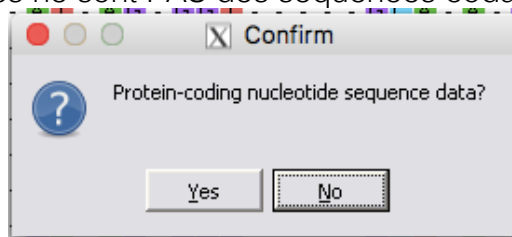
2.4 Attendez jusqu'à ce que l'alignement soit finit (plus il y a des séquences et plus les séquences sont longues, plus cela prend du temps). Une boîte de dialogue va montrer l'avancement de l'alignement.




2.5 Une fois que l'alignement est fait, vous pouvez exporter l'alignement dans un format MEGA ou FASTA en cliquant sur “Data => Export alignment” et en sélectionnant ces options de formats dans la barre d'outils.

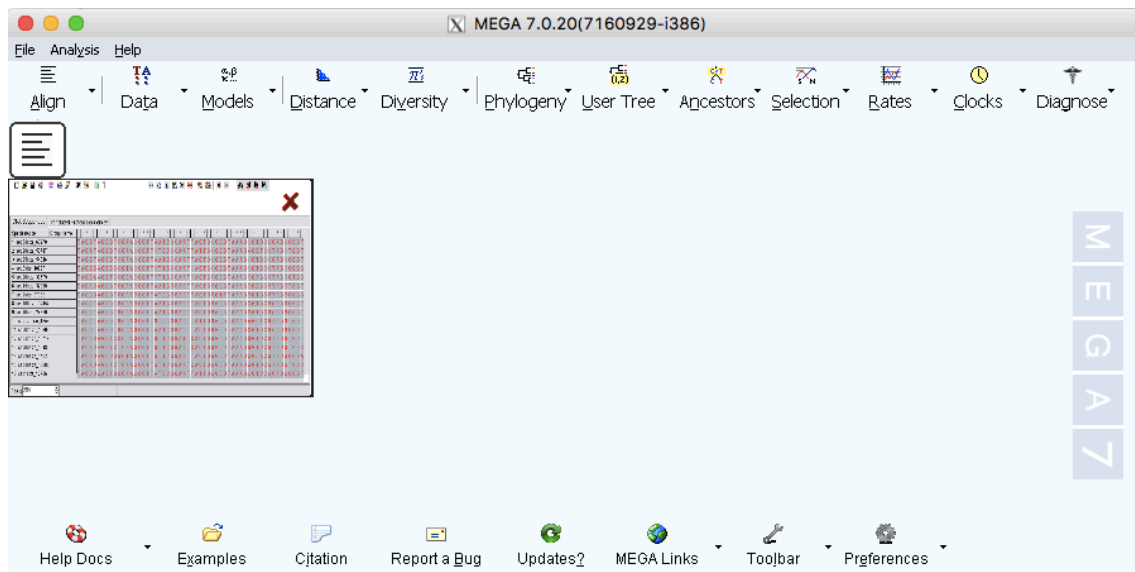


2.6 Choisissez où sauvegarder sur votre ordinateur, créez un titre pour vos données et confirmez que les séquences ne sont PAS des séquences codant pour des protéines.

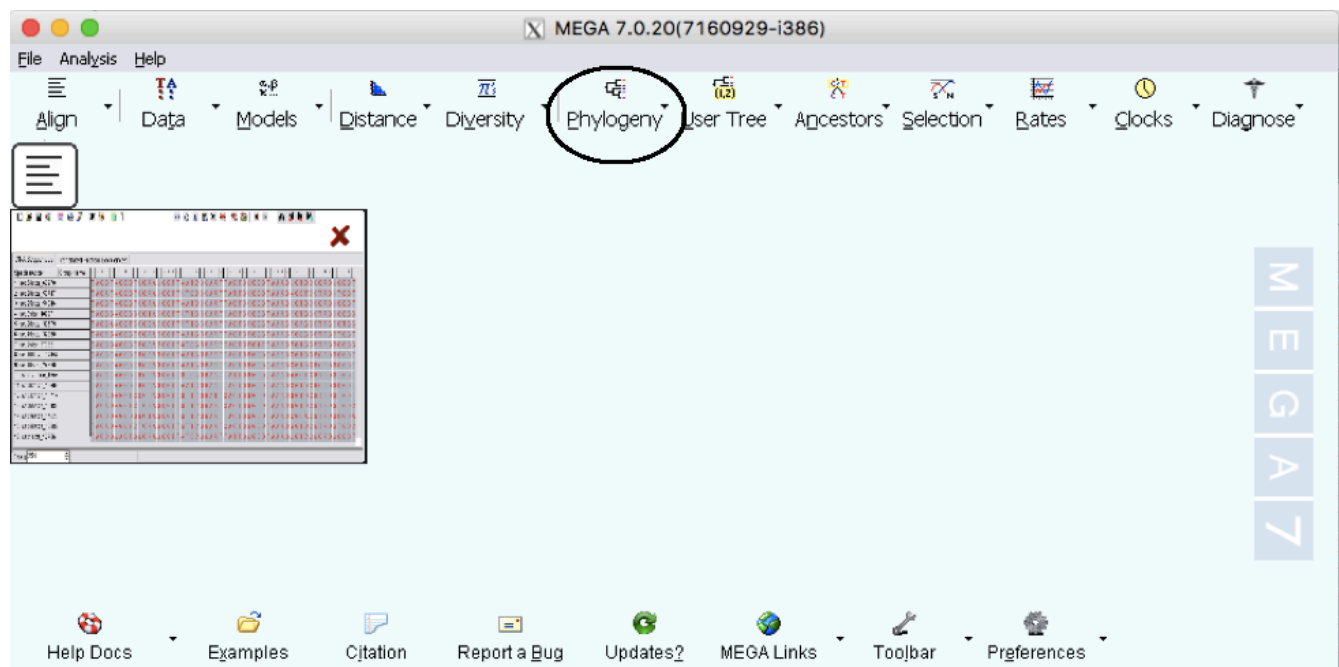


### Troisième étape : construisez l'arbre phylogénétique

3.0 Retournez sur l'écran principal de MEGA et, si vous n'avez pas fermé votre session d'alignement, vous la verrez en cliquant sur le lien de l'alignement 

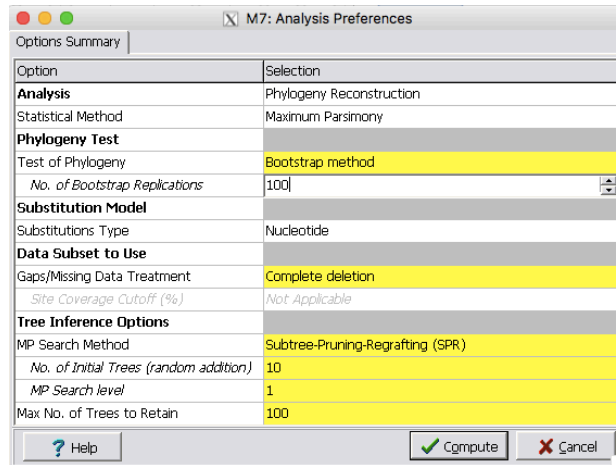


3.1 Sélectionnez l'option "Phylogeny" dans la barre d'outils et choisissez la méthode avec laquelle vous voulez construire votre arbre phylogénétique. À la fin de ce document, il y a une brève explication à propos des différentes méthodes. Dans ce cas-ci, nous utilisons la méthode de Maximum Parsimony Tree.

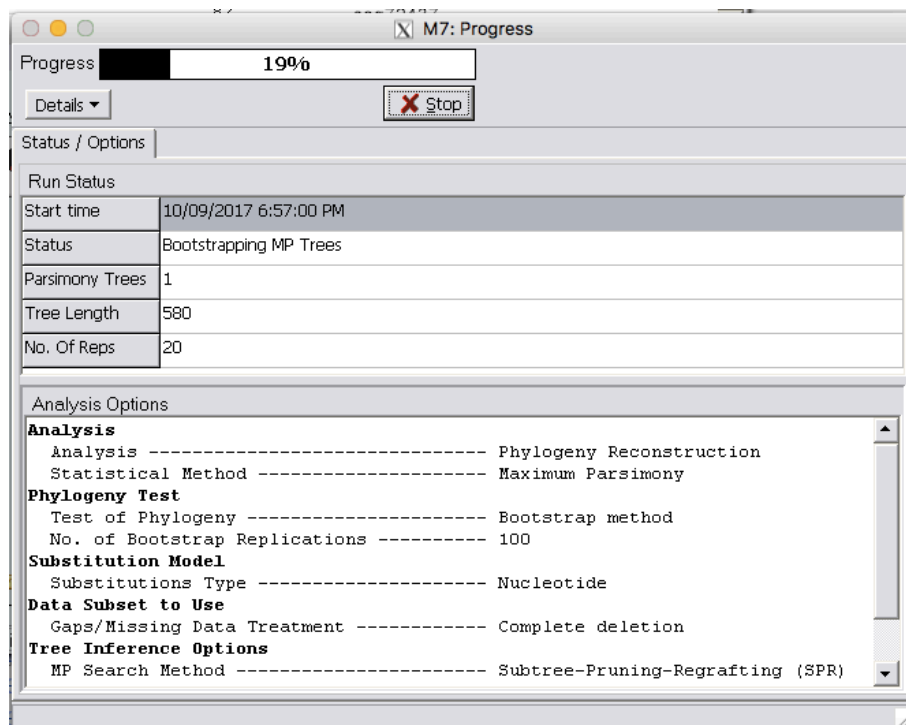


3.2 Le programme va vous demander de choisir le fichier (en format FASTA ou MEGA) à analyser. Chargez le fichier que vous avez sauvegardé à l'étape 2.6

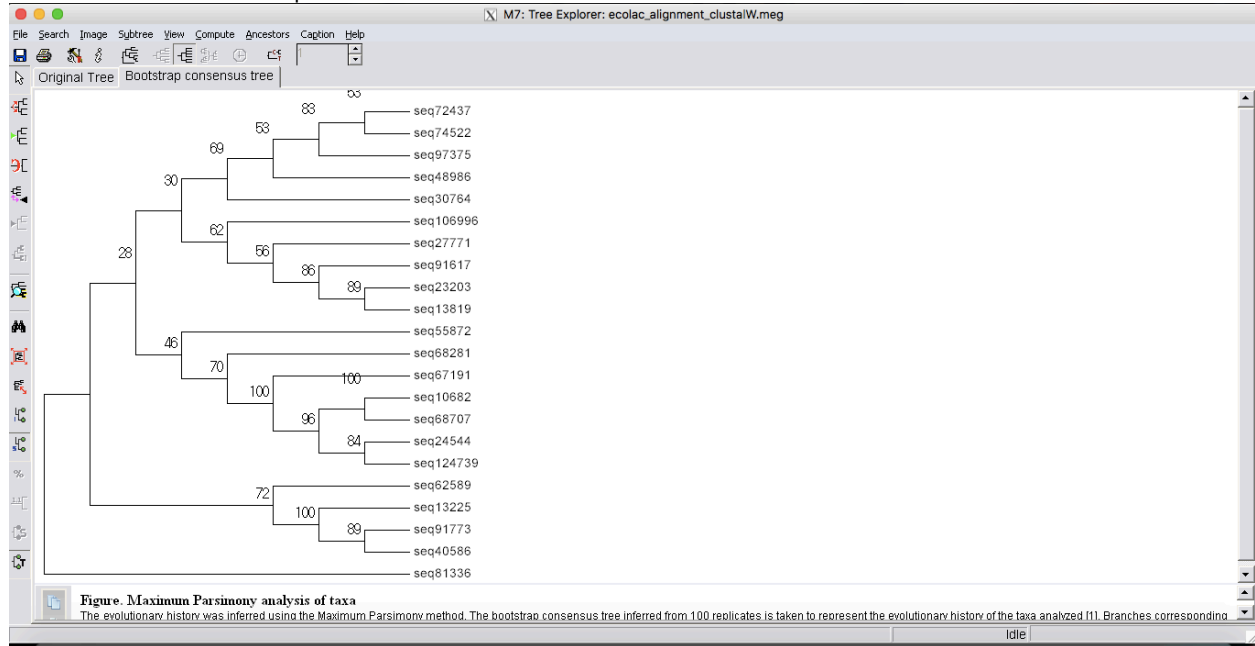
3.3 Après avoir chargé les données, vous aurez accès à la boîte de préférence d'analyse de la méthode que vous avez choisie. Dans l'option "Test of Phylogeny", choisissez la méthode de bootstrap et dans "No. of bootstrap replications" écrivez "100". En choisissant ces options, le programme va lancer 100 itérations d'arbres phylogénétiques basés sur les séquences que vous avez chargées et les paramètres que vous avez choisis, et vous retournez l'arbre le plus probable avec les pourcentages de confiance du bootstrap pour chaque branche. Nous choisissons 100 itérations pour une question de temps, mais le plus d'itérations choisies, le mieux sera votre certitude dans votre arbre phylogénétique. Vous pouvez garder ces autres paramètres comme paramètres par défaut.



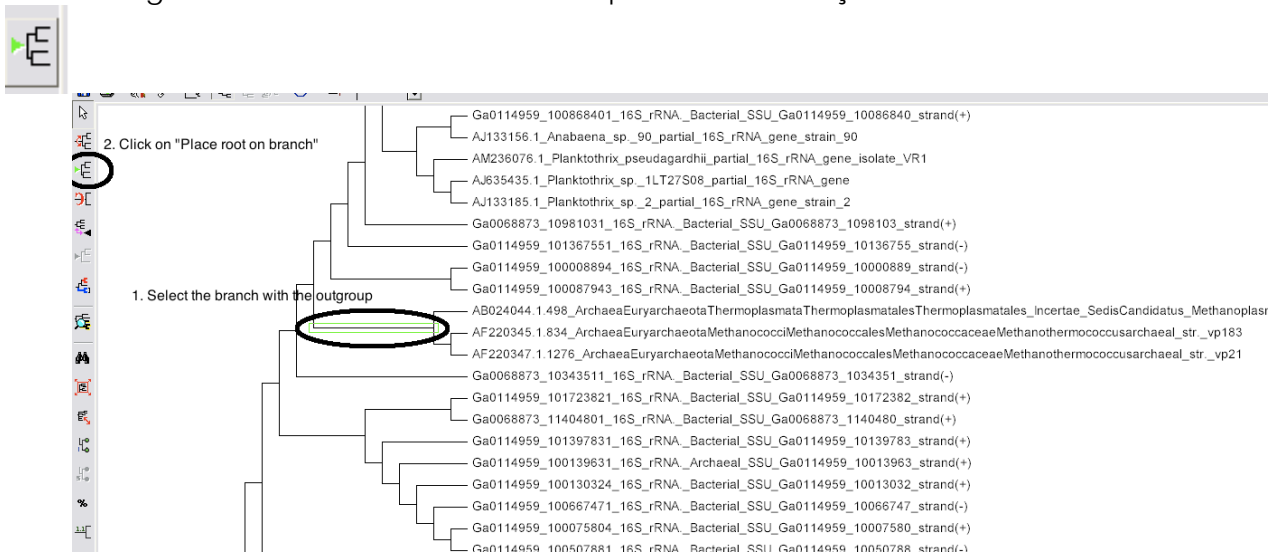
3.4 Attendez jusqu'à ce que l'analyse soit finie. Vous pouvez vérifier la progression de l'analyse dans la barre de progrès et dans la boîte "Analysis options" se trouve le résumé des paramètres que vous avez choisis pour l'analyse. Ça va prendre du temps, allez prendre un café!



3.5 Une fois que l'analyse est faite, le programme va vous montrer l'arbre original ainsi que les valeurs de bootstrap de l'arbre consensus.



3.6 Bravo! Vous avez construit un arbre phylogénétique à partir de séquences du gène 16S rRNA ! À présent vous pouvez formater l'arbre en utilisant la barre d'outils à gauche et répondre à votre question de recherche. N'oubliez pas de placer la racine de l'arbre : sélectionnez la branche qui contient l'outgroup et cliquez sur la troisième option de la barre d'outil de gauche "Place root on branch" qui ressemble à ça :



#### Quatrième étape : enregistrer et exporter l'arbre phylogénétique

4.1 Vous pouvez enregistrer votre arbre phylogénétique en PDF en cliquant sur "Image → Save as PDF file" et également enregistrer dans un format MEGA (pour pouvoir l'ouvrir à nouveau dans le programme) en cliquant sur "File → Save current session". Afin de pouvoir « éditer » votre arbre dans la prochaine partie du tutoriel, veuillez sauvegarder en format Newick (File → Export as NEWICK) (extension .nwk).



## Informations additionnelles

- Un **fichier FASTA** ou en format FASTA est un fichier en format texte qui représente des séquences nucléotidiques ou des séquences de peptides, dans lesquelles les nucléotides ou les acides aminés sont représentés en utilisant un code de lettre unique. Ce format permet d'avoir le nom de la séquence et des commentaires (ce qui suit le signe ">") avant les séquences.

Exemple de format FASTA pour une séquence d'ADN:

>séquence\_1

```
TACGAAGGGGGCTAGCGTTACTCGGAATTACTGGGCGTAAAGGGCGCGTAGGCGGCTCT
GTAAGTCGGCGTGAAATTCCTGGGCTCAACCTGGGGGCTGCGCTTGAGACTGTGGGGCT
AGAGGATGGAAGAGGGTCGTGGAATTCCCAGTGTAGAGGTGAAATTCGTAGATATTGGGA
AGAACACCGGTGGCGAAGGCGGCGACCTGGTCCATTACT
```

- Un **alignement de séquences** est un moyen d'arranger les séquences d'ADN, ARN ou de protéines pour identifier les régions similaires qui peuvent être la conséquence de relations fonctionnelles, structurelles ou évolutives entre les séquences. Des séquences très courtes ou très similaires peuvent être alignées à la main. Cependant, la plupart des études phylogénétique requièrent l'alignement de séquences longues, hautement variables ou extrêmement nombreuses qui ne peuvent pas être alignées manuellement. Pour cette raison, différents algorithmes, tel que ClustalW et Muscle, qui sont présent dans MEGA7, réalisent ces alignements de séquences.

- Un **arbre phylogénétique**, aussi appelé phylogénie, est un diagramme qui illustre les lignes de descendance évolutives entre différentes espèces, organismes ou gènes venant d'un ancêtre commun (Baum 2008). Il existe différentes méthodes pour l'estimation de la phylogénie. La méthode de Neighbor-Joining, par exemple, est basée sur une matrice de distance de toutes les paires possibles de taxa. Il existe aussi une méthode basée sur le concept de parcimonie, dont le but est de trouver une phylogénie qui implique le moins de changement nécessaires pour expliquer les différences entre les séquences. Finalement, la méthode du maximum likelihood fournit des probabilités à plusieurs arbres phylogénétiques possibles selon un certain modèle d'évolution et choisit l'arbre le plus probable (source: NCBI Advanced Workshop).

### Références:

- Baum, D. 2008 Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. Nature Education 1(1):190

- Kumar et al. 2016 MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016 Jul. 33(7):1870-4

- Wikipedia:

[https://en.wikipedia.org/wiki/Sequence\\_alignment](https://en.wikipedia.org/wiki/Sequence_alignment) and

[https://en.wikipedia.org/wiki/FASTA\\_format](https://en.wikipedia.org/wiki/FASTA_format)

- NCBI Advanced Workshop for Bioinformatics Information Specialists, Module title: Phylogenetic resources. Available on

<https://www.ncbi.nlm.nih.gov/Class/NAWBIS/Modules/Phylogenetics/phylo12.html>