

## Quelques indications pour l'utilisation du logiciel GénieGen

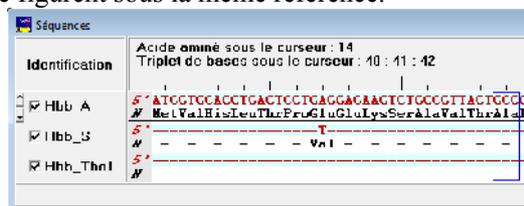
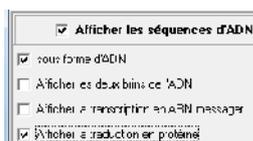
Le logiciel permet de travailler sur les séquences génétiques ADN, ARN protéines, un peu comme Anagène, avec quelques différences :

- les choix d'affichage sont plus nombreux.
- Une méthode d'alignement adaptée à la recherche des exons (alignement sur la plus longue séquence) a été ajoutée.
- Il lit des fichiers de différents types, y compris les fichiers pdb des représentations 3d des protéines (mais sans afficher la 3d par lui-même). Il peut produire des fichiers .rsm pour Rastop où les molécules 3d ont des couleurs définies à partir des comparaisons faites dans GénieGen.
- Il permet d'entrer manuellement (ou par copier-coller) des séquences. Par exemple des séquences provenant de la banque de données NCBI. Il conserve alors le lien avec la page internet du site NCBI.
- Le logiciel traite les séquences circulaires (plasmide, ADNmt) et contient 2 codes génétiques (le code ordinaire et le code de l'ADNmt des Mammifères).
- Il permet d'obtenir des schémas synthétiques des séquences et traite plus complètement les enzymes de restriction.
- Par contre il n'y a pas d'outil DotPlot contrairement à Anagène. Il ne gère ni les thèmes ni les documents d'Anagène.

### Choix d'affichage :

1 – Premier exemple : On cherche à comparer la séquence d'un gène normal (en fait un CDS) et de ses allèles mutés, et à prévoir la conséquence sur la protéine (ex : fichier Hbb\_H\_mut du sous\_répertoire betaglob)

- Charger le fichier : seules les séquences d'ADN s'affichent (le fichier ne contient pas les séquences protéiques).
- Pour faciliter le repérage, utiliser le mode d'affichage comparaison (menu *Affichage/Comparaison*). Les bases identiques à celle de référence sont remplacées par des tirets. En explorant toute la longueur des séquences, on trouve facilement les différences.
- Pour voir la conséquence sur les protéines, on utilisera la traduction automatique (ce qui est possible ici puisque seules les parties codantes sont reprises dans les séquences. Utiliser le menu *Affichage/Choix d'affichage/choix globaux*. Cocher au moins ADN5'3' et ...traduction en protéine. Il est possible aussi d'afficher la chaîne complémentaire de l'ADN et/ ou l'ARN messenger. A l'affichage on remarquera que l'ADN et sa traduction automatique figurent sous la même référence.



On pourrait obtenir le même résultat à partir des choix par catégories

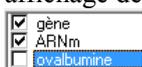
Avec les *choix par séquence*, on peut masquer provisoirement une séquence en décochant sa case.

2 – Deuxième exemple : Comparaison entre gène et ARN messenger. Dans ce cas, il faut afficher les séquences suivant leur nature. On a alors, si on le souhaite, la possibilité d'afficher les deux brins de l'ADN (ce choix est maintenant conservé à l'enregistrement du fichier .edi).

### Recherche des parties communes entre un gène et son ou ses ARNm (recherche des exons)

L'exemple historique, qui a l'avantage de pouvoir être illustré par une électrographie de l'appariement ADN-ARN, est celui de l'ovalbumine.

- Charger le fichier ovalbumine.edi (dans le répertoire ADN\_ARN\_pro)
- On peut désactiver l'affichage de la protéine (utiliser les choix d'affichage par séquence). Elle n'apporte rien à ce stade de l'étude
- Utiliser l'alignement sur la plus longue séquence qui convient dans ce cas<sup>1</sup> : le logiciel propose alors de conserver les parties identiques (exons) comme annotations, ce qui permettra d'en obtenir une représentation graphique.

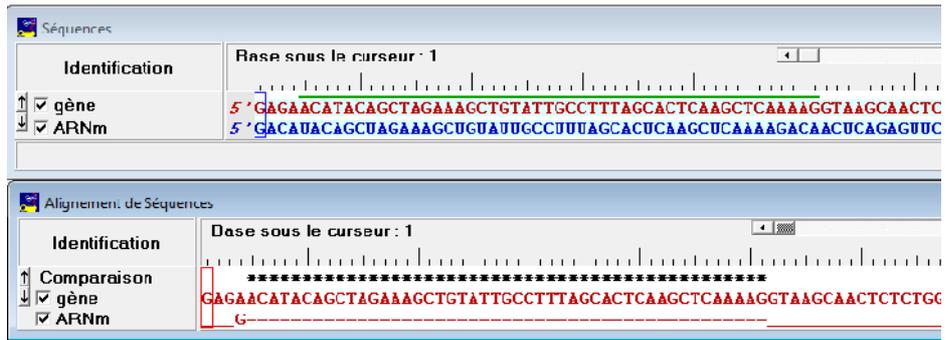


1 Les méthodes d'alignement multiple, destinées à comparer des séquences pour établir les phylogénies ne donnent pas toujours de bons résultats dans ce type d'alignement où il y a à la fois des portions identiques et de grands « trous ».

Valider en cliquant sur OK.

Dans la fenêtre d'alignement, l'ARN a été transformé en son équivalent ADN. Dans la fenêtre initiale, un trait vert signale les exons (portions de l'ADN qui correspondent à l'ARNm).

Après avoir cliqué une fois sur la fenêtre des séquences, on peut utiliser le menu *Edition/Dessiner la séquence* pour obtenir un schéma où il faut repositionner les étiquettes de légende avant de cliquer hors de la zone sélectionnée pour faire disparaître les marques de sélection.

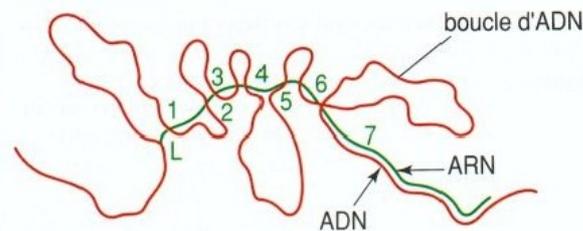
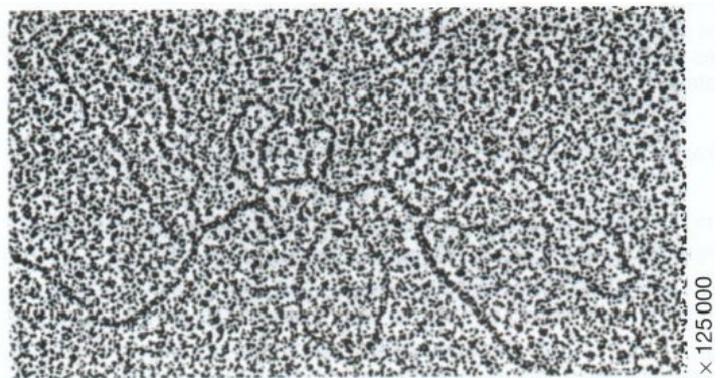


Le schéma obtenu correspond bien à l'image historique d'hybridation (L correspond à l'exon 1 du schéma... 7 correspond à l'exon 8).

- Dans le cas où il y a un épissage alternatif (fichier CGRP\_gene\_ARN.edi), il faut faire les alignements en 2 fois : l'ARN1 avec le gène puis l'ARN2 avec le gène (en changeant la couleur de représentation et le type de numérotation).



Au final on peut obtenir un schéma de ce type :



**Utilisation d'un fichier pdb :** Comparaison de séquences d'immunoglobulines pour montrer où se trouvent les parties variables :

1. Appeler le logiciel GénieGen et charger le fichier *iggtotal.pdb* qui comporte les séquences des 4 chaînes de la molécule étudiée..
2. Sélectionner les séquences des 2 chaînes lourdes (H et I) et utiliser le menu *Affichage/Comparaison* et comparer les deux séquences. Opérer de même pour les chaînes légères (L et M).
3. **Ajouter** les séquences du fichier *Igg-comp.edi*
4. *Sélectionner toutes les séquences des chaînes lourdes (h..., H et I) et utiliser le menu Action/Alignement et noter les numéros de début (nd) et de fin (nf) de la partie variable de ces chaînes. Utiliser le menu Edition/Conserver comme annotation* et cliquer sur OK. Opérer de même pour les chaînes légères (l..., L et M). Choisir un vert sombre pour ce qui est identique et en rose ce qui ne l'est pas.
5. Utiliser le menu *Fichier/Montrer les comparaisons...*, choisir le nom *IggVar.rsm* et enregistrer dans votre répertoire. Choisir de masquer les hétéroatomes et de masquer l'eau.
6. Ouvrir le fichier de la molécule colorée avec le logiciel Rastop.

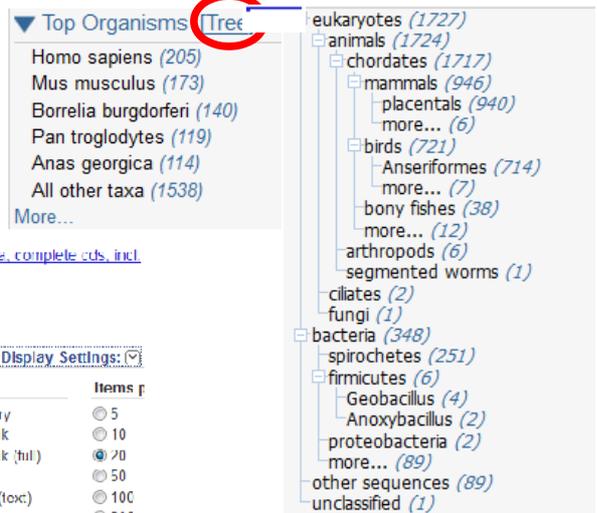
**Obtenir une ou des séquence à partir de la banque de données NCBI :**

Cela donne aussi la méthode pour entrer une séquence « à la main ».

- Aller sur le site NCBI : <http://www.ncbi.nlm.nih.gov/> et choisir la base de données (Database) « Nucleotid » pour ADN ou ARN ou « Protein » suivant ce que l'on cherche.

- Taper le nom du gène ou de la protéine à rechercher : par exemple HBB pour la chaîne β de l'hémoglobine.
- A droite, on peut faire un tri parmi les organismes

Dans l'exemple ci-dessous, après avoir sélectionné Homo sapiens, les séquences 2 (partie codante : cDNA avec le codon stop) et 4 (ARNm) ont été sélectionnées.



2. [Homo sapiens full open reading frame cDNA clone RZPDu834D0222D for gene HBB, hemoglobin, beta, complete cds, incl.](#)

stop:codon

444 bp linear mRNA

Accession: CR536530.1 GI: 49168513

[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)

4. [Homo sapiens hemoglobin, beta \(HBB\), mRNA](#)

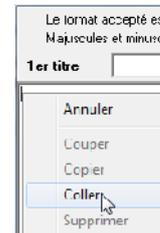
626 bp linear mRNA

Accession: NM\_000518.4 GI: 28302128

[GenBank](#) [FASTA](#) [Graphics](#) [Related Sequences](#)



- Choisir l'affichage (Display Settings) de type fasta (text) et valider avec le bouton Apply.
- Sélectionner le texte et faire un clic droit pour faire apparaître le menu où l'on choisit Copier.
- Dans GenieGen, après avoir choisi le menu *Fichier/Entrée d'une séquence*, Faire un clic droit sur la zone de texte et choisir coller.
- **Attention** les séquences ne seront correctement entrées que si la ligne de titre de chaque séquence (celle qui commence par >) n'est pas trop longue et tient bien sur une ligne. Sinon ce qui est sur la deuxième ligne est considéré comme faisant partie de la séquence (en éliminant les caractères non conformes au codage). Agrandir la fenêtre ou raccourcir le texte d'en-tête.



**Un autre aménagement utile** est de mettre le nom de l'espèce entre crochets dans la ligne d'en-tête

[Homo sapiens] | Le logiciel l'utilisera alors comme nom d'espèce. Il est bon de conserver le code figurant en début d'en-tête > | 49168543 | le logiciel le conservera dans le fichier edi enregistré. On aura alors un accès à la page correspondante de la base de données NCBI.

- Vérifier aussi que c'est le bon type de séquence qui est coché (ici ADN5'3')  Circulaire  ADN 5'3' puis cliquer sur le bouton OK.
- Une fenêtre de vérification apparaît. Choisir l'onglet Données pour obtenir un tableau où l'on peut modifier le nom proposé pour chaque séquence ainsi que les autres colonnes. Valider pour entrer définitivement la ou les séquences.

Composition		Texte	Co
Numéro	Nom		
0	ADNc		A
0	Homo sapiens		A

- Le tableau sera par la suite consultable avec le menu Informations de la fenêtre des séquences. La fiche NCBI sera consultable (s'il y a une liaison internet active) avec le menu [Accès à la fiche NCBI \(sur internet\)](#). La fiche qui s'ouvrira dans le navigateur sera celle qui correspond à la ligne du tableau qui est sélectionnée. Il faut pour cela que le code NCBI de la séquence ait été enregistré.
- L'enregistrement des séquences se fait avec le menu Fichier/Enregistrer sous... **cet enregistrement ne concerne que les séquences sélectionnées**. Les fichiers enregistrés au format .edi (format Anagène) sont compatibles avec toutes les versions actuelles d'Anagène.