

Collection « Analyse de données, simulation... »

ANAGÈNE

Analyse de séquences nucléiques et protéiques

version 1.8

Sélection en sciences de la vie et de la Terre

À consulter dans les logithèques des CRDP et CDDP
et à découvrir dans la librairie sur le site du CNDP
<http://www.cndp.fr>

COLLECTION « ANALYSE DE DONNÉES, SIMULATION... »

- Anagène* analyse de séquences nucléiques et protéiques
Écol 1 sensibilisation à la notion d'équilibre écologique
En Jeu Ma Planète découverte ludique des enjeux de l'environnement
Génie génétique II simulation d'une manipulation génétique
Géocéan : vol. 1, 2, 3 géodynamique océanique, subduction et expansion
Léman un écosystème en chiffres
Magma simulation de la cristallisation magmatique
Nut équilibre de la ration alimentaire
Réhor régulation de la sécrétion des hormones sexuelles
Rouge-gorge étude expérimentale d'un comportement

COLLECTION « IMAGERIE SCIENTIFIQUE »

- Explorer le corps* surfaces d'échange
Explorer le corps communication nerveuse
Information génétique reproduction conforme
Information génétique brassage
Observ'AO observer avec l'aide de l'ordinateur
Explorer la Terre la télédétection

COLLECTION « DÉCOUVERTE DE LA NATURE »

- Les Oiseaux d'Europe* encyclopédie multimédia interactive permettant de gérer ses observations
Découvrir la planète Terre notre environnement en 7 thèmes : la planète, l'eau, l'air, la Terre, la vie, les Hommes, les défis

COLLECTION « PROGRÈS »

- Infogène* ressources pour l'étude de l'information génétique
L'Étang découvrir un écosystème
Communication et systèmes nerveux explorer, identifier, analyser des documents
Mitose-Méiose conservation et brassage de l'information génétique

Collection « Analyse de données, simulation... »

ANAGÈNE

Analyse de séquences nucléiques et protéiques

Centre national de documentation pédagogique
Institut national de recherche pédagogique

L'édition de logiciels et multimédias
est coordonnée par le CNDP,
Service national des productions imprimées et numériques,
29, rue d'Ulm, 75005 Paris.
Chef de service :
Annie Leprince

Chefs de Projet Maurice Bouyssou (coordination)
Catherine Bouyssou
Daniel Buret
Sylvie Fumel
Marie-Line Périllat-Mercerot

Développement Érik Boucher (coordination)
David Elvecque
Robert Lagoutte
Yves Lévine

Micro-édition Fabien Biglione
Sylvie Morel
Ghislaine Rochette

Graphisme Marie-Claude Barbier

Fabrication Monique Amoura

ISBN 2-240-00530-0 (2^e édition, © CNDP, 2003)
ISBN 2-240-00479-7 (1^{re} édition, © CNDP, 1997)

Collection « Analyse de données, simulation... »

ANAGÈNE

Analyse de séquences nucléiques et protéiques

Discipline Biologie

Niveau Lycée, Enseignement supérieur.

Sujet Expression de l'information génétique, polymorphisme des gènes, famille multigénique, prévisions en génétique humaine.

Conception CNDP, Service national des productions imprimées et numériques,
MAURICE BOUYSSOU,
professeur de sciences naturelles
INRP, Département Technologies nouvelles et éducation,
NAOUM SALAMÉ,
ingénieur de recherche

*Documentation scientifique
et pédagogique* JEAN-CLAUDE HERVÉ,
IPR-IA de sciences de la vie et de la Terre.
BERNARD THERRIÉ,
professeur de sciences naturelles associé à l'INRP

*Adaptation
et développement* CNDP, Service national des productions imprimées et numériques,
ÉRIK BOUCHER,
ingénieur en informatique et biologie

Aide en ligne MAÏTÉ DEFAYSSSE, YVAN VINCENT,
bio-informaticiens

*Adaptation de l'aide et de la
documentation à la version 1.8* JEAN-CLAUDE LE HIR,
professeur de sciences naturelles et formateur académique

Conduite du projet MAURICE BOUYSSOU,
professeur de sciences naturelles et d'option informatique,
coordinateur des collections hors ligne au SnPIN

Configuration minimale

Compatible PC *Pentium*, 32 Mo de RAM, carte vidéo 256 couleurs,
Windows 95 et suivants, disque dur avec 10 Mo.

Remerciements

BERNARD THERRIÉ, lycée Parc-des-Loges à Évry (91),
MICHEL COSTE, lycée Camille-Claudel à Vauréal (95),
JACQUES BARRÈRE ET JEAN-YVES DUPONT, lycée Paul-Louis-Courier à Tours (37),
enseignants associés aux recherches de l'INRP

JEAN-CLAUDE LE HIR, lycée Henri-Moissan à Meaux (77),
MARTINE PERNODET, lycée Romain-Rolland à Ivry (94),
membres du groupe de recherche disciplinaire « Biologie informatique et géologie » (BIG)
de l'académie de Créteil

SOMMAIRE

| | |
|--|------------|
| PRÉSENTATION..... | 9 |
| PRISE EN MAIN..... | 11 |
| L'environnement d' <i>Anagène</i> | 11 |
| L'interface d' <i>Anagène</i> | 12 |
| Les données fournies avec <i>Anagène</i> | 14 |
| Les données personnelles de l'utilisateur | 17 |
| L'édition et les traitements de séquences | 21 |
| Le classeur d' <i>Anagène</i> | 28 |
| GUIDE DE RÉFÉRENCE..... | 29 |
| Les commandes des menus..... | 29 |
| Les icônes de la barre d'outils..... | 30 |
| Les raccourcis clavier | 31 |
| Le visionneur d'animations..... | 32 |
| SUGGESTIONS PÉDAGOGIQUES..... | 33 |
| Introduction | 33 |
| Expression de l'information génétique..... | 35 |
| Les relations génotype-phénotype | 43 |
| Polymorphisme des gènes | 52 |
| Familles multigéniques..... | 74 |
| Détermination des génotypes et prévision en génétique humaine..... | 86 |
| BIBLIOGRAPHIE | 117 |
| INDEX DE LA BANQUE DE SÉQUENCES | 119 |
| INDEX DES THÈMES D'ÉTUDE | 127 |
| INDEX DES PROGRAMMES ET DOCUMENTS | 131 |

PRÉSENTATION

Anagène est un logiciel d'analyse de séquences nucléiques et protéiques qui propose dans un environnement *Windows*, quelques uns des traitements présents dans les logiciels *SEQAID II* (D.D. Rhoads et D.J. Rouffa, Université du Kansas, versions 1987, 1989, 1992 pour MS-DOS) et *CLUSTAL W* (D. Higgins, European Molecular Biology Laboratory (EMBL), version 1.8 1998, pour MS-DOS). Le premier comporte une gamme de traitements simples pour l'édition et l'analyse de séquences nucléiques et protéiques, alors que le second est spécialisé dans l'alignement multiple de séquences.

CLUSTAL figure au *Biocatalogue* de l'Institut Européen de Bioinformatique (EBI, Royaume Uni), station du EMBL. Des versions de ce logiciel peuvent être téléchargées à l'adresse suivante :

http://corba.ebi.ac.uk/Biocatalog/biocatalog_algt.html

La version initiale d'*Anagène*, diffusée en 1997, a constitué une première phase dans la construction d'un nouvel outil qui assure une continuité avec les logiciels précités. Il convient de souligner en particulier que les traitements de *SEQAID II* utilisés à des fins pédagogiques sont conservés dans *Anagène*. À ces traitements s'ajoute l'alignement multiple de séquences réalisé par *Clustal*. Les séquences associées aux deux versions de *SEQAID II* sont également intégrées dans *Anagène* de telle sorte que toutes les démarches pédagogiques proposées précédemment soient aisément réalisables.

Cette documentation emprunte à deux publications de l'INRP — « Analyse de séquences nucléiques et protéiques avec le logiciel *SEQAID II* » (1993) et « Thèmes d'étude en génétique humaine » (1995) — de larges extraits en ce qui concerne les informations scientifiques et les propositions pédagogiques, la manière de réaliser ces dernières étant adaptée pour tenir compte des facilités apportées par *Anagène*.

Cette nouvelle version 1.8 d'*Anagène*, diffusée sur cédérom, a été totalement réécrite pour s'intégrer aux nouveaux environnements *Windows* et apporter des améliorations d'ordre fonctionnel et ergonomique, notamment en ce qui concerne l'accès aux thèmes d'étude, la puissance des outils de traitement, la navigation dans les séquences et le repérage des positions des éléments, la gestion des nouveaux formats de médias dont ceux accessibles en ligne.

La version 2, en cours de développement, disposera en plus d'un module graphique de ressemblance, outil de traitement que possédait *SEQAID II* et qui est déjà intégré dans la version 1.8 d'*Anagène* sans être encore fonctionnel. D'autre part, cette nouvelle version 2 sera enrichie d'une banque de séquences constituée en fonction des nouveaux programmes et qui fait l'objet actuellement d'expérimentations. Dès leur disponibilité, ce nouveau moteur logiciel, les nouvelles séquences et les suggestions pédagogiques produites par les équipes associées à l'INRP seront téléchargeables à partir du site du CNDP à l'adresse de l'espace dédié à *Anagène* :

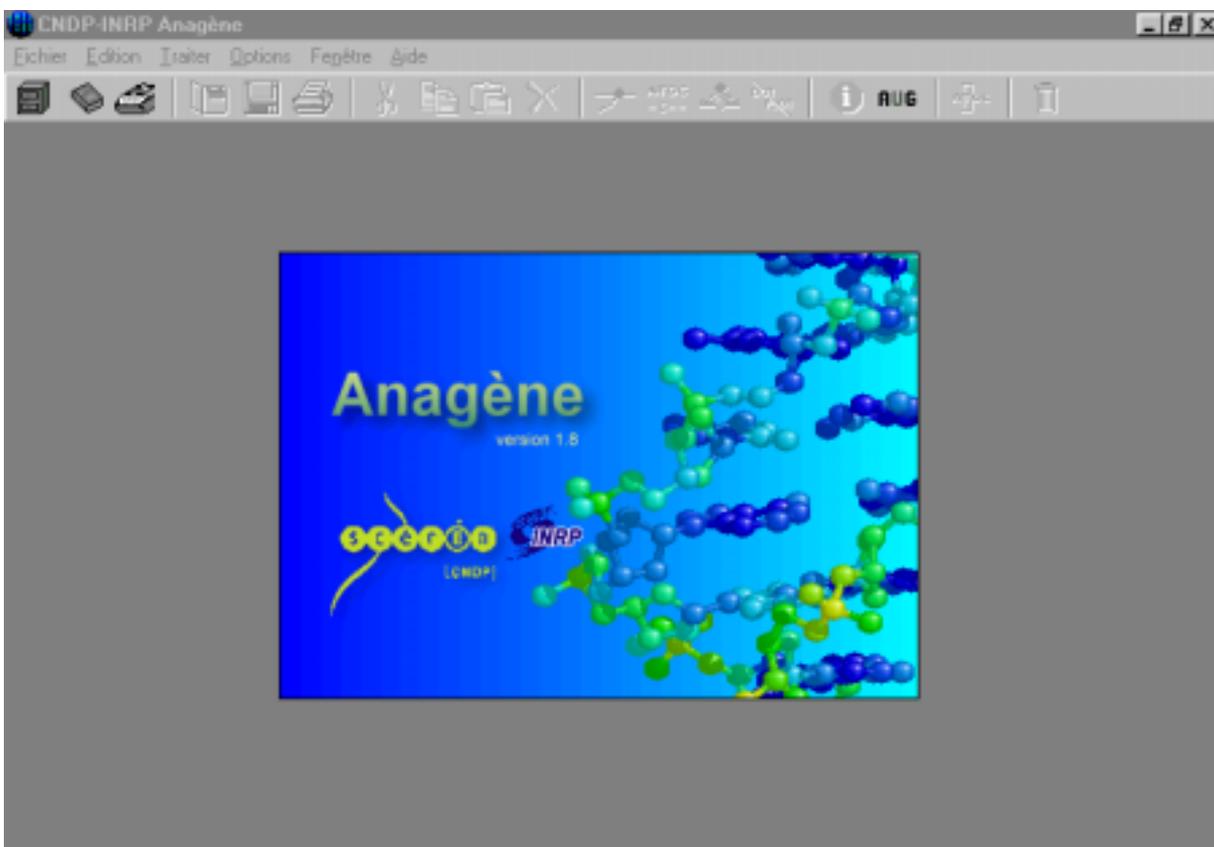
<http://www.cndp.fr/svt/anagene/>

PRISE EN MAIN

L'ENVIRONNEMENT D'ANAGÈNE

Anagène fonctionne dans l'environnement *Windows*. Les dimensions de la fenêtre d'exécution du programme s'adaptent automatiquement aux paramètres d'affichage définis dans la configuration. Néanmoins, pour améliorer le confort visuel, et si la carte graphique le permet, il est préférable de travailler avec une résolution de 800 x 600 pixels. On se référera au manuel d'utilisation de *Windows* ou à celui de la carte graphique pour effectuer les adaptations nécessaires.

Sous *Windows 95* et versions suivantes, lors de l'installation, *Anagène* est ajouté à la liste des programmes accessibles par le menu **Démarrer**.

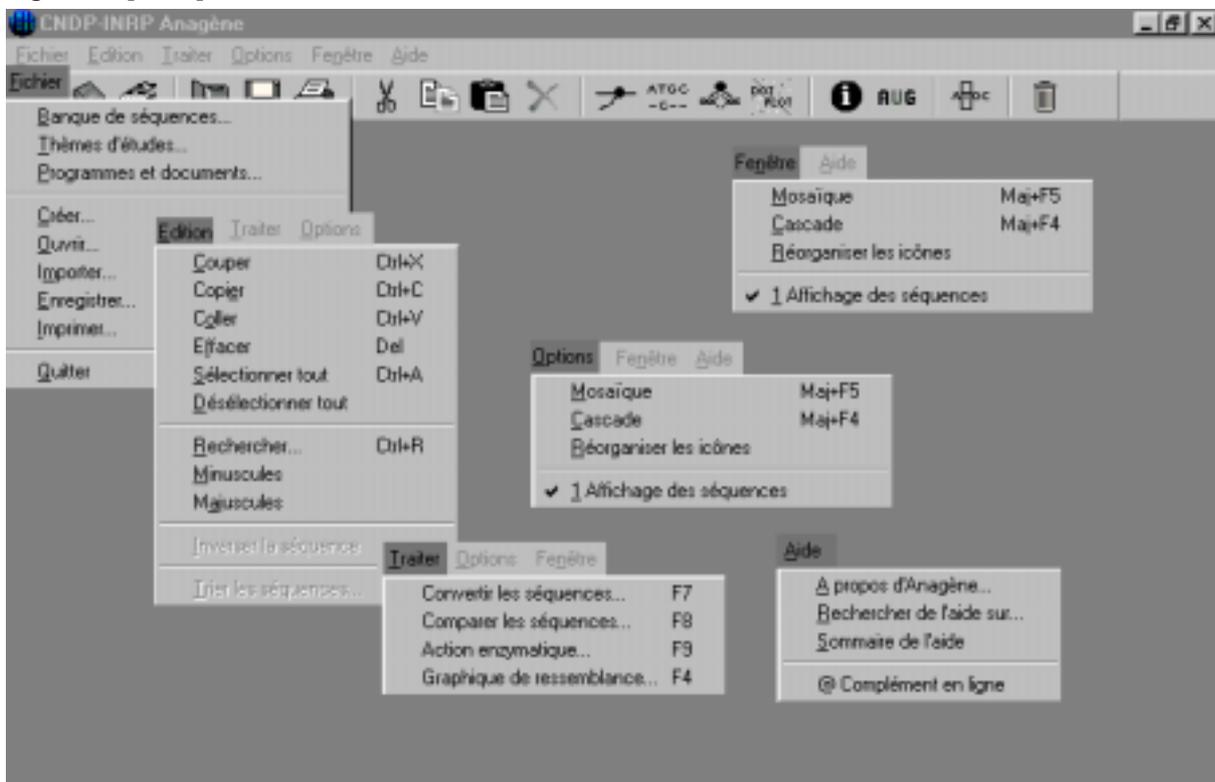


Après un temps de chargement d'autant plus rapide que le micro-ordinateur est de type récent et une phase d'initialisation nécessaire, l'interface d'*Anagène* apparaît.

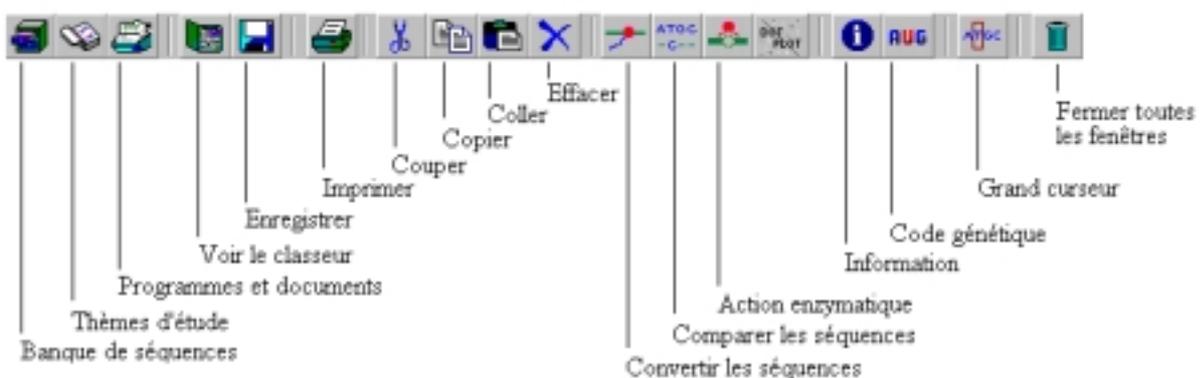
L'INTERFACE D'ANAGÈNE

Anagène est constitué d'une fenêtre-mère présentant sous la barre de titre, une barre de menus et une barre d'outils.

Les modes de déplacement et de sélection pilotés à l'aide de la souris ou du clavier obéissent aux règles adoptées par *Windows*.



Par ailleurs, lorsque le pointeur de la souris est maintenu quelques instants en place sur l'un des objets de l'écran tels par exemple un bouton de la barre d'outils, une bulle d'aide apparaît et rappelle la fonction qui lui est associée.



Ainsi les fonctions d'*Anagène* sont accessibles soit à travers les différentes options de la barre des menus soit directement pour la plupart par l'intermédiaire des boutons de la barre d'outils. Certaines

commandes des menus sont aussi directement applicables par l'utilisation de raccourcis clavier ; le détail en est précisé dans le Guide de référence (p. 29 et suiv.) et accompagne les commandes dans les menus déroulants du logiciel.

Menu Fichier

On y trouve les options classiques de gestion de fichiers, à savoir l'accès aux banques de données du logiciel (séquences, programmes et documents), la création, l'ouverture, l'importation, l'enregistrement, l'impression de fichiers de séquences, et la possibilité de quitter le logiciel.

Menu Edition

Il permet d'accéder aux commandes présentes dans la plupart des logiciels d'édition ou de traitement de texte, à savoir les commandes liées au presse-papiers de *Windows* (**Couper**, **Copier**, **Coller**), les commandes d'effacement, de sélection, de recherche, d'inversion et de tri de séquences. Certaines de ces commandes ne sont applicables aux séquences que si la commande **Protéger les données** du menu **Options** n'est plus cochée. En effet, au lancement du logiciel, les séquences chargées dans la fenêtre d'édition sont par défaut ainsi protégées.

Menu Traiter

Il contient les commandes assurant les trois grands types de traitements qu'il est possible d'effectuer sur les séquences, à savoir :

- les conversions de séquences nucléiques en séquences peptidiques et inversement pour les séquences peptidiques l'affichage des différents codons possibles ;
- les comparaisons de séquences qu'elles soient simples ou par alignement mettant en évidence les discontinuités ;
- l'action d'enzymes de restriction sur des séquences d'ADN ;
- le graphique de ressemblance représente un quatrième traitement à venir et qui sera disponible à partir de la version 2.

Menu Options

Il permet de déprotéger les séquences et d'en autoriser la modification, de définir le type de curseur (simple ou grand curseur) et de règle graduée (en caractères ou en triplets) utilisés, d'afficher ou non les bulles d'aide.

Menu Fenêtre

On y retrouve de classiques options propres à Windows, à savoir les commandes d'organisation des fenêtres et des icônes. La liste des fenêtres-filles présentes dans la fenêtre mère d'*Anagène* y est aussi accessible.

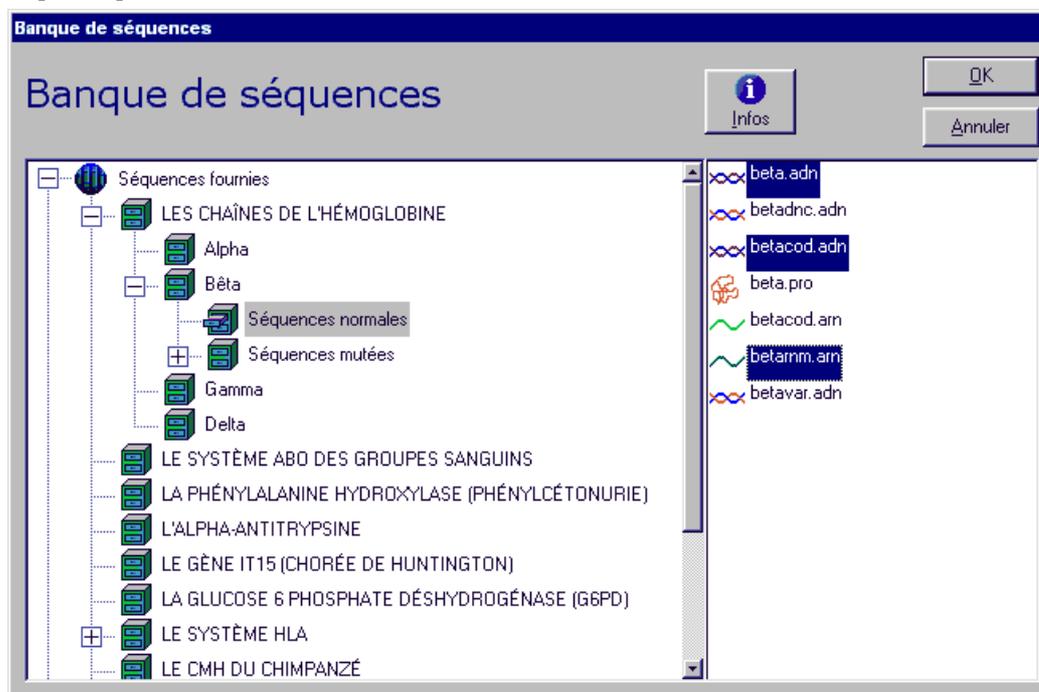
Menu Aide

Il contient la commande **A propos d'Anagène** qui renseigne sur la version du logiciel et sur son édition. L'aide en ligne accessible avec la touche de fonction F1 est aussi disponible dans ce menu. Si *Anagène* fonctionne sur un micro-ordinateur connecté à Internet, des informations et données complémentaires disponibles sur les sites Internet du CNDP et de l'INRP peuvent être consultées dans un navigateur et téléchargées. Un lien mène également vers un site dédié à Anagène. L'accès à ces sites est réalisé à partir de l'option @ **Complément en ligne**.

LES DONNÉES FOURNIES AVEC ANAGÈNE

La banque de séquences

Destinée en priorité à l'enseignant, cette banque comporte toutes les séquences utilisées pour la constitution de thèmes d'étude. Chaque séquence y figure en un seul exemplaire, associée à un pictogramme spécifique en relation avec sa nature (séquence nucléique de type ADN, ARN ou séquence protéique) et est assortie d'un commentaire.



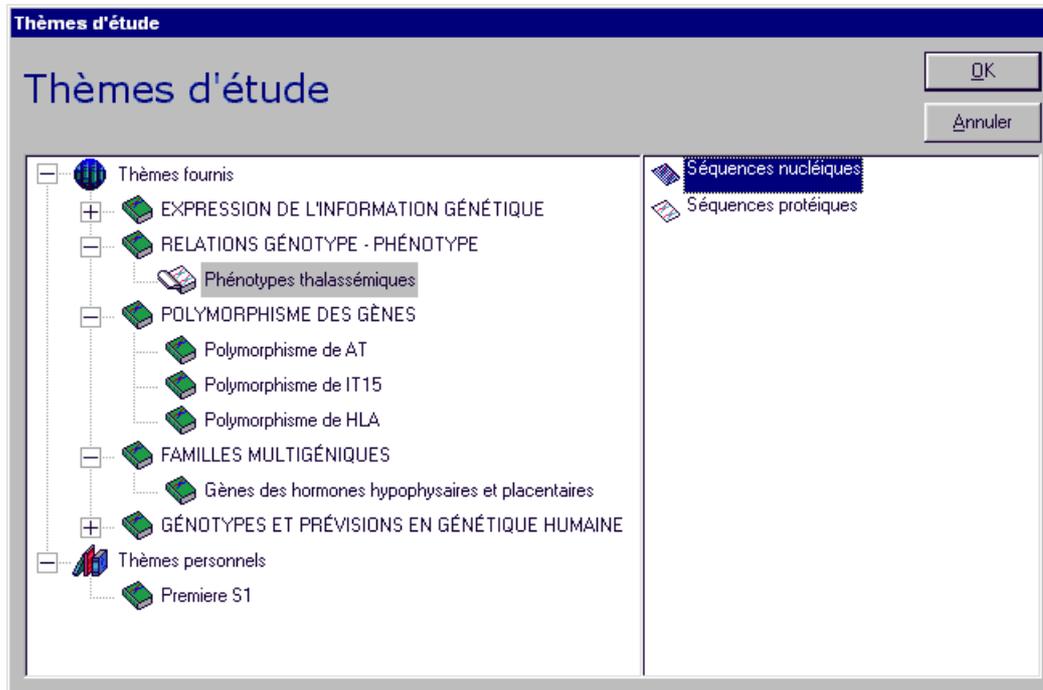
Pour charger une ou plusieurs séquences de la banque, il faut ouvrir la boîte de dialogue intitulée **Banque de séquences**. Elle est accessible par le menu **Fichier** ou en cliquant sur le premier bouton de la barre d'outils. Les séquences sont classées par dossiers, eux-mêmes ordonnés de manière arborescente. Pour les atteindre, il convient de déployer plus ou moins cette arborescence et de sélectionner le dossier désiré. Dans la partie droite de la boîte de dialogue s'affiche alors la liste des séquences disponibles dans ce dossier. Une ou plusieurs séquences peuvent être sélectionnées, elles apparaissent alors en inversion vidéo. Un clic porté sur le bouton **Infos** renseigne sur la nature de la dernière séquence pointée qui apparaît alors encadrée. Un clic porté sur une séquence déjà sélectionnée ou la fermeture de la branche correspondante de l'arborescence la désélectionne.

La validation de la sélection d'une ou plusieurs séquences effectuées est obtenue en cliquant sur le bouton **OK** ou en appuyant sur la touche **Entrée**. Elle provoque la fermeture de la boîte de dialogue et le chargement des séquences sélectionnées dans la fenêtre d'**Affichage des séquences** qui apparaît alors à l'écran. La répétition de cette procédure permet d'ajouter de nouvelles séquences dans cette fenêtre. Le bouton **Annuler** provoque la fermeture de la boîte de dialogue et les séquences qui ont pu être sélectionnées ne sont pas chargées.

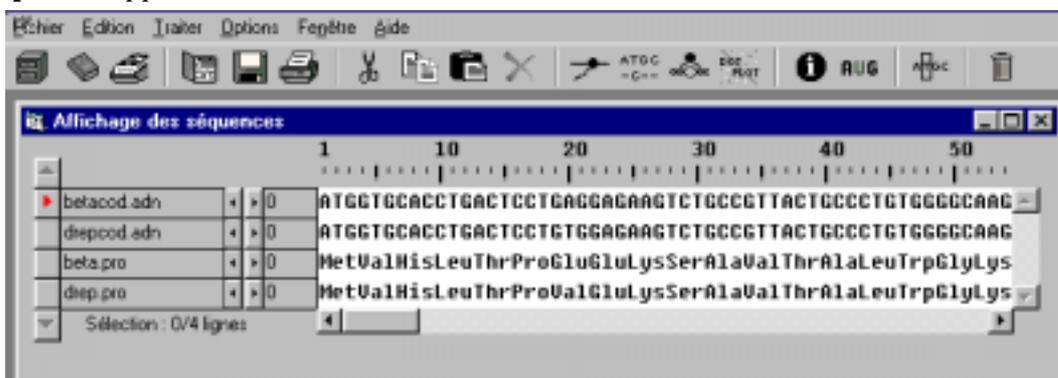
L'enseignant peut ainsi constituer des regroupements de séquences pour constituer des thèmes d'étude personnalisés.

Les thèmes d'étude

C'est l'entrée normale pour les élèves. Les thèmes sont regroupés par grandes sections correspondant à différents contenus d'apprentissage en relation avec les programmes. Dans chaque section on trouve les problématiques introduites par chaque groupe de séquences et si nécessaire un découpage plus fin en thèmes d'étude.



Les séquences sont regroupées et classées par dossiers thématiques, eux-mêmes ordonnés de manière arborescente. Pour les atteindre, il convient de déployer plus ou moins cette arborescence et de sélectionner le dossier désiré; il n'est possible de sélectionner qu'un seul thème d'étude à la fois. Dans la partie droite de la boîte de dialogue s'affiche alors la liste des groupes de séquences disponibles dans ce dossier. Le bouton **Annuler** de la boîte de dialogue provoque sa fermeture et le thème qui a pu être sélectionné n'est pas chargé. La validation de la sélection d'un ou de plusieurs groupes de séquences d'un même dossier thématique s'obtient, ici encore, en cliquant sur le bouton **OK** ou en appuyant sur la touche **Entrée**. Cette validation provoque la fermeture de la boîte de dialogue et le chargement de l'ensemble des séquences que le thème d'étude peut contenir. La fenêtre d'**Affichage des séquences** apparaît alors à l'écran.

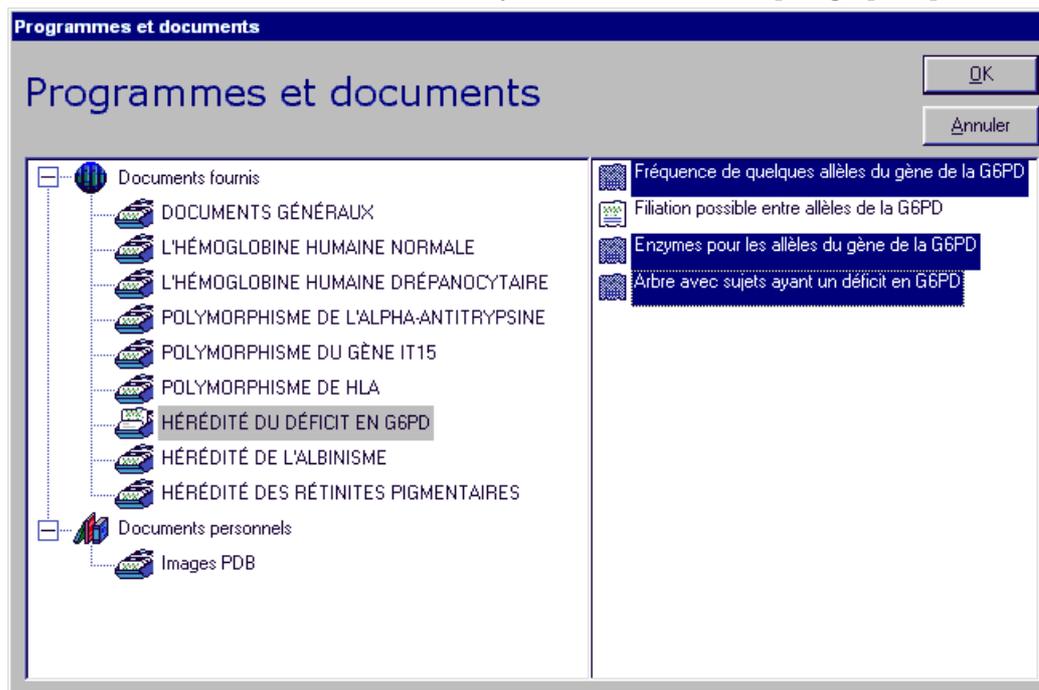


La répétition de cette procédure permet de charger d'autres thèmes d'étude. Un message signalant le changement de thème d'étude demande alors s'il n'est pas souhaitable de fermer les autres fenêtres ouvertes lors de l'étude du thème précédent de façon à vider totalement l'espace de travail avant de charger un nouveau thème d'étude.

Les programmes et documents

À chaque thème d'étude, une série de documents peut être associée (images fixes, tableaux, schémas, séquences animées). Il appartient à l'enseignant de les exploiter dans l'ordre qui lui convient.

L'exploration de l'arborescence des **Programmes et documents** disponibles est identique à celle de la banque de séquences. Le libellé des titres sélectionnés apparaît toujours en inversion vidéo et les boutons **OK** et **Annuler** exercent les fonctions déjà rencontrées dans les paragraphes précédents.



Les fenêtres des documents chargés s'empilent dans l'espace de travail d'*Anagène* mais comme toutes les fenêtres d'*Anagène*, elles disposent d'un bouton de réduction en icône et de la possibilité de modifier leurs dimensions tout en respectant les proportions du document.

Il est aussi possible de charger des documents textuels, iconographiques (images aux formats BMP, WMF, RLE, DIB, GIF ou JPG) et animations de type AVI, MOV, MPG, MP2, MPEG, ou ASF) et même des programmes exécutables propres à l'utilisateur qui pourront enrichir les possibilités d'*Anagène*. Cette fonction est assurée par la commande **Ouvrir** du menu **Fichier** qui assure aussi l'ouverture de fichiers de séquences, de résultats de traitements, et de thèmes d'étude enregistrés en d'autres lieux sur différents supports informatiques (disquette, disque dur, etc.). Il est à noter que la lecture des animations et vidéos numériques n'est possible que si les pilotes correspondants sont correctement installés.

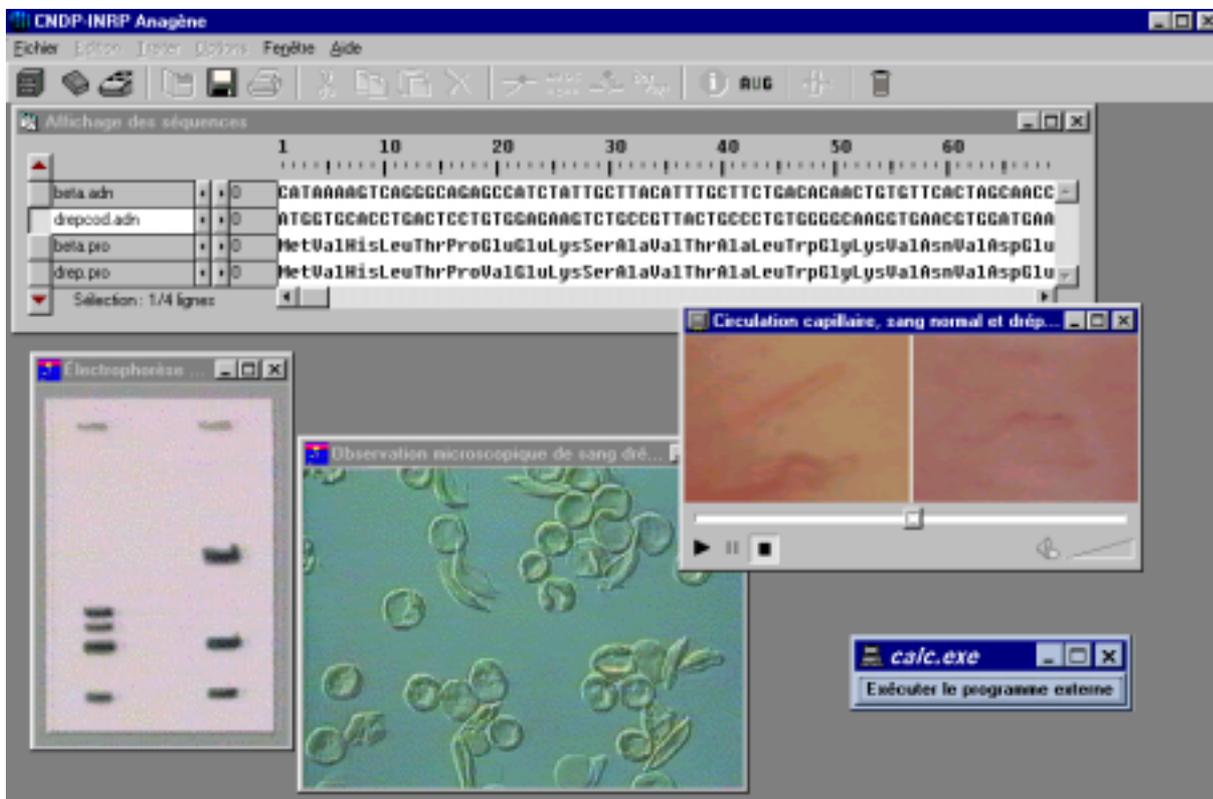
Quant aux programmes externes (type exe ou com) ainsi associés à *Anagène*, il est possible de les exécuter à la condition que la mémoire vive du micro-ordinateur soit suffisante. Si des paramètres, voire des fichiers sont à passer au programme externe, il est nécessaire de les déclarer dans la ligne de commande accessible par le menu **Édition**. Cette ligne d'édition n'apparaît que lorsque la barre de titre du programme externe révèle par sa couleur que la fenêtre est activée. Le lancement d'un programme externe, par exemple la calculatrice de *Windows* présentée dans la copie d'écran, s'obtient

par un clic de la souris pointée sur le bouton intitulé **Exécuter le programme externe**. Le retour dans *Anagène* s'obtient en quittant le programme externe selon la procédure qui lui est propre. Plusieurs programmes peuvent ainsi être associés à *Anagène* mais ils peuvent être aussi répertoriés dans les données propres à l'utilisateur.

LES DONNÉES PERSONNELLES DE L'UTILISATEUR

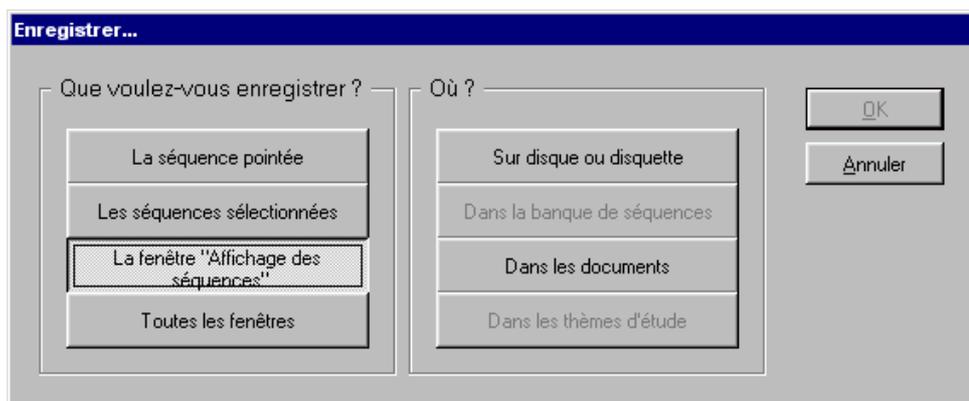
Anagène est un logiciel ouvert qui permet d'intégrer, par le biais de menus personnalisés, les séquences créées ou chargées par l'utilisateur, les documents textuels ou iconographiques dont il dispose ainsi que les nouveaux thèmes d'étude qu'il peut souhaiter concevoir pour le public auquel il s'adresse. *Anagène* constitue ainsi un outil de travail d'autant plus performant que des programmes complémentaires peuvent lui être associés.

La méthode consiste à choisir dans l'espace de travail les éléments qui viendront enrichir les banques de données personnelles. Ces éléments peuvent être de différents types : séquences isolées, fenêtre de séquences, résultat d'un traitement, fenêtres avec les résultats de plusieurs traitements, documents textuels, images fixes, animations, programmes, ensemble des fenêtres présentes dans l'espace de travail.

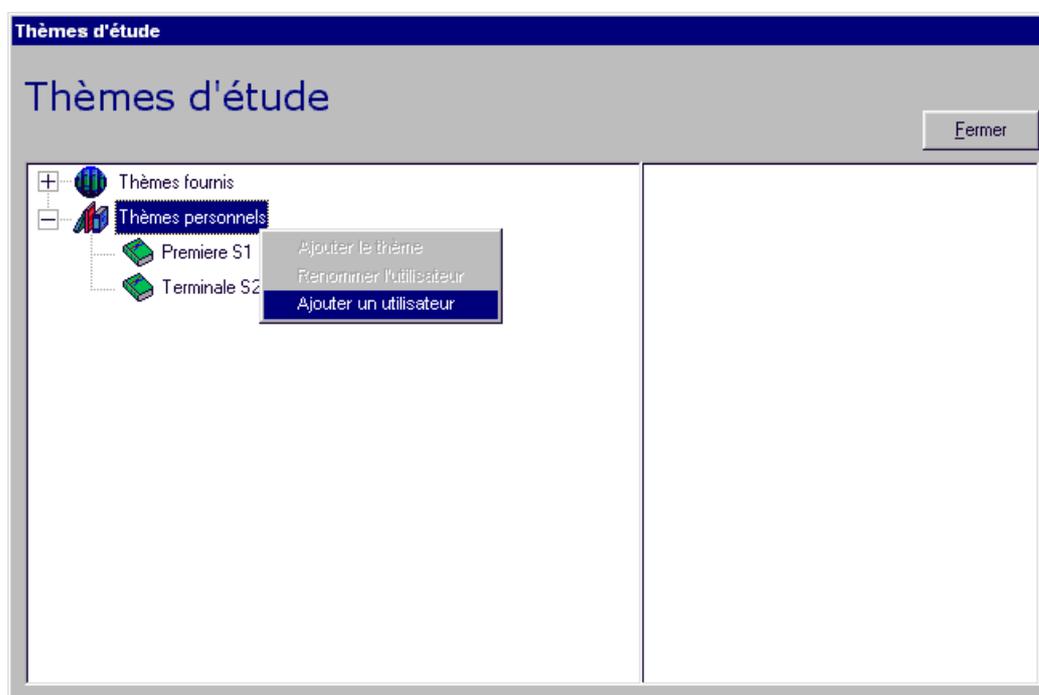


Pour sauvegarder l'un ou l'autre de ces éléments, il suffit d'utiliser la commande **Enregistrer** accessible dans le menu **Fichier** ou plus simplement de cliquer sur le bouton disquette de la barre d'outils.

La boîte de dialogue **Enregistrer...** apparaît alors. Après avoir sélectionné ce qui doit être enregistré, vous devez alors choisir la destination des données. Ce choix dépend de la nature des données et des objectifs.

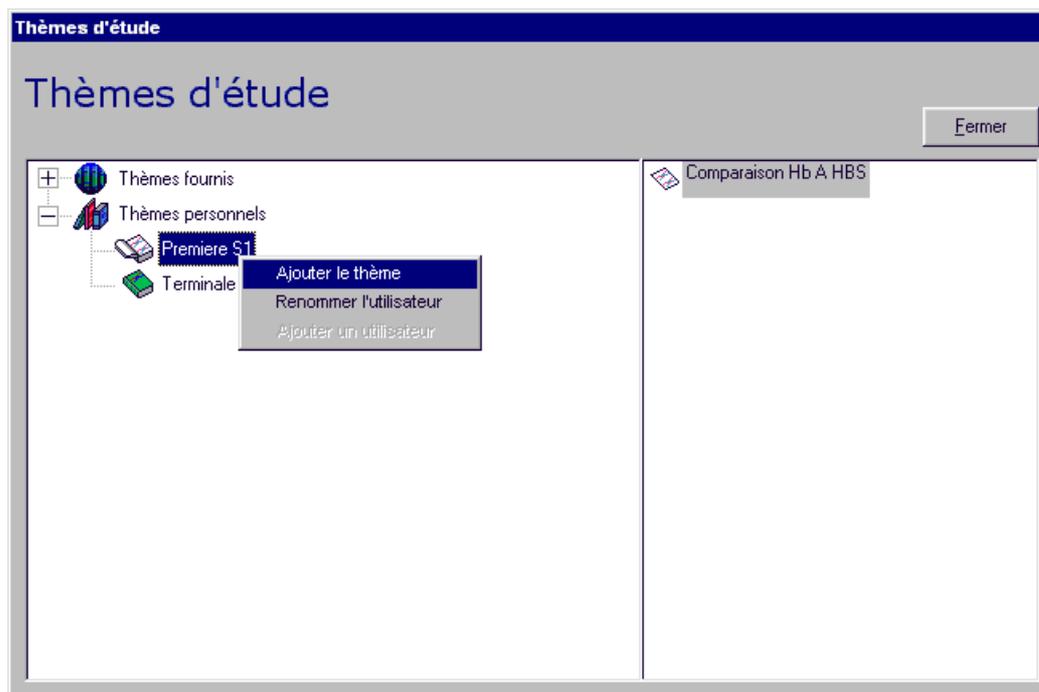


Après définition du choix de l'une des trois banques de données - un thème d'étude par exemple - et confirmation obtenue après validation, la fenêtre du menu correspondant est ouverte.



Il faut commencer par ajouter un utilisateur et lui donner un nom. Ensuite on pourra associer un thème d'étude qui recevra aussi un intitulé.

Sur la copie d'écran suivante, on voit que deux noms d'utilisateurs ont déjà été enregistrés. À la classe de première S1 est déjà associé un thème et on s'apprête à y associer un nouveau.



Les libellés des thèmes et des utilisateurs sont modifiables. La suppression d'un utilisateur n'est possible que s'il ne contient aucun thème. La suppression d'un thème s'obtient en effaçant le libellé dans la ligne d'édition.

L'ajout de séquences

Les séquences que l'on souhaite ajouter au menu des séquences personnelles peuvent avoir trois origines différentes.

Les séquences fournies

Ces séquences, peuvent être chargées à partir de la banque de séquences ou des thèmes d'étude fournis. Pour mettre ces séquences dans le dossier d'un utilisateur, il convient de les sélectionner dans la fenêtre d'affichage/édition d'*Anagène* et d'en demander l'enregistrement dans la banque de séquences personnelles.

Les séquences créées avec l'éditeur d'*Anagène*

La commande **Créer** du menu **Fichier** fait passer la fenêtre d'affichage de séquences en mode édition. Au moment de la création, choisir l'un des quatre types possibles : ADN, ARN, Peptidique ou Texte sachant que l'on peut modifier le nom proposé par défaut.

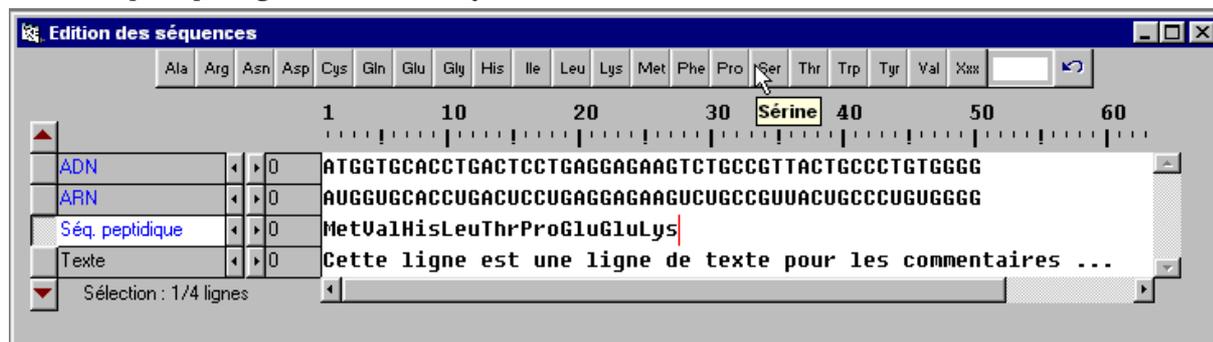
La validation en cliquant sur le bouton **OK** ou en appuyant sur la touche **Entrée** déclenche l'ouverture de la fenêtre **Édition des séquences**.

Selon les cas, « l'alphabet » disponible est limité aux caractères autorisés :

- A, T, G, C, N pour les séquences d'ADN, la lettre N désignant les bases non définies ;
- A,U, G, C, N pour les séquences d'ARN, la lettre N désignant les bases non définies ;
- le code 3 lettres ou 1 lettre adopté au lancement d'*Anagène* pour désigner les acides aminés des séquences peptidiques, les lettres Xxx ou X désignant les acides aminés non définis ;
- les caractères alphabétiques, numériques et les signes de ponctuation pour les lignes de texte servant de commentaire.

En cas d'erreur de saisie lors de la création de séquences nucléiques ou peptidiques, l'appui sur le bouton situé à l'extrémité droite de « l'alphabet » efface la dernière phase de saisie. Il est donc

recommandé, lors de la saisie de longues séquences, de cliquer périodiquement hors de la ligne d'écriture pour protéger les données déjà saisies.



Les séquences importées

La commande **Ouvrir** du menu **Fichier** permet de charger les séquences sauvegardées sur disque, disquette ou autre support informatique. Ces séquences peuvent avoir été créées avec un éditeur ou un traitement de texte et sauvegardées au format texte avec pour extension ADN, COD, ARN, PRO. Si *Anagène* n'a pas pu déterminer la nature d'un fichier de séquence, il propose à l'utilisateur de l'identifier et même de le modifier s'il le souhaite. Il reconnaît en particulier les séquences au format du logiciel *SEQAID*.

La suppression d'une séquence ajoutée par l'utilisateur dans la banque de séquences personnelles s'effectue en six étapes :

- afficher la séquence,
- demander son enregistrement grâce à la commande **Enregistrer** du menu **Fichier**,
- sélectionner son nom,
- demander à **Renommer la séquence** en cliquant sur le bouton droit de la souris,
- effacer le libellé de la séquence et valider.
- accepter la suppression en cliquant sur le bouton OK.

La suppression d'un utilisateur ne se fait que si la liste des séquences personnelles de cet utilisateur est vide.

L'ajout de thèmes d'étude

Malgré la variété des thèmes d'étude fournis, il est possible d'en créer de nouveaux adaptés à une stratégie pédagogique personnelle ou à un public déterminé.

Un thème d'étude est constitué par l'enregistrement de toutes les fenêtres présentes dans l'espace de travail d'*Anagène*, même si elles sont réduites en icônes. Ces fenêtres peuvent provenir des chargements effectués à partir des banques de séquences, des thèmes d'études ou des programmes et documents fournis ou personnels, des compléments en ligne téléchargés. Ces fenêtres peuvent aussi résulter des traitements effectués sur les séquences. Enfin, elles peuvent être importées grâce à la commande **Ouvrir** du menu **Fichier** qui propose types de fichiers suivants :

- **Séquences** nucléiques et protéiques qui seront chargés dans la fenêtre d'affichage de séquences d'*Anagène* ;
- **Traitements**, fenêtres de résultats des conversions, comparaisons, actions enzymatiques effectuées sur des séquences et sauvegardées lors de précédentes sessions de travail avec *Anagène* ;
- **Textes**, fichiers créés avec des éditeurs ou des traitements de texte et enregistrés au format texte (extension TXT) ;
- **Internet**, fichiers de pages au format HTM ou HTML en téléchargeant sur le réseau Internet ;

- **Images**, provenant en particulier de banques de la collection Imagerie Scientifique éditée par le CNDP ou issues d'autres sources ; les formats reconnus sont : BMP, WMF , RLE, DIB, GIF ou JPG ;
- **Vidéos**, au format AVI, MOV, MPG, MP2, MPEG ou ASF, qu'elles soient d'origine infographique ou obtenues par numérisation de vidéogrammes ;
- **Programmes**, qu'il s'agisse des fichiers de commandes BAT ou PIF ou de programmes exécutables au format EXE ou COM.
- **Thèmes d'étude**, créés au cours d'autres sessions avec *Anagène* et enregistrés sur disque dur , disquette ou autre support informatique.

La première étape de la création d'un nouveau thème d'étude personnel, consiste donc à ne conserver dans l'espace de travail d'*Anagène* que les fenêtres souhaitées et en éliminant de la fenêtre d'affichage/édition, les séquences inutiles. La seconde étape a pour but d'enregistrer le thème d'étude.

La suppression d'un thème d'étude créé par l'utilisateur se fait de la même manière que la suppression d'une séquence.

L'ajout de programmes et de documents

Toute fenêtre présente dans l'espace de travail d'*Anagène*, quel que soit son contenu (programme exécutable, document iconographique ou textuel, séquences ou résultat de traitement de séquences), peut être ajoutée dans la liste des programmes et documents personnels. Il est à noter que c'est la fenêtre active, c'est à dire celle dont la barre de titre est en inversion vidéo, qui constituera, selon sa nature, le programme ou le document à enregistrer. On utilisera à cet effet la commande **Enregistrer** du menu **Fichier**.

Ici encore, ce programme ou ce document aura pour origine l'une des banques fournies avec le logiciel ou pourra être importé grâce à la commande **Ouvrir** du menu **Fichier**.

La suppression d'un programme ou d'un document ajouté par l'utilisateur s'effectue de la même manière que la suppression d'une séquence.

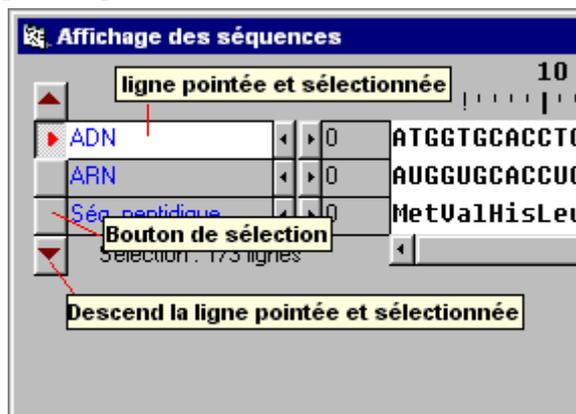
L'ÉDITION ET LES TRAITEMENTS DE SÉQUENCES

L'édition de séquences

Il convient de bien noter au préalable que :

- les banques de séquences et de thèmes fournies ne sont pas modifiables par l'utilisateur enseignant ou élève avec le logiciel *Anagène* lui-même ;
- au début d'une session de travail, les séquences sont protégées contre une éventuelle altération, et il convient donc de les déprotéger dans le menu **Options** si on veut utiliser des traitements qui modifient ces séquences ;
- la modification de séquences ne peut avoir lieu que dans la fenêtre d'édition des séquences ; le seul fait de décocher « Protéger les données » transforme la fenêtre **Affichage des séquences** en **Édition des séquences** et réciproquement ;
- les commandes d'édition sont regroupées dans le menu **Edition** et directement accessibles pour certaines à partir de leurs icônes ; noter la présence dans ce groupe de la commande d'**inversion de séquences** ;
- il est recommandé de dupliquer toute séquence avant de la modifier par l'usage du presse-papiers (**Copier** et **Coller**) ; les modifications étant limitées par l'« alphabet » propre à chaque type de séquence, l'espace n'étant autorisé que dans les lignes de commentaires ;

- l'ordre vertical dans lequel se trouvent les séquences dans la fenêtre **Affichage des séquences** ou **Edition des séquences** peut être modifié :



- en triant (commande **Trier** du menu **Fichier**) les séquences sélectionnées par ordre alphabétique croissant, inverse, ou par succession de types de séquences (ADN-ARN-Peptidique ou Peptidique-ARN-ADN) ;
- en cliquant sur les flèches situées en haut et en bas des boutons de sélection de la liste des séquences, la ligne pointée et sélectionnée pourra monter ou descendre dans cette liste de séquences ;
- le décalage d'une séquence est possible en cliquant sur les boutons de décalage.

Les traitements de séquences

- Les commandes des traitements se trouvent dans le menu **Traiter** et sont regroupées en quatre catégories : conversion de séquences, comparaison de séquences, action enzymatique et graphique de ressemblance.
- *Remarque* : le module graphique de ressemblance est en cours de développement et sera présent dans la version 2 d' *Anagène*.
- Un traitement n'est applicable que si un nombre suffisant de séquences déjà sélectionnées est atteint et si ces séquences sont de même nature ; pour sélectionner une séquence, il faut cliquer sur son bouton de sélection, ce qui provoque un affichage en inversion vidéo de son libellé.
- Après traitement, les séquences perdent leur sélection et il convient d'effectuer une nouvelle sélection pour appliquer de nouveau un traitement ; chaque traitement donne lieu à un choix d'options à valider dans une boîte de dialogue spécifique ; ces options restent mémorisées au cours d'une session de travail.
- Les résultats d'un traitement sont par défaut dirigés vers la ou les fenêtres spécifiques du traitement ; lorsqu'un traitement engendre de nouvelles séquences qui peuvent à leur tour être l'objet d'un traitement, celles-ci peuvent être placées dans la fenêtre d'**Affichage des séquences** ou d' **Edition des séquences** à la suite des séquences déjà présentes.

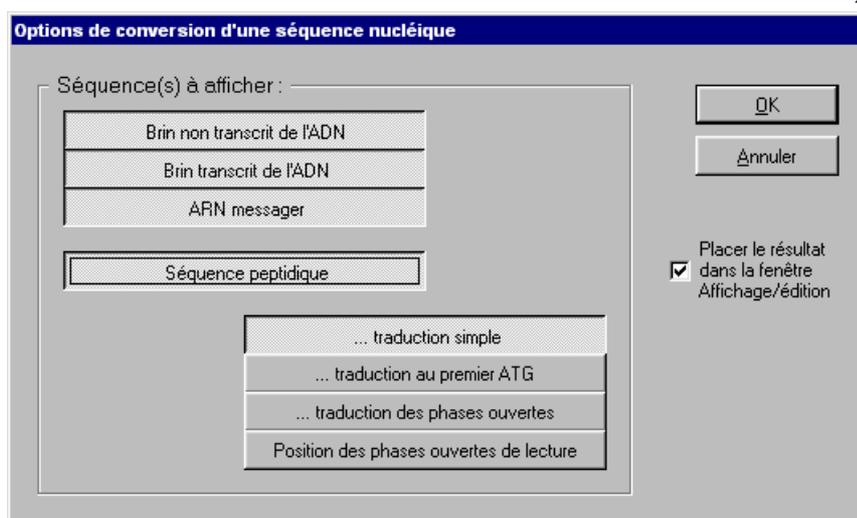
La conversion de séquences

Cette fonction peut être appliquée sur les séquences nucléiques ou peptidiques et simultanément sur plusieurs séquences si elles sont de même nature.

Si les séquences sélectionnées appartiennent à de l'ADN ou de l'ARN, une boîte de dialogue spécifique propose de cocher les conversions souhaitées et la fenêtre de destination.

Pour une conversion en séquence peptidique la traduction simple est proposée par défaut, mais on peut faire un autre choix. Il est important de noter qu'il ne s'agit pas de « traduction biologique » dans

la mesure où cette conversion s'effectue du premier triplet jusqu'au premier codon Stop rencontré en respectant le code génétique mais sans tenir compte de la présence d'un codon d'initiation. Pour que la traduction commence effectivement au codon initiation, choisir alors **Traduction au premier ATG**.

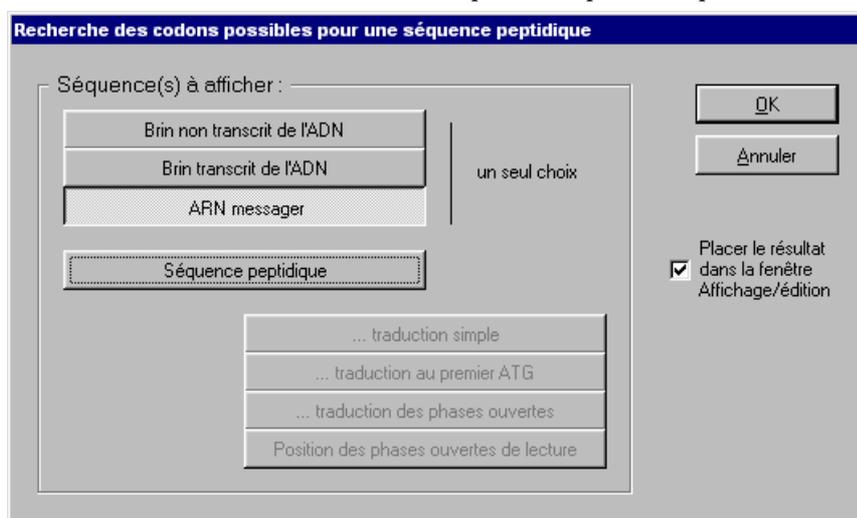


L'option **Traduction des phases ouvertes** assure les conversions correspondantes d'une séquence du premier codon d'initiation rencontré jusqu'au premier codon de terminaison, en commençant respectivement par la première base, la seconde et la troisième (les trois cadres de lecture).

L'option **Position des phases ouvertes de lecture** permet de repérer les codons spécifiques de l'initiation (symbolisés par >>>) et de la terminaison (symbolisés par ~~~).

On peut aussi, avec les commandes du menu **Edition, Copier** une sélection effectuée dans une séquence et la **Coller** dans la fenêtre d'édition en tant que nouvelle séquence à traiter. Il faut garder à l'esprit que le logiciel *Anagène* est un outil qui assure des traitements définis à l'initiative de l'utilisateur et que l'interprétation des résultats lui incombe.

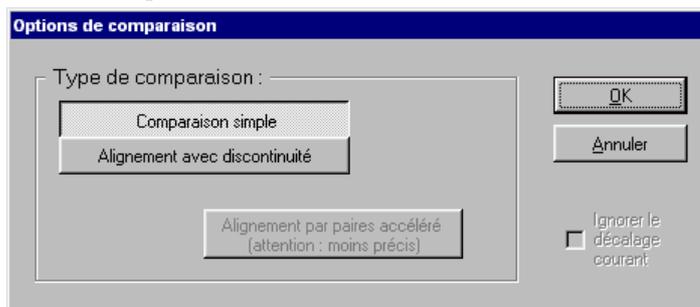
Si les séquences sélectionnées sont de nature peptidique, la seule option de conversion autorisée consiste à déterminer en ADN ou en ARN, les codons possibles pour chaque acide aminé.



La comparaison de séquences

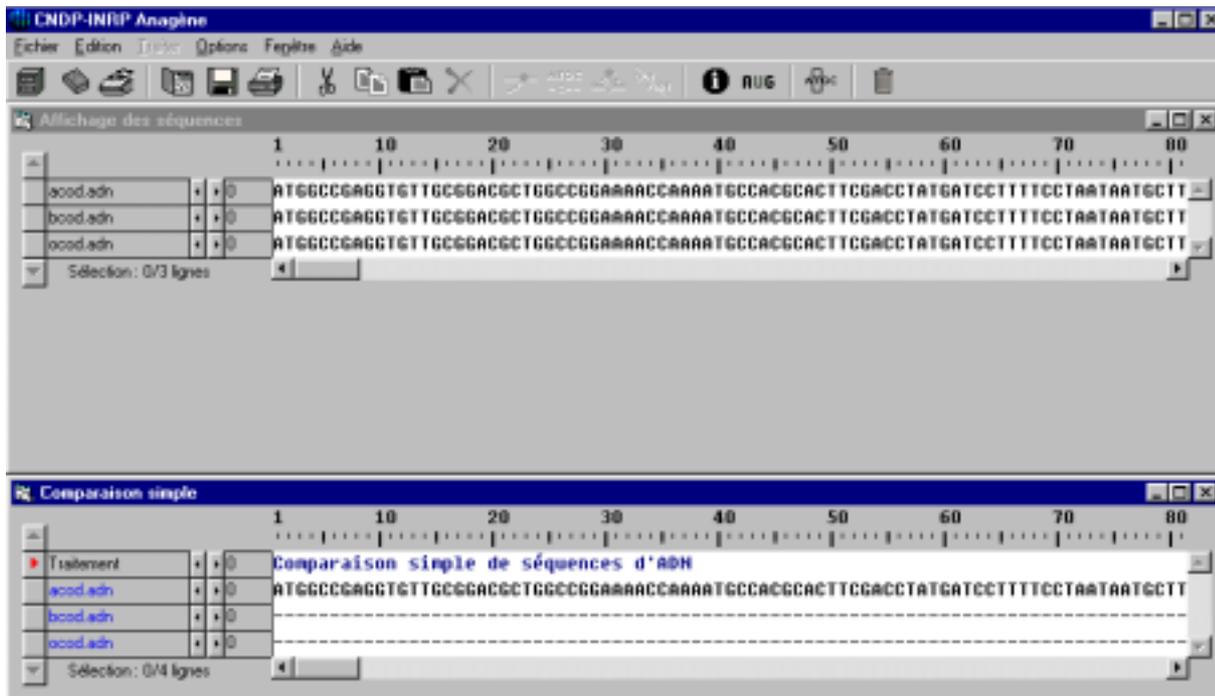
Pour être appliqué, ce traitement nécessite que deux séquences au moins, toutes les deux de nature nucléique ou peptidique, soient au préalable sélectionnées dans la fenêtre d'affichage/édition.

Deux types de comparaison sont possibles :



- **Comparaison simple** à partir de la première position dans chacune des séquences ; ce mode est plus particulièrement indiqué pour comparer par exemple des allèles d'un même gène comportant le même nombre de bases.
- Remarque : si un décalage des séquences a été introduit avant la comparaison, il peut être ignoré en cochant la case **Ignorer le décalage courant** dans la boîte de dialogue **Options de comparaison**.
- **Alignement avec discontinuité** : l'algorithme de comparaison utilisé par les chercheurs en biologie moléculaire fait intervenir plusieurs paramètres définis par défaut ; le principe consiste à arranger les séquences de manière à obtenir le plus grand nombre d'identités ; ce mode est indiqué lorsque les séquences ne sont pas de même longueur ;
- **Alignement avec discontinuités par paires accéléré** : l'algorithme de comparaison utilisé avec cette option est plus ancien, plus rapide mais moins précis.

Le résultat du traitement s'affiche dans une fenêtre spécifique qui se place, comme pour les autres traitements appliqués, en dessous de la fenêtre d'affichage des séquences.

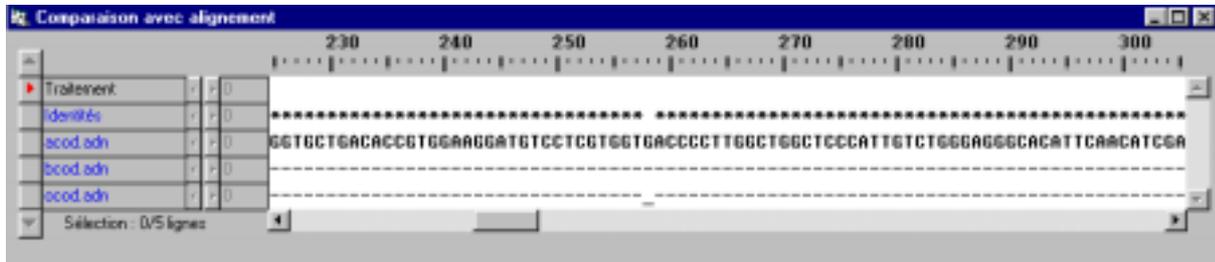


La première des séquences sert de référence pour la comparaison. L'information accessible par le bouton correspondant de la barre d'outils reconnaissable à son « Point i » fournit à propos de la ligne pointée, le degré de similitude entre les séquences et la signification des symboles utilisés.

Dans le cas d'une comparaison simple, il est possible, dans la fenêtre de résultat, de provoquer le décalage d'une séquence en cliquant sur les boutons de décalage et d'obtenir immédiatement les modifications du résultat de la comparaison à la séquence de référence.

L'alignement avec discontinuités nécessite des temps de calcul d'autant plus importants que les séquences sont nombreuses, longues et différentes.

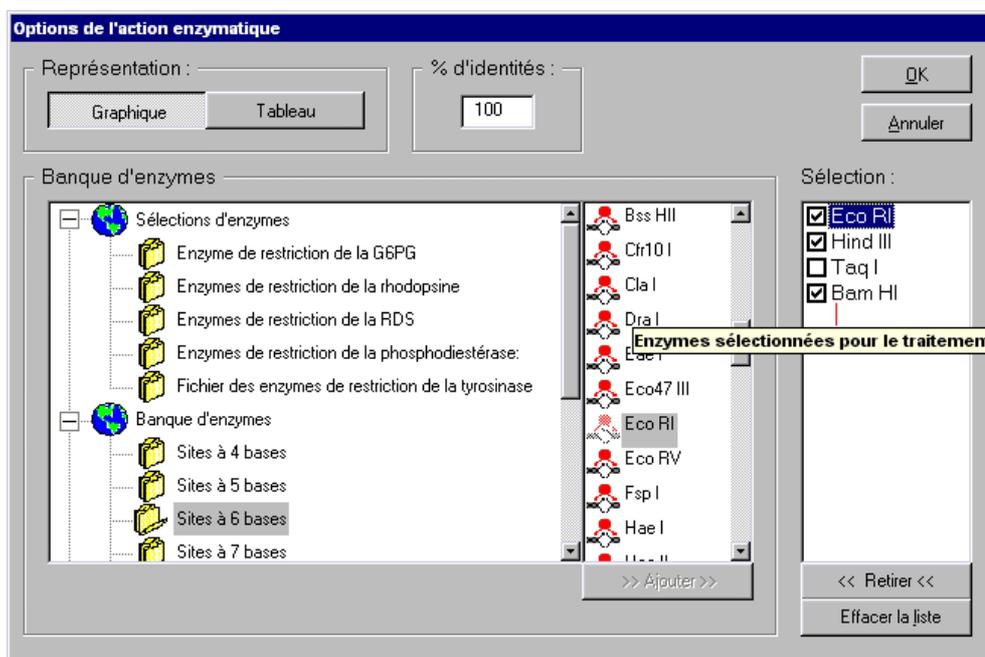
Il convient d'explorer les séquences avec la barre de défilement horizontal pour découvrir les régions plus ou moins semblables.



Remarque : si l'on souhaite, pour une comparaison, changer de séquence de référence, il suffit avec les flèches situées en haut et en bas des boutons de sélection de la fenêtre d'affichage des séquences, de remonter ou de descendre la séquence pointée et sélectionnée. On a intérêt à bien choisir l'ordre des séquences en fonction du degré de ressemblance entre elles. Si la comparaison porte sur trois séquences et que deux d'entre elles sont semblables, leur ressemblance sera plus évidente si l'une des deux sert de référence.

L'action enzymatique

L'analyse avec des enzymes de restriction s'applique aux séquences d'ADN sélectionnées au préalable dans la fenêtre d'affichage/édition. Le choix des enzymes s'effectue dans la boîte de dialogue suivante.



Cette boîte de dialogue se présente comme une série de sélections préétablies d'enzymes de restriction utiles au traitement de certaines séquences dans le cadre d'études de prévisions en génétique humaine et d'une banque d'enzymes organisée en dossiers dans lesquels les enzymes se trouvent classées en

fonction de la longueur du site de coupure qu'elles reconnaissent. Quand un dossier est ouvert, les enzymes sont à sélectionner dans la liste qui s'affiche au milieu et à **Ajouter** dans la fenêtre de droite. Dans cette fenêtre, à droite, avec **Sélection** d'enzymes, il est possible :

- d'ajouter de nouvelles enzymes présentes dans d'autres dossiers ;
- de décocher chacune d'elle en cliquant sur la coche ;
- de retirer l'enzyme sélectionnée en cliquant sur le bouton **Retirer** ;
- de retirer la totalité des enzymes de cette liste en cliquant sur le bouton **Effacer la liste**.

De plus, cette sélection peut être mémorisée en cliquant sur le bouton droit de la souris au niveau de **Lots personnels** (situé tout en bas de l'arborescence), puis en cliquant sur **Ajouter un lot personnel** et en lui donnant un nom.

Ainsi ces fichiers d'enzymes pourront être ouverts lors d'une nouvelle session de travail en cliquant sur le bouton **Fichier**.

Deux modes de représentation de l'action enzymatique sont prévus : la visualisation sous la forme d'un tableau ou d'un graphique, les deux modes pouvant être retenus simultanément.

Après validation de la liste des enzymes sélectionnées, les fenêtres de résultat se superposent en dessous de la fenêtre d'affichage / édition des séquences.

Action enzymatique

Alu I (100% d'identités) : 5 sites de coupure

| Enzymes (100%) | Alu I | Hpa II | Mnl I | Fsp I |
|----------------|-------|--------|-------|-------|
| acod.adn | 5 | 6 | 12 | 0 |
| bcod.adn | 6 | 5 | 12 | 0 |
| ocod.adn | 5 | 6 | 12 | 0 |

Sélect : 0/4 lignes

Carte de restriction correspondant à la sélection

Carte de restriction

acod.adn | Alu I | % d'identités : 100

Carte de restriction correspondant à la sélection

1 1062

TTCACCGACCAGCTGCCCGCGGTGCCCGCGTGACGCTGGGGACCGGTCGGCAGCTGTCAGTCTGGAGGTGCCGCCTACAAGCGC
 AAGTGGCTGGTCGACCGCGGCCACGGGGCGCACTGCGACCCCTGGCCAGCCGTCGACAGTCACGACCTCCACGCGCGGATGTTCCGG

460 470 480 490 500 510 520 530

Loupe sur la portion de séquence encadrée

On peut voir les deux fenêtres avec les commandes du menu **Fenêtre** d'*Anagène* qui permettent divers affichages dont le mode **Mosaïque**. Il est à noter qu'en présence de fenêtres contenant des documents iconographiques, ce mode ne peut être parfait, les proportions des images étant toujours respectées.

La fenêtre **Action enzymatique** présente le mode d'affichage **Tableau** et montre le nombre de sites reconnus par chaque enzyme de la sélection active (enzymes cochées dans la liste **Sélection d'enzymes**) agissant sur chaque séquence nucléique sélectionnée.

La fenêtre **Carte de restriction** présente le mode d'affichage **Graphique** et illustre l'action d'une enzyme de restriction sur une des séquences traitées. On peut faire varier soit la séquence choisie, soit

l'enzyme. Il faut explorer la carte de restriction en déplaçant la « loupe » le long de la séquence pour y découvrir les sites de coupure.

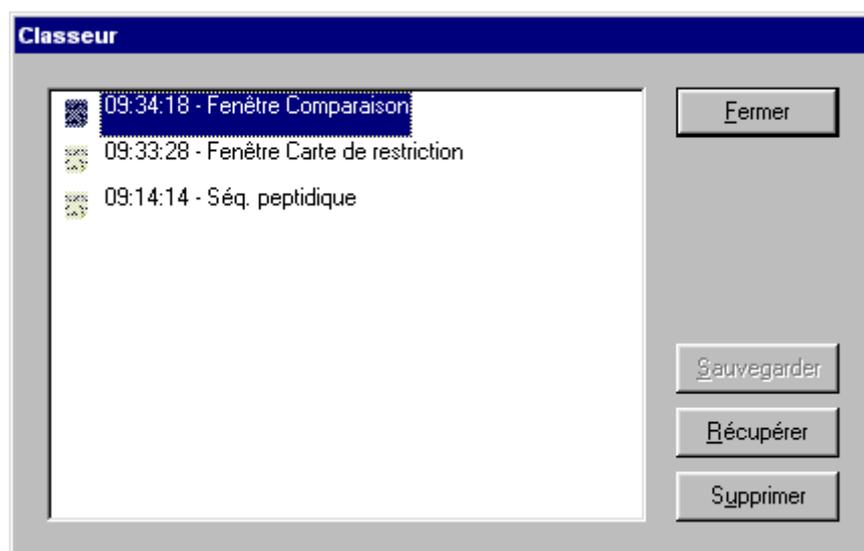
La valeur du % **d'identités** permet de rechercher des sites plus ou moins strictement identiques à celui que chaque enzyme détecte.

Bien d'autres données associées à ces tableaux de résultats et accessibles par le bouton **Informations sur la ligne pointée...** sont disponibles tant sur les sites d'action que sur les fragments de séquences d'ADN obtenus. Comme pour les autres fenêtres d'informations, il est possible d'en imprimer le contenu ou de le copier dans le presse-papiers de *Windows*.

LE CLASSEUR D'ANAGÈNE

Il s'agit d'un utilitaire qui permet à l'élève de sauvegarder simplement et temporairement les séquences qui, trop nombreuses, viennent à encombrer la fenêtre d'affichage des séquences ou les résultats des traitements effectués.

Sa simplicité tient au fait qu'il n'y a pas à parcourir l'arborescence du disque dur pour y enregistrer des fichiers. Après avoir sélectionné les séquences de la fenêtre d'affichage ou la fenêtre de résultat à sauvegarder, il suffit de cliquer sur le bouton correspondant de la barre d'outils pour l'ouvrir. Les données pourront alors être intégrées au classeur par l'ajout de leur libellé et de l'heure correspondante.



En fin de session, ce classeur est vidé après affichage d'un message d'avertissement invitant à en visiter le contenu pour récupérer si besoin des données. Le disque dur n'est pas ainsi surchargé avec des fichiers inutiles.

En cas de coupure inopinée de l'électricité ou de réinitialisation du micro-ordinateur, le contenu du classeur est récupérable dès le lancement d'*Anagène*.

GUIDE DE RÉFÉRENCE

LES COMMANDES DES MENUS

Menu Fichier

- Banque de séquences** affiche sous forme arborescente la liste des séquences disponibles.
- Thèmes d'étude** affiche sous forme arborescente la liste des thèmes d'étude prédéfinis.
- Programmes et documents** affiche sous forme arborescente la liste correspondante.
- Créer** permet de créer une séquence nucléique, protéique ou une ligne de texte.
- Ouvrir** permet de choisir un fichier dans l'arborescence des répertoires et l'ouvre dans l'espace de la fenêtre mère d'*Anagène*.
- Intégrer** permet d'entrer les données chargées directement dans la banque.
- Enregistrer** assure la sauvegarde de séquences, d'une fenêtre ou de l'ensemble.
- Imprimer** imprime les séquences choisies et les résultats de traitements appliqués.
- Quitter** permet, après confirmation, de quitter définitivement le logiciel *Anagène*.

NB : Il est toujours proposé d'enregistrer une fenêtre dont le contenu a changé avant de la fermer.

Menu Edition

- Couper** permet de mémoriser une sélection dans le presse-papiers ; de plus il efface la sélection.
- Copier** permet simplement de mémoriser une sélection dans le presse-papiers.
- Coller** permet de recopier à l'emplacement du curseur, la sélection mémorisée dans le presse-papiers.
- Effacer** permet de supprimer définitivement une sélection.
- Sélectionner tout** assure la sélection de l'ensemble des séquences d'une fenêtre.
- Désélectionner tout** annule les sélections effectuées dans une fenêtre.
- Rechercher** permet de rechercher dans la séquence où se situe le curseur, les occurrences d'une chaîne de caractères.
- Minuscules** convertit en minuscules la sélection effectuée dans la fenêtre d'édition des séquences.
- Majuscules** convertit en majuscules la sélection effectuée dans la fenêtre d'édition des séquences.
- Inverser la séquence** provoque l'inversion de la séquence pointée dans la fenêtre d'édition des séquences.
- Trier les séquences** organise dans l'ordre choisi les séquences de la fenêtre d'affichage/édition des séquences.

Menu Traiter

- Convertir les séquences** affiche ADN brin non transcrit, ADN brin transcrit, ARNm, séquence peptidique, codons possibles pour un polypeptide.
- Comparer les séquences** compare des séquences à une séquence de référence à partir de la première position et réalise l'alignement avec discontinuités.
- Action enzymatique** révèle les sites d'action d'enzymes de restriction sur des séquences d'ADN.
- Graphique de ressemblance** compare deux séquences en produisant une matrice rectangulaire. (*Remarque* : ce traitement est en cours de développement et sera présent dans la version 2 d'*Anagène*).

Menu Options

- ✓ **Protéger les données** empêche toute modification de séquence dans la fenêtre d'affichage.
- ✓ **Grand curseur** allonge le curseur sur la hauteur de l'ensemble des séquences jusqu'à l'échelle et sur la largeur de la sélection en cours.
- ✓ **Règle en triplets** affiche la règle graduée en triplets plutôt qu'en caractères.
- ✓ **Bulles d'aide** affiche les bulles d'aide associées aux objets de l'interface.

Menu Fenêtre

- Mosaïque** place les fenêtres côte à côte tout en respectant les proportions des documents iconographiques.
- Cascade** organise les fenêtres en les superposant légèrement décalées.
- Réorganiser les icônes** aligne les fenêtres réduites en icônes grâce au bouton de réduction situé à droite de la barre de titre.
- Les titres des fenêtres** permettent d'accéder directement aux fenêtres désirées.

Menu Aide

- À propos d'Anagène** affiche les informations relatives à l'édition du logiciel.
- Recherche de l'aide** assure la recherche sur les mots clés indexés de l'aide.
- Sommaire de l'aide** permet d'accéder au sommaire de l'aide en ligne.
- @ Complément en ligne** affiche une page HTML si un navigateur récent est fonctionnel sur le micro-ordinateur. Cette page comporte des liens vers les sites du CNDP et de l'INRP ainsi qu'un lien vers un site dédié à *Anagène*. En l'absence de navigateur récent, le micro-ordinateur affiche une page comportant les adresses de ces sites.

LES ICÔNES DE LA BARRE D'OUTILS

Première série : banques de données



- 1- Afficher l'arborescence de la banque de séquences.
- 2- Afficher l'arborescence de la liste des thèmes d'étude.
- 3- Afficher l'arborescence de la liste des programmes et documents.

Deuxième série : outils de sauvegarde



- 1- Voir le classeur.
- 2- Enregistrer les séquences sélectionnées, la fenêtre active ou l'ensemble des fenêtres.
- 3- Imprimer les séquences choisies et les résultats des traitements appliqués.

Troisième série : outils d'édition



- 1- Couper une sélection et la mémoriser dans le presse-papiers.
- 2- Copier une sélection et la mémoriser dans le presse-papiers.
- 3- Coller à l'emplacement du curseur, la sélection mémorisée dans le presse-papiers.
- 4- Effacer définitivement une sélection.

Quatrième série : outils de traitement



- 1- Convertir les séquences : affiche ADN brin non transcrit, ADN brin transcrit, ARNm, séquence peptidique, codons possibles pour un polypeptide.
 - 2- Comparer les séquences à une séquence de référence à partir de la première position et réaliser l'alignement avec discontinuités.
 - 3- Révéler les sites d'action d'enzymes de restriction sur des séquences d'ADN.
 - 4- Afficher le graphique de ressemblance des deux séquences sélectionnées.
- Remarque* : l'icône est active mais n'affiche qu'un message rappelant que le module graphique de ressemblance est en cours de développement.

Cinquième série : accès aux informations



- 1- Renseigner la séquence pointée ou d'un résultat d'un traitement pointé.
- 2- Afficher le tableau interactif du code génétique en ARN.

Avant-dernière icône : gestion du curseur



Allonger le curseur sur la hauteur de l'ensemble des séquences et de la règle et sur la largeur de la sélection en cours ou rétablir le curseur par défaut.

Dernière icône



Après confirmation, fermer toutes les fenêtres.

LES RACCOURCIS CLAVIER

Alt F4 permet, après confirmation, de fermer toutes les fenêtres puis de quitter *Anagène*.

Ctrl F4 provoque la fermeture de la fenêtre active.

Ctrl C permet de copier une sélection et de la mémoriser dans le presse-papiers.

Ctrl X permet de couper une sélection et de la mémoriser dans le presse-papiers.

Ctrl V permet de coller à l'emplacement du curseur, la sélection mémorisée dans le presse-papiers.

Ctrl A permet de sélectionner toutes les séquences de la fenêtre active.

Ctrl B supprime ou rétablit les bulles d'aide.

Ctrl R permet de rechercher dans la séquence une suite de caractères préalablement sélectionnée ou saisie au clavier.

Ctrl I permet d'imprimer les séquences choisies et les résultats de traitements appliqués.

F1 affiche l'aide en ligne du logiciel.

F3 permet d'atteindre l'occurrence suivante dans une recherche sur une suite de caractères.

F4 permettra d'activer le graphique de ressemblance.

F5 modifie l'aspect du curseur.

F6 modifie l'aspect de la règle graduée.

F7 assure les traitements de conversion de séquences.

F8 assure les traitements de comparaison de séquences.

F9 assure les traitements enzymatiques sur les séquences d'ADN.

F12 permet d'enregistrer les séquences sélectionnées, la fenêtre active ou l'ensemble des fenêtres.

Maj F4 organise les fenêtres en cascade.

Maj F5 organise les fenêtres en mosaïque tout en respectant les proportions des documents iconographiques.

Suppr efface définitivement une sélection.

Entrée provoque le même effet qu'un clic sur le bouton **OK** ou sur tout autre bouton de validation de choix.

Echap provoque le même effet qu'un clic sur le bouton **Annuler** pour abandonner un choix.

Alt Lettre soulignée équivaut à un clic sur la commande correspondante d'un menu déroulant.

LE VISIONNEUR D'ANIMATIONS

Le visionneur d'animations permet d'agir sur le déroulement de l'animation ou de la vidéo chargée.



-  Lancement de la lecture de l'animation ou de la vidéo.
-  Pause dans la lecture de l'animation ou de la vidéo.
-  Arrêt de la lecture et repositionnement au début de l'animation ou de la vidéo.
-  Positionnement par déplacement du curseur la lecture à un moment précis.
-  Activation ou désactivation du son qui accompagne l'animation ou la vidéo.
-  Règlage du volume sonore par déplacement du curseur le long de la barre.

SUGGESTIONS PÉDAGOGIQUES

INTRODUCTION

Au lycée, tous les programmes de sciences de la vie et de la Terre en vigueur depuis la rentrée de septembre 2002 accordent une place importante aux relations existant entre le génotype d'un organisme et son phénotype.

En seconde, l'accent est mis sur l'universalité et la variabilité de la molécule d'ADN.

En première S, l'ensemble du programme de la partie sciences de la vie s'articule autour des relations existant entre le génotype d'un organisme et son phénotype. L'étude de la synthèse protéique permet d'établir le lien entre gènes et protéines ; l'élève doit donc savoir expliquer comment la diversité phénotypique résulte d'interactions complexes entre la variabilité génétique et l'environnement.

En première L et ES, les applications biotechnologiques sont le prolongement de la mise en évidence de la relation génotype-phénotype.

En terminale S, il s'agit de faire comprendre l'apparente contradiction entre stabilité et variabilité, tant au niveau de la géosphère que de la biosphère. L'établissement de phylogénies moléculaires rend compte de la stabilité et de la variabilité des génomes. L'évolution des espèces sera envisagée au travers des processus de formation des nouveaux allèles et de formation des nouveaux gènes par mutation et duplication des gènes. L'adaptabilité et la variabilité du système immunitaire et son évolution résultent de l'interaction entre le génotype et l'environnement.

Un des trois thèmes enseignés en spécialité de terminale S traite des débuts de la génétique aux enjeux actuels des biotechnologies. L'utilisation des enzymes de restriction a ouvert la voie au clonage des gènes et de leur séquençage permettant de connaître un grand nombre de séquences de gènes dont on saisit mieux la répartition des différents allèles au sein des populations humaines.

On dispose donc dans la littérature scientifique et les banques de données internationales d'informations qui peuvent servir de support pour traiter les sujets inscrits aux programmes. Reste donc le problème de l'approche pédagogique permettant aux élèves de se les approprier.

L'édition, la conversion, la comparaison de séquences, ainsi que leur étude avec les enzymes de restriction constituent des outils d'analyse simples et particulièrement adaptés pour mettre les élèves en situation de résolution de problèmes. Les stratégies pédagogiques proposées ont pour dénominateur commun de solliciter l'initiative des élèves dans le choix des séquences pertinentes et des traitements à réaliser, ainsi que dans l'enchaînement des actions adéquates en fonction des résultats obtenus.

Les fonctions du logiciel *Anagène* sont suffisamment simples et les thèmes d'étude assez variés pour permettre aux enseignants des choix alternatifs et leur laisser toute liberté dans la conception d'autres approches.

La banque de séquences et les thèmes d'étude fournis avec cette version 1.8 d'*Anagène* ont été initialement établis pour la version 1 diffusée en 1997. Depuis septembre 2002, tous les programmes des sciences de la vie et de la Terre au lycée ont changé. Cependant, la majorité des thèmes d'étude proposés ci-après conservent leur intérêt pédagogique et peuvent toujours s'intégrer dans les nouvelles orientations et attentes des nouveaux programmes. Le téléchargement de séquences, de nouveaux thèmes d'étude proposés sur le site de l'INRP dédié à *Anagène* et orientées vers les nouveaux programmes enrichiront et diversifieront les démarches pédagogiques.

Le tableau ci-dessous établit la correspondance possible entre les thèmes d'étude et la banque de séquences livrés avec la version 1.8 et leur exploitation dans le cadre des nouveaux programmes en vigueur en 2002.

| | Seconde | Première L | Première ES | Première S | Terminale S | Spécialité TS |
|--|---------|------------|-------------|------------|-------------|---------------|
| EXPRESSION DE L'INFORMATION GÉNÉTIQUE | ✓ | ✓ | ✓ | ✓ | | |
| RELATIONS PHÉNOTYPE-GÉNOTYPE | | | | | | |
| Phénotype drépanocytaire | | ✓ | ✓ | ✓ | | ✓ |
| Phénotype thalassémique | | ✓ | ✓ | ✓ | | |
| Phénotype « groupe sanguin ABO » | | ✓ | ✓ | ✓ | ✓ | |
| Phénotype clinique phénylcétonurique | | | | | | |
| POLYMORPHISME DES GÈNES | | | | | | |
| Polymorphisme de ABO | | | | | ✓ | |
| Polymorphisme de AT | | | | | ✓ | |
| Polymorphisme de IT15 | | | | | ✓ | |
| Polymorphisme de HLA | | | | | ✓ | |
| Polymorphisme du CMH | | | | | ✓ | |
| FAMILLES MULTIGÉNIQUES | | | | | | |
| Gènes des globines | | | | | ✓ | |
| Gènes des HLA | | | | | ✓ | |
| Gènes des hormones hypophysaires et placentaires | | | | | ✓ | |
| GÉNOTYPES ET PRÉVISIONS EN GÉNÉTIQUE HUMAINE | | | | | | |
| Déficit en G6PD | | | | ✓ | ✓ | |
| Albinisme | | | | ✓ | ✓ | |
| Rétinites pigmentaires | | ✓ | | ✓ | ✓ | ✓ |

EXPRESSION DE L'INFORMATION GÉNÉTIQUE

Vue d'ensemble

Les programmes

Les documents d'accompagnement apportent quelques précisions sur l'application des programmes :

- En première S, il s'agit de mettre en place les deux grandes étapes de la synthèse : transcription et traduction. Le brin transcrit sert de matrice à la construction de l'ARN messager (ARNm), construction catalysée par l'ARN polymérase. La synthèse protéique est localisée dans le cytoplasme et décrite de manière très simple : les ribosomes sont présentés comme des outils de cette synthèse ; le codon initiateur oriente la lecture du message contenu dans l'ARNm et la traduction s'arrête au codon STOP. Les élèves apprennent à utiliser le code génétique. Ce code génétique est universel, non ambigu et dégénéré (le doublet initial peut être seul déterminant) ; la nature dégénérée du code a pour conséquence sa redondance.
- En première L et ES, il s'agit uniquement de montrer que la séquence de nucléotides dans l'ADN gouverne la séquence des acides aminés dans la protéine selon un système de correspondance, le code génétique.
- Les notions de gène morcelé, d'exon et d'intron ne sont pas au programme de première S ni de TS.

Les données

Dans cette perspective des programmes, on peut envisager l'étude de l'expression de l'information génétique en utilisant comme support les séquences des gènes codant pour les différentes chaînes de globine (alpha, bêta, delta et gamma) et les séquences d'acides aminés de ces chaînes. À vrai dire, ce sont les séquences des allèles les plus répandus dans les populations humaines qui sont ici retenues.

Pour chaque gène sont fournies :

- les séquences des deux brins de l'ADN (appelées brin1 et brin2) correspondant à un ARN messager commençant au codon d'initiation AUG et se terminant par un codon stop ; cet ARN également fourni est dénommé ARNm_{cod} ;
- la séquence du gène (brin non transcrit, appelé ADN_c) correspondant à l'ARN messager réellement présent dans le hyaloplasme (sans la coiffe et le poly A), appelé ici ARNm ; cet ARNm comprend une région non codante en amont de AUG et une autre en aval du codon stop ;
- la protéine codée par ce gène.

D'autres séquences pourront être introduites par les utilisateurs grâce aux fonctions d'édition du logiciel.

Les prérequis

Les élèves doivent savoir au préalable que :

- l'ADN, localisé au niveau chromosomique, est le support de l'information génétique ;
- une molécule d'ADN renferme des milliers de gènes, la notion élémentaire du gène étant à ce stade un « bout » d'ADN ;
- la structure de l'ADN, son caractère informatif et son mode de réplication ;
- l'ADN régit le phénotype à travers la synthèse des polypeptides ;
- la fonction d'une protéine dépend de sa structure spatiale laquelle est conditionnée par la séquence des acides aminés.



Convertir les séquences

- comparer le polypeptide résultant de cette traduction avec le polypeptide de la banque et constater ainsi leur identité ;

- inverser la séquence ; traduire la séquence inversée et constater que le polypeptide obtenu est très différent du polypeptide réel. À l'issue de ces opérations, l'unidirectionnalité de la traduction est corroborée ;
- traduire l'ARNm codant à partir des positions 4, 7, 10, etc. (en dupliquant la séquence et en supprimant le premier triplet, le second, etc.) et constater qu'on obtient toujours la même séquence polypeptidique diminuée à chaque fois d'un acide aminé. A contrario, la traduction de l'ARNm codant à partir du 2^e, 3^e, 5^e, 6^e... nucléotide, conduit à un polypeptide différent du polypeptide réel. Ainsi est confirmée l'idée d'une traduction de l'ARNm triplet par triplet, sans chevauchement.

Le code génétique

Une première approche

On peut en construire le tableau de façon théorique et voir émerger des questions relatives au nombre des triplets possibles par rapport au nombre d'acides aminés : des triplets différents peuvent-ils coder pour le même acide aminé ? Existe-t-il beaucoup de triplets ne codant pour aucun acide aminé ? Quelle est leur signification biologique ?

L'analyse de la séquence nucléique et du polypeptide mis en vis-à-vis (résultat de la conversion), permet de trouver des éléments de réponse à la première question : pour cela, il suffit de relever les triplets de l'ARNm codant pour chaque acide aminé.

À l'issue de ce relevé (qui peut être enrichi par le travail des élèves sur les ARNm codant de globines différentes) le caractère dégénéré du code apparaît, mais il reste des triplets (non utilisés dans les séquences étudiées) pour lesquels on ne dispose pas d'informations. Pour en savoir plus à leur sujet, on peut créer une séquence d'ARN (option **Créer** dans le menu **Fichier**) comportant la suite de ces triplets et en demander la conversion en polypeptide.



La création d'une séquence est facilitée par les icônes des bases : il suffit de cliquer dessus.

Les résultats seront variables en fonction de la séquence de triplets constituée, mais dans tous les cas, les élèves seront confrontés à la signification des triplets qui ne codent pour rien. Ils devraient être capables d'apporter la preuve qu'ils ont valeur de codons d'arrêt de la traduction. Ce travail est important pour la compréhension des conséquences d'une mutation entraînant l'apparition d'un codon stop anticipé (mutation non sens).

Une deuxième approche

On peut également partir de la visualisation du code génétique proposé par *Anagène*. Le caractère univoque mais redondant du code génétique ressort avec évidence ; chaque acide aminé peut être désigné de manière active par la succession des trois bases qui le codent, et chaque triplet introduit montre l'acide aminé codé ce qui pourrait faciliter la compréhension et du code génétique.

AUG Code génétique

| Tableau du code génétique | | | | | |
|---------------------------|-----------|-----|-----|-----|-----------|
| 1ère base | 2ème base | | | | 3ème base |
| | U | C | A | G | |
| U | Phe | Ser | Tyr | Cys | U |
| | Phe | Ser | Tyr | Cys | C |
| | Leu | Ser | ~~~ | ~~~ | A |
| | Leu | Ser | ~~~ | Trp | G |
| C | Leu | Pro | His | Arg | U |
| | Leu | Pro | His | Arg | C |
| | Leu | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | U |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | U |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val | Ala | Glu | Gly | G |

Fermer

CAC

His

Histidine

On est là aussi conduit à s'interroger sur les triplets qui ne codent pour aucun acide aminé. L'introduction de séquences d'ARN (option **Créer**, dans le menu **Fichier**) contenant l'un ou l'autre de ces triplets, puis la conversion de ces séquences en polypeptides permettent de tester les hypothèses formulées.

Au terme de ces activités, les notions suivantes sont établies :

- un ARN messager est synthétisé à partir d'un des brins du gène à la suite d'un processus basé sur la complémentarité des nucléotides ;
- la séquence de nucléotides de l'ARNm est la même que celle du brin non transcrit, mais les nucléotides à uracile remplacent ceux à thymine ;
- la traduction de l'ARNm en polypeptide est unidirectionnelle et séquentielle (triplet par triplet) ;
- un triplet de nucléotides de l'ARNm code pour un seul acide aminé ;
- plusieurs triplets codent pour le même acide aminé (ou tous les acides aminés sauf deux, tryptophane et méthionine, sont codés par plus d'un codon) ;
- trois triplets UAA, UAG, UGA ne codent pour aucun acide aminé, mais leur insertion dans une séquence d'ARNm entraîne un arrêt de la traduction : ce sont des codons Stop ;
- une délétion ou une addition d'un nombre de nucléotides qui n'est pas un multiple de 3 dans une séquence d'ADN ou d'ARN, entraîne un décalage dans la lecture et un changement presque complet du polypeptide.

Si on considère que les connaissances sur la machinerie impliquée dans la traduction (ARNt, enzymes aminoacyl-ARNt synthétase) ne sont pas au programme de première S, on voit que le logiciel permet d'atteindre des points significatifs de l'expression des gènes.

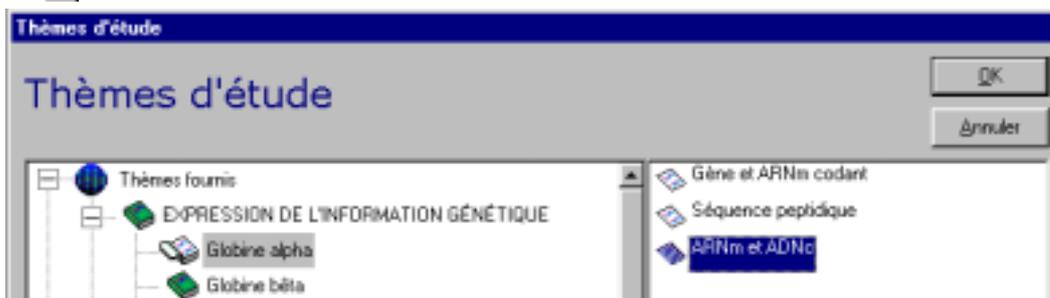
Fermer toutes les fenêtres



Une vision un peu plus complexe du gène et de l'ARNm

L'ARN messager réellement traduit dans la cellule est plus complexe que l'ARNm codant considéré jusqu'ici car il comprend en amont et en aval de la région strictement codante des parties non traduites. Il est possible de fournir cette vision plus complète en alignant (avec discontinuités) l'ARNm complet de alpha avec l'ARNm codant correspondant strictement au polypeptide traduit.

Attention : la comparaison doit se faire avec la séquence la plus longue (Alpha ARNm) en premier, sinon l'algorithme de comparaison place le premier A des deux AUG à l'extrémité gauche de l'alignement. Pour remonter une séquence, il faut la désigner d'abord en pointant la ligne puis utiliser le bouton ▲.



Avec Alpha ARNm (donc l'ARNm complet) le résultat est le suivant :



Convertir les séquences

La conversion de Alpha ARNm dans les trois cadres de lecture possibles (en commençant par la première base, la seconde et la troisième), fait apparaître le premier AUG dans le troisième cadre de lecture. On peut vérifier que cette séquence correspond bien au polypeptide réel fourni dans la banque de données.

La vision du gène et celle de son expression qui découlent de cette étude peuvent être schématisées de la manière suivante :

| | Début | Codon d'initiation | Codon de terminaison | Fin |
|-------------|--------|--------------------|----------------------|-------|
| ADN | [GENE | ATG | TAA (ou TGA) | |
| | | TAC | ATT (ou ACT) | |
| | ARNm | AUG | UAA (ou UGA) | |
| POLYPEPTIDE | | MET | | |

Il est utile de préciser que la méthionine mise en place en premier est ensuite éliminée, ce qui fait que les polypeptides matures ne commencent généralement pas par une méthionine.

Fermer toutes les fenêtres



LES RELATIONS GÉNOTYPE-PHÉNOTYPE

Les documents d'accompagnement des programmes précisent (pour la première S) : un même gène peut exister sous la forme de nombreux allèles. Les notions de dominance et de récessivité (envisagées en troisième) peuvent être discutées ; seules les notions d'homozygotie, d'hétérozygotie, d'allèle conduisant à une protéine fonctionnelle ou non sont nécessaires.

Avec un nombre limité d'exemples, on montre que le phénotype macroscopique dépend non seulement des gènes mais aussi de facteurs environnementaux ; il s'agit seulement de dresser un constat, les phénomènes explicatifs ne sont pas évoqués mais l'élève comprend qu'un même génotype peut conduire à des phénotypes macroscopiques différents. On remet ainsi en question le caractère univoque du lien direct qui pourrait être encore fait entre la possession de deux allèles d'un gène et le phénotype macroscopique.

Les données

Les phénotypes relatifs aux hémoglobinopathies

Il s'agit du phénotype drépanocytaire et des phénotypes thalassémiques. Dans ces exemples le polypeptide résultant de l'expression du gène est le phénotype moléculaire dont dépendent le phénotype cellulaire et le phénotype clinique de l'organisme.

Les phénotypes du groupe sanguin A, B, O

Dans cet exemple, le polypeptide produit de l'expression du gène est une enzyme qui intervient dans la dernière étape de la synthèse des marqueurs membranaires de ce système de groupe sanguin. Le phénotype moléculaire, c'est à dire le marqueur membranaire, résulte de l'activité catalytique du polypeptide expression du gène.

Le phénotype phénylcétonurique

Là aussi, le produit de l'expression du gène est une enzyme qui catalyse la transformation d'un acide aminé, la phénylalanine, en tyrosine. Cet exemple présente surtout un double intérêt :

- sensibiliser à l'idée que le phénotype ne dépend pas que du génotype mais résulte d'une interaction entre génotype et milieu ;
- sensibiliser à l'idée de la diversité des phénotypes phénylcétonuriques en relation avec la diversité génétique (nombreux allèles à l'origine de la phénylcétonurie).

Le phénotype drépanocytaire

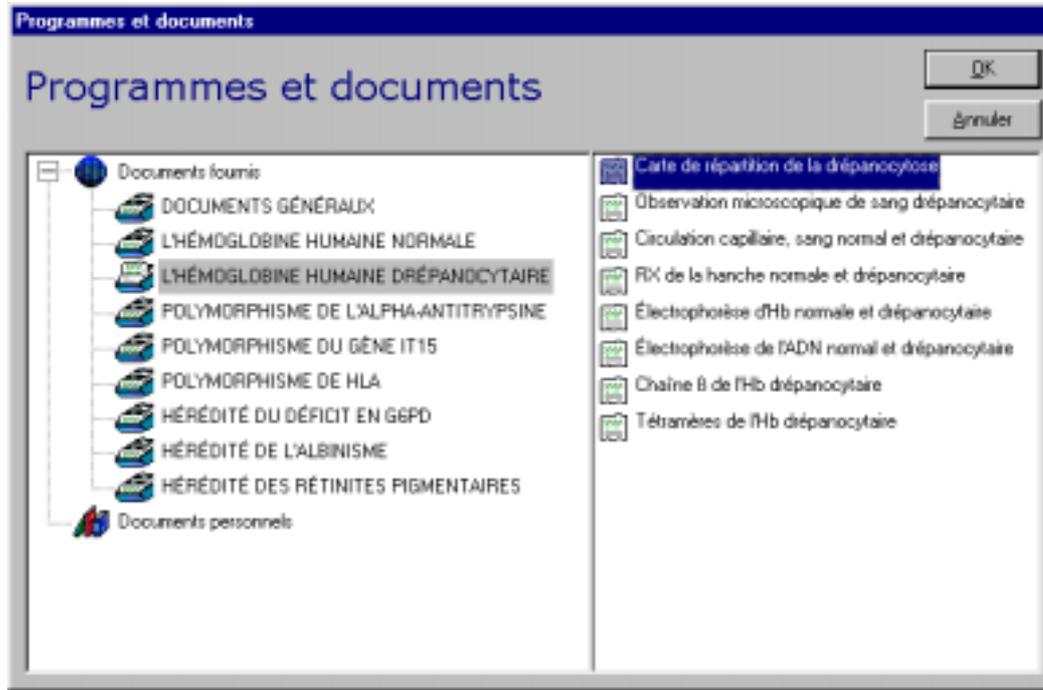
À vrai dire, on compare deux phénotypes : le phénotype drépanocytaire et un phénotype non drépanocytaire et on montre comment les différences entre ces deux phénotypes résultent des différences entre les génotypes. C'est un modèle incontournable car :

- c'est le meilleur exemple pour expliquer la relation entre le phénotype clinique macroscopique et le phénotype moléculaire. Pour l'élève, il s'agit de comprendre que **le génotype ne régit que le phénotype moléculaire** et qu'une intervention agissant sur les conséquences de ce phénotype moléculaire peut modifier plus ou moins fortement le phénotype clinique. Tout phénomène macroscopique résulte en réalité de l'interaction du phénotype moléculaire et du milieu ;
- c'est un exemple simple mais quasi unique où un phénotype macroscopique correspond à **un seul allèle**.

La cassette vidéo, éditée par le CNDP (*La drépanocytose, une maladie héréditaire*) introduit les notions de phénotype clinique, de phénotype cellulaire et de phénotype moléculaire. Ainsi les signes cliniques multiples de la maladie seront associés à un comportement anormal des molécules d'hémoglobine au sein des hématies drépanocytaires. Cette cassette permet aussi de réfléchir à la notion de maladie

héréditaire et d'établir dans l'esprit des élèves une liaison entre la synthèse d'une chaîne bêta anormale et la possession d'un allèle particulier du gène qui code pour cette chaîne (la notion d'homozygotie doit être précisée).

Plusieurs images extraites de cette cassette figurent dans la banque de documents pour illustrer ce thème. Ces documents sont à visionner dans l'ordre qui semble nécessaire. Il est possible d'y ajouter tout autre document personnel utile pour les élèves.

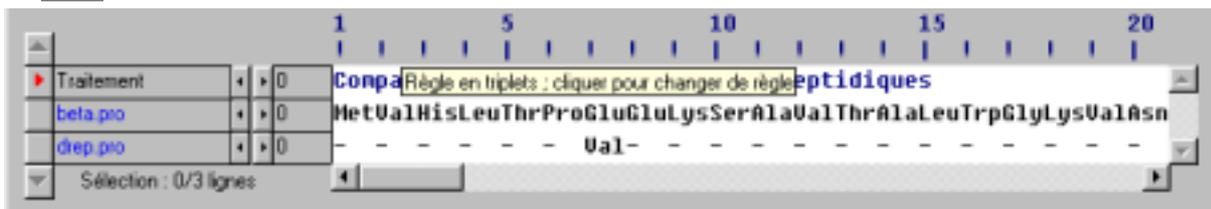


Une fois établie la liaison entre les aspects macroscopiques des phénotypes drépanocytaire et non drépanocytaire et la tendance ou non des molécules d'hémoglobine à former des polymères fibreux au sein des hématies, on peut demander aux élèves :

- de comparer les séquences des deux séquences protéiques drépanocytaire et non drépanocytaire ;

ATGC
-C--

Comparer les séquences



- de comparer les séquences des deux allèles correspondants ;

- de montrer en réinvestissant les notions vues sur la synthèse des protéines comment les différences entre les deux allèles entraîne des différences au niveau des séquences protéiques.

Fermer toutes les fenêtres 

Les phénotypes thalassémiques

Il existe en réalité de nombreux phénotypes thalassémiques. Ceux envisagés dans cette banque de données sont en relation avec **l'absence de synthèse d'une chaîne bêta fonctionnelle** de l'hémoglobine. Il s'agit de thalassémies bêta majeures dont les manifestations cliniques débute dès l'enfance. Les signes hématologiques sont marqués par une anémie hémolytique grave, une morphologie des hématies très irrégulière et par un retard staturo-pondéral et des modifications du squelette en rapport avec l'hémolyse chronique. L'évolution spontanée est constamment mortelle en quelques années. Les malades arrivent à survivre quelques années car l'absence d'hémoglobine A1 fonctionnelle (hémoglobine formée de deux chaînes alpha et de deux chaînes bêta) est partiellement compensée par une synthèse accrue d'hémoglobine A2 (deux chaînes alpha et deux chaînes delta) et la persistance de synthèse d'hémoglobine fœtale (deux chaînes alpha et deux chaînes gamma).

Il n'existe pas de traitement curatif des thalassémies. Les traitements proposés sont basés sur la transfusion mensuelle d'hématies qui, outre l'intérêt d'augmenter le taux d'hémoglobine, diminuent l'hyperplasie de la moelle osseuse compensatrice et par là préviennent un peu les anomalies du squelette. Ces transfusions fréquentes ont cependant pour complication une surcharge inéluctable en fer dommageable.

L'intérêt de cet exemple est que le même phénotype macroscopique dû à l'absence de chaîne bêta (phénotype moléculaire) peut être engendré par des génotypes différents. Il y a plusieurs allèles du gène bêta qui peuvent être à l'origine du phénotype thalassémique possible.

Là encore, une démarche pédagogique consiste à :

- comparer les phénotypes macroscopiques thalassémique et non thalassémique et à les relier au phénotype moléculaire : absence ou présence de chaîne bêta fonctionnelle ;
- comparer les chaînes bêta protéiques synthétisées par les diverses thalassémies à la chaîne bêta normale et à noter les différences dans un tableau faisant ressortir la **position** et la **nature** du changement constaté, ou bien sur un graphique.

Les séquences polypeptidiques fournies sont celles **initialement synthétisées** au niveau des ribosomes. Elles possèdent donc toutes, en premier, une méthionine. Malgré leur durée de vie parfois très brève, les polypeptides à l'origine des thalassémies envisagées figurent dans la banque de données.

La banque renferme les séquences de 8 allèles à l'origine de thalassémies. On peut répartir le travail entre les groupes : il est intéressant que chaque groupe étudie 3 allèles différents par la nature de la mutation à leur origine (substitution, addition, délétion).

Pour les polypeptides, la **comparaison simple** permet de visualiser immédiatement les différences. Il en est de même pour les séquences des allèles des thalassémies 1 à 3, dont la longueur est identique à celle de l'allèle normal.

ATGC Comparer les séquences

30 40 50 60 70

Traitement 0

Hb A nucléique 0

Tha1 nucléique 0

Sélection : 0/3 lignes

séquences d'ADN

AAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGT

-----T-----

10 15 20

Traitement 0

Hb A protéique 0

Tha1 protéique 0

Sélection : 0/3 lignes

séquences peptidiques

GluLysSerAlaValThrAlaLeuTrpGlyLysValAsnValAspGluVal

- - - - -

Les allèles à l'origine des thalassémies 4 à 8 ont un nombre de nucléotides différent de celui de l'allèle normal. La comparaison révèle des séquences différentes à partir d'une certaine position. L'analyse fine de cette région conduit à découvrir l'existence d'un décalage entre les deux séquences, dû à l'insertion ou la délétion de 1 ou plusieurs nucléotides.

30 40 50 60 70

Traitement 0

Hb A nucléique 0

Tha6 nucléique 0

Sélection : 0/3 lignes

nces d'ADN

TCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGT

-----A-G-TGA-CGTG-ATGA-GT-G-TG-TG-----

Comparaison simple de séquences montrant que le décalage est dû à la délétion dans la séquence Tha6 nucléique de la base C en position 51.

Il est néanmoins préférable de recourir à l'**alignement avec discontinuités** qui montre mieux la mutation à l'origine du décalage.

30 40 50 60 70

Traitement 0

Identités 0

Hb A nucléique 0

Tha6 nucléique 0

Sélection : 0/4 lignes

équences d'ADN

AAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGT

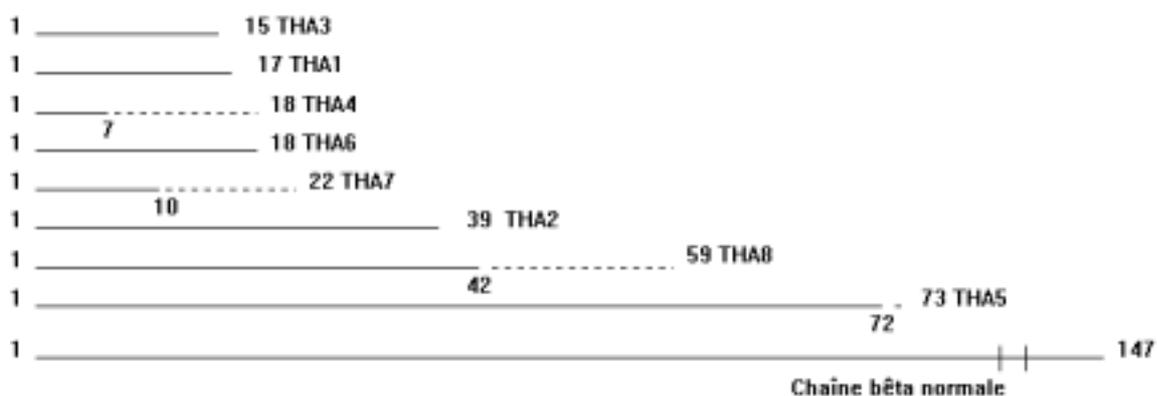
Alignement avec discontinuités des deux séquences. Le tiret bas (_) montre bien où se situe la délétion dans la séquence Tha6 nucléique (C en moins en position 51).

Représentation des résultats

Les élèves doivent construire un tableau récapitulatif des informations extraites à partir des traitements et schématiser les diverses chaînes thalassémiques par rapport à la chaîne normale.

| | ADN | | Polypeptide | |
|----------------|-------------------------------------|----------------------|-------------------------------------|---|
| | Position dans la séquence nucléique | Nature du Changement | Position dans la séquence protéique | Changement global |
| Tha1 nucléique | 52 | A → T | 18 | chaîne de 17 a.a. |
| Tha2 nucléique | 118 | C → T | 40 | chaîne de 39 a.a. |
| Tha3 nucléique | 47 | G → A | 16 | chaîne de 15 a.a. |
| Tha4 nucléique | 20 | manque A | 7 | chaîne de 18 a.a. différences à partir du 7 |
| Tha5 nucléique | 218 | A en plus | 73 | chaîne de 73 a.a. le dernier différent |
| Tha6 nucléique | 51 | manque C | 18 | chaîne de 18 a.a. le dernier différent |
| Tha7 nucléique | 28 | C en plus | 10 | chaîne de 22 a.a. différences à partir du 10 |
| Tha8 nucléique | 124 | manque TCTT* | 42 | chaîne de 59 a.a. différences à partir du 42 |

*Le logiciel affiche une discontinuité qui correspond à l'absence de **CTTT** à cause de la présence d'un T de part et d'autre du groupement TCTT. Cette différence est due à l'algorithme d'alignement.



Les polypeptides codés par les divers allèles du gène bêta sont représentés à l'aide de segments de droites parallèles de longueurs proportionnelles à leurs nombres d'acides aminés. Les parties identiques sont indiquées avec un trait continu et les parties différentes avec un trait en pointillé.

L'arrêt anticipé de la traduction de l'ARNm, par suite de l'apparition d'un codon Stop plus ou moins précoce dans la séquence, est un dénominateur commun à l'origine du phénotype thalassémique. Mais ce codon Stop est dû à des mécanismes différents selon les allèles :

- substitution transformant un triplet codant pour un acide aminé en un codon Stop (thalassémies 1, 2, 3) ;

- délétion entraînant un décalage dans la lecture et provoquant l'apparition d'un codon Stop plus ou moins distant de la position de la délétion (thalassémies 4, 6, 8) ;
- addition d'un nucléotide entraînant un décalage de la lecture et provoquant l'apparition d'un codon Stop plus ou moins distant dans la séquence (thalassémies 5,7).

Fermer toutes les fenêtres



Le phénotype « groupe sanguin ABO »

Les phénotypes sanguins A, B et O sont définis respectivement par la possession des antigènes membranaires A, B et H. Cependant il est possible de retrouver les divers niveaux d'expression du phénotype vus avec les hémoglobinopathies :

- phénotype macroscopique : il ne se révèle pas spontanément mais il se manifeste si on met les globules rouges d'un individu en contact avec des sérums tests anti A et anti B et aussi en cas de transfusions incompatible ;
- phénotype cellulaire : c'est la panoplie des marqueurs relatifs à ce système de groupe sanguin présents dans la membrane des hématies mais aussi à la surface des autres cellules de l'organisme ;
- phénotype moléculaire : les marqueurs membranaires résultent de l'action de l'enzyme produit de l'expression du gène ; cette enzyme est en somme le phénotype moléculaire primaire.

Une démarche pédagogique possible consiste à :

- présenter ce groupe sanguin en envisageant les divers niveaux du phénotype et en indiquant où intervient l'enzyme dans la chaîne de biosynthèse ;
- comparer les séquences des différents allèles (prendre A comme référence) ; identifier la nature et la position des mutations ;
- rechercher les différences au niveau des polypeptides enzymatiques ;
- demander de réinvestir les connaissances sur les mécanismes d'action des enzymes et notamment sur le site actif pour expliquer comment les différences au niveau de l'enzyme peuvent entraîner des différences au niveau des marqueurs.

Commentaire

La comparaison des allèles A et B montre qu'ils diffèrent par 4 nucléotides situés aux positions 523, 700, 793 et 800. Il en résulte aussi quatre différences au niveau de la séquence d'acides aminés du polypeptide enzymatique. On peut penser qu'elles sont à l'origine des différences de spécificité vis-à-vis des substrats de l'enzyme et que les quatre acides aminés en jeu participent directement ou indirectement à la constitution du site actif de l'enzyme. Dans un article de La Recherche de mai 1993, Hakomori précise que deux des acides aminés (leucine 265 et glycine 267) rendent l'enzyme A capable de reconnaître la N-acétylgalactosamine, les deux autres acides aminés (méthionine 265 et alanine 267) rendent l'enzyme B capable de reconnaître le galactose. (Sans qu'on sache pourquoi, l'article de La Recherche attribue à ces acides aminés les positions 266 et 268). L'allèle O diffère fondamentalement des allèles A et B par une délétion en position 258 (G en moins) qui entraîne un décalage du cadre de lecture et l'apparition d'un Stop au codon 117. Il en résulte que toute la partie du polypeptide contenant les acides aminés précédents n'est pas synthétisée chez les individus du groupe O, ce qui peut être relié à l'inactivité de l'enzyme.

Comparer les séquences

The screenshot shows a sequence comparison interface. At the top, a scale indicates positions 75, 80, 85, and 90. Below the scale, there are four rows: 'Traitement', 'Identités', 'ocod.adn', and 'ocod.adn'. The 'ocod.adn' row contains the DNA sequence: AAGGTCCTGACACCGTGGAAAGGATGTCCTCGTGGTGACCCCTTGGCTGGCTCCATT. A red box highlights a deletion of a 'G' base at position 86, where the sequence is 'GTCCTCGTGGT' instead of 'GTCCTCGTGGTG'.

L'allèle O est identique à l'allèle A sauf au codon 86 où on observe la délétion d'une base (-G).

The screenshot shows a sequence comparison interface. At the top, a scale indicates positions 110, 115, 120, and 125. Below the scale, there are two rows: 'ocod.adn' and 'Pro-ocod.adn'. The 'ocod.adn' row contains the DNA sequence: GGCTCCAGAACACCACCATTGGGTAACTGTGTTGCCATCAAGAAATACGTGG. The 'Pro-ocod.adn' row contains the protein sequence: GlySerArgThrProProLeuGly. A red box highlights the stop codon 'TAA' at position 117, which corresponds to the 'Gly' in the protein sequence.

La mutation au codon 86 entraîne l'apparition d'un codon stop (TAA) après le 117^e codon.

Fermer toutes les fenêtres



Le phénotype clinique phénylcétonurique

Informations scientifiques

La phénylalanine hydroxylase (PAH) est une enzyme hépatique qui intervient dans la transformation de la phénylalanine en tyrosine au niveau du foie. En cas d'inactivité enzymatique, la phénylalanine s'accumule dans le sang (hyperphénylalaninémie) et se transforme en acide cétonique (acide phénylpyruvique et ses dérivés) ce qui provoque une intoxication cellulaire, notamment des cellules nerveuses. Les symptômes cliniques de la maladie sont une grave arriération psychique et des troubles caractériels. L'acide phénylpyruvique et la phénylalanine en excès sont éliminés par les reins dans les urines : l'enfant est atteint de phénylcétonurie. Le test de Guthrie permet de dépister une hyperphénylalaninémie à la naissance. Ce test consiste à déposer un papier filtre imbibé d'une goutte de sang du nourrisson sur une culture de *Bacillus subtilis* dont la croissance est inhibée par manque de phénylalanine. Selon la croissance ou non de la culture bactérienne, on estimera s'il y a hyperphénylalaninémie ou pas. En cas de résultat positif, un dosage sanguin précis de la phénylalanine sera effectué. L'hyperphénylalaninémie dépend de l'intensité du déficit enzymatique.

La banque renferme la séquence du gène normal de la phénylalanine hydroxylase et celles de plusieurs allèles mutés. Le tableau joint donne les caractéristiques de ces allèles et des polypeptides qui leur correspondent, ainsi que les phénotypes cliniques des homozygotes.

| Séquence nucléique | | | Polypeptide | | |
|--------------------|--|-----------------------------------|--------------------------------|------------------------|------------------------------------|
| Noms des allèles | Nucléotides changés Nature-Position | Codons changés Nature-Position | A.A changés Nature-Position | Type de mutation | Phénotype clinique de l'homozygote |
| Phenorm | (référence) | (référence) | (référence) | | Normal |
| Phe1 | -T165 | TTTGAGGAG → TTGA.....TAA60 | Phe55Leu....60X | Délétion décalante | Sévère |
| Phe2 | - ATC283-284-285 | -ATC 95 | - Ile95 | Délétion non décalante | Peu sévère |
| Phe3 | C331→T | CGA111→TGA | Arg111X | Substitution non sens | Sévère |
| Phe4 | G473→A | CGG158→CAG | Arg158Gln | Substitution faux sens | Assez sévère |
| Phe5 | - AG663-664 | GAT222→TAA | Asp222X | Délétion décalante | Sévère |
| Phe6 | A696→G | CAA232→CAG | Gln232Gln | Substitution neutre | Normal |
| Phe7 | C727→T | CGA243→TGA | Arg243X | Substitution non sens | Sévère |
| Phe8 | G735 →A | GTG245→GTA | Val245Val | Substitution neutre | Normal |
| Phe9 | G782→A | CGA261→CAA | Arg261Gln | Substitution faux sens | Peu sévère |
| Phe10 | G814→T | GGA272→TGA | Gly272X | Substitution non sens | Sévère |
| Phe11 | G838→A | GAA280→AAA | Glu280Lys | Substitution faux sens | Sévère |
| Phe12 | T896→G | TTT299→TGT | Phe299Cys | Substitution faux sens | Sévère |
| Phe13 | G1038 | GGG346→GGC...TAA398 | Leu347Ser....398X | Délétion décalante | Sévère |
| Phe14 | CTC1093-1095 | CTC365 | Leu365 | Délétion non décalante | Sévère |
| Phe15 | G1223→A | CGG408→CAG | Arg408Gln | Substitution faux sens | Peu sévère |
| Phe16 | C1222→T | CGG408→TGG | Arg408Trp | Substitution faux sens | Sévère |
| Phe17 | A1241→G | TAC414→TGC | Tyr414Cys | Substitution faux sens | Peu sévère |

Quelques allèles du gène de la phénylalanine hydroxylase.

Arg408Gln : au 408^e acide aminé, l'arginine est remplacée par la glutamine

X : absence d'acide aminé suite à la rencontre d'un codon STOP lors de la traduction

Arg111X : l'arginine 111 n'existe pas dans le polypeptide codé par l'allèle (chaîne de 110 a.a.)

Phe55Leu....60X : à partir du 55^eme a.a. le polypeptide codé par l'allèle change et sa chaîne n'a que 59 a.a.

On a estimé par ailleurs l'activité de la PAH en % par rapport à l'activité normale selon les allèles présents chez les individus homozygotes ou hétérozygotes.

| Allèle2 Allèle1 | Phe7 | Phe11 | Phe4 | Phe9 | Phe17 |
|--------------------|----------|------------|------------|-------------|-------------|
| Phe7 | 0 | 1,5 | 5 | 15 | 25 |
| Phe11 | | 3 | 6,5 | 16,5 | 26,5 |
| Phe4 | | | 10 | 20 | 30 |
| Phe9 | | | | 30 | 40 |
| Phe17 | | | | | 50 |

Activité de la phénylalanine hydroxylase en % par rapport à l'activité normale.

Du point de vue phénotype clinique, les individus ayant une activité PAH égale ou inférieure à 10% présentent une phénylcétonurie sévère. Au dessus de 15%, la maladie est atténuée. Le dépistage précoce de l'hyperphénylalaninémie (supérieure à 20 mg.dl⁻¹) permet la mise en place d'un régime alimentaire pauvre en phénylalanine ; cela constitue le traitement diététique de la phénylcétonurie, **exemple d'une influence du milieu** sur l'apparition ou non des symptômes de la maladie ! Les régimes sont établis à partir de tables de composition en phénylalanine des divers aliments.

La démarche pédagogique à adopter doit veiller à bien faire réfléchir sur :

- la multiplicité des allèles à l'origine des phénotypes phénylcétonuriques ;
- la variété des polypeptides produits de l'expression de ces allèles ; bien faire comprendre comment des substitutions différentes d'acides aminés peuvent affecter l'activité enzymatique et conduire à des phénotypes phénylcétonuriques plus ou moins accusés ;
- le fait que le phénotype macroscopique clinique ne dépend presque pas du génotype à partir du moment où les conséquences du phénotype moléculaire sont quasi abolies par l'action sur le milieu (régime alimentaire).

POLYMORPHISME DES GÈNES

Le document d'accompagnement du programme de terminale S fixe les orientations à privilégier.

- Le polymorphisme des gènes et les familles de gènes au sein d'une espèce doivent être compris comme le résultat d'une accumulation des innovations génétiques au cours des générations successives. Des liens entre génétique et mécanismes de l'évolution deviennent compréhensibles : trois exemples sont abordés.
- L'étude n'est plus centrée sur l'individu mais sur les populations au sein des espèces. La notion de polymorphisme génique a un caractère quantitatif et implique la prise de conscience que, pour beaucoup de gènes de l'espèce, il existe plusieurs allèles répandus dans la population. La notion de fréquence génique est associée à celle de polymorphisme.
- L'origine du polymorphisme, à savoir les mutations, est envisagée dans la perspective d'une étude au niveau des populations.
- Seuls les *principes* des innovations génétiques sont à connaître : duplication du gène ancestral, divergence plus ou moins grande de copies.



Les séquences du gène qui code pour l'alpha-antitrypsine

Informations scientifiques

La protéine

L'alpha-antitrypsine est une **protéine plasmatique** constituée par une chaîne de 394 acides aminés et de trois chaînes latérales glucidiques. Ces chaînes latérales se branchent sur la chaîne peptidique au niveau de résidus asparagine en positions 46, 83 et 247. C'est une glycoprotéine globulaire dont la masse molaire est de 52 kilodaltons ; au cours d'une électrophorèse du sérum, elle fait partie du pic des alpha1 globulines d'où le début de son nom. La concentration plasmatique d'alpha-antitrypsine est généralement comprise entre 150 et 350 mg.dl⁻¹. Cette protéine diffuse dans le liquide interstitiel où sa concentration est toutefois plus faible, environ le dixième de ce qu'elle est dans le plasma. Les variations de sa concentration dans le liquide interstitiel suivent celles du plasma.

La demi-vie de l'alpha-antitrypsine est de 4 à 5 jours. Cette protéine est produite et sécrétée par les **cellules hépatiques**. Elle est synthétisée au niveau du réticulum endoplasmique granuleux des cellules hépatiques sous forme d'un précurseur de 418 acides aminés. Le peptide signal formé par les 24 premiers résidus de ce précurseur est éliminé durant le passage dans les cavités du réticulum. Là, en même temps que la protéine prend sa configuration tridimensionnelle, les trois chaînes glucidiques y sont ajoutées. La protéine ainsi glycosylée est transférée aux saccules golgiens, complétée dans sa partie glucidique, puis ensuite sécrétée (exocytose des vésicules golgiennes). On estime à 34 mg par kilo de masse corporelle la production journalière d'alpha-antitrypsine.

L'alpha-antitrypsine est un **inhibiteur des protéases** à sérine, protéases dont le site actif comprend la triade d'acides aminés catalytiques : acide aspartique - histidine - sérine. La trypsine, la chymotrypsine et l'élastase des granulocytes sont des protéases de ce type. L'action inhibitrice de l'alpha-antitrypsine sur ces protéases a été mise en évidence pour la première fois sur la trypsine, d'où son nom. In vivo, cependant, le seul substrat réel pour l'alpha-antitrypsine est l'élastase, une endopeptidase extrêmement puissante capable de cliver la plupart des protéines de la matrice extracellulaire, l'élastine et les divers collagènes, notamment. Cette élastase est libérée par les granulocytes à leur mort. Au niveau du conjonctif des alvéoles pulmonaires, cette élastase est libérée à des bas taux en permanence. **L'alpha-antitrypsine protège ainsi la matrice extracellulaire** des divers organes, en particulier celle des alvéoles pulmonaires.

L'alpha-antitrypsine exerce son action inhibitrice en se liant fortement et de façon quasi irréversible au site actif de l'élastase. Le site de liaison de l'alpha-antitrypsine est localisé au niveau des résidus méthionine 382 - sérine 383. La méthionine 382 peut être facilement oxydée, ce qui réduit fortement l'affinité de l'alpha-antitrypsine pour l'élastase. Il semble que la fumée de cigarette entraîne l'oxydation de la méthionine, ce qui expliquerait, au moins en partie, l'aggravation des symptômes chez le fumeur en cas de déficience en alpha-antitrypsine.

Le gène et ses allèles

Situé sur le chromosome 14, le gène comprend 12 200 paires de bases. Il est composé de 7 exons et 6 introns ; la région strictement codante pour la protéine se trouve au niveau des 4 derniers exons. On connaît 75 allèles dont plusieurs avec une fréquence supérieure à 1% ; ce gène est donc très polymorphique dans les populations humaines. La figure 1 renseigne sur la fréquence des allèles les plus répandus et sur les concentrations plasmatiques d'alpha-antitrypsine chez les personnes homozygotes pour chacun de ces allèles.

| Allèles | Fréquence * estimée (en %) | Taux d'alpha-antitrypsine dans le sang (en % par rapport au normal) | Quantité d'alpha- antitrypsine dans le sang (mg/dl ⁻¹) ** | Risques de maladie chez l'homozygote |
|-------------|----------------------------------|---|--|--|
| M'1(Ala237) | 20-23 | 100 | 150-350 | non |
| M1 | 44-49 | 100 | 150-350 | non |
| M2 | 14-19 | 100 | 150-350 | non |
| M3 | 10-11 | 100 | 150-350 | non |
| S | 2-4 | 40-70 | 100-200 | non |
| Z | 1-2 | 10-15 | 15-50 | oui |
| NULL1 | Rare | 0 | 0 | oui et précoces |
| NULL2 | Rare | 0 | 0 | oui et précoces |

Fréquence des allèles de l'alpha-antitrypsine et concentrations plasmatiques.

** Fréquence estimée pour la population blanche des États-Unis d'Amérique*

*** Quantité d'alpha-antitrypsine chez l'homozygote*

Le tableau ci-dessous indique les différences dans les séquences de ces allèles en prenant comme référence l'allèle **M'1** (Ala237). Ces séquences du brin non transcrit qui commencent par le codon ATG et se terminent par un codon STOP **incluent la région correspondant au peptide signal** (24 acides aminés). D'autres publications faisant référence à l'alpha antitrypsine ne prennent en compte que la séquence codante correspondant à la protéine mature qui n'inclut donc pas la partie codant pour le peptide signal.

| Séquence nucléique | | | Polypeptide | |
|-----------------------------|--|-------------------------------------|-----------------------------------|------------------|
| Noms des allèles | Nucléotides changés Nature - Position | Codons changés Nature - Position | A.A. changés Nature - Position | Type de mutation |
| M'1 (Ala237) (Référence) | C710 | GCG237 | Ala237 | |
| M1 | C710 → T | GCG237 → GTG | Ala237Val | Substitution |
| M2 | C710 → T | GCG237 → GTG | Ala237Val | Substitution |
| | A1200 → C | GAA400 → GAC | Glu400Asp | Substitution |
| | G374 → A | CGT125 → CAT | Arg125His | Substitution |
| M3 | C710 → T | GCG237 → GTG | Ala237Val | Substitution |
| | A1200 → C | GAA400 → GAC | Glu400Asp | Substitution |
| S | C710 → T | GCG237 → GTG | Ala237Val | Substitution |
| | A863 → T | GAA288 → GTA | Glu288Val | Substitution |
| Z | G1096 → A | GAG366 → AAG | Glu366Lys | Substitution |
| NULL1 | - C552 | TAC184 → TAG | Tyr184X | Délétion |
| NULL2 | C710 → T | GCG237 → GTG | Ala237Val | Substitution |
| | A721 → T | AAG241 → TAG | Lys241X | Substitution |

Caractéristiques des allèles de l'alpha-antitrypsine.

Ces allèles peuvent être regroupés en 4 ensembles : les variants normaux, les variants déficients, les variants « NULL » et les variants dysfonctionnels.

Les variants normaux (M'1, M1, M2, M3)

Ces allèles codent pour des molécules d'alpha-antitrypsine différentes mais également fonctionnelles et sécrétées de façon équivalente. Les concentrations plasmatiques de la protéine sont du même ordre de grandeur chez les homozygotes pour chacun de ces allèles, comprises entre 150 et 350 mg.dl⁻¹. L'allèle le plus fréquent est M1 (Val237) ce qui signifie que le 237^e triplet code pour la valine. Les allèles variants normaux diffèrent les uns des autres par 1, 2 ou 3 substitutions de nucléotides, entraînant 1, 2 ou 3 substitutions d'acides aminés dans la protéine sans modification de ses propriétés (mutations neutres).

Il existe beaucoup d'autres variants normaux que ceux présentés ici, généralement rares ou très rares ; M4, toutefois, a une fréquence comprise entre 1 et 5% dans les populations européennes.

Les variants déficients

Les deux seuls ayant une fréquence supérieure à 1% sont les allèles S et Z.

L'allèle S code pour une protéine inhibant correctement l'activité de l'élastase, donc fonctionnelle, mais sécrétée en plus faible quantité que les protéines codées par les variants normaux. En conséquence, les homozygotes pour cet allèle ont une concentration plasmatique de l'alpha-antitrypsine comprise entre 100 et 200 mg.dl⁻¹. L'allèle S diffère de l'allèle M1 (Val237) par **une seule substitution** de nucléotide entraînant la présence de valine en position 288 à la place d'acide glutamique. Il semble que cette mutation entraîne une relative instabilité de la protéine, cause d'une "destruction" précoce à l'intérieur même des cellules hépatiques. Cela est à l'origine de la déficience de sécrétion. Une fois sécrétée, l'alpha-antitrypsine a une durée de vie normale.

L'allèle Z code pour une protéine ayant une activité inhibitrice de l'élastase réduite ; de plus, elle est sécrétée en faible quantité de sorte que les concentrations plasmatiques chez les homozygotes Z sont comprises entre 15 et 50 mg.dl⁻¹. Cet allèle Z diffère de l'allèle normal M1 (Ala237) par **une seule substitution** de nucléotide entraînant la présence de lysine à la place d'acide glutamique en position 366. Chez les homozygotes Z, les niveaux d'ARNm sont normaux ainsi que la traduction. Cependant, après la mise en forme des chaînes glucidiques dans le réticulum, les molécules d'alpha-antitrypsine Z s'agrègent limitant ainsi le transfert vers l'appareil de Golgi et donc la sécrétion. La structure tridimensionnelle de la protéine est affectée par la substitution.

Les variants « NULL »

Un homozygote pour un allèle « NULL » a une concentration plasmatique nulle d'alpha-antitrypsine. Ces allèles sont toujours rares avec une fréquence inférieure à 0,1%. Ils diffèrent des variants normaux soit par une substitution non sens (allèle *Bellingham* appelé NULL2 dans la banque) soit par une délétion entraînant un décalage du cadre de lecture et l'apparition prématurée d'un triplet STOP (allèle *granite falls* appelé NULL1 dans la banque). Les protéines codées par ces allèles sont raccourcies, très instables et rapidement détruites.

Le variant dysfonctionnel

Un seul allèle de ce type est connu et il a été trouvé chez un seul individu... Il diffère des allèles normaux par une substitution qui entraîne le remplacement de la méthionine du site actif de l'alpha-antitrypsine par l'arginine. Il en résulte que ce type d'alpha-antitrypsine a une activité inhibitrice de l'élastase faible, mais les concentrations plasmatiques sont normales.

Les relations de dominance entre les allèles

Les deux allèles présents chez un individu s'expriment indépendamment l'un de l'autre. On trouve donc dans le sang, en cas d'hétérozygotie, les deux types d'alpha-antitrypsine. Les spécialistes de ce gène disent qu'il y a codominance. Cela est vrai au niveau biochimique mais ne l'est plus au niveau macroscopique comme le montrent les phénotypes cliniques associés aux divers génotypes.

Les phénotypes cliniques associés aux divers génotypes

Les études épidémiologiques montrent que les concentrations plasmatiques d'alpha-antitrypsine inférieures à 80 mg.dl⁻¹ sont associées à un risque d'emphysème pulmonaire. À ces concentrations, l'alpha-antitrypsine inhibe insuffisamment l'élastase et celle-ci détruit peu à peu le tissu conjonctif, en particulier au niveau des alvéoles pulmonaires, ce qui perturbe les échanges gazeux et entraîne l'emphysème.

Les personnes possédant deux allèles « NULL » présentent les symptômes d'emphysème avant l'âge de 30 ans et ne vivent que rarement au-delà de 40 ans. Les individus de génotype ZZ (concentration

plasmatique comprise entre 15 et 50 mg.dl⁻¹) souffrent aussi d'emphysème mais plus tardivement et la maladie est généralement fatale vers 60 ans. Par suite des modalités d'expression des allèles, la possession d'un seul allèle normal suffit pour avoir une concentration plasmatique d'alpha-antitrypsine protectrice vis-à-vis de l'élastase et cela même si le second allèle est un allèle « NULL ». Les individus S/Z ont une concentration plasmatique à la limite de risque d'emphysème. Ce risque peut être fortement augmenté par des facteurs d'environnement comme la fumée de tabac.

Évolution du gène qui code pour l'alpha-antitrypsine

Bien que l'allèle M1 (Val237) soit le plus répandu dans les populations, les spécialistes s'accordent pour dire que l'allèle le plus ancien et seul présent dans les premières populations humaines est M'1 (Ala237). Ils s'appuient pour cela sur la séquence codante du gène connue chez le Chimpanzé qui est identique à celle de M'1 sauf à un site (ce qui entraîne une différence entre les protéines en position 409 : méthionine chez l'Homme, valine chez le Chimpanzé). En particulier, le codon 237 est identique dans l'allèle M'1 et chez le Chimpanzé.

Si M'1 est bien l'allèle initial, tous les autres en dérivent par mutations. En appliquant le principe de parcimonie, on peut dire que deux allèles sont d'autant plus apparentés que le nombre de différences entre eux est réduit. La figure 6 montre les filiations entre allèles reconnues par les spécialistes.

Utilisations pédagogiques

Analyse du tableau de fréquence des allèles

Après une présentation du rôle joué par l'alpha-antitrypsine dans l'organisme ainsi que celle du phénotype morbide associé à sa déficience, on peut proposer l'analyse du tableau des fréquences des allèles du gène qui code pour l'alpha-antitrypsine. Les élèves doivent discuter :

- du polymorphisme du gène ;
- des rapports entre différences au niveau du génotype et différences au niveau du phénotype ; cette discussion peut déboucher sur une classification des allèles semblable à celle indiquée dans les informations scientifiques (variants normaux, variants déficients, variants « NULL »).

Recherche des différences entre les allèles

Les élèves doivent comparer les séquences et noter les différences constatées dans un tableau. Une fois rempli, ce tableau doit conduire à un premier bilan des différences entre les divers allèles et l'allèle de référence : substitution d'un ou plusieurs nucléotides, délétion d'un nucléotide (NULL1).

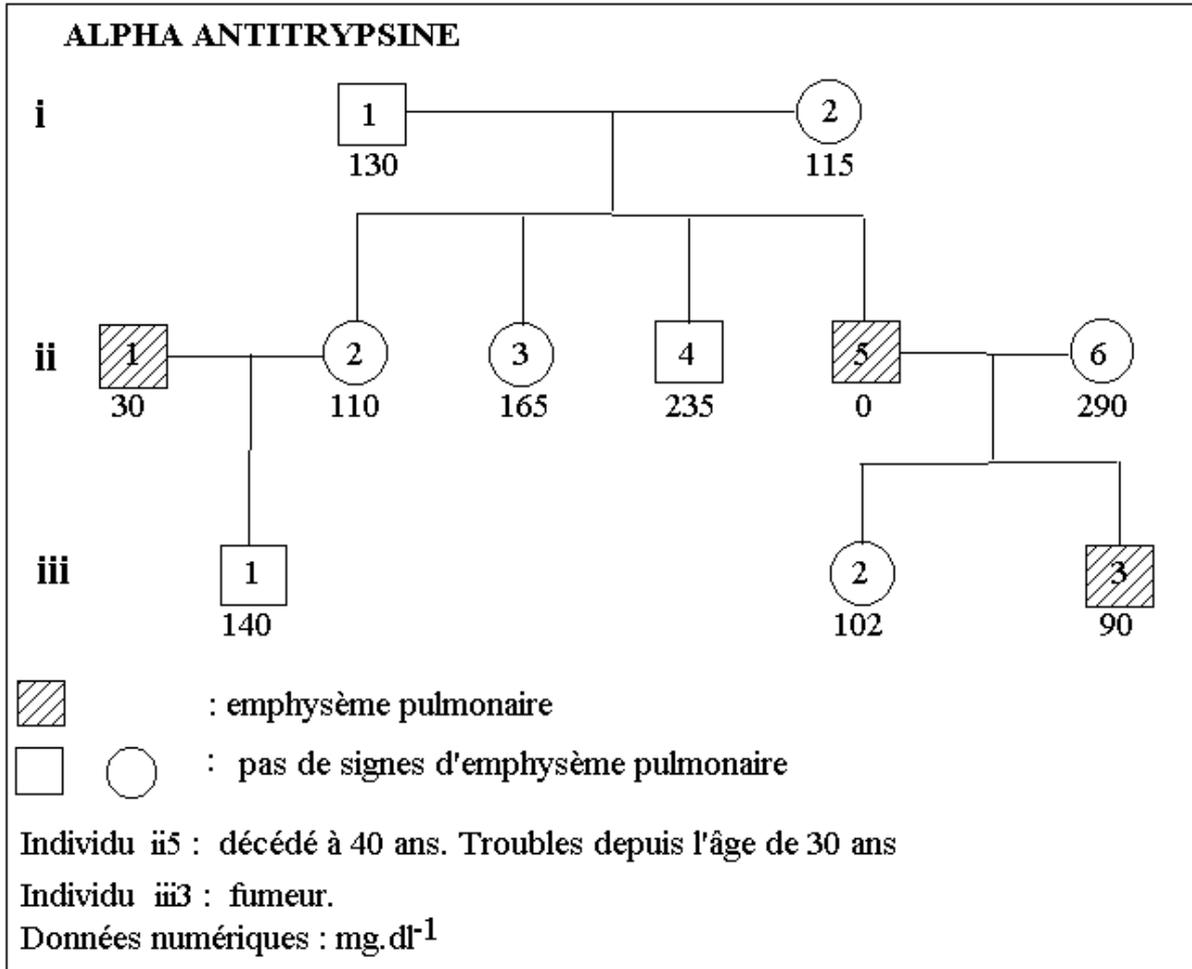
Remarque : il est indiqué de choisir comme allèle de référence non pas M1 qui est le plus fréquent, mais M'1 que les généticiens estiment l'allèle initialement présent dans les premières populations humaines.

Recherche des conséquences des différences entre les allèles sur le polypeptide

Les élèves peuvent soit se référer au code génétique, soit utiliser la fonction de conversion de séquences, et les comparer au polypeptide de référence. Les informations obtenues sont intégrées dans un tableau qui doit permettre de discuter des conséquences variables des substitutions de nucléotides sur le phénotype (distinction phénotype moléculaire, phénotype cellulaire, phénotype macroscopique ou clinique) ainsi que de dégager l'importance du décalage du cadre de lecture en cas de délétion.

Analyse des relations génotype-phénotype chez les hétérozygotes

Depuis la classe de seconde, les élèves connaissent la localisation chromosomique des gènes et savent que pour les gènes autosomaux, un individu possède deux allèles du même gène. Le problème de la relation entre le phénotype et le génotype chez un organisme hétérozygote peut donc être posé. Pour le résoudre, on peut utiliser la banque de séquences des allèles du gène de l'alpha-antitrypsine que possèdent les individus de l'arbre ci-après.



Arbre généalogique d'une famille avec troubles de l'alpha-antitrypsine.

Il ne s'agit pas de faire une analyse classique d'un arbre généalogique mais d'associer son exploitation à la banque de données pour **centrer l'étude sur les notions de dominance et de récessivité**. Pour cela, on peut demander aux élèves d'indiquer le génotype possible de l'individu ii5 (absence d'alpha-antitrypsine) puis de tester leurs idées grâce au logiciel. Ils trouvent ainsi que ii5 est homozygote NULL1/NULL1. À partir de là, les élèves sont invités à indiquer les génotypes possibles des parents i1 et i2 et à les tester avec le logiciel. La connaissance de ces génotypes, NULL1/M'1 (pour i1) et NULL1/M1 (pour i2) permet de travailler sur la notion de dominance en considérant le phénotype macroscopique, puis d'en montrer la relativité en prenant en compte le phénotype biochimique.

Cet arbre peut être aussi utilisé pour renforcer l'idée qu'un phénotype macroscopique peut être dû à des génotypes différents (individus ii1 et ii5) et que le phénotype dépend en réalité de l'interaction du génotype et de l'environnement (individus iii2 et iii3 de même génotype S/NULL1 mais de phénotype macroscopique différent en liaison avec le comportement de fumeur de iii3). Cela implique la recherche des génotypes des individus de cette famille. Le tableau ci-dessous présente les résultats de cette recherche.

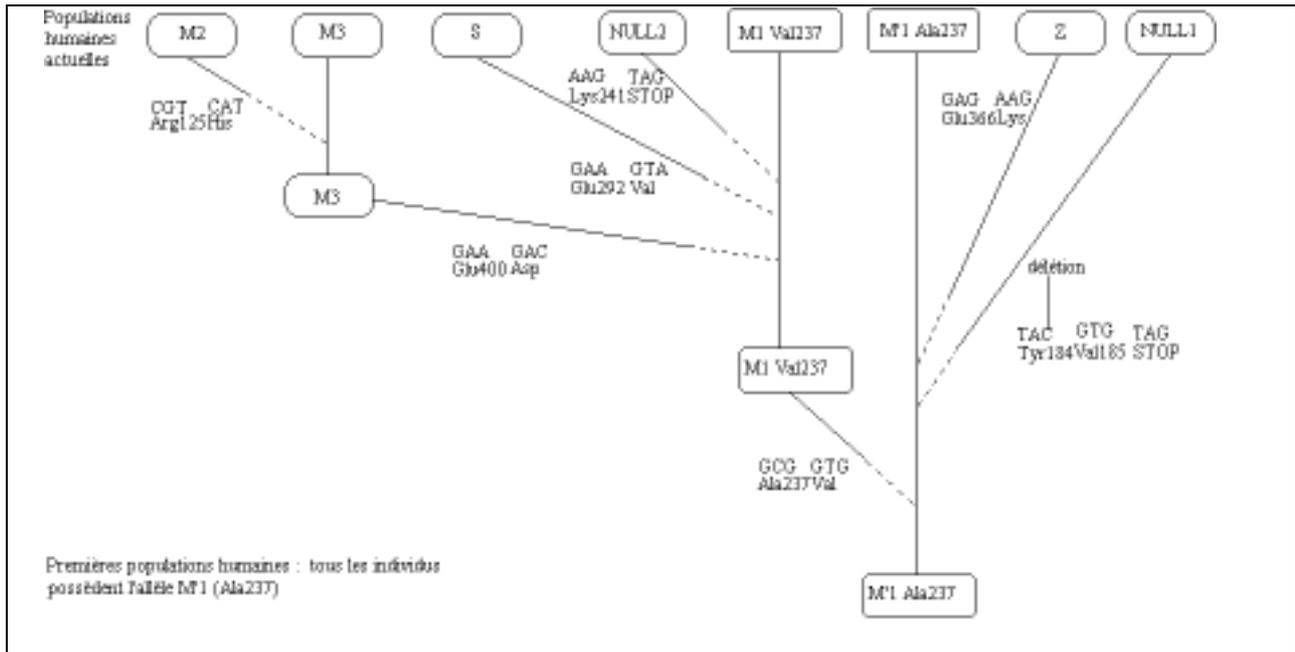
| Allèles des individus | Nucléotides changés Nature - Position | Allèles du gène |
|------------------------------|--|------------------------|
| F1-i1all1.at | Référence | M'1 |
| F1-i1all2.at | -C552 | NULL1 |
| F1-i2all1.at | C710 → T | M1 |
| F1-i2all2.at | -C552 | NULL1 |
| F1-ii1all1.at | G1096 → A | Z |
| F1-ii1all2.at | G1096 → A | Z |
| F1-ii2all1.at | Référence | M'1 |
| F1-ii2all2.at | -C552 | NULL1 |
| F1-ii3all1.at | -C552 | NULL1 |
| F1-ii3all2.at | Référence | M'1 |
| F1-ii4all1.at | Référence | M'1 |
| F1-ii4all2.at | C710 → T | M1 |
| F1-ii5all1.at | -C552 | NULL1 |
| F1-ii5all2.at | -C552 | NULL1 |
| F1-ii6all1.at | C710 → T | M1 |
| F1-ii6all2.at | C710 → T | M1 |
| F1-iii1all1.at | G1096 → A | Z |
| F1-iii1all2.at | Référence | M'1 |
| F1-iii2all1.at | C710 → T | M1 |
| F1-iii2all2.at | -C552 | NULL1 |
| F1-iii3all1.at | -C552 | NULL1 |
| F1-iii3all2.at | C710 → T | M1 |

Allèles portés par les individus de la famille avec troubles de l'alpha-antitrypsine.

Établissement des filiations entre les allèles du gène

La discussion sur l'origine des différents allèles du gène d'alpha-antitrypsine conduit à la notion de mutation. En admettant que l'allèle M'1 était seul présent dans les premières populations humaines, on peut demander aux élèves de travailler sur les informations réunies dans le tableau 3 pour imaginer l'histoire évolutive du gène qui code pour l'alpha-antitrypsine depuis les premiers hommes jusqu'à aujourd'hui.

Leur analyse doit les conduire à proposer un schéma de filiation entre les différents allèles. Plusieurs représentations sont possibles, celle de la figure 6 présente l'intérêt de situer cette filiation dans le temps. Bien sûr, une indétermination totale règne sur le moment d'apparition des différents allèles.



Filiations entre les allèles de l'alpha-antitrypsine. Les pointillés indiquent seulement que le moment et l'ordre d'apparition des allèles par mutation au cours de l'histoire de l'humanité sont inconnus.

À travers ce travail, les élèves sont amenés à réfléchir sur :

- le fait qu'un nouvel allèle apparaît par mutation d'un allèle préexistant chez un individu et se répand plus ou moins dans les populations ;
- le degré d'apparentement des divers allèles.



Les séquences du gène qui code pour la huntingtine

Informations scientifiques

Le gène et ses allèles

Ce gène, dont certains allèles sont à l'origine de la chorée de Huntington, a été isolé et séquencé en 1993. Il code pour une protéine, la huntingtine, dont le rôle est en 1995 encore inconnu. Appelé IT15, il est situé sur le bras court du chromosome 4 et comprend 67 exons. Il s'exprime dans de nombreux tissus, en particulier les neurones cérébraux et notamment ceux des corps striés. La région codante du gène est caractérisée par la présence d'un triplet CAG (brin non transcrit), près de l'extrémité 5' (qui correspond à l'extrémité NH₂ de la huntingtine) répété plusieurs fois à partir de la position 52. Dans toutes les populations humaines, il existe de nombreux allèles qui diffèrent par le nombre de répétitions de ce triplet CAG.

Les variants normaux

Ces variants ont un nombre de répétitions compris entre 9 et 39 ; les allèles les plus fréquents correspondent à un triplet répété entre 17 et 21 fois. La fréquence de la majorité de ces allèles est supérieure à 1% ce qui indique un gène très polymorphe.

| Nombre de répétitions du triplet CAG de l'allèle | Nombre de fois où l'allèle a été trouvé | Fréquence de l'allèle (en %) |
|--|---|------------------------------|
| 11 | 2 | 0,98 |
| 12 | 0 | 0 |
| 13 | 2 | 0,98 |
| 14 | 2 | 0,98 |
| 15 | 4 | 1,96 |
| 16 | 6 | 2,94 |
| 17 | 41 | 20,01 |
| 18 | 28 | 13,73 |
| 19 | 17 | 8,33 |
| 20 | 22 | 10,78 |
| 21 | 30 | 14,71 |
| 22 | 14 | 6,86 |
| 23 | 11 | 5,39 |
| 24 | 6 | 2,94 |
| 25 | 8 | 3,92 |
| 26 | 0 | 0 |
| 27 | 3 | 1,47 |
| 28 | 0 | 0 |
| 29 | 1 | 0,49 |
| 30 | 2 | 0,98 |
| 31 | 2 | 0,98 |
| 33 | 1 | 0,49 |

Fréquence des allèles normaux du gène IT15 dans une population européenne (allemande). 202 allèles analysés. D'après Hum. Mol. Gen., 1993, vol 2, n°12, p. 2063-2067.

Les variants morbides

Causes de la chorée de Huntington, ils possèdent un triplet CAG répété entre 36 et 121 fois, avec une fréquence maximale comprise entre 42 et 46 répétitions.

| Nombre de répétitions du triplet CAG de l'allèle | Nombre de fois où l'allèle a été trouvé | Fréquence de l'allèle (en %) |
|--|---|------------------------------|
| 40 | 21 | 6,69 |
| 41 | 35 | 11,15 |
| 42 | 36 | 11,46 |
| 43 | 34 | 10,83 |
| 44 | 39 | 12,42 |
| 45 | 38 | 12,10 |
| 46 | 28 | 8,91 |
| 47 | 22 | 7 |
| 48 | 16 | 5,10 |
| 49 | 7 | 2,23 |
| 50 | 6 | 1,91 |
| 51 | 5 | 1,59 |
| 52 | 9 | 2,87 |
| 53 | 3 | 0,96 |
| 54 | 2 | 0,64 |
| 55 | 4 | 1,27 |
| 56 | 0 | 0 |
| 57 | 2 | 0,64 |
| 58 | 3 | 0,96 |
| 61 | 1 | 0,32 |
| 63 | 1 | 0,32 |
| 73 | 1 | 0,32 |
| 75 | 1 | 0,32 |

Fréquence des allèles morbides dans une population de personnes atteintes de chorée de Huntington. D'après Hum. Mol. Gen., 1993, vol. 2, n°12, p. 2063-2067. Toutes les personnes atteintes étaient hétérozygotes.

La très grande majorité des personnes atteintes de chorée ont un allèle du gène où le triplet est répété plus de 40 fois. Il reste une indétermination pour un triplet répété entre 36 et 39 fois car, dans certains cas, la possession d'un tel allèle est associée aux signes de la maladie et dans d'autres cas non.

Les phénotypes associés aux divers génotypes. Dominance et récessivité

Le codon CAG de l'ARNm - correspondant au CAG du brin non transcrit de l'ADN - code pour la glutamine. En conséquence, les protéines codées par les différents allèles diffèrent par le nombre de répétitions de cet acide aminé. Chez un individu hétérozygote, les deux allèles s'expriment et on trouve donc deux types de huntingtine dans les cellules.

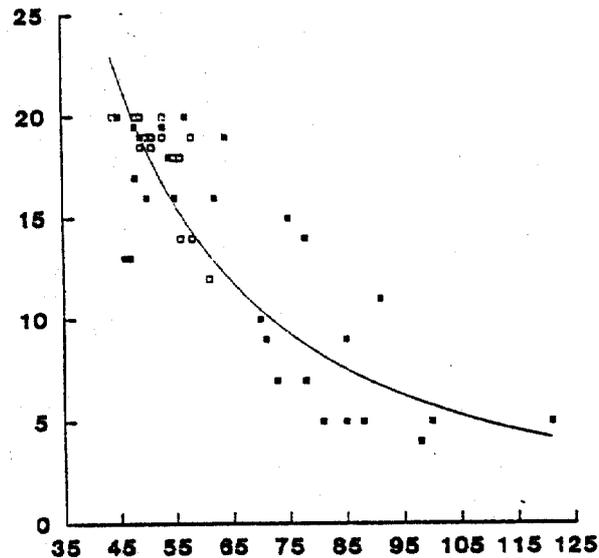
En ce qui concerne le phénotype macroscopique, les individus hétérozygotes possédant un seul allèle morbide présentent les mêmes symptômes que ceux ayant deux allèles morbides : mouvements involontaires désordonnés, perturbations du psychisme, perte de mémoire, troubles du langage et finalement démence. Ces troubles sont liés à une destruction progressive des neurones cérébraux, notamment de ceux des corps striés impliqués dans la motricité.

Le phénotype morbide et par extension l'allèle morbide est dominant. L'explication de cette dominance reste très hypothétique d'autant plus que le rôle physiologique de la huntingtine est inconnu. On a identifié des individus possédant une délétion de la région d'un chromosome 4 où est localisé le gène IT15. Ces personnes ne possèdent qu'un seul allèle de ce gène et on constate qu'elles ne

présentent pas de signes de chorée. Il semble donc que chez l'individu choréique hétérozygote, la huntingtine anormale a des propriétés qui entravent celles des molécules protéiques « normales ». On ignore lesquelles et pourquoi leur effet s'avère particulièrement drastique sur les neurones des corps striés.

L'âge d'apparition des premiers signes de la maladie varie en fonction de la nature de l'allèle morbide. En règle générale, il est d'autant plus précoce que le nombre de répétitions du triplet CAG est élevé. En revanche, il y a peu de corrélation avec la gravité des symptômes présentés par le patient une fois la maladie exprimée.

En x, nombre de répétitions du triplet CAG dans l'allèle de la personne malade. En y, âge d'apparition des premiers signes de la maladie.



Une fréquence de mutations assez élevée ?

On a constaté l'apparition de cas de chorée de Huntington dans des familles où la maladie n'avait pas été identifiée précédemment. Etant donné la dominance du phénotype morbide, ces cas semblent dus à des néomutations. L'analyse au niveau moléculaire du gène IT15 chez de tels malades et chez leurs parents a révélé qu'un des parents, très généralement le père, possédait un allèle avec une **répétition du triplet CAG comprise entre 30 et 38**. Cet allèle paternel avait subi une expansion au cours de la spermatogenèse conduisant à un allèle où le triplet CAG est répété plus de 40 fois. En outre, dans des familles où le père était atteint, on a constaté une apparition plus précoce de la maladie chez les enfants atteints, liée à la possession d'un allèle morbide ayant plus de triplets CAG que l'allèle paternel. Il semble donc qu'au delà de 30 répétitions, l'allèle IT15 ait tendance à muter avec expansion du triplet au cours de la spermatogenèse. Cette instabilité n'a pas été trouvée pour les allèles dont le nombre de répétitions est inférieur à 30. Ainsi, les études au niveau moléculaire de ce gène semblent, pour un certain type d'allèles, révéler une fréquence de mutations assez élevée alors qu'auparavant elle était estimée très basse.

Utilisations pédagogiques

Analyse du tableau de fréquence des allèles

Il est souhaitable de prendre d'abord contact avec le gène à partir de la description du phénotype morbide qui lui est associé, celui de la chorée de Huntington. On peut alors proposer l'analyse du tableau sur la fréquence des allèles du gène en demandant de discuter du polymorphisme d'IT15 et d'indiquer en quoi les données corroborent l'idée que ce gène est bien la cause de la chorée.

| Noms des allèles | Fréquence des allèles dans une population témoin (en %) | Fréquence des allèles dans une population atteinte de la maladie de Huntington (en %) |
|------------------|---|---|
| C11 | 0,98 | 0,33 |
| C13 | 0,98 | 0,81 |
| C16 | 2,94 | 0,33 |
| C17 | 20,01 | 9,06 |
| C20 | 10,78 | 7,6 |
| C21 | 14,71 | 8,09 |
| C22 | 6,86 | 4,2 |
| C23 | 5,39 | 2,25 |
| C33 | 0,49 | 0,16 |
| C40 | 0 | 3,35 |
| C43 | 0 | 5,42 |
| C45 | 0 | 6,05 |
| C48 | 0 | 2,05 |
| C50 | 0 | 0,95 |
| C75 | 0 | 0,16 |

Fréquence de quelques allèles contenus dans la banque. Dans la population étudiée, on a trouvé d'autres allèles ; il est donc normal que la somme des pourcentages ne soit pas égale à 100. Le nom de chaque allèle indique le nombre de fois où le triplet CAG est répété.

Recherche des différences entre les allèles

La séquence du gène IT15 qui correspond à la région codante est assez longue (9435 nucléotides), le triplet répété débutant à la position 52. On s'est donc limité à une portion de la séquence (brin non transcrit) commençant au triplet ATG et encadrant la partie répétée.

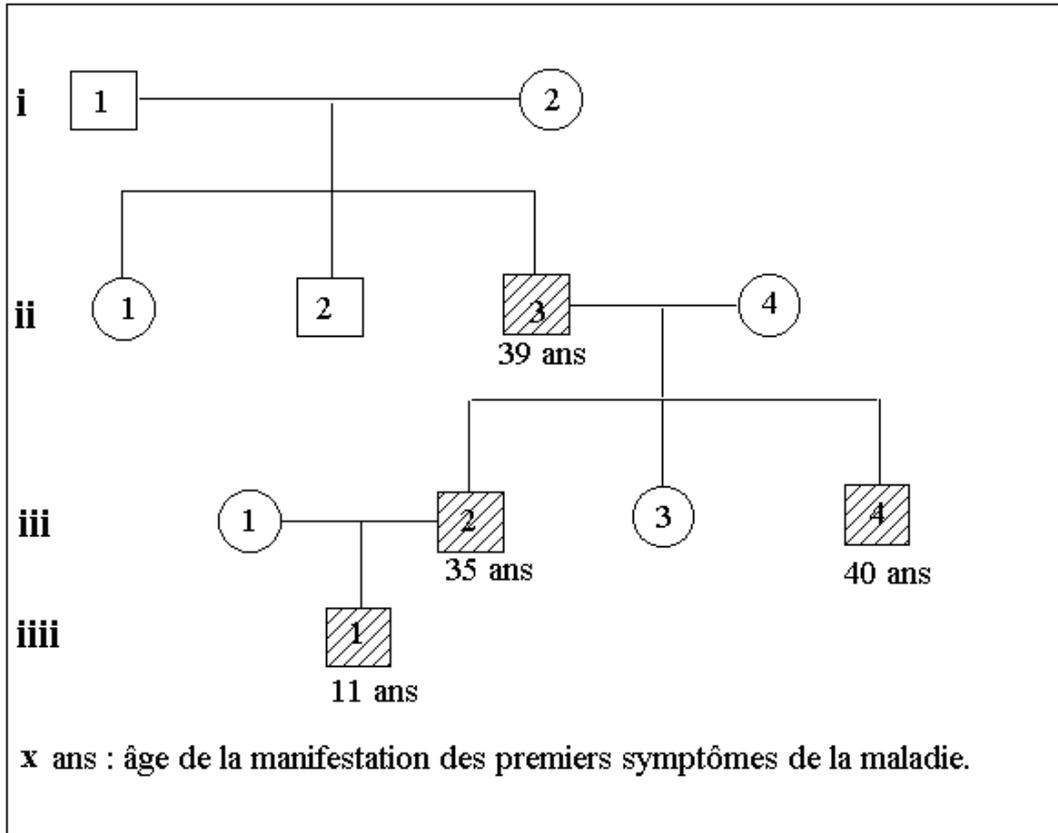
On demande aux élèves de rechercher les différences entre les allèles du gène et de dégager les caractéristiques de ceux qui sont à l'origine de la chorée. On peut prendre comme référence l'allèle dont la fréquence est la plus élevée dans les deux populations, à savoir C17. Les élèves utiliseront la modalité de comparaison dite « **Alignement avec discontinuités** » mieux adaptée que la modalité de « comparaison simple » dans ce cas (les séquences n'ont pas la même longueur). Ils remplissent le tableau résumant les différences entre allèles qui leur sera utile par la suite.

Conséquences des différences entre allèles sur la huntingtine

Diverses stratégies sont possibles. L'une d'elles consiste à rechercher, a priori, la conséquence de la variabilité du nombre de répétitions du triplet. La question de la position du triplet répété par rapport au cadre de lecture du gène est alors soulevée et les élèves doivent identifier les trois solutions imaginables et leurs conséquences au niveau de la protéine à l'aide du tableau du code génétique (CAG → glutamine, AGC → sérine, GCA → alanine).

Recherche sur les relations de dominance et récessivité entre allèles

Comme pour l'alpha-antitrypsine, on peut approcher ces notions grâce à l'arbre généalogique en demandant aux élèves de rechercher le génotype de l'individu ii3 qui souffre de chorée. Pour cela, il faut comparer les séquences des deux allèles de cette personne à un allèle de référence (C21) et les identifier à l'aide du tableau précédemment construit. ii3 est hétérozygote (C21/C43) et possède un allèle morbide.



Arbre généalogique d'une famille avec individus atteints de chorée.

Recherche d'explications sur l'évolution des phénotypes dans la famille

Les élèves doivent d'abord exprimer la contradiction existant entre les phénotypes des individus de première génération et le génotype de ii3. Pour résoudre cette contradiction, il faut rechercher les génotypes de i1 (C22/C33) et de i2 (C21/C16). ii3 possède un des deux allèles de sa mère mais aucun allèle identique à ceux portés par son père. Cela doit conduire à l'idée que l'autre allèle de ii3 (C43) provient d'une **mutation** subie par un des allèles du père lors de la spermatogenèse (sans doute l'allèle possédant le plus grand nombre de répétitions du triplet CAG, l'allèle C33). On peut alors demander aux élèves de réfléchir sur l'âge d'apparition des signes de la maladie dans les générations successives, de proposer une explication et de la tester avec le logiciel. Ils sont ainsi amenés à construire un tableau semblable au suivant dont l'analyse renforce la notion de mutation par expansion du triplet. La comparaison des génotypes des individus iii2 et iii4 met l'accent sur le caractère aléatoire des mutations : les spermatozoïdes à l'origine de ces individus possédaient respectivement un allèle muté (C48 au lieu de C43) et un allèle identique à l'allèle paternel (C43).

| Allèles des individus | Nucléotides changés | Allèles du gène |
|------------------------------|----------------------------|------------------------|
| F1-i1all1 | 22 CAG répétés | C22 |
| F1-i1all2 | 33 CAG répétés | C33 |
| F1-i2all1 | 21 CAG répétés | C21 |
| F1-i2all2 | 16 CAG répétés | C16 |
| F1-ii1all1 | 22 CAG répétés | C22 |
| F1-ii1all2 | 16 CAG répétés | C16 |
| F1-ii2a1 | 21 CAG répétés | C21 |
| F1-ii2all2 | 22 CAG répétés | C22 |
| F1-ii3all1 | 21 CAG répétés | C21 |
| F1-ii3all2 | 43 CAG répétés | C43 |
| F1-ii4all1 | 13 CAG répétés | C13 |
| F1-ii4all2 | 20 CAG répétés | C20 |
| F1-iii1all1 | 23 CAG répétés | C23 |
| F1-iii1all2 | 17 CAG répétés | C17 |
| F1-iii2all1 | 20 CAG répétés | C20 |
| F1-iii2all2 | 48 CAG répétés | C48 |
| F1-iii3all1 | 21 CAG répétés | C21 |
| F1-iii3all2 | 20 CAG répétés | C20 |
| F1-iii4all1 | 13 CAG répétés | C13 |
| F1-iii4ALL2 | 43 CAG répétés | C43 |
| F1-iiii1ALL1 | 17 CAG répétés | C17 |
| F1-iiii1ALL2 | 75 CAG répétés | C75 |

Allèles des individus de la famille comportant des sujets atteints de chorée.



Les séquences des gènes qui codent pour les glycoprotéines membranaires HLA, HLAB, HLAC

Informations scientifiques et contenu de la banque de données

Généralités

Chez l'Homme, les gènes du CMH sont situés très près les uns des autres sur le bras court du chromosome 6. Ils codent pour des glycoprotéines membranaires appelées antigènes majeurs d'histocompatibilité. Celles-ci doivent leur nom à leur capacité d'induire une forte réponse immunitaire allogénique. Introduites chez une personne qui ne les possède pas, elles provoquent une réaction immunitaire puissante se marquant par la destruction des cellules qui les portent en 7 à 10 jours (exemple : rejet de greffe de peau de première intention). Plus tard, on a élucidé leur fonction biologique essentielle qui est de présenter des peptides antigéniques aux récepteurs des lymphocytes T.

On distingue les antigènes HLA de classe I exprimés à la surface des membranes de toutes les cellules nucléées de l'organisme, et les antigènes HLA de classe II présents uniquement sur les membranes des cellules du système immunitaire : macrophages et plus généralement cellules présentatrices d'antigènes (CPA), lymphocytes B et lymphocytes T activés. Les antigènes de classe II jouent un rôle fondamental dans les communications entre cellules immunitaires indispensables au déclenchement des réactions immunitaires. Les antigènes de classe I signalent aux lymphocytes T cytotoxiques les cellules ayant un « soi modifié », étape indispensable à leur destruction.

Les informations les plus riches sur les séquences des gènes qui codent pour les protéines HLA concernent celles relatives aux antigènes de classe I ; aussi, la banque ne renferme que des séquences de ces gènes.

La structure des gènes et des protéines HLA de classe I

Les antigènes majeurs d'histocompatibilité de classe I sont codés par 3 gènes : HLA, HLAB et HLAC. Chaque gène comprend 8 exons et 7 introns. À vrai dire, un antigène de classe I est constitué de deux molécules : la chaîne polypeptidique dite alpha, codée par un gène HLA, et la bêta microglobuline ; le rôle de cette dernière est de maintenir la structure de la chaîne alpha par des liaisons chimiques faibles et sans doute d'en permettre l'expression extracellulaire.

La chaîne alpha présente :

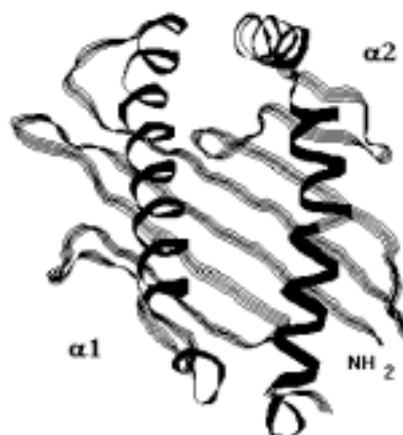
- une partie extracellulaire constituée par 3 domaines de 90 acides aminés chacun : les domaines alpha1, alpha2 et alpha3 ;
- une partie intramembranaire riche en acides aminés hydrophobes permettant l'ancrage solide de la molécule dans la membrane ;
- une partie intracytoplasmique constituant l'extrémité carboxyterminale de la molécule.

L'exon 1 correspond à l'extrémité 5' non traduite du gène et au peptide signal de 24 acides aminés ; ce peptide signal synthétisé lors de la lecture de l'ARNm par les ribosomes permet le passage de la chaîne polypeptidique dans le réticulum et est ensuite éliminé. Il ne se retrouve donc pas dans le polypeptide mature.

Les séquences strictement codantes des gènes de la banque (brin non transcrit) commencent au codon d'initiation de la traduction ATG et se terminent à un codon STOP. Elles comportent donc la partie de l'exon 1 correspondant au peptide signal. Les exons 2, 3 et 4 du gène codent respectivement pour les domaines alpha1, alpha2 et alpha3. L'exon 5 code pour la partie intramembranaire ; les exons 6, 7 ainsi que le tout début de l'exon 8 codent pour la région intracytoplasmique. Les séquences de la banque ne comprennent pas la partie de l'exon 8 non traduite située en arrière du codon STOP. Pour le locus HLAB, l'exon 8 se trouve entièrement dans la partie non traduite du gène.

La structure tridimensionnelle d'une chaîne alpha de classe I, celle de HLAA2, a été établie par Bjorkman en 1987. Sa caractéristique la plus remarquable est la corbeille présentatrice de peptides antigéniques dessinée par les domaines alpha1 et alpha2.

Représentation en ruban des 181 premiers acides aminés (sans le peptide signal). Obtenue à partir du fichier PDB 1HHH (HLAA0201) avec le logiciel RASMOL (Roger Sayle, Glaxo-Wellcome Research and Development Centre).



Les allèles des gènes HLA de classe I

Ce sont des gènes extraordinairement polymorphes puisqu'on reconnaît plus de 40 allèles HLAA, plus de 60 allèles HLAB et une quinzaine d'allèles HLAC. Dans la majorité des populations étudiées, la fréquence des divers allèles répertoriés dépasse 1% mais varie d'une population à l'autre.

| Allèles | Français | Japonais |
|---------|----------|----------|
| N | 355 | 939 |
| HLAA1 | 14,6 | 0,7 |
| HLAA2 | 20,9 | 24,1 |
| HLAA3 | 13,6 | 0,6 |
| HLAA11 | 5,7 | 10,4 |
| HLAA23 | 2,8 | - |
| HLAA24 | 9,5 | 35,6 |
| HLAA25 | 2,5 | 0 |
| HLAA26 | 4,2 | 10,9 |
| HLAA28 | 5,5 | 0 |

Fréquence en % de quelques allèles du locus HLAA dans deux échantillons de populations française et japonaise.

N = nombre d'individus testés dans chaque population.

D'après J. Colombani, "HLA. Fonctions immunitaires et applications médicales", John Libbey, Eurotext, p. 205.

Les différents allèles d'un locus ont d'abord été définis sérologiquement grâce à des sérums tests contenant des anticorps réagissant spécifiquement avec un antigène HLA produit de l'expression de l'allèle. La technique de référence est la lymphotoxicité dépendante du complément ; c'est aussi la technique utilisée pour le typage HLA de routine. Les sérums tests et les cellules d'une personne sont mis en présence ; on ajoute le complément et on estime ensuite le nombre de cellules lysées. Cette technique a permis de définir les spécificités HLAA1, HLAA2, HLAA3, etc.

Au cours des années 80, on a commencé à cloner et à séquencer divers allèles HLA. Les allèles HLAB27 et HLAA2 ont reçu beaucoup d'attention en raison de l'association du premier avec la

spondylarthrite ankylosante (inflammation et blocage des articulations vertébrales ; 80 à 90 % des malades possèdent l'allèle HLAB27) et parce que le second est très répandu. Les chercheurs ont ainsi découvert que l'allèle HLAB27, défini sérologiquement, correspondait à 6 séquences nucléotidiques différentes, donc à 6 allèles, et l'allèle HLAA2 à 12 séquences. Par la suite, les mêmes faits ont été découverts pour d'autres allèles (HLAA11, HLAA24, HLACW1, etc.). Le répertoire des allèles HLA est ainsi en constante évolution. La nécessité d'une nouvelle nomenclature s'est imposée à l'Organisation mondiale de la santé chargée d'homologuer les différentes spécificités. Désormais, chaque allèle est défini par quatre chiffres : les deux premiers correspondent à la spécificité repérée sérologiquement, les deux derniers à une séquence précise. Ainsi, l'ancien allèle HLAB27 est actuellement scindé en 6 allèles : HLAB2701 à HLAB2706.

Outre l'importance de leur polymorphisme, les gènes HLA sont remarquables par les caractéristiques des différences entre les allèles d'un même locus. Pour la plupart des autres gènes les différences alléliques se limitent à des substitutions de un à quelques nucléotides, ou à des délétions ou insertions d'un petit nombre de nucléotides. Il en est de même pour les allèles HLA qui résultent de la subdivision d'une spécificité allélique définie sérologiquement (exemple : HLAB2701, HLAB2702...). En revanche, la comparaison d'allèles d'un locus appartenant à des groupes antigéniques repérés sérologiquement (HLAA01, HLAA02, HLAA24, etc.) révèle que ces allèles diffèrent par beaucoup de nucléotides, 40 à 50 en moyenne. Les sites de variabilité sont dispersés sur toute la chaîne mais sont surtout situés dans les régions qui codent pour les domaines alpha1 et alpha2 de la protéine.

Les chaînes polypeptidiques codées par les différents allèles d'un locus ont toutes la même longueur (365 acides aminés avec le peptide signal pour HLAA) ce qui signifie qu'aucune substitution n'entraîne l'apparition anticipée d'un codon STOP. Etant donné le rôle fondamental des antigènes HLA dans les réactions immunitaires, il est probable que tout allèle entraînant la formation d'une chaîne raccourcie serait rapidement éliminé par la sélection naturelle.

Les polypeptides alléliques correspondant à une même spécificité sérologique ne diffèrent que par un à quelques acides aminés ce qui explique qu'ils aient été confondus par les méthodes immunologiques. Les polypeptides alléliques correspondant à des spécificités sérologiques différentes ont en moyenne 20 à 25 acides aminés différents.

Puisque dans les deux cas la variabilité est surtout concentrée au site de liaison du peptide antigénique, il en résulte que les antigènes HLA codés par les divers allèles diffèrent dans leur capacité à présenter des peptides. Certains antigènes lient certains peptides et non d'autres. Chaque organisme répond plus ou moins efficacement à des antigènes viraux, bactériens, vaccinaux, en fonction de l'assemblage d'antigènes HLA qu'il possède. Mais le polymorphisme au niveau de l'espèce permet la présentation d'un répertoire étendu de peptides et lui confère une meilleure capacité de réponse immunitaire contre les substances étrangères.

Aspects de l'évolution des gènes HLA

Considérons en premier les allèles correspondant à la spécificité HLAB27. L'allèle HLAB2705 est répandu dans toutes les populations étudiées (90% des allèles HLAB27 chez les blancs européens et 45% chez les extrême-orientaux). Les allèles HLAB2701 (rare) et HLAB2702 (près de 10%) n'ont été trouvés que chez les blancs. HLAB2704 et HLAB2706 sont restreints aux extrême-orientaux (les allèles HLAB2704 représentent près de 55% des allèles HLAB27 dans ces populations). L'allèle HLAB2703, très rare, n'a été trouvé que chez les noirs américains. Les associations entre tels allèles et telles populations indiquent que les allèles HLAB2701 à HLAB2706 sont apparus après diversification des populations. On peut penser que l'allèle HLAB2705 présent avec une fréquence élevée dans toutes les populations, était l'allèle initial dans les premières populations humaines.

En estimant que deux allèles sont d'autant plus apparentés que les différences entre leurs séquences sont minimales, on peut construire une représentation graphique traduisant les filiations entre les divers allèles qui s'accorde bien avec les données de génétique des populations résumées plus haut.

Envisageons maintenant le problème de façon plus générale en considérant l'ensemble des spécificités d'un locus. Comment rendre compte d'un tel polymorphisme et du grand nombre de différences entre les séquences nucléiques de beaucoup de ces allèles ? Puisqu'un allèle provient d'une mutation d'un allèle préexistant, on pourrait supposer que les caractéristiques précédentes résultent d'une fréquence de mutation élevée des gènes HLA. Diverses données laissent à penser qu'il n'en est rien. En particulier, au cours des analyses génétiques familiales, on n'a jamais constaté le cas d'un enfant ayant un allèle HLA que ne posséderait pas l'un de ses parents.

Klein a proposé une autre explication : une grande partie du polymorphisme HLA n'aurait pas été acquise après la spéciation de l'*Homo sapiens sapiens*, mais serait héritée des espèces ancestrales. Seul le polymorphisme correspondant aux séquences relatives à une même spécificité sérologique (HLA-B*27 par exemple) serait post-spécifique. Cela revient à dire que les premières populations d'*Homo sapiens sapiens* possédaient plusieurs allèles à chaque locus HLA hérités d'*Homo erectus*. Suivant cette hypothèse trans-spécifique, les allèles HLA de l'Homme moderne peuvent remonter à des espèces ancestrales relativement reculées.

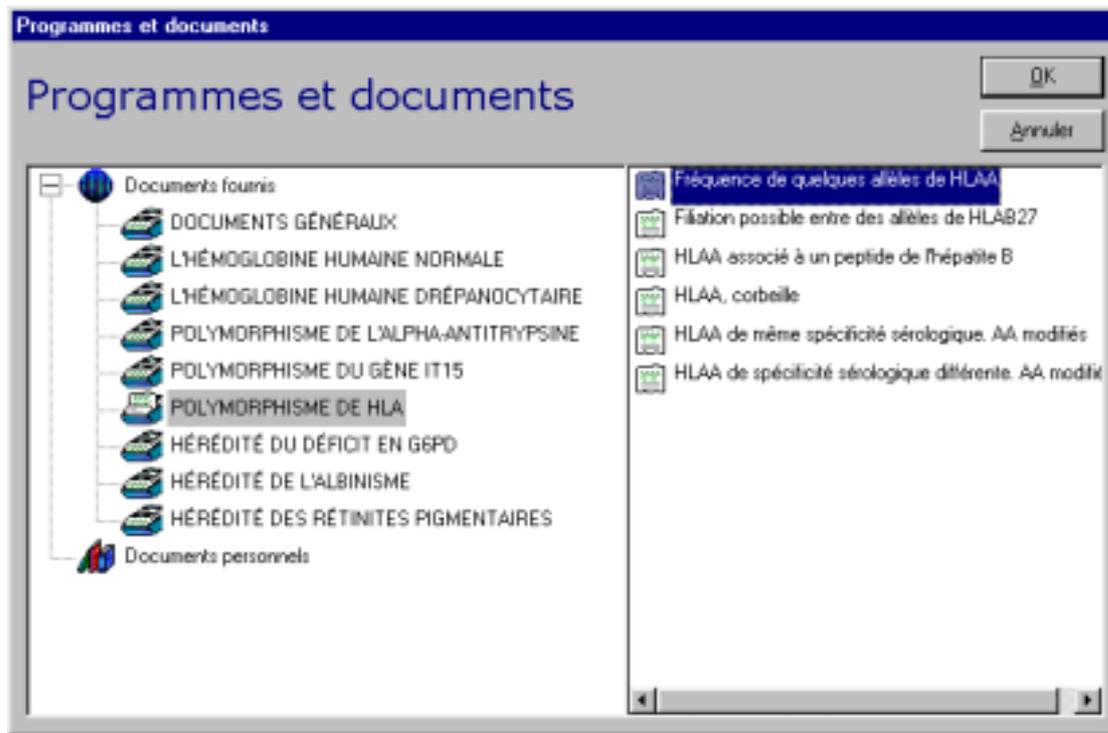
Récemment, l'analyse au niveau moléculaire du CMH du Chimpanzé a fourni des données corroborant cette hypothèse. Chez le Chimpanzé comme chez l'Homme, il y a trois locus polymorphiques pour les antigènes de classe I. On s'est aperçu que **le degré de similitude entre des allèles CHLAA du Chimpanzé (CH pour Chimpanzé) et des allèles HLAA humains était plus grand qu'entre certains allèles humains**. Ainsi, la séquence codante du CHLA108 du Chimpanzé ne diffère que par 15 nucléotides de celle de l'allèle humain HLAA11, alors que celles de HLAA11 et HLAA2 présentent 50 différences. On est amené à conclure que l'origine des allèles HLAA11 et HLAA2 est antérieure à la divergence de la lignée humaine et de celle du Chimpanzé.

Utilisations pédagogiques

L'extraordinaire polymorphisme des gènes HLA en fait un exemple incontournable pour approcher cette notion. Le rôle précis des antigènes HLA est envisagé plus tard avec l'étude des mécanismes de l'immunité. Il est néanmoins possible de fournir simplement les informations sur les gènes HLA qui rendent significative l'étude de leur polymorphisme. Ces informations portent sur :

- la localisation des locus des gènes HLA de classe I ;
- les fréquences des allèles de chaque locus dans une ou deux populations ;
- le rôle des marqueurs du soi des antigènes HLA en liaison avec la forte réponse immunitaire allogénique qu'ils induisent lors de greffe ou d'implantation ;
- une évocation de leur fonction biologique de présentation de peptides antigéniques aux cellules du système immunitaire.

Plusieurs documents peuvent venir en appui à cette approche.



Recherche des différences entre les allèles d'un gène HLA

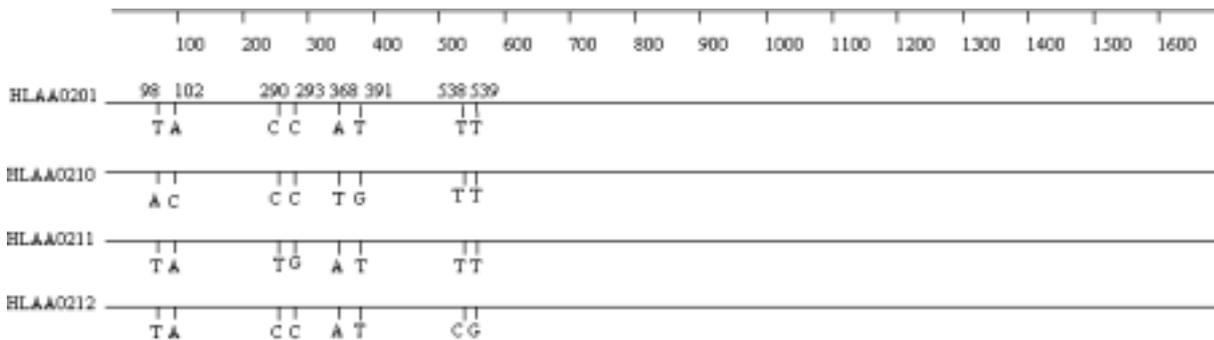
Les seules séquences sur les allèles HLA A suffisent pour dégager les idées essentielles. On peut proposer de rechercher séparément, en utilisant la modalité de « comparaison simple » :

- Les différences entre les allèles correspondant à une même spécificité sérologique : ceux du groupe HLA A2 ou ceux du groupe HLA A24.

| | | | | |
|----------|----------|----------|----------|----------|
| HLAA0201 | 0 | | | |
| HLAA0210 | 4 | 0 | | |
| HLAA0211 | 2 | 6 | 0 | |
| HLAA0212 | 2 | 6 | 4 | 0 |
| | HLAA0201 | HLAA0210 | HLAA0211 | HLAA0212 |

| | | | |
|----------|----------|----------|----------|
| HLAA2401 | 0 | | |
| HLAA2402 | 5 | 0 | |
| HLAA2403 | 7 | 2 | 0 |
| | HLAA2401 | HLAA2402 | HLAA2403 |

Matrices des différences entre quelques allèles de HLA A2 et HLA A24.



Représentation des différences entre quelques allèles de HLA A2

- Les différences entre allèles appartenant à des spécificités différentes

| | | | | |
|----------|-----------|-----------|-----------|----------|
| HLAA0101 | 0 | | | |
| HLAA0201 | 50 | 0 | | |
| HLAA1101 | 12 | 50 | 0 | |
| HLAA2401 | 47 | 49 | 48 | 0 |
| | HLAA0101 | HLAA0201 | HLAA1101 | HLAA2401 |

Matrice des différences entre HLA A0101, HLA A0201, HLA A1101 et HLA A2401.

Les élèves doivent rassembler les informations obtenues et dégager l'originalité des différences alléliques des gènes HLA par rapport à celles vues sur d'autres gènes : uniquement des substitutions ; grand nombre de différences entre allèles de spécificité sérologique différente ; nombre de différences beaucoup plus faible entre allèles de même spécificité sérologique. L'analyse plus fine de la comparaison multiple des allèles montre qu'il existe des sites qui sont conservés et d'autres qui sont variables.

Conséquences des différences alléliques au niveau protéique

Après traduction et comparaison des séquences protéiques, les élèves doivent dresser un tableau des résultats.

| | | | | |
|----------|----------|----------|----------|----------|
| HLAA0201 | 0 | | | |
| HLAA0210 | 3 | 0 | | |
| HLAA0211 | 2 | 5 | 0 | |
| HLAA0212 | 1 | 4 | 3 | 0 |
| | HLAA0201 | HLAA0210 | HLAA0211 | HLAA0212 |

| | | | |
|----------|----------|----------|----------|
| HLAA2401 | 0 | | |
| HLAA2402 | 4 | 0 | |
| HLAA2403 | 6 | 2 | 0 |
| | HLAA2401 | HLAA2402 | HLAA2403 |

| | | | | |
|----------|-----------|-----------|-----------|----------|
| HLAA0101 | 0 | | | |
| HLAA0201 | 33 | 0 | | |
| HLAA1101 | 11 | 27 | 0 | |
| HLAA2401 | 33 | 30 | 33 | 0 |
| | HLAA0101 | HLAA0201 | HLAA1101 | HLAA2401 |

Matrices des différences entre polypeptides codés par divers allèles de HLAA.

La comparaison du nombre de différences au niveau nucléique et au niveau protéique conduit à la notion de mutation silencieuse en rapport avec la redondance du code génétique. Une estimation de la fréquence des mutations silencieuses peut être faite. Là aussi, une analyse plus fine des différences entre les séquences peptidiques correspondant aux divers allèles peut être réalisée. Il est intéressant, notamment, de localiser les régions de l'antigène HLA où la variabilité allélique est la plus importante (domaines alpha1 et alpha2) en liaison avec la fonction biologique de présentation du peptide.

Recherche de l'histoire évolutive des gènes HLA de classe I

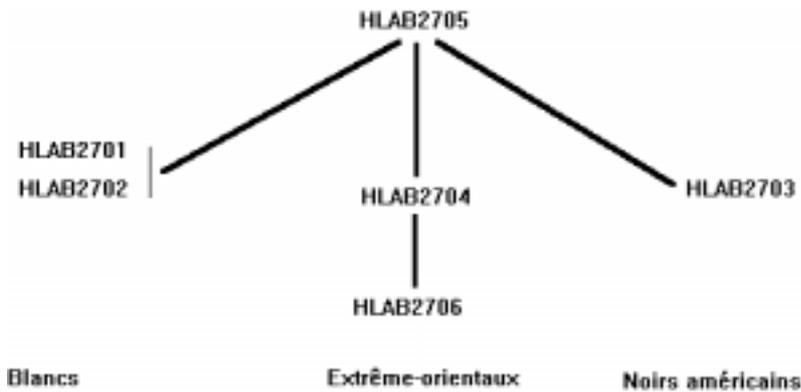
Comme pour le gène codant pour l'alpha-antitrypsine, l'objectif de cette étude est de faire prendre conscience que les allèles présents dans les populations humaines sont l'aboutissement actuel des mutations survenues dans le passé.

Les séquences relatives à la spécificité HLAB27 sont propices pour approcher ce problème et il est plus aisé de travailler au niveau protéique qu'au niveau nucléique. Les élèves doivent traduire les séquences HLAB2701 à HLAB2706 et garder en mémoire les séquences protéiques pour les comparer. À partir des informations fournies dans les pages précédentes sur les fréquences des différents allèles HLAB27, ils doivent choisir HLAB2705 comme référence (allèle le plus fréquent, présent avant la diversification des populations humaines). Un repérage précis de la localisation et de la nature des différences par rapport à la séquence de référence peut aboutir à une représentation du type suivant.

| | -----I----- | | | | | -----I----- | | -----I----- | |
|----------|-------------|----|--------|-----|--------|-------------|-----|-------------|--|
| | alpha1 | | alpha2 | | alpha3 | | | | |
| | 83 | 98 | 101 | 104 | 105 | 138 | 140 | 176 | |
| HLAB2705 | Y | D | D | T | L | H | D | V | |
| HLAB2701 | | Y | N | | A | | | | |
| HLAB2702 | | | N | I | A | | | | |
| HLAB2703 | H | | | | | | | | |
| HLAB2704 | | | S | | | | | E | |
| HLAB2706 | | | S | | | D | Y | E | |

Différences entre le polypeptide codé par l'allèle HLAB2705 et les polypeptides codés par divers allèles HLAB27. Y=Tyrosine ; D=acide aspartique ; L=Leucine ; I=Isoleucine ; H=Histidine ; S=Sérine ; T=Thréonine ; E=Acide glutamique ; A=Alanine ; V=Valine ; N=Asparagine.

Il reste à travailler sur ces constats pour regrouper les allèles les plus étroitement apparentés et proposer un schéma de filiation entre ces différents allèles.



Filiation possible entre les divers allèles de HLAB27.

On peut ensuite faire prendre conscience des limites des modalités précédentes (apparition des allèles par mutations à partir d'un allèle ancestral unique) pour expliquer l'étendue des différences entre les allèles HLAA1 ET HLAA2 et proposer l'hypothèse trans-spécifique de Klein. Les élèves sont invités alors à comparer l'allèle CHLA108 du chimpanzé avec les allèles humains HLAA et à indiquer en quoi les informations tirées de cette comparaison corroborent l'hypothèse trans-spécifique.

| | | | | | |
|----------|-----------|-----------|-----------|-----------|----------|
| HLAA0101 | 0 | | | | |
| HLAA0201 | 50 | 0 | | | |
| HLAA1101 | 12 | 50 | 0 | | |
| HLAA2401 | 47 | 49 | 48 | 0 | |
| CHLA108 | 24 | 48 | 15 | 53 | 0 |
| | HLAA0101 | HLAA0201 | HLAA1101 | HLAA2401 | CHLA108 |

Matrice des différences entre des allèles de HLAA et l'allèle CHLA108 du Chimpanzé.

FAMILLES MULTIGÉNIQUES

L'étude de l'origine du polymorphisme des gènes, en particulier celle des allèles des gènes qui codent pour l'alpha-antitrypsine ou la G6PD, a révélé l'importance des mutations ponctuelles dans la diversification des génomes au cours de l'évolution. Mais ce mécanisme est loin d'être le seul. Dans le génome de tous les organismes existent des gènes dont les similitudes sont telles qu'elles impliquent une parenté entre eux : on dit que ce sont des gènes homologues paralogues. Ils proviennent tous d'un gène ancestral par duplications successives suivies d'une divergence plus ou moins importante des copies. L'ensemble des gènes ainsi apparentés forme une famille multigénique dont le modèle classique est celui de la famille des gènes qui codent pour les chaînes alpha, bêta, gamma et delta de l'hémoglobine humaine. D'autres exemples sont apportés qui illustrent le fait que la duplication génique fait partie du processus général d'expansion du génome qui va de la répétition d'un triplet au sein d'un gène (cas du gène qui code pour la huntingtine) jusqu'à la polyploïdie.



La famille des globines

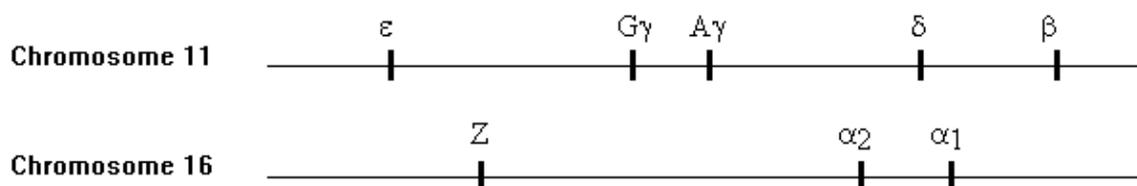
Informations scientifiques

L'élève doit être sensibilisé à l'idée que l'analyse en cours des génomes des espèces actuelles peut fournir des données informatives sur les processus qui, dans le passé, ont contribué à la genèse de nouveaux gènes. Il en est ainsi en ce qui concerne les séquences des gènes qui, dans l'espèce humaine, codent pour les chaînes de l'hémoglobine.

Pour aborder cette étude avec les globines, il est nécessaire d'apporter des informations sur la synthèse des différentes chaînes au cours de la vie de l'individu :

| Hémoglobine | Type de chaînes | Moment de synthèse |
|-----------------|---------------------|------------------------|
| Hb embryonnaire | $Z_2 \epsilon_2$ | avant 9 semaines |
| Hb foétale | $\alpha_2 \gamma_2$ | 9 semaines---naissance |
| Hb A | $\alpha_2 \beta_2$ | après naissance |
| Hb D | $\alpha_2 \delta_2$ | après naissance |

Après la naissance il y a synthèse de 97% Hb A, 2% Hb D et 1% Hb foétale. L'intérêt du changement de chaîne après la naissance pourra être évoqué. Il importe que l'élève saisisse que ces différentes chaînes sont codées par des gènes différents et non par les allèles d'un même gène, ce qui peut être facilité par la localisation chromosomique de ces gènes.



Ce schéma révèle l'existence de deux gènes gamma ($G\gamma$ et $A\gamma$) et de deux gènes alpha (α_1 et α_2). Les deux gènes gamma ne diffèrent, dans leur région codante, qu'au niveau du codon 136 (GGA = Glycine pour $G\gamma$, et GCA = Alanine pour $A\gamma$). Dans la banque, on a retenu un seul gène $G\gamma$. De même, l'identité

des séquences codantes des deux gènes alpha fait qu'on a retenu un seul des deux. De ce fait, le schéma ci-dessus est à simplifier pour les élèves.

Une grande ressemblance dans la structure des chaînes alpha, bêta, gamma et delta a été mise en évidence : elles possèdent :

- 3 exons et 2 introns dans tous les cas ;
- des longueurs de séquences codantes identiques ou très voisines.

Structures des gènes codant pour les chaînes alpha, bêta, gamma et delta de l'hémoglobine humaine

| CHAÎNE | EXON 1 | intron 1 | EXON 2 | intron 2 | EXON 3 |
|--------|------------------|-------------------|-------------------|--------------------|---------------------|
| bêta | 1--140 (140)* | 141--270 (130) | 271--494 (224) | 495--1344 (850) | 1345--1606 (262) |
| delta | 1--144 (144) | 145--272 (128) | 273--493 (221) | 494--1382 (889) | 1383--1641 (259) |
| gamma | 1--144 (145) | 145--267 (123) | 268--488 (223) | 489--1376 (887) | 1377--1593 (217) |
| alpha | 1--128 (128) | 129--245 (117) | 246--450 (205) | 451--590 (140) | 591--831 (241) |

* Longueur de la séquence

Utilisation pédagogique

Comparaison des chaînes protéiques

Toutes ces chaînes d'hémoglobine participent au transport du dioxygène, et sont synthétisées dans les mêmes cellules, les erythroblastes. On peut se demander si, à la similitude de fonction de ces chaînes, correspondent des ressemblances dans leur structure et leurs séquences d'acides aminés. Pour obtenir des informations, les élèves doivent noter le nombre d'acides aminés de chaque chaîne, comparer les séquences deux à deux, relever le nombre d'acides aminés identiques ou le pourcentage d'identité (attention le logiciel donne le nombre de différences). Ces comparaisons donnent les résultats suivants concernant le nombre d'acides aminés identiques entre les séquences :

| | | | | |
|------------------------|-------|------------|------------|-----------|
| | delta | | | |
| | bêta | 137 | | |
| | gamma | 106 | 108 | |
| | alpha | 65 | 65 | 60 |
| | | delta | bêta | gamma |
| Longueur des séquences | | 147 | 147 | 147 |
| | | | | alpha |
| | | | | 142 |

On constate que β et δ sont très semblables, que le nombre d'éléments identiques est pratiquement le même entre β et γ et entre δ et γ . La chaîne α montre le plus de différences par rapport aux trois autres.

On peut aussi leur demander de rechercher les sites conservés dans les quatre chaînes : pour cela, ils réalisent une comparaison simultanée (alignement avec discontinuités) des quatre séquences d'hémoglobine. Il est recommandé d'utiliser comme séquence de référence une des séquences les plus longues (bêta, gamma ou delta longues de 444 bases chacune).

Le décompte du nombre de positions identiques dans les quatre chaînes ainsi que les identités entre chacune des séquences et la référence est fourni par l'information (cliquer sur la fenêtre de résultat de l'alignement pour l'activer ; cliquer sur l'icône d'information).

Les résultats de cet alignement multiple peuvent être très légèrement variables suivant la séquence choisie comme référence. Dans l'exemple ci-dessous montrant les 24 premiers acides aminés alignés, c'est la séquence de la globine bêta qui est la référence ; les positions identiques dans les quatre séquences sont indiquées par des étoiles (les points indiquent les positions comportant des acides aminés similaires, acides ou basiques).

| | 1 | 5 | 10 | 15 | 20 | |
|-------------------|-----|-----|-----|-----|-----|-----|
| Traitement | < | > | < | > | < | > |
| Identities | : | : | : | * | : | : |
| Polypeptide bêta | Met | Val | His | Leu | Thr | Pro |
| Polypeptide delta | - | - | - | - | - | - |
| Polypeptide gamma | Gly | Phe | Glu | Asp | Ala | Thr |
| Polypeptide alpha | Met | Val | Ser | Ala | Asp | Thr |



L'alignement comprend 149 acides aminés

-> 51 acides aminés identiques (représentés par le signe *)
soit 34,2 % d'identité, et 68,5 % de ressemblance

Bêta protéique

longueur : 147 acides aminés (sans compter les discontinuités)
> référence pour la comparaison

Delta protéique

longueur : 147 acides aminés (sans compter les discontinuités)
-> 137 a.a. identiques à la séquence de référence Bêta protéique,
soit 93,2 % d'identité

Gamma protéique

longueur : 147 acides aminés (sans compter les discontinuités)
-> 108 a.a. identiques à la séquence de référence Bêta protéique,
soit 73,5 % d'identité

Alpha protéique

longueur : 142 acides aminés (sans compter les discontinuités)
-> 65 a.a. identiques à la séquence de référence Bêta protéique,
soit 45,8 % d'identité

le signe - représente les identités

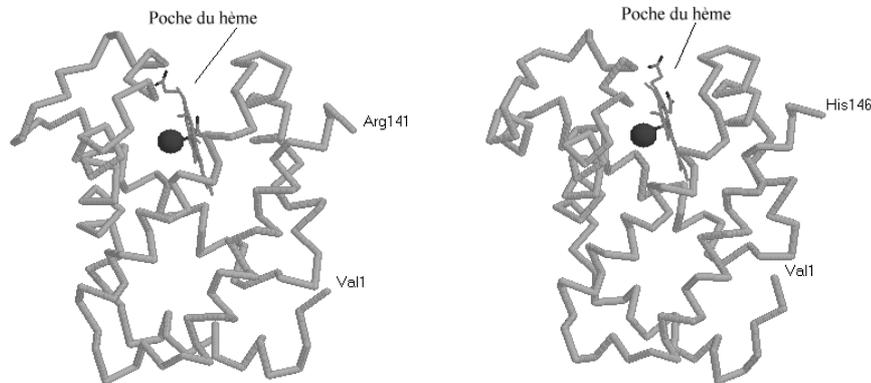
le signe _ représente les discontinuités

le signe . représente la proximité chimique d'acides aminés

La réflexion sur la signification des similitudes constatées conduit à rechercher si elles peuvent être dues au hasard ou si elles témoignent d'une « parenté » entre les chaînes des globines. Les élèves peuvent approcher cette question à partir :

- de **considérations théoriques** : si les 20 acides aminés étaient également utilisés dans les séquences protéiques, la probabilité d'avoir, entre deux séquences non apparentées (constituées au hasard), le même acide aminé à un site déterminé, serait de 5 %. Comme certains acides aminés sont plus fréquemment utilisés que d'autres, on considère que les ressemblances n'indiquent une parenté qu'au delà de 20 %. Or, les ressemblances entre les chaînes de l'hémoglobine sont nettement supérieures à ces valeurs ;
- de la **comparaison de deux polypeptides ayant des fonctions différentes**, par exemple, une chaîne de globine et l'enzyme qui intervient dans la synthèse de l'antigène A du système du groupe sanguin ABO. La comparaison de la séquence de la chaîne de globine avec des segments de même longueur dans la séquence de l'enzyme, montre que le pourcentage des acides aminés identiques ne dépasse pas 20 %.

Alors que les globines bêta, delta et gamma sont assez proches, la globine alpha apparaît comme la plus différente des autres. Cependant, malgré un grand nombre d'acides aminés différents, les globines alpha ont une structure tridimensionnelle semblable. En particulier, la poche accueillant le hème est conservée.



Les globines alpha (à gauche) et bêta (à droite). Visualisation en squelette carboné ; le hème est en bâtonnets et l'oxygène en sphère. Visualisation obtenue avec le logiciel RASMOL (Roger Sayle, Glaxo Wellcome Research & Development, Stevenage, Hertfordshire, UK). Coordonnées spatiales provenant de Protein Data bank.

Il importe maintenant de voir si la « parenté » mise en évidence entre les chaînes des globines peut être confirmée au niveau des gènes qui les codent.

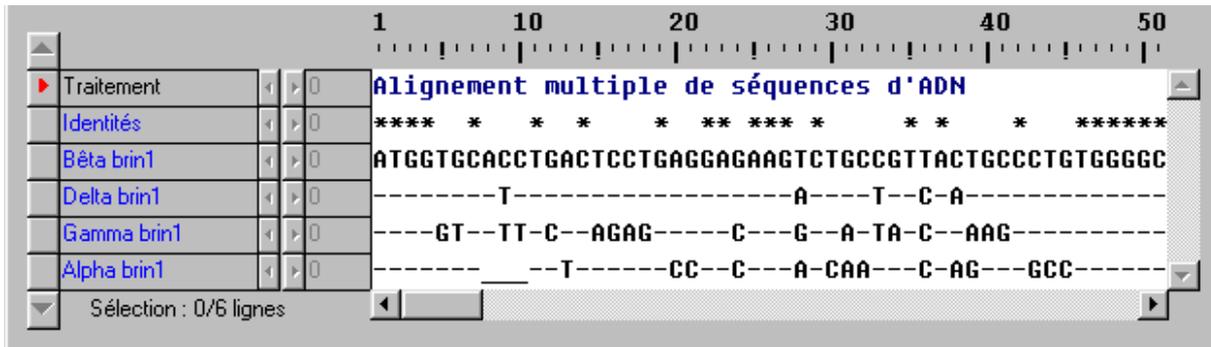
Comparaison des chaînes nucléiques

Le tableau qui résulte de la comparaison des séquences deux par deux est le suivant :

| | | | | | |
|------------------------|-------|------------|------------|------------|-------|
| | delta | | | | |
| | bêta | 411 | | | |
| | gamma | 340 | 339 | | |
| | alpha | 258 | 267 | 238 | |
| | delta | | bêta | gamma | alpha |
| Longueur des séquences | 444 | 444 | 444 | 444 | 429 |

L'alignement multiple avec discontinuité de ces quatre séquences est réalisé en prenant comme référence la séquence bêta.

ATGC Comparer les séquences



Alignement avec discontinuité montrant les 51 premières bases des séquences des globines humaines alpha, bêta, gamma et delta. Le signe - représente les identités, le signe _ les discontinuités.

i

Alignement multiple de séquences d'ADN :
 -> 213 bases identiques (représentées par le signe *)
 soit 47,3 % d'identité

Bêta
 longueur : 444 bases (sans compter les discontinuités)
 -> référence pour la comparaison

Delta
 longueur : 444 bases (sans compter les discontinuités)
 -> 411 bases identiques à la séquence de référence Bêta,
 soit 92,6 % d'identité

Gamma
 longueur : 444 bases (sans compter les discontinuités)
 -> 339 bases identiques à la séquence de référence Bêta,
 soit 76,4 % d'identité

Alpha
 longueur : 429 bases (sans compter les discontinuités)
 -> 264 bases identiques à la séquence de référence Bêta,
 soit 61,5 % d'identité

Comme attendu, les similitudes constatées à propos des protéines se retrouvent au niveau des parties codantes des gènes. Notons toutefois que de nombreuses différences au niveau des gènes ne se traduisent pas par des différences entre les polypeptides, à cause du caractère dégénéré du code génétique.

Le bilan de tout ce travail conduit à l'idée qu'il existe dans le génome humain des gènes apparentés car présentant des ressemblances qui ne peuvent s'expliquer par des processus aléatoires. Il reste à rechercher, ce qui ne peut être fait avec le logiciel :

- une explication à cette parenté (notion de duplication des gènes) ;
- une explication aux différences constatées entre ces gènes et à leur degré plus ou moins important (mutations des gènes après leur duplication) ;

- une explication à l'importance des différences constatées au niveau des introns (absence de pression sélective).

Les considérations sur l'importance des différences entre les gènes ou les polypeptides pourra permettre d'établir une généalogie des diverses duplications.

La famille des gènes HLA de classe I

Informations scientifiques

Ce système de gènes permet de cibler deux aspects de l'évolution des génomes :

- la duplication génique révélée par la similitude des séquences aux trois locus ;
- le polymorphisme révélé par le grand nombre d'allèles présents dans les populations avec une fréquence élevée à chaque locus.

Ainsi, les trois gènes HLAA, HLAB et HLAC ont la même organisation d'ensemble : 8 exons et 7 introns, comme résumé dans la figure ci-après.

| | HLAA0201 | HLAB2705 | HLACW301 |
|-----------|-----------------|-----------------|-----------------|
| Promoteur | 0-27 | 0-27 | 0-27 |
| Exon 1 | 28-125 | 28-124 | 28-124 |
| Intron 1 | 126-264 | 125-253 | 125-255 |
| Exon 2 | 265-524 | 254-623 | 256-525 |
| Intron 2 | 525-765 | 524-764 | 526-767 |
| Exon 3 | 766-1041 | 765-1040 | 768-1043 |
| Intron 3 | 1042-1640 | 1041-1615 | 1044-1631 |
| Exon 4 | 1641-1926 | 1616-1891 | 1632-1906 |
| Intron 4 | 1927-2015 | 1892-1983 | 1907-2027 |
| Exon 5 | 2016-2130 | 1984-2103 | 2028-2147 |
| Intron 5 | 2131-2568 | 2104-2554 | 2148-2588 |
| Exon 6 | 2569-2604 | 2555-2576 | 2589-2622 |
| Intron 6 | 2605-2746 | 2577-2682 | 2623-2776 |
| Exon 7 | 2747-2794 | 2683-3334 | 2777-2824 |
| Intron 7 | 2795-2963 | | 2825-2988 |
| Exon 8 | 2964-3528 | | 2989-3408 |

L'alignement multiple des trois allèles HLAA0201, HLAB2705 et HLACW301 montre l'importance des sites identiques dans les séquences nucléiques (brin non transcrit). La similitude moyenne entre les gènes HLAA et HLAB est de 82,5%, entre HLAA et HLAC de 83,5%, entre HLAB et HLAC de 87%. La différence moyenne entre les trois gènes est environ trois fois plus importante qu'entre allèles de même locus.

On s'est interrogé sur la signification évolutive de cette famille multigénique. Les trois gènes s'expriment dans la quasi totalité des cellules nucléées de l'organisme en codant pour des polypeptides dont la fonction est de présenter les peptides antigéniques aux lymphocytes T. Par suite du polymorphisme, la plupart des individus sont hétérozygotes pour chacun de ces gènes. Comme il y a codominance, cela fait que chaque personne possède six types d'antigènes HLA de classe I ce qui augmente la capacité à présenter des peptides et sans doute celle de réagir à une multiplicité d'agents pathogènes.

Utilisations pédagogiques

Recherche des similitudes entre les séquences nucléiques des gènes HLAA, HLAB et HLAC

L'étude du polymorphisme a familiarisé les élèves avec les gènes du système HLA. La recherche d'une parenté éventuelle entre ces gènes peut être posée à partir de l'identité des fonctions qu'exercent les polypeptides qu'ils codent (notion envisagée lors de l'étude des mécanismes immunitaires). Le seul moyen de tester cette hypothèse est de comparer les trois séquences codantes des allèles de ces gènes (HLAA0201, HLAB2705, HLACW301). Puisqu'elles n'ont pas la même longueur, on peut utiliser la modalité " **Alignement avec discontinuités** ". La comparaison de ces séquences deux à deux permet de dresser le tableau suivant :

| | | | |
|------------------------|------------|------------|----------|
| HLAA0201 | | | |
| HLAB2705 | | 976 | |
| HLACW301 | 952 | 993 | |
| | HLAA0201 | HLAB2705 | HLACW301 |
| Longueur des séquences | 1098 | 1089 | 1101 |

Plus long, l'alignement simultané des trois séquences donne les résultats suivants :



Comparer les séquences

60 70 80 90 100

Traitement

Identités

hlaa0201.adn

hlab2705.adn

hlacw301.adn

Sélection : 0/5 lignes



L'alignement comprend 1101 bases

-> 919 bases identiques (représentées par le signe *)

soit 83,5 % d'identité

HLAA0201 codant

longueur : 1098 bases (sans compter les discontinuités)

-> référence pour la comparaison

HLAB2705 codant

longueur : 1089 bases (sans compter les discontinuités)

-> 974 bases identiques à la séquence de référence HLAA0201 codant, soit 89,4 % d'identité

HLACW301 codant

longueur : 1101 bases (sans compter les discontinuités)

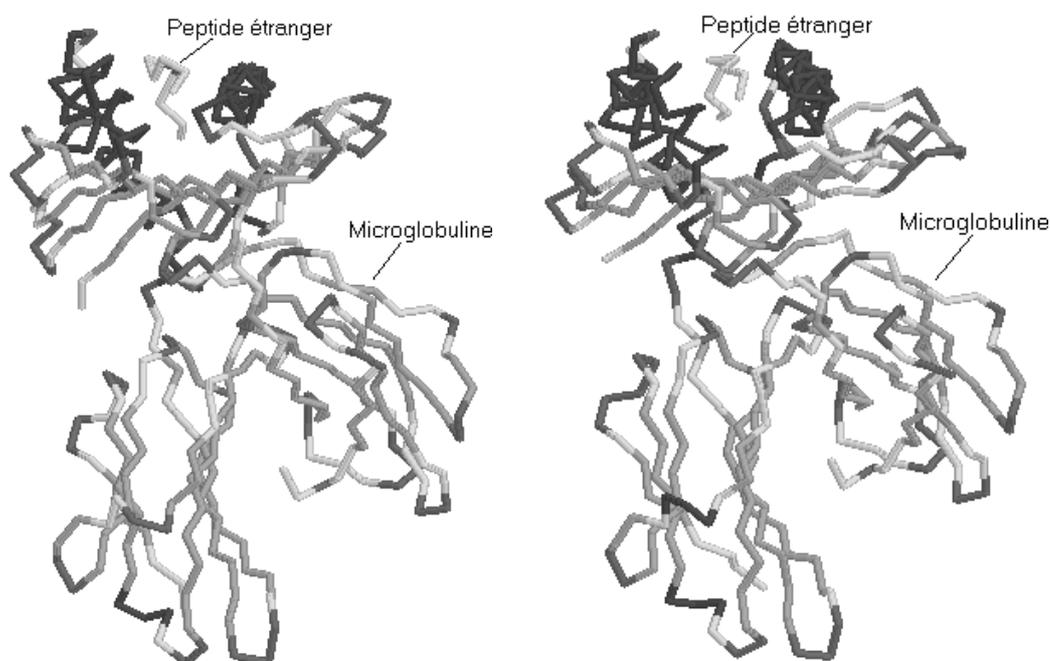
-> 952 bases identiques à la séquence de référence HLAA0201 codant, soit 86,5 % d'identité

Recherche des similitudes entre les séquences protéiques des gènes HLAA, HLAB et HLAC

Un travail similaire peut être réalisé avec les séquences des polypeptides codés par ces allèles, ce qui donne les résultats suivants :

| | | | |
|------------------------|------------|------------|------------|
| HLAA0201 | | | |
| HLAB2705 | | 304 | |
| HLACW301 | 291 | | 305 |
| | HLAA0201 | HLAB2705 | HLACW301 |
| Longueur des séquences | 365 | 362 | 366 |

Malgré les différences dans leurs séquences, HLAA02 (à gauche) et HLAB27 (à droite) ont des structures spatiales similaires, avec une corbeille de présentation de peptides antigéniques.



Les deux molécules, flanquées sur leur droite de la microglobuline, sont montrées en squelette carboné avec le logiciel RASMOL (Roger Sayle, Glaxo Wellcome Research & Development, Stevenage, Hertfordshire, UK). Coordonnées tridimensionnelles provenant de Protein Data Bank.

L'existence d'une famille multigénique étant corroborée, il reste à faire la synthèse sur l'évolution de cette partie du génome : duplication de gènes et nombreuses mutations à chaque locus.



Les hormones hypophysaires et placentaires LH, FSH, TSH et HCG

Informations scientifiques

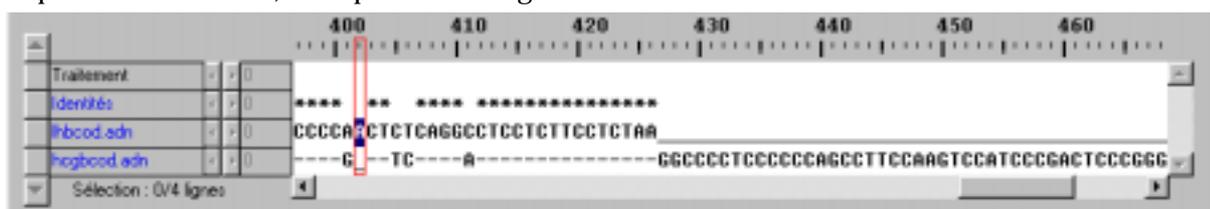
La famille des hormones LH, FSH, TSH et HCG

Ces hormones sont des glycoprotéines formées par l'assemblage de deux chaînes polypeptidiques dénommées alpha et bêta. La chaîne alpha est la même pour les quatre hormones ; celles-ci diffèrent par leurs chaînes bêta. La spécificité fonctionnelle de ces hormones est liée à la chaîne bêta.

Ces chaînes bêta sont codées par des gènes qui possèdent la même structure globale : 3 exons et deux introns, mais sont situés sur des chromosomes différents : chromosome 1 pour TSH bêta, chromosome 11 pour FSH bêta, et chromosome 19 pour LH bêta et HCG bêta.

Les gènes **LH bêta** et **HCG bêta** sont ceux qui présentent le plus de similitude dans leurs séquences et dans celles des polypeptides qu'ils codent. Ainsi, au niveau polypeptidique, la similitude des séquences des chaînes bêta de LH et HCG est de 82 %. L'homologie entre les deux séquences est donc très claire. Cependant, la chaîne bêta de HCG comprend 165 acides aminés (y compris le peptide signal), alors que celle de LH n'en possède que 141 (avec le peptide signal). La différence résulte d'une extension de 24 acides aminés dans la région carboxyterminale de HCG par rapport à LH.

La comparaison des séquences nucléiques permet d'expliquer l'origine de cette extension (utiliser la modalité « alignement avec discontinuités »). Le gène HCG bêta diffère de LH bêta par une délétion qui entraîne un décalage du cadre de lecture et par là un changement dans les sept derniers acides aminés et la disparition du codon STOP (un autre codon STOP apparaît un peu plus loin dans la séquence de HCG bêta). Une partie de la région non traduite de LH bêta l'est dans HCG bêta.



Alignement de HCG β et LH β montrant une délétion dans HCG en position 402 (-A).

Le gène LH bêta existe chez tous les Vertébrés alors que HCG bêta n'est présent que chez quelques ordres de Mammifères placentaires notamment les Primates et les Equidés. On pense donc que le gène HCG bêta provient d'une duplication d'un gène LH ancestral suivie d'une divergence marquée notamment par la délétion. Mais les propriétés des deux hormones, bien que les deux gènes s'expriment dans des tissus différents (hypophyse et placenta), sont restées voisines puisqu'elles agissent sur les mêmes cibles ovariennes.

Les similitudes des chaînes bêta de **FSH** et **TSH** entre elles d'une part, et avec celles de **LH** et **HCG** sont beaucoup moins évidentes sur un plan quantitatif car elles sont comprises entre 25% et 40 % seulement.

L'alignement multiple des séquences polypeptidiques de ces chaînes révèle toutefois la **présence de douze résidus "cystéine"** à des sites identiques. Or, ces résidus, en permettant la formation des ponts disulfure, jouent un rôle essentiel dans la stabilisation de la structure tridimensionnelle de ces chaînes et donc de leurs propriétés.

Ces données qualitatives sont des arguments puissants en faveur de l'idée que les gènes FSH bêta, LH bêta, HCG bêta et TSH bêta forment une famille multigénique. Ils résultent de duplications successives d'un gène ancestral présent chez les premiers Vertébrés. Le caractère remarquable de cette famille multigénique est que la divergence des copies dupliquées a conduit à des gènes codant pour des polypeptides ayant des fonctions différentes.

Les informations obtenues sont :



-> 394 bases identiques (représentées par le signe *)
soit 79,0 % d'identité

LH β codant

longueur : 426 bases (sans compter les discontinuités)

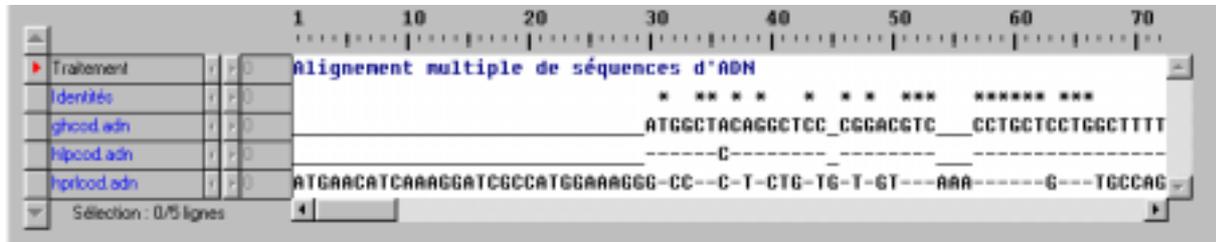
-> référence pour la comparaison

HCG β codant

longueur : 498 bases (sans compter les discontinuités)

-> 394 bases identiques à la séquence de référence LH β codant,
soit 79,1 % d'identité

En revanche, il est difficile de conclure pour le gène de la prolactine. On peut toutefois constater que ce gène présente un peu plus de similitude avec GH et HLP qu'avec les hormones de l'autre famille.



L'alignement comprend 695 bases

-> 291 bases identiques (représentées par le signe *)
soit 41,9 % d'identité

GH codant

longueur : 654 bases (sans compter les discontinuités)

-> référence pour la comparaison

HLP codant

longueur : 654 bases (sans compter les discontinuités)

-> 706 bases identiques à la séquence de référence GH codant,
soit 92,7 % d'identité

HPRL codant

longueur : 684 bases (sans compter les discontinuités)

-> 306 bases identiques à la séquence de référence GH codant,
soit 44,7 % d'identité

le signe - représente les identités

le signe _ représente les discontinuités

Il est possible de réaliser l'alignement multiple des chaînes bêta de FSH, LH, TSH et HCG et repérer l'existence de sites identiques remarquables pour les quatre chaînes (en particulier les cystéines).

| | 95 | 100 | 105 | 110 | 115 |
|------------|--|-----|-----|-----|-----|
| Traitement | | | | | |
| Identités | * | - | - | ; | * |
| hcg.pro | CysProArgGlyValAsnProValValSerTyrAlaValAlaLeuSerCysGlnCysAlaLeuCysArgArg | | | | |
| lh.pro | - - - - Asp- - - - PhePro- - - - Arg- GlyPro- - - | | | | |
| fsh.pro | - AlaHisHisAlaAspSerLeuTyrThr- Pro- - ThrGln- His- GlyLys- AspSer | | | | |
| tsh.pro | - - LeuHis- Ala- TyrPhe- - Pro- - - - Lys- GlyLys- AsnThr | | | | |

Alignement multiple des séquences peptidiques des hormones HCG, LH, FSH et TSH. Quelques-unes des cystéines alignées sont visibles en positions 94, 110, 112 et 115.

Cela conduit à l'idée que la recherche de similitudes repose sur des analyses plus fines que sur le simple dénombrement de sites identiques. Ce sont des analyses de ce type qui amènent les spécialistes à l'idée que le gène HPRL fait partie de la même famille multigénique que GH et HLP. Il reste à envisager la signification des différences de similitude entre gènes faisant partie d'une même famille multigénique pour amorcer une histoire des duplications géniques successives.

En conclusion, cette activité doit permettre une sensibilisation à l'importance évolutive des duplications géniques et à l'idée qu'elles peuvent aboutir à des gènes codant pour des polypeptides aux fonctions différentes.

DÉTERMINATION DES GÉNOTYPES ET PRÉVISION EN GÉNÉTIQUE HUMAINE

Une banque de données géniques sur les individus des arbres généalogiques

Dans les arbres généalogiques relatifs à des phénotypes cliniques, on est amené classiquement à considérer, pour chaque gène en cause, un allèle normal et un allèle muté. À partir de l'analyse de l'arbre, on détermine si les individus sont homozygotes ou hétérozygotes. En réalité, les conclusions tirées de cette analyse au niveau macroscopique sont **hypothétiques** car, comme l'a montré l'analyse des gènes, plusieurs allèles peuvent être à l'origine du phénotype normal ou à l'origine d'un même phénotype clinique, ou encore des allèles de deux gènes différents peuvent entraîner deux phénotypes cliniques voisins. **L'analyse génétique des individus d'un arbre généalogique permet de confirmer ou non les conclusions tirées de l'analyse classique.** C'est ce que nous proposons de faire à partir d'une banque contenant les séquences géniques des individus des arbres généalogiques.

Contenu de la banque

La banque renferme :

- les gènes, ou plus précisément le brin d'ADN non transcrit de chacun d'eux limité à la partie strictement « codante ». Chaque séquence commence donc par le triplet ATG et se termine par un des trois triplets STOP. Pour chaque gène, les *principaux allèles* répertoriés, en particulier tous ceux impliqués dans les arbres généalogiques proposés, sont donnés en référence dans un thème (ALL suivi du nom du gène) ;
- des fichiers d'enzymes de restriction correspondant à chaque gène étudié. Leur utilisation est détaillée dans la deuxième démarche ci-dessous ;
- le génotype (pour le gène étudié) des membres de chaque famille. L'extension de chaque nom de séquence indique le gène en cause. L'individu est repéré dans l'arbre par sa Famille (F1, F2, etc.) sa génération (i, ii, iii, iiiii) et sa position (1, 2, 3,...). Par exemple, F1-i1all1.tyr (all = allèle) et F1-ii1all2.tyr sont dans le premier arbre avec des individus albinos les 2 allèles du gène de la tyrosinase que possède la personne en position 1 à la première génération.

Principe de la recherche du génotype d'un individu d'un arbre généalogique avec le logiciel

L'élève doit utiliser les éléments de cette banque pour déterminer les génotypes des membres (ou de certains membres bien choisis) de la famille étudiée.

Deux démarches sont possibles :

- la première ressemble à l'utilisation vue à propos du polymorphisme génique. Pour comparer les allèles d'un individu avec un allèle choisi comme référence, on utilise les options **Comparaison simple** ou **Alignement avec discontinuités**. Les informations obtenues sont exploitées grâce à un tableau des différences entre l'allèle de référence et les autres allèles ce qui permet d'identifier précisément le génotype des individus. Cette démarche suppose le parcours de l'ensemble des deux séquences comparées qui peuvent être assez longues ;
- la deuxième est basée sur une utilisation simplifiée des méthodes de diagnostic génotypique à l'aide des **enzymes de restriction** : une enzyme de restriction reconnaît un motif spécifique dans une séquence d'ADN et coupe cette séquence au niveau du motif reconnu (ou à proximité). Pour chaque gène considéré, un fichier d'enzymes permettant d'identifier les allèles impliqués dans les arbres généalogiques et une grille de référence montrant l'action de ces enzymes sur tous les allèles de ce gène sont fournis.

Cette deuxième démarche est la seule illustrée dans l'analyse génétique des différents cas d'hérédité humaine envisagés dans cette brochure. L'encart ci-après précise les principes de cette démarche.

Considérons l'exemple théorique suivant :

Un gène G possède 4 allèles GA1, GA2, GA3 et GA4. Les enzymes du fichier d'enzymes (correspondant à G) coupent chaque allèle toutes les fois qu'elles rencontrent leur motif spécifique. Dans le diagnostic génotypique réel, les fragments sont séparés par électrophorèse puis hybridés avec une sonde spécifique et ainsi repérés. Le logiciel, lui, donne pour chaque enzyme la position des sites reconnus et la longueur des fragments obtenus après coupure par l'enzyme. En toute rigueur il faudrait comparer la longueur des fragments pour identifier les allèles les uns par rapport aux autres ; cela serait fort long et l'on peut se contenter du nombre de sites reconnus dans chaque allèle par les enzymes. Le tableau ci-dessous indique, dans notre cas théorique, le nombre de sites obtenus après avoir fait agir séparément 4 enzymes sur les 4 allèles du gène G. Ce tableau sert de référence.

| Allèles | Enzymes | | | |
|---------|---------|----|----|----|
| | E1 | E2 | E3 | E4 |
| GA1 | 2 | 3 | 2 | 2 |
| GA2 | 3 | 3 | 2 | 2 |
| GA3 | 2 | 3 | 3 | - |
| GA4 | 2 | 2 | 2 | - |

La lecture de ce tableau montre que :

- l'enzyme E4 a reconnu sur les allèles GA1 et GA2 2 sites et qu'elle n'a reconnu aucun site sur GA3 et GA4,
- l'enzyme E3 permet d'identifier GA3,
- l'enzyme E2 permet d'identifier GA4,
- l'enzyme E1 permet d'identifier GA2,
- les enzymes E4 et E1 combinées permettent d'identifier GA1.

En soumettant un allèle inconnu à l'action de ces enzymes on voit comment il est découpé par chacune d'entre elles ce qui permet son identification grâce au tableau de référence. Ainsi, si l'enzyme E2 ne reconnaît que deux sites sur cet allèle cela signifie que ce dernier est GA4.

Cette technique suppose que toutes les différences entre les allèles sont repérables grâce au panel d'enzymes de restriction dont on dispose. En réalité, il n'en est pas toujours ainsi. Les séquences des allèles des gènes contenus dans la banque de données concernent uniquement les régions strictement codantes. Lorsque pour un allèle n'existait pas d'enzymes reconnaissant sa spécificité, nous avons choisi de définir des enzymes théoriques qui reconnaissent des motifs permettant de faire le diagnostic de ces allèles. Ces « enzymes » sont appelées ENZINV et concernent uniquement quelques allèles de la tyrosinase impliqués dans l'albinisme.



Hérédité de l'albinisme

Informations scientifiques

Généralités

L'albinisme humain est la plus fréquente des anomalies héréditaires de pigmentation. Le phénotype macroscopique visible (hypopigmentation de la peau, des poils, des yeux) fut pendant longtemps le seul critère retenu pour détecter cette anomalie due à une réduction (ou absence) de synthèse de mélanine. Si cette hypopigmentation est générale, on parle d'albinisme oculo-cutané (OCA). Si elle est localisée aux yeux, on la nomme albinisme oculaire (OA). OCA correspond généralement à de l'hérédité autosomale récessive. OA est lié à un gène situé sur une région propre au chromosome X (Nature genetics, mai 1995).

Le seul critère de la pigmentation est insuffisant pour diagnostiquer l'albinisme. Des changements dans les yeux et le système visuel ont été trouvés dans tous les cas d'albinisme oculo-cutané (OCA) et d'albinisme oculaire (OA). L'iris et la rétine étant dépigmentés, les pupilles apparaissent rouges du fait de la réflexion de la lumière sur la choroïde très riche en vaisseaux sanguins. D'autre part, le développement des voies optiques est altéré : anomalies de decussation des fibres nerveuses au niveau du chiasma optique faisant que le nombre de fibres ipsilatérales est réduit ; anomalie de l'organisation des corps genouillés latéraux. Tout cela fait que beaucoup d'albinos sont photophobes et la plupart présentent un nystagmus résultant d'une hypoplasie fovéale (associée à une réduction de l'acuité visuelle qui peut varier de 0,5 dixième à 4 dixièmes) et aussi de la malformation des voies optiques. Ils ont de plus une diminution de l'acuité stéréoscopique et un strabisme alternant. Chez une personne qui pourrait avoir une pigmentation presque normale, l'examen ophtalmologique complet permet donc de dépister un albinisme.

Dans la suite, seul sera envisagé l'albinisme oculo-cutané (OCA) qui représente les formes les plus fréquentes (1 cas sur 16 000 à 20 000 aux USA).

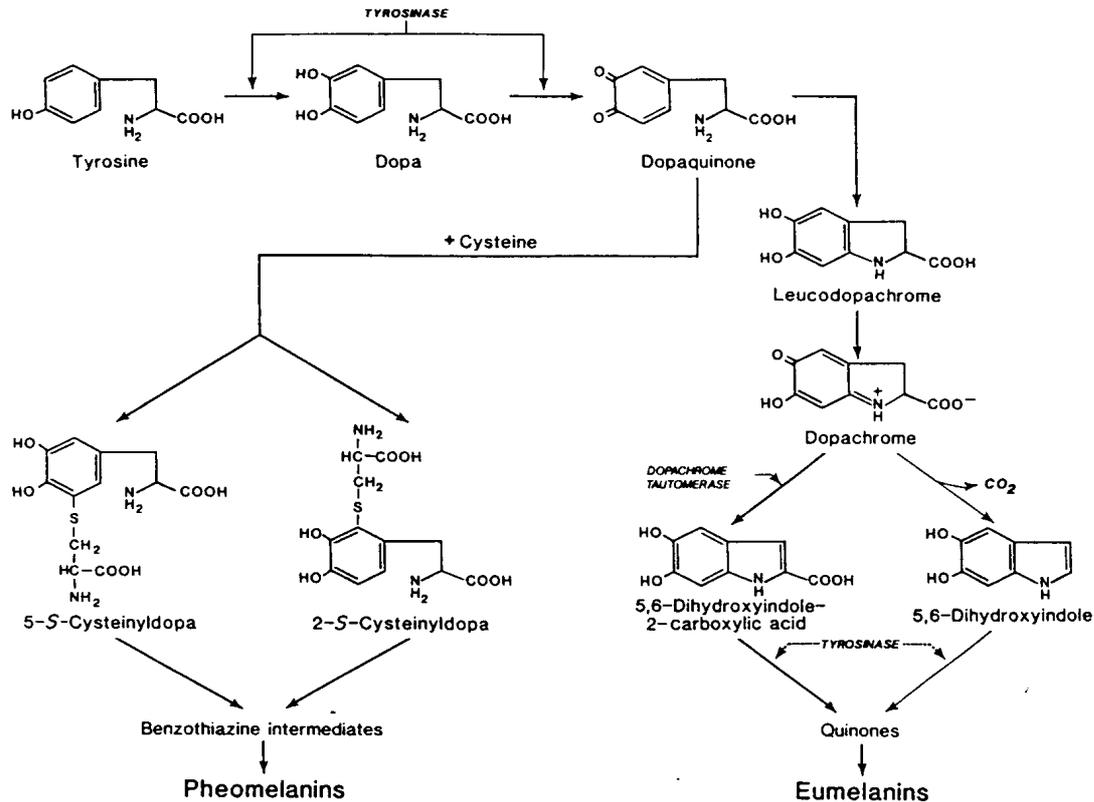
Pigmentation de la peau et synthèse de la mélanine

La pigmentation est un processus complexe qui résulte de la synthèse de pigments spécifiques, les mélanines. Les mélanocytes sont des cellules épidermiques capables de synthétiser des pigments, les mélanines, au sein d'organites spécialisés, les mélanosomes. Ces mélanosomes sont transférés dans les kératinocytes et transportés par eux à la surface de l'épiderme. Les mélanocytes synthétisent la mélanine ; les kératinocytes la stockent et l'éliminent lors de leur desquamation.

La pigmentation résulte donc de processus complexes. Chez les albinos, l'insuffisance de mélanine peut être due soit à une anomalie dans la chaîne de biosynthèse (anomalie enzymatique) soit une anomalie non spécifique à la chaîne de synthèse (par exemple, membrane défectueuse du mélanosome).

La tyrosinase est une enzyme cruciale de la chaîne de biosynthèse des mélanines car elle catalyse 4 réactions : l'hydroxylation de la tyrosine en dopa, l'oxydation de la dopa en dopaquinone et les deux réactions aboutissant aux quinones. La banque renferme uniquement les séquences des allèles du gène qui code pour la tyrosinase. Certains de ces allèles sont à l'origine de l'albinisme dit oculo-cutané de type 1 (OCA-1).

Une autre forme d'albinisme oculo-cutané est due à un ou plusieurs gènes autre (s) que celui qui code pour la tyrosinase. Elle est dite albinisme oculo-cutané de type 2. En 1995, les séquences de ce (s) gènes (s) ne sont pas connues.



Biosynthèse de la mélanine.

L'albinisme oculo-cutané de type 1

Phénotypes cliniques

Tyrosinase négative (cas OCA-1A)

Ce phénotype a été trouvé dans tous les groupes ethniques étudiés à travers le monde. Il est lié à une activité nulle de la tyrosinase. Les enfants naissent avec les cheveux blancs, la peau blanche, les yeux bleus, et ces caractéristiques persistent pendant toute la vie. La peau ne bronze pas et reste très sensible au soleil tout au long de la vie. On peut détecter cette anomalie en pratiquant le test du bulbe pileux qui consiste à incuber des bulbes pileux d'une personne pendant 24 heures à 37°C dans une solution de tyrosine. On constate qu'il n'y a aucune synthèse de mélanine.

Type jaune (cas OCA-1B)

Il est reconnu dans beaucoup de populations et a été découvert dans la population Amish (Etat de l'Indiana aux Etats-Unis). Les enfants naissent avec les cheveux blancs et la peau blanche. Les cheveux se colorent en blond au cours des premières années de la vie et peuvent devenir blond foncé chez l'adulte. La peau devient crème et peut bronzer au soleil. L'iris, bleu gris, peut aussi se pigmenter davantage.

Type thermo-sensible (cas OCA-1TS)

Nées avec une peau blanche, des poils blancs et des yeux bleus, certaines personnes, au moment de la puberté, développent une pigmentation sur les zones les plus froides de leur corps comme les bras et les jambes. Selon la température de l'endroit du corps considéré, on aura une pigmentation plus ou moins développée.

Phénotypes moléculaires

Les allèles du gène

La tyrosinase résulte de l'expression d'un gène situé sur le chromosome 11. Le séquençage du gène a montré que l'allèle normal présentait un polymorphisme : au codon 192 TAT (sérine) ou TCT (tyrosine), avec une fréquence égale chez les « caucasiens » alors que l'on trouve 100% de TCT chez les orientaux ; au codon 402 CGA (arginine) ou CAA (glutamine) avec 100% de CGA pour les orientaux alors que pour les caucasiens on a 85% de CGA et 15% de CAA. Dans la banque, nous n'avons retenu que TAT/TCT, les deux allèles normaux étant appelés TyrCod1 avec TAT et TyrCod2 avec TCT. Le tableau suivant fournit les caractéristiques des allèles à l'origine de l'albinisme retenus dans la banque.

| Séquence nucléique | | | Polypeptide | |
|--------------------------|--|-------------------------------------|-----------------------------------|------------------------|
| Noms des allèles | Nucléotides changés Nature - Position | Codons changés Nature - Position | A.A. changés Nature - Position | Type de mutation |
| TyrCod1 (Référence 1) | A575 | TAT192 | Tyr192 | |
| TyrCod2 (Référence 2) | C575 | TCT192 | Ser192 | Substitution neutre |
| (1) TyrAlba1 | G1147 A | GAT383 AAT | Asp383Asn | Substitution faux sens |
| (2) TyrAlba2 | - TG732-733 | TGT244 TGA | Cys244X | Délétion |
| (2) TyrAlba3 | G533 A | TGG178 TAG | Trp178X | Substitution non sens |
| (1) TyrAlba4 | C242 T | CCT81 CTT | Pro81Leu | Substitution faux sens |
| (2) TyrAlba5 | C242 T | CCT81 CTT | Pro81Leu | Substitution faux sens |
| (2) TyrAlbb1 | C1217 T | CCT406 CTT | Pro406Leu | Substitution faux sens |
| (1) TyrAlbts | G1265 A | CGG422 CAG | Arg422Gln | Substitution faux sens |

Quelques allèles du gène qui code pour la tyrosinase.

Caractéristiques des enzymes codées par ces allèles

Ces mutations déterminent des enzymes ayant des caractéristiques diverses :

- enzymes totalement inactives (à l'origine du phénotype OCA-1A) : ce sont notamment celles codées par les allèles TyrAlba1, TyrAlba2, TyrAlba3, TyrAlba4 et TyrAlba5 dans la banque ;
- enzyme partiellement active (à l'origine du phénotype OCA-1B) : l'allèle muté est appelé TyrAlbb1 dans la banque ; la tyrosinase possède une activité résiduelle et la dopaquinone formée qui a une forte affinité pour les groupements sulfhydryl (présents dans la cystéine) est transformée en phéomélanine (pigment jaunâtre). La voie métabolique vers la phéomélanine est donc privilégiée aux dépens de celle de l'eumélanine (cf. chaîne de biosynthèse) ;
- enzyme thermo-sensible (à l'origine du phénotype OCA-1TS) : l'allèle muté est appelé TyrAlbts dans la banque : la tyrosinase est inactive au dessus de 35°C. La mutation serait semblable à celle qui existe chez le chat siamois.

L'albinisme oculo-cutané de type 2

Il est dit tyrosinase-positif car le test du bulbe pileux est positif, ce qui laisse supposer que le gène de la tyrosinase est normal. Cette forme d'albinisme est la plus commune et la plus variable. Sa distribution est mondiale avec une forte fréquence en Afrique, notamment équatoriale.

Phénotype clinique

Il est globalement moins sévère que l'OCA de type 1. L'hypopigmentation cutanée peut varier suivant l'origine ethnique et l'hétérozygotie éventuelle de l'individu. Cependant, certains caractères paraissent communs : les individus ont des cheveux légèrement pigmentés à la naissance, la peau est blanche, les yeux sont bleus et surtout la pigmentation augmente avec l'âge.

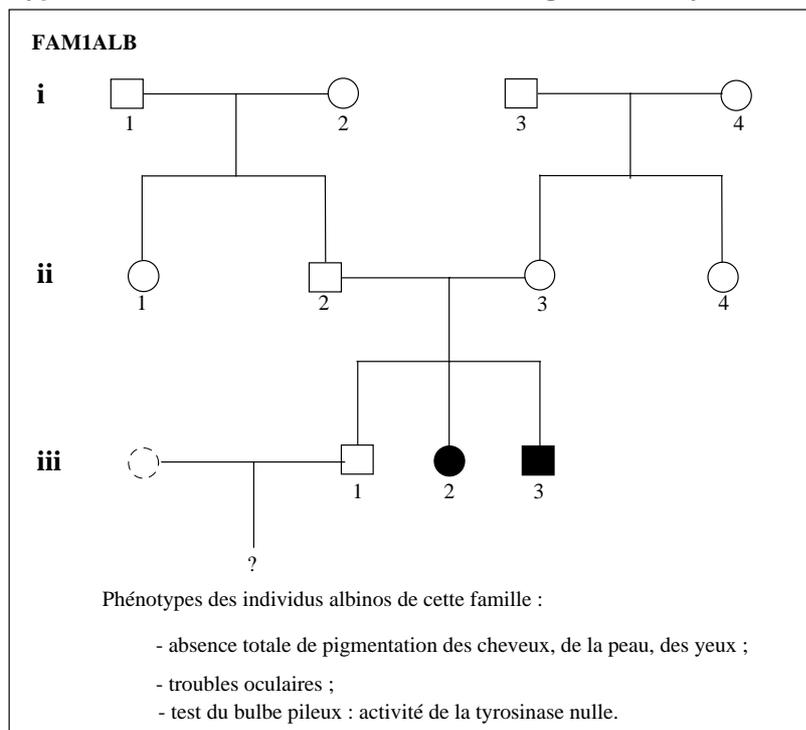
Phénotype moléculaire

Le ou les gènes responsables n'ont pas encore été séquencés et la pathologie moléculaire n'a pas été déterminée.

Utilisations pédagogiques

Deux aspects sont envisagés pour chaque exemple : l'un très sommairement car correspondant à une étude classique en génétique humaine, l'autre plus en détail car reposant sur l'utilisation du logiciel pour effectuer un diagnostic génotypique.

Albinisme de type OCA-1 dû à des allèles différents du gène de la tyrosinase



Famille 1 (à partir de *Am. J. Hum. Gen.*, 1993, V52, p. 406-413).

Analyse classique

Après la présentation des données de base sur la synthèse de mélanine et le rôle joué par la tyrosinase, l'analyse de cet arbre débouche sur :

- le constat qu'il s'agit d'une hérédité autosomale récessive ;
- les génotypes possibles des individus (en utilisant le symbolisme " a+ " pour désigner tout allèle fonctionnel, et " a " pour désigner tout allèle à l'origine de l'albinisme) ;
- le calcul de la probabilité qu'a un couple d'avoir un enfant albinos.

Par exemple, on peut rechercher la probabilité qu'a iii1 d'avoir un enfant albinos sachant que sa femme ne présente aucun antécédent d'albinisme dans sa famille. iii1 est normal mais il a 2 chances sur 3 de porter l'allèle muté responsable de l'albinisme de sa sœur et de son frère. Sachant que dans la population la fréquence des allèles de l'albinisme est de 1%, ce couple a une probabilité de $2/3 \times 1/100 = 2/300$ d'être formé de deux hétérozygotes ; si c'est le cas, ils ont 25% de chances d'avoir un enfant albinos. La probabilité qu'a le couple d'avoir un enfant albinos est donc de : $2/300 \times 1/4 = 1/600$.

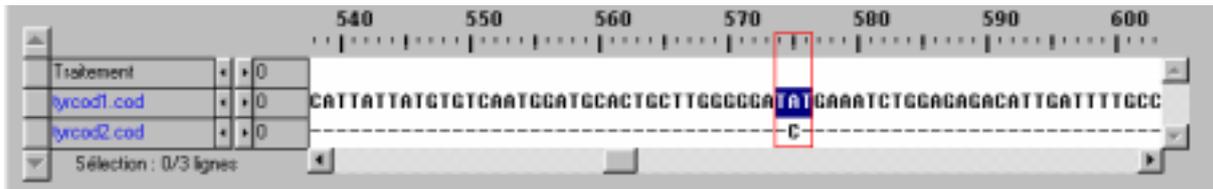
Analyse génotypique

Elle peut permettre de tester les hypothèses émises sur le génotype des individus, par exemple le génotype de iii1, et d'affiner en conséquence le calcul de probabilité de naissance d'un individu albinos.

Reconnaissance des allèles *Tyrcod1* et *Tyrcod2* par la méthode des enzymes de restriction

Dans cet exercice préliminaire, il s'agit d'amener les élèves à s'approprier la technique d'identification d'un allèle grâce aux enzymes de restriction afin qu'ils puissent par la suite utiliser librement le logiciel pour déterminer le génotype d'un individu. Pour cela :

- comparer ces séquences et repérer l'existence d'une seule différence au niveau de la base 575 : **A** dans *Tyrcod1* devient **C** dans *Tyrcod2* ;



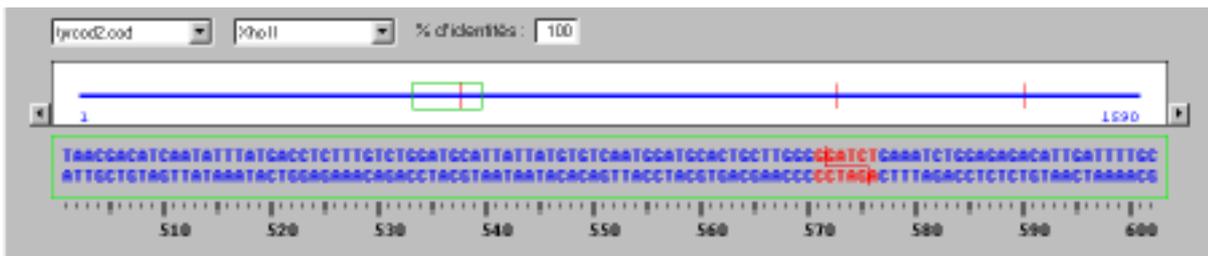
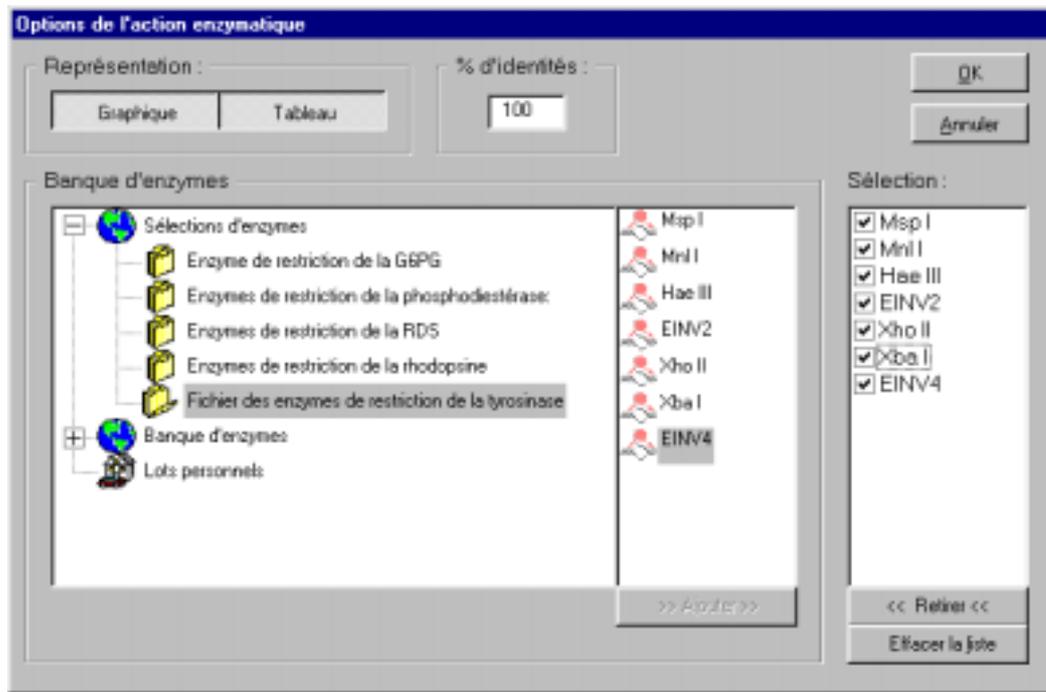
- introduire la notion d'enzyme de restriction et utiliser la banque des enzymes de restriction : repérer dans la liste des enzymes une enzyme susceptible de différencier les deux allèles *Tyrcod1* et *Tyrcod2*. L'élève doit arriver à la conclusion que *XhoII* (GGATCT) coupe *Tyrcod2* dans la région « 575 » mais pas *Tyrcod1* ;

Enzymes du fichier **des enzymes de restriction de la tyrosinase**

| Nom | Motif (s) reconnu (s) |
|---------|-----------------------|
| Msp I | CCGG |
| Mnl I | CCTC |
| Hae III | GGCC |
| EINV2 | CAATC |
| Xho II | GGATCT |
| Xba I | TCTAGA |
| EINV4 | AGTGTC |

- vérifier cette conclusion en faisant agir cette enzyme et d'autres appartenant au fichier des enzymes de restriction de la tyrosinase sur *Tyrcod1* et *Tyrcod2*.

Action enzymatique



Visualisation de l'action enzymatique sous la forme d'un graphique. *XhoII* coupe 2 fois *TyrcoD1* (en 1135 et 1418) et 3 fois *TyrcoD2* (en 571, 1135 et 1418). Le résultat obtenu signifie aussi qu'il y a plusieurs motifs GGATCT dans les séquences.

Pour les autres enzymes, il n'existe aucune différence dans le nombre de coupures entre les deux séquences.

| Enzymes (100%) | Hsp I | Mnl I | Hae III | EINu2 | Xho II | Xba I | EINu4 |
|----------------|-------|-------|---------|-------|--------|-------|-------|
| tyrcod1.cod | 2 | 11 | 7 | 0 | 2 | 0 | 1 |
| tyrcod2.cod | 2 | 11 | 7 | 0 | 3 | 0 | 1 |

Visualisation de l'action enzymatique sous la forme d'un tableau.

Pour bien faire saisir la signification des informations fournies par le logiciel (nombre de sites et dimensions des fragments) on peut leur demander de reporter sur une représentation schématique de chaque allèle les sites de coupure et les longueurs des fragments. Il peut être intéressant de faire l'analogie avec la séparation des fragments de restriction par électrophorèse.

*Reconnaissance de tous les allèles du gène de la tyrosinase ;
construction du tableau de référence*

Pour appliquer cette technique à tous les allèles du gène de la tyrosinase, il convient de les analyser avec le fichier des enzymes de restriction de la tyrosinase. Les élèves doivent obtenir le tableau ci-dessous qui sert de référence pour la détermination des génotypes de tous les individus.

| Enzymes (100%) | Hsp I | Mnl I | Hae III | EINu2 | Xho II | Xba I | EINu4 |
|----------------|-------|-------|---------|-------|--------|-------|-------|
| tyrcod2.cod | 2 | 11 | 7 | 0 | 3 | 0 | 1 |
| tyralba1.cod | 2 | 11 | 7 | 1 | 2 | 0 | 1 |
| tyralba2.cod | 2 | 11 | 7 | 0 | 3 | 0 | 0 |
| tyralba3.cod | 2 | 11 | 7 | 0 | 3 | 1 | 1 |
| tyralba4.cod | 2 | 11 | 6 | 0 | 2 | 0 | 1 |
| tyralba5.cod | 2 | 11 | 6 | 0 | 3 | 0 | 1 |
| tyralbb1.cod | 2 | 10 | 7 | 0 | 3 | 0 | 1 |
| tyralbt.cod | 1 | 11 | 7 | 0 | 2 | 0 | 1 |

Nombre de sites reconnus dans les allèles du gène de la tyrosinase.

À partir de ce tableau, les élèves doivent déterminer l'enzyme ou les enzymes permettant d'identifier tel ou tel allèle. On remarquera que Tyralba4 et Tyralba5 présentent la même mutation à l'origine de l'albinisme mais cette mutation est portée dans un cas par Tyrcod1 et dans l'autre par Tyrcod2.

Détermination des génotypes de l'arbre généalogique

La technique étant acquise, les élèves peuvent maintenant **tester les hypothèses** sur les génotypes et faire des prévisions sur le phénotype d'un descendant.

On peut cibler l'analyse génotypique sur l'individu iii1 prolongeant ainsi les suppositions émises à partir d'une analyse classique de l'arbre généalogique. Pour cela, il convient d'accéder aux séquences de la famille 1. L'analyse des allèles de l'individu iii1 (iii1ALL1.TYR et iii1ALL2.TYR) avec le fichier **enztyr** révèle que cet individu a comme génotype Tyrcod2/Tyrcod2. Il est donc homozygote et ne risque donc pas d'avoir un enfant albinos. Ainsi l'analyse génétique permet d'affiner les prévisions faites sur des bases phénotypiques. On peut aussi chercher à identifier les allèles portés par les individus albinos de cette famille (iii2 et iii3). L'analyse montre qu'ils sont hétérozygotes Tyralba1/Tyralba2 et remet en cause partiellement le génotype a/a établi à partir des phénotypes.

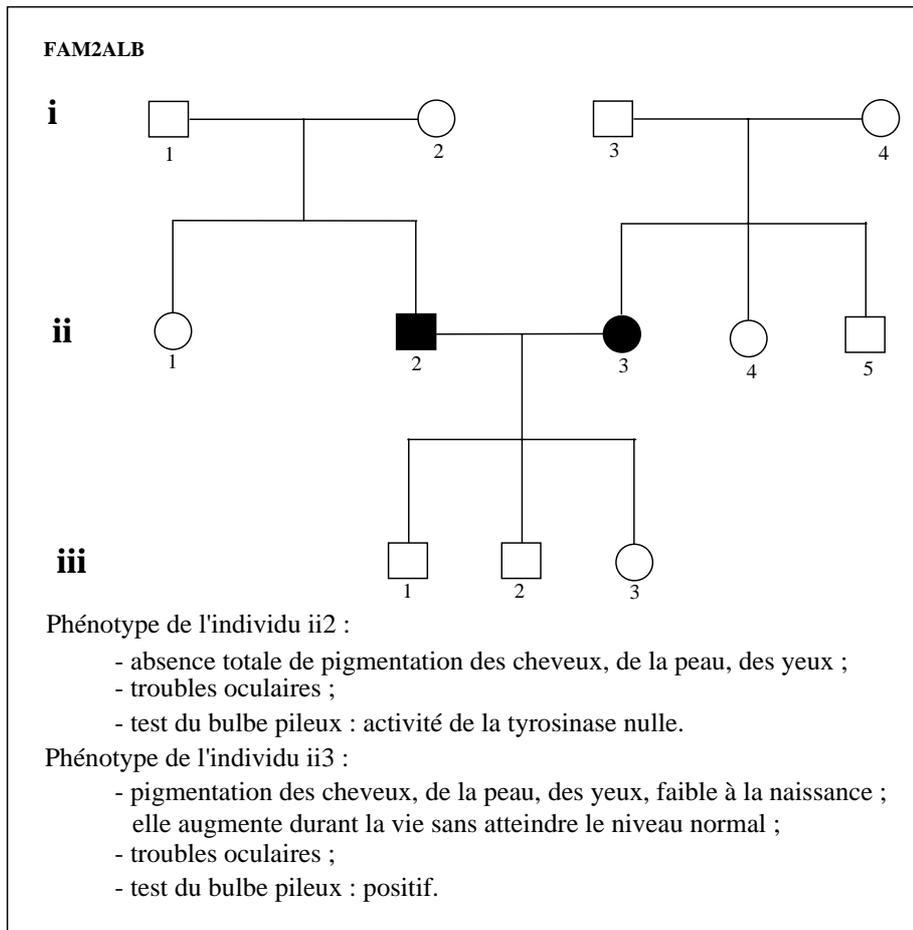
Il reste à préciser les caractéristiques de ces allèles morbides par rapport aux allèles de référence et d'en étudier les conséquences au niveau polypeptidique.

Le tableau ci-dessous renseigne sur les génotypes de tous les individus de la famille FAM1ALB.

| Allèles des individus | Nombre de sites reconnus par | | | | | | | Allèles du gène |
|-----------------------|------------------------------|-------|---------|-------|--------|-------|-------|-----------------|
| | Msp I | Mnl I | Hae III | EINV2 | Xho II | Xba I | EINV4 | |
| F1-i1all1 | 2 | 11 | 7 | | 3 | | - | Tyralba2 |
| F1-i1all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-i2all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-i2all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-i3all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F1-i3all2 | 2 | 11 | 7 | 1 | 2 | | 1 | Tyralba1 |
| F1-i4all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-i4all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F1-ii1all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-ii1all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-ii2all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-ii2all2 | 2 | 11 | 7 | | 3 | | - | Tyralba2 |
| F1-ii3all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-ii3all2 | 2 | 11 | 7 | 1 | 2 | | 1 | Tyralba1 |
| F1-ii4all1 | 2 | 11 | 7 | 1 | 2 | | 1 | Tyralba1 |
| F1-ii4all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F1-iii1all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-iii1all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F1-iii2all1 | 2 | 11 | 7 | | 3 | | - | Tyralba2 |
| F1-iii2all2 | 2 | 11 | 7 | 1 | 2 | | 1 | Tyralba1 |
| F1-iii3all1 | 2 | 11 | 7 | 1 | 2 | | 1 | Tyralba1 |
| F1-iii3all2 | 2 | 11 | 7 | | 3 | | - | Tyralba2 |

Allèles du gène de la tyrosinase chez les individus de FAM1ALB.

Albinisme dû à des gènes différents



Famille 2 (à partir de Brit. J. Ophtal., 1952, V 36, p. 37).

Analyse classique

L'analyse classique doit déboucher sur l'apparente contradiction soulevée par la descendance du couple ii2 et ii3 : alors qu'ils sont albinos et que le phénotype albinos est récessif, leurs trois enfants ont un phénotype non albinos. Les données sur la tyrosinase conduisent à proposer une interprétation de cette contradiction : l'albinisme du père serait dû à des allèles morbides du gène de la tyrosinase et l'albinisme de la mère serait dû à des allèles d'un ou d'autres gènes. Cette hypothèse peut être testée grâce à l'analyse génotypique en utilisant le fichier des enzymes de restriction enztyr.

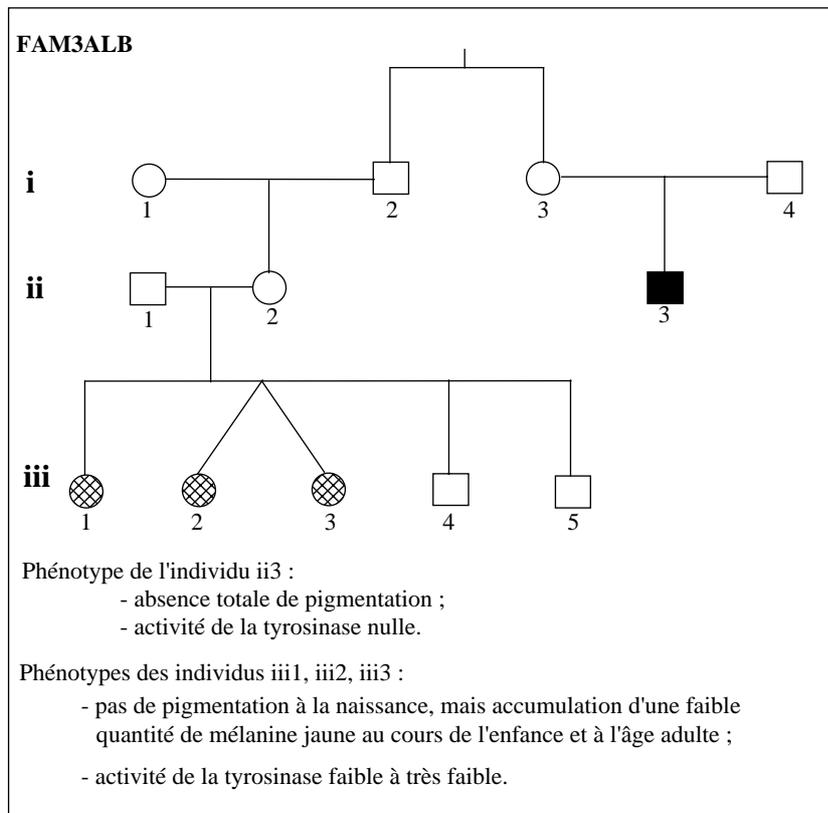
Analyse génotypique

La démarche est identique au cas précédent. L'analyse génotypique confirme cette interprétation puisque l'individu ii2 possède deux allèles morbides de la tyrosinase et ii3 deux allèles fonctionnels Tyr_{cod1}/Tyr_{cod2}. Cette analyse est néanmoins incomplète car on ne dispose pas d'informations sur le ou les autres gènes responsables de l'albinisme de l'individu ii3. On peut aussi s'intéresser au génotype d'autres individus de cette famille pour corroborer cette interprétation et préciser les caractéristiques des allèles morbides.

| Allèles des individus | Nombre de sites reconnus par | | | | | | | Allèles du gène |
|-----------------------|------------------------------|-------|---------|--------|--------|-------|--------|-----------------|
| | Msp I | Mnl I | Hae III | EINV 2 | Xho II | Xba I | EINV 4 | |
| F2-i1all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F2-i1all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F2-i2all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F2-i2all2 | 2 | 11 | 7 | | 3 | 1 | 1 | Tyralba3 |
| F2-i3all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F2-i3all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F2-i4all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F2-i4all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F2-ii1all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F2-ii1all2 | 2 | 11 | 7 | | 3 | 1 | 1 | Tyralba3 |
| F2-ii2all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F2-ii2all2 | 2 | 11 | 7 | | 3 | 1 | 1 | Tyralba3 |
| F2-ii3all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F2-ii3all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F2-ii4all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F2-ii4all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F2-ii5all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F2-ii5all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F2-iii1all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F2-iii1all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod2 |
| F2-iii2all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F2-iii2all2 | 2 | 11 | 7 | | 3 | 1 | 1 | Tyralba3 |
| F2-iii3all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F2-iii3all2 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |

Allèles du gène de la tyrosinase chez les individus de FAM2ALB.

Phénotypes différents de l'albinisme OCA-1 dus à des allèles différents du gène de la tyrosinase



Famille 3 (d'après Am. J. Hum. Gen., 1980, V32, p. 387-395).

Analyse classique

Deux phénotypes albinos sont présents dans l'arbre généalogique de cette famille, l'un traduisant un albinisme oculo-cutané sévère, l'autre marqué par une pigmentation résiduelle augmentant au cours de la vie. L'analyse classique conduit à soulever le problème de l'origine de cette diversité phénotypique : est-elle due à des allèles de deux gènes différents ou des allèles d'un même gène. L'analyse génétique classique n'apporte pas de réponse (il n'y a pas dans la famille une union entre deux albinos de phénotypes différents pouvant révéler une complémentation). Toutefois, puisque le test du bulbe pileux révèle que l'activité de la tyrosinase est affectée chez les deux phénotypes, la deuxième hypothèse est la plus probable.

L'analyse des allèles du gène qui code pour la tyrosinase possédés par les membre de cette famille permet de tester cette dernière hypothèse. S'il en est bien ainsi, les individus à phénotype albinisme moins sévère doivent posséder au moins un allèle à la fois différent des allèles normaux Tyr_{cod1} et Tyr_{cod2} et des allèles morbides de la personne ii3.

Analyse génotypique

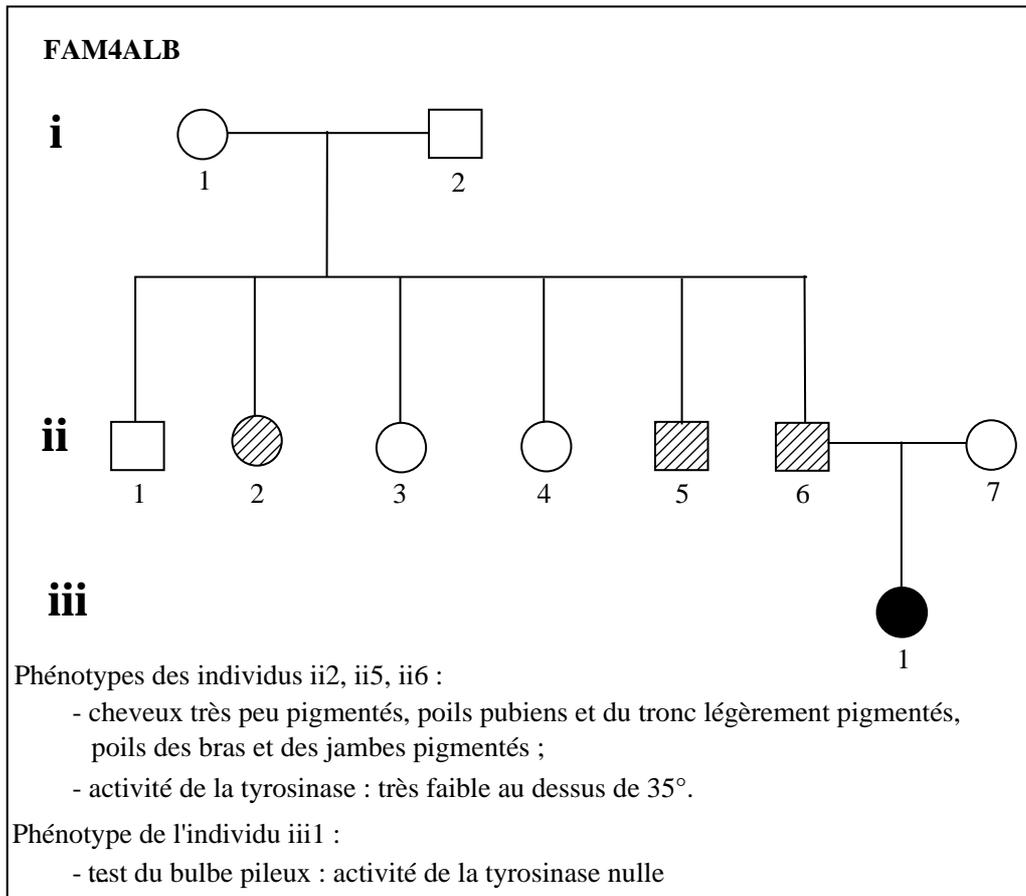
C'est ce que confirme l'analyse génotypique des personnes iii1, iii2, iii3 et ii3 réalisée avec le fichier des enzymes de restriction enztyr. Le tableau résume les différents génotypes des individus de l'arbre.

| Allèles des individus | Nombre de sites reconnus par | | | | | | | Allèles du gène |
|-----------------------|------------------------------|-------|---------|--------|--------|-------|--------|-----------------|
| | Msp I | Mnl I | Hae III | EINV 2 | Xho II | Xba I | EINV 4 | |
| F3-i1all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F3-i1all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F3-i2all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F3-i2all2 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-i3all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-i3all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F3-i4all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F3-i4all2 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-ii1all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F3-ii1all2 | 2 | 10 | 7 | | 3 | | 1 | Tyralbb1 |
| F3-ii2all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F3-ii2all2 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-ii3all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-ii3all2 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-iii1all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-iii1all2 | 2 | 10 | 7 | | 3 | | 1 | Tyralbb1 |
| F3-iii2all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-iii2all2 | 2 | 10 | 7 | | 3 | | 1 | Tyralbb1 |
| F3-iii3all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F3-iii3all2 | 2 | 10 | 7 | | 3 | | 1 | Tyralbb1 |
| F3-iii4all1 | 2 | 10 | 7 | | 3 | | 1 | Tyralbb1 |
| F3-iii4all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F3-iii5all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F3-iii5all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |

Allèles du gène de la tyrosinase chez les individus de FAM3ALB.

La personne ii3 est homozygote Tyralba4/Tyralba4. Les individus iii1, iii2 et iii3 ont hérité de leur mère, cousine germaine de ii3, l'allèle Tyralba4. Ils possèdent en commun l'allèle Tyralbb1 qui provient de leur père. Comme leur phénotype est différent de celui de ii3, cette différence ne peut être attribuée qu'à la présence de l'allèle Tyralbb1 ; cela corrobore l'idée que cet allèle doit coder pour une tyrosinase à activité réduite à l'origine de leur faible pigmentation. Tyralba4 et Tyralbb1 sont tous les deux des allèles résultant d'une substitution faux sens qui entraîne le remplacement d'un résidu proline par de la leucine, l'une en position 81, l'autre en 406, occasion de rappeler que les conséquences d'une mutation au niveau polypeptidique dépendent non seulement de la nature de la substitution mais aussi de son emplacement.

Phénotype albinos température dépendant



Famille 4 (d'après J. Clin. Inv., 1991, V87, p. 1119-1122).

Analyse classique

L'analyse de cet arbre généalogique enrichit le spectre des phénotypes de l'albinisme. Les données laissent à penser que le phénotype composite observé est dû à un allèle codant pour une tyrosinase dont l'activité dépend de la température. La naissance d'un enfant albinos iii1 à partir du couple ii6 et ii7 peut paraître étonnante. Elle signifie en effet que iii1 n'a reçu ni l'allèle fonctionnel de sa mère ni un allèle de tyrosinase thermo-dépendante de son père. Cela ne se comprend que si ii6 et ii7 possèdent tous deux un allèle morbide codant pour une tyrosinase à activité nulle.

Analyse génotypique

La détermination des génotypes des individus cités permet de corroborer ces hypothèses.

| Allèles des individus | Nombre de sites reconnus par | | | | | | | Allèles du gène |
|-----------------------|------------------------------|-------|---------|--------|--------|-------|--------|-----------------|
| | Msp I | Mnl I | Hae III | EINV 2 | Xho II | Xba I | EINV 4 | |
| F4-i1all1 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F4-i1all2 | 1 | 11 | 7 | | 2 | | 1 | Tyralbts |
| F4-i2all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F4-i2all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F4-ii1all1 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F4-ii1all2 | 2 | 11 | 7 | | 3 | | 1 | Tyrcod2 |
| F4-ii2all1 | 1 | 11 | 7 | | 2 | | 1 | Tyralbts |
| F4-ii2all2 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F4-ii3all1 | 1 | 11 | 7 | | 2 | | 1 | Tyralbts |
| F4-ii3all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F4-ii4all1 | 1 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F4-ii4all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F4-ii5all1 | 1 | 11 | 7 | | 2 | | 1 | Tyralbts |
| F4-ii5all2 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F4-ii6all1 | 1 | 11 | 7 | | 2 | | 1 | Tyralbts |
| F4-ii6all2 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F4-ii7all1 | 2 | 11 | 6 | | 3 | | 1 | Tyralba5 |
| F4-ii7all2 | 2 | 11 | 7 | | 2 | | 1 | Tyrcod1 |
| F4-iii1all1 | 2 | 11 | 6 | | 2 | | 1 | Tyralba4 |
| F4-iii1all2 | 2 | 11 | 6 | | 3 | | 1 | Tyralba5 |

Allèles du gène de la tyrosinase chez les individus de FAM4ALB.

Conclusions

Ces arbres généalogiques sur l'albinisme permettent de faire percevoir la complexité des maladies héréditaires traduite par une hétérogénéité phénotypique. Cette dernière est due à une hétérogénéité génétique liée à des allèles différents d'un même gène ou due à des allèles de gènes occupant des loci différents. Cette complexité génétique fait saisir toute la difficulté du diagnostic génétique. Les données de la banque et les fonctions du logiciel aident à comprendre les principes d'utilisation des enzymes de restriction pour la détermination du génotype et montrent tout l'intérêt dans le domaine prévisionnel de la détermination des génotypes par rapport à l'analyse classique.



Hérédité des rétinites pigmentaires

Informations scientifiques

Maladies héréditaires, les rétinites pigmentaires sont la cause principale des malvoyances (30% des déficiences visuelles). La maladie débute par une atteinte de la vision nocturne avec réduction du champ visuel causée par la dégénérescence progressive des bâtonnets. La dégénérescence de la rétine s'accroît et finit par atteindre la région de la fovéa (cellules en cônes) ce qui entraîne une baisse de l'acuité visuelle et une cécité presque complète.

Le phénotype rétinite pigmentaire est très hétérogène sur le plan clinique, se traduisant par une variabilité de la précocité des premiers signes et de la durée d'extension de la maladie. Cette diversité phénotypique a pour support une diversité génétique liée au fait que plusieurs gènes peuvent être impliqués (trois gènes ont été clairement identifiés) et qu'il existe pour chacun de ces gènes plusieurs allèles. Les gènes connus à l'origine des rétinites codent pour des protéines intervenant dans des processus de phototransduction (genèse d'un potentiel de récepteur à partir d'un stimulus lumineux). Ce sont les gènes qui codent pour :

- la rhodopsine : pigment situé dans la membrane des disques des segments externes des bâtonnets et qui capte les photons ;
- la périphérine : protéine de la membrane des disques des bâtonnets responsable du maintien de la structure des disques membranaires ;
- la phosphodiesterase : enzyme directement impliquée dans la genèse du potentiel de récepteur après absorption de photons par la rhodopsine.

L'analyse génétique a montré que certaines formes de rétinite sont dues à d'autres gènes que les trois cités ; ainsi, 15% des rétinites pigmentaires sont dues à des gènes situés sur le chromosome X. En fonction du mode de transmission, on reconnaît les rétinites pigmentaires dominantes, les rétinites pigmentaires autosomales récessives et les rétinites récessives liées au sexe.

Gènes des rétinites autosomales dominantes

Gène de la rhodopsine situé sur le chromosome 3

Plusieurs allèles de ce gène sont à l'origine de **rétinites sévères** apparaissant au cours de la **première décennie** et entraînant des atteintes graves de la fonction visuelle vers la fin de la deuxième décennie, la gravité variant selon la nature de l'allèle.

| Séquence nucléique | | | Polypeptide | |
|--------------------|--|-------------------------------------|-----------------------------------|------------------------|
| Noms des allèles | Nucléotides changés Nature - Position | Codons changés Nature - Position | A.A. changés Nature - Position | Type de mutation |
| Rhonorm | (Référence) | (Référence) | (Référence) | (Référence) |
| Rhoret1 | C68 →A | CCC23 →CAC | Pro23His | Substitution faux sens |
| Rhoret2 | C1040 →T | CCG347 →CTG | Pro347Leu | Substitution faux sens |
| Rhoret3 | GG404-405 →TT | CGG135 →CTT | Arg135Leu | Substitution faux sens |
| Rhoret4 | C403 →T | CGG135 →TGG | Arg135Trp | Substitution faux sens |

| | | | | |
|---------|-------------------------------|------------------------------|---------------------------|---------------------------|
| Rhoret5 | - CTGCGCACGCCT 202 →213 | - CTG68 CGC69 ACG70 CCT71 | -Leu68Arg69 Thr70Pro71 | Délétion |
| Rhoret6 | A886 →G | AAG296 →GAG | Lys296GLU | Substitution faux sens |
| Rhoret7 | - ATC763-764-765 | - ATC255 | - Ile255 | Délétion |
| Rhoret8 | C403 →G | CGG135→GGG | Arg135Gly | Substitution faux sens |
| Rhoret9 | G745 →T | GAG249 →TAG | Glu249X | Substitution non sens |

Caractéristiques de quelques allèles du gène de la rhodopsine.

Ce tableau résume les caractéristiques des principaux allèles morbides du gène de la rhodopsine qui résultent de mutations diverses (substitutions faux sens, non sens, délétions). L'incorporation des rhodopsines mutées dans la membrane, au cours du renouvellement des disques, est plus ou moins entravée selon les mutations, ce qui entraîne une dégénérescence plus ou moins rapide des photorécepteurs par des mécanismes non encore élucidés.

Gène de la périphérine

Situé sur le chromosome 6, ce gène s'exprime dans les cellules en cônes et en bâtonnets. Plusieurs allèles morbides de ce gène ont été identifiés.

| Séquence nucléique | | | Polypeptide | |
|--------------------|--|-------------------------------------|-----------------------------------|---------------------------|
| Noms des allèles | Nucléotides changés Nature - Position | Codons changés Nature - Position | A.A. changés Nature - Position | Type de mutation |
| Pernorm | (Référence) | (Référence) | (Référence) | (Référence) |
| Perret1 | -TGC355-356-357 | -TGC119 | - Cys119 | Délétion |
| Perret2 | -TG73-74 | TGG25 →GTT..... TGA43 | Trp25Val..... 43X | Délétion |
| Perret3 | C647 →T | CCT216 → CTT | Pro216Leu | Substitution faux sens |
| Perret4 | - CCA655-656- 657 | - CCA219 | - Pro219 | Délétion |
| Perret5 | T554 →C | CTG185 →CCG | Leu185Pro | Substitution faux sens |

Caractéristiques de quelques allèles du gène de la périphérine.

Ces allèles sont à l'origine de rétinites pigmentaires modérées qui se manifestent au cours de la deuxième décennie. L'acuité visuelle est généralement conservée pendant 20 à 30 ans d'évolution car la dégénérescence des cellules visuelles est lente (plus de 40 ans).

Gène des rétinites autosomales récessives

- Gène de la sous-unité bêta de la phosphodiesterase (PDE du guanine monophosphate cyclique - GMPc)

Ce gène est situé sur le chromosome 4. Plusieurs allèles morbides ont été identifiés : les mutations affectent l'activité catalytique de l'enzyme et empêchent la genèse du potentiel de récepteur.

| Séquence nucléique | | | Polypeptide | |
|--------------------|--|-------------------------------------|-----------------------------------|------------------------|
| Noms des allèles | Nucléotides changés Nature - Position | Codons changés Nature - Position | A.A. changés Nature - Position | Type de mutation |
| Pdenorm | (Référence) | (Référence) | (Référence) | |
| Pderet1 | C892 →T | CAG298 →TAG | Gln298X | Substitution non sens |
| Pderet2 | C1591 →T | CGA531 →TGA | Arg531X | Substitution non sens |
| Pderet3 | - C1486 → | CCC497CCA.....TGA574 | Thr497Pro.... X574 | Délétion décalante |
| Pderet4 | C1669 →T | CAC557 →TAC | His557Tyr | Substitution faux sens |

Caractéristiques de quelques allèles du gène de la PDE.

Utilisations pédagogiques

La banque de données contient, pour les trois gènes, les séquences relatives aux génotypes des individus appartenant à 5 arbres généalogiques. Cette richesse permet des démarches très diverses recoupant plus ou moins celles vues dans l'étude de l'albinisme. Néanmoins, cette banque a surtout été conçue en vue de mettre les élèves en situation d'autonomie au cours d'un travail de recherche impliquant le réinvestissement de méthodes de raisonnement rencontrées précédemment. **Le problème à résoudre consiste à identifier le ou les gènes pouvant être à l'origine des rétinites pigmentaires des différentes familles généalogiques.** La recherche peut être répartie entre plusieurs groupes d'élèves, chaque groupe s'occupant d'un ou de deux arbres généalogiques. Les informations suivantes peuvent être fournies au préalable :

- les caractéristiques des phénotypes rétinites pigmentaires ;
- les noms et localisations des gènes connus pouvant être à l'origine de ce phénotype, avec indication succincte du rôle des protéines qu'ils codent ;
- la dominance ou récessivité de ces gènes ; certaines formes de rétinites récessives sont dues à un (ou des) gène (s) situé (s) sur un chromosome X mais non encore identifié (s) ;
- les tableaux de référence permettant l'identification de chaque allèle grâce à un fichier d'enzymes de restriction propre à chaque gène (**enzrho**, **enzpde**, **enzper**).

| Enzymes du fichier enzrho.zym | |
|--------------------------------------|----------------------------|
| Nom | Motif(s) reconnu(s) |
| Alu I | AGCT |
| Hpa II | CCGG |
| Mnl I | CCTC,GAGG |
| Fsp I | TGCGCA |

| | Alu I | Hpa II | Mnl I | Fsp I |
|------------------------|--------------|---------------|--------------|--------------|
| Rhonorm (Référence) | 2 | 4 | 12 | 1 |
| Rhoret2 | 2 | 3 | 12 | 1 |
| Rhoret3 | 3 | 4 | 12 | 1 |
| Rhoret5 | 2 | 4 | 11 | - |

Nombre de sites reconnus dans quelques allèles du gène qui code pour la rhodopsine.

| Enzymes du fichier enzpde.zym | |
|--------------------------------------|----------------------------|
| Nom | Motif(s) reconnu(s) |
| Mae I | CTAG |
| Ava II | GGACC,GGTCC |

| | Mae I | Ava II |
|------------------------|--------------|---------------|
| Pdenorm (Référence) | - | 10 |
| Pderet1 | 1 | 10 |
| Pderet2 | - | 9 |

Nombre de sites reconnus dans quelques allèles du gène qui code pour la PDE.

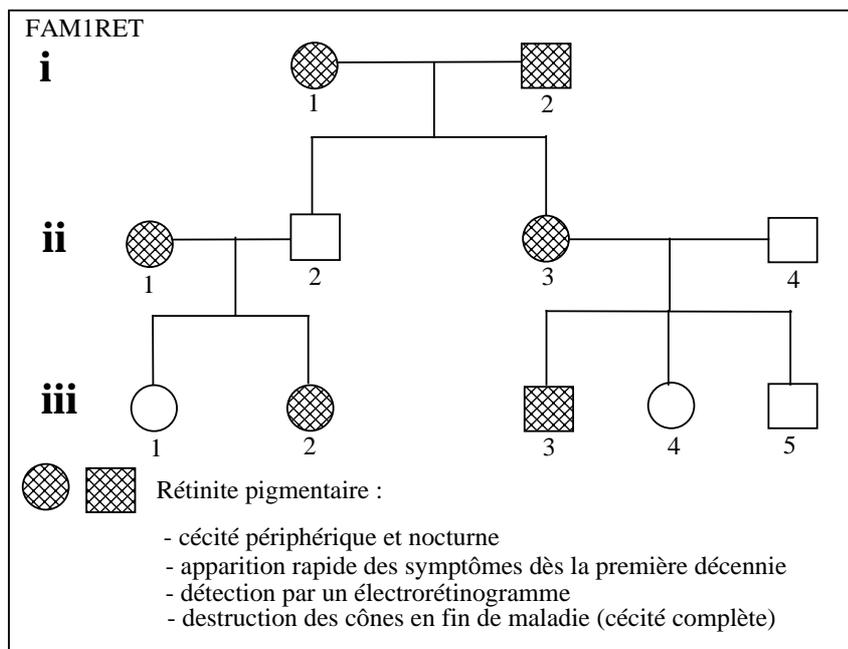
| Enzymes du fichier enzper.zym | |
|--------------------------------------|-------------|
| Nom reconnu(s) | Motif(s) |
| Bbv I | GCAGC,GCTGC |

| | |
|---------------------|-------|
| | Bbv I |
| Pernorm (Référence) | 10 |
| Perret1 | 9 |

Nombre de sites reconnus dans deux allèles du gène de la périphérine.

Munis de ces renseignements, les élèves doivent exploiter les informations tirées de l'analyse des arbres généalogiques pour argumenter sur le ou les gènes en cause possibles et tester leurs suppositions par la détermination des génotypes d'individus bien choisis.

Rétinite due à des allèles du gène de la rhodopsine



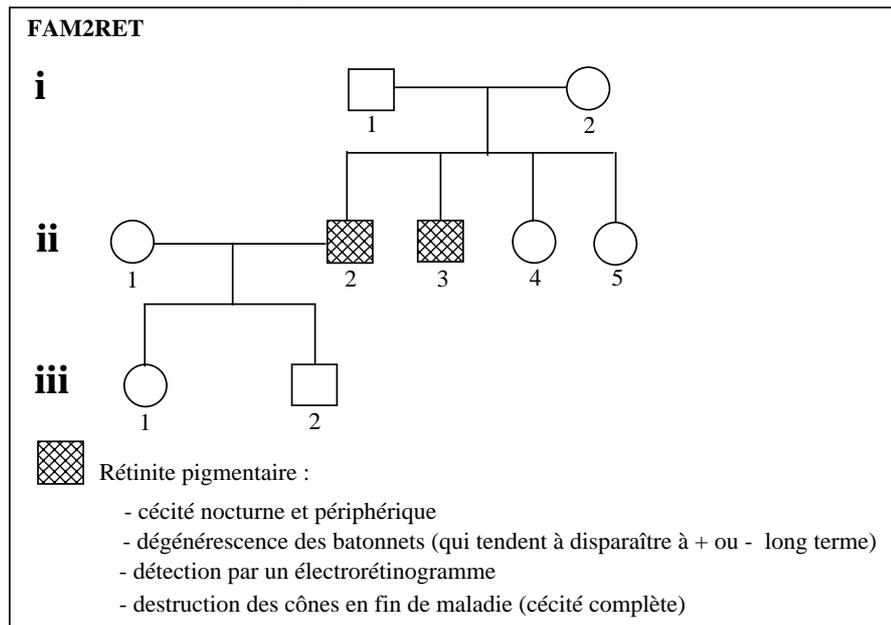
Famille 1 (d'après Genomics, 11, 1991).

De l'analyse de cet arbre l'élève doit conclure que le phénotype rétinite pigmentaire est dominant et que la transmission peut s'expliquer par un gène autosomal ou par un gène porté par X. Etant donné les informations fournies au préalable, il doit songer soit au gène de la rhodopsine soit à celui de la périphérine, mais plus probablement au gène de la rhodopsine vu la gravité des symptômes dans cette famille. En identifiant les génotypes des individus malades, il constate que tous ces individus possèdent des allèles normaux du gène de la pésiphérine ; en revanche, ils ont pour le gène de la rhodopsine un allèle connu comme étant à l'origine d'une rétinite et un allèle normal.

On pourrait prolonger l'étude en demandant en quoi les données sur les génotypes de cet arbre confirment que les allèles Rhoret3 et Rhoret5 sont bien causes de rétinites : on ne les retrouve pas chez les individus de la famille à phénotype normal.

| Allèles des individus | Nombre de sites reconnus par | | | | Allèles du gène |
|-----------------------|------------------------------|--------|-------|-------|-----------------|
| | Alu I | Hpa II | Mnl I | Fsp I | |
| F1-i1all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-i1all2 | 3 | 4 | 12 | 1 | Rhoret3 |
| F1-i2all1 | 2 | 4 | 11 | - | Rhoret5 |
| F1-i2all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-ii1all1 | 3 | 4 | 12 | 1 | Rhoret3 |
| F1-ii1all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-ii2all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-ii2all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-ii3all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-ii3all2 | 2 | 4 | 11 | - | Rhoret5 |
| F1-ii4all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-ii4all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-iii1all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-iii1all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-iii2all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-iii2all2 | 3 | 4 | 12 | 1 | Rhoret3 |
| F1-iii3all1 | 2 | 4 | 11 | - | Rhoret5 |
| F1-iii3all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-iii4all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-iii4all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-iii5all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F1-iii5all2 | 2 | 4 | 12 | 1 | Rhonorm |

Allèles du gène de la rhodopsine chez les individus de la famille FAM1RET.

Rétinite due à des allèles du gène de la phosphodiesterase*Famille 2 (d'après Nature Genetics, vol. 4, juin 1993).*

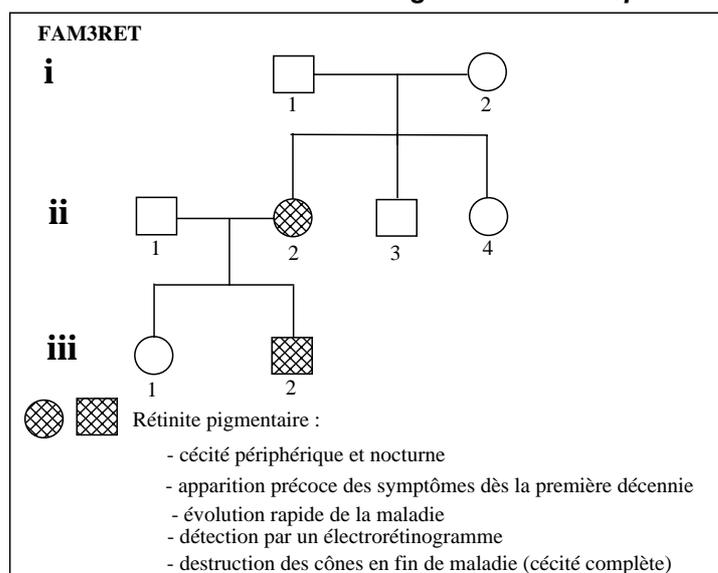
L'analyse classique conduit à conclure à un cas d'hérédité autosomale récessive ou à un cas d'hérédité liée au sexe. D'après les données fournies, c'est plutôt le gène de la phosphodiesterase qui pourrait être impliqué. La détermination des génotypes de ii2 et ii3 montre qu'ils possèdent deux allèles morbides du gène de la PDE. On peut poursuivre l'étude en mettant en évidence les conséquences des mutations sur les polypeptides : dans les deux cas, les polypeptides sont tronqués (cf. tableau des caractéristiques des allèles de ce gène).

| Allèles des individus | Nombre sites reconnus par | | Allèles du gène |
|-----------------------|---------------------------|-------|-----------------|
| | Ava II | Mae I | |
| F2-i1all1 | - | 10 | Pdenorm |
| F2-i1all2 | 1 | 10 | Pderet1 |
| F2-i2all1 | - | 9 | Pderet2 |
| F2-i2all2 | - | 10 | Pdenorm |
| F2-ii1all1 | - | 10 | Pdenorm |
| F2-ii1all2 | - | 10 | Pdenorm |
| F2-ii2all1 | 1 | 10 | Pderet1 |
| F2-ii2all2 | - | 9 | Pderet2 |
| F2-ii3all1 | - | 9 | Pderet2 |
| F2-ii3all2 | 1 | 10 | Pderet1 |
| F2-ii4all1 | - | 10 | Pdenorm |
| F2-ii4all2 | 1 | 10 | Pderet1 |
| F2-ii5all1 | - | 9 | Pderet2 |

| | | | |
|-------------|---|----|---------|
| F2-ii5all2 | - | 10 | Pdenorm |
| F2-iii1all1 | - | 10 | Pdenorm |
| F2-iii1all2 | 1 | 10 | Pderet1 |
| F2-iii2all1 | - | 10 | Pdenorm |
| F2-iii2all2 | 1 | 10 | Pderet1 |

Allèles du gène de la PDE chez les individus de la famille FAM2RET.

Rétinite résultant d'une néomutation dans le gène de la rhodopsine



Famille 3 (d'après Proc. Nat. Acad. Sci. USA, vol. 88, oct. 1991).

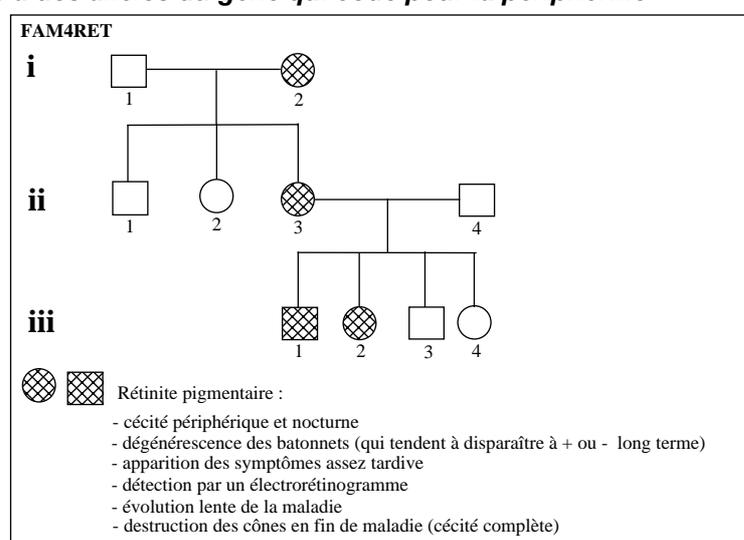
L'analyse classique laisse penser à un cas d'hérédité récessive et autosomale ce qui implique que ii1 soit hétérozygote comme i1 et i2. D'après les informations dont on dispose, on peut soupçonner le gène de la phosphodiésterase. L'analyse génotypique concernant ii2 et iii2 indique qu'ils possèdent 2 allèles normaux de ce gène. Pour rechercher d'autres solutions, l'élève peut examiner le génotype de ii2 dans les autres gènes disponibles et découvrir qu'il possède un allèle morbide du gène de la rhodopsine. Cette conclusion est en contradiction avec la récessivité apparente de la transmission de la rétinite dans cette famille. En analysant les génotypes de i1 et i2, l'élève doit conclure que la rétinite de ii2 s'explique par une mutation nouvelle lors de la formation des gamètes de i1 ou i2.

| Allèles des individus | Nombre de sites reconnus par | | | | Allèles du gène |
|-----------------------|------------------------------|--------|-------|-------|-----------------|
| | Alu I | Hpa II | Mnl I | Fsp I | |
| F3-i1all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-i1all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-i2all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-i2all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-ii1all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-ii1all2 | 2 | 4 | 12 | 1 | Rhonorm |

| | | | | | |
|-------------|---|---|----|---|---------|
| F3-ii2all1 | 2 | 3 | 12 | 1 | Rhoret2 |
| F3-ii2all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-ii3all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-ii3all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-ii4all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-ii4all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-iii1all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-iii1all2 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-iii2all1 | 2 | 4 | 12 | 1 | Rhonorm |
| F3-iii2all2 | 2 | 3 | 12 | 1 | Rhoret2 |

Allèles du gène de la rhodopsine chez les individus de la famille FAM3RET.

Rétinite due à des allèles du gène qui code pour la périphérine



Famille 4 (d'après Nature Genetics, vol. 3, mars 1993)

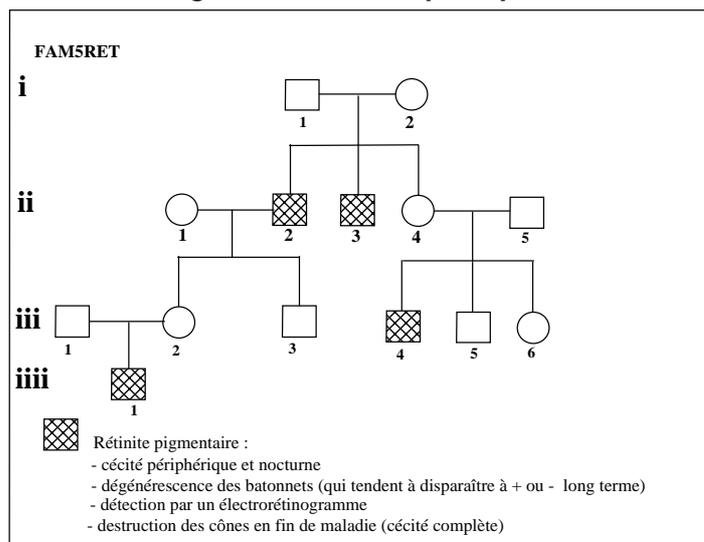
L'analyse classique de cet arbre exclut une hérédité récessive liée au sexe ; elle laisse possible, avec une probabilité faible, une hérédité autosomale récessive. Les interprétations les plus probables sont soit une hérédité autosomale dominante, soit une hérédité liée au sexe dominante. Les symptômes orientent plutôt vers la périphérine. L'analyse génotypique corrobore ce point de vue : tous les individus s'avèrent normaux pour la rhodopsine.

| Allèles des individus | Nombre de sites reconnus par Bbv I | Allèles du gène |
|-----------------------|------------------------------------|-----------------|
| F4-i1all1 | 10 | Pernorm |
| F4-i1all2 | 10 | Pernorm |
| F4-i2all1 | 10 | Pernorm |
| F4-i2all2 | 9 | Perret1 |
| F4-ii1all1 | 10 | Pernorm |

| | | |
|-------------|----|---------|
| F4-ii1all2 | 10 | Pernorm |
| F4-ii2all1 | 10 | Pernorm |
| F4-ii2all2 | 10 | Pernorm |
| F4-ii3all1 | 9 | Perret1 |
| F4-ii3all2 | 10 | Pernorm |
| F4-ii4all1 | 10 | Pernorm |
| F4-ii4all2 | 10 | Pernorm |
| F4-iii1all1 | 10 | Pernorm |
| F4-iii1all2 | 9 | Perret1 |
| F4-iii2all1 | 9 | Perret1 |
| F4-iii2all2 | 10 | Pernorm |
| F4-iii3all1 | 10 | Pernorm |
| F4-iii3all2 | 10 | Pernorm |
| F4-iii4all1 | 10 | Pernorm |
| F4-iii4all2 | 10 | Pernorm |

Allèles du gène de la périphérine chez les individus de la famille FAM4RET.

Rétinite liée à un allèle d'un gène non identifié porté par un chromosome X



Famille 5 avec rétinite.

L'analyse classique conduit à une hérédité récessive soit autosomale soit liée au sexe. On peut suspecter le gène de la phosphodiesterase ; par ailleurs, l'apparition d'une néomutation expliquant le phénotype malade chez deux individus de la deuxième génération peut être exclue (très faible probabilité de deux mutations).

L'analyse génétique des individus malades infirme l'hypothèse du gène de la phosphodiesterase. Les autres gènes pourraient-ils avoir des allèles récessifs à l'origine de rétinites ? L'analyse génétique des individus malades montre qu'ils possèdent des allèles normaux pour les gènes de la rhodopsine et pour celui de la périphérine.

La seule conclusion possible est que cette rétinite est due à un ou des allèles d'un gène autre que ceux de la rhodopsine, de la péréphérine et de la phosphodiesterase. Ce gène est probablement situé sur le chromosome X étant donné l'aspect de l'arbre généalogique.

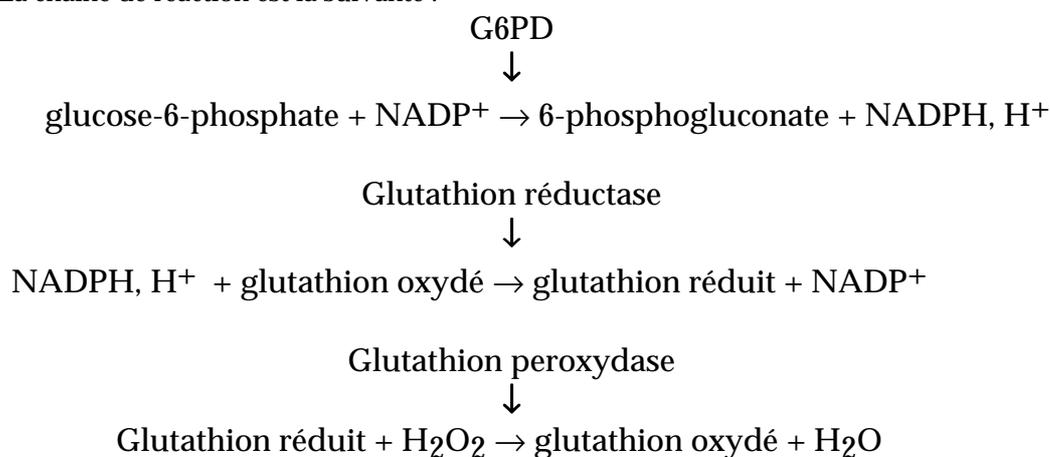


Les déficiences en glucose 6 phosphate déshydrogénase (G6PD)

Données scientifiques

Rôle de l'enzyme G6PD

C'est une enzyme cytoplasmique présente dans toutes les cellules. Elle catalyse la première réaction de la voie des pentoses phosphates. Cette autre voie du catabolisme glucidique produit du ribose 5 phosphate (qui servira ultérieurement à la synthèse des nucléotides) et du NADPH, coenzyme qui est le principal donneur d'hydrogène dans de nombreuses réactions de biosynthèse. NADPH est aussi indispensable pour que se réalise la destruction du peroxyde d'hydrogène hautement toxique pour la cellule. La chaîne de réaction est la suivante :



La catalase catalyse la réaction : $\text{H}_2\text{O}_2 \rightarrow \text{H}_2\text{O} + 1/2 \text{O}_2$

Ces quelques réactions nous montrent que pour la destruction du peroxyde d'hydrogène (H_2O_2), la catalase et le glutathion sont indispensables. Avec une G6PD inactive ou de très peu active, il y a un arrêt de production de NADPH par la voie des pentoses phosphates. Cela empêche la réduction du glutathion et par là la destruction de H_2O_2 . On sait d'autre part que le NADPH « stabilise » la catalase. Donc sans NADPH, H_2O_2 ne sera pas détruit et la cellule sera tuée. Dans les globules rouges cette situation est d'autant plus dramatique que d'autres enzymes permettant la production de NADPH manquent.

Le polymorphisme du gène G6PD

Le gène codant pour la G6PD est situé dans la partie télomérique du bras long du chromosome X ; il est formé par 13 exons et mesure 18 kilo paires de bases environ. Toutefois, sa région codante ne comprend que 1545 paires de bases ce qui correspond à une protéine enzymatique formée par 515 acides aminés. On connaît de très nombreux allèles (plus d'une centaine), dont certains ont une fréquence supérieure à 1%.

| Allèles | Fréquence | | | Activité enzymatique (% par rapport au normal) | Manifestations cliniques |
|----------|-----------|--------|--------------|---|---|
| | Afrique | Europe | Méditerranée | | |
| G6pdb | 65 % | 99,7 % | 90-99 % | 100 | Aucune |
| G6pda | 20 % | | < 1 % | 85 | Aucune |
| G6pda-1 | - | - | - | 12 | Jaunisse néo-natale ; anémie hémolytique aiguë (médicaments, infection) |
| G6pda-2 | 15 % | - | < 12 % | 12 | Jaunisse néo-natale ; anémie hémolytique aiguë (médicaments, infection) |
| G6pda-3 | - | - | - | 12 | Jaunisse néo-natale ; anémie hémolytique aiguë (médicaments, infection) |
| G6pdm | < 0,1 % | - | 1 - 8 % | 3 | Jaunisse néo-natale ; anémie hémolytique aiguë (médicaments, ingestion de fèves, infection) |
| G6pdseat | - | - | - | 25 | Rares |

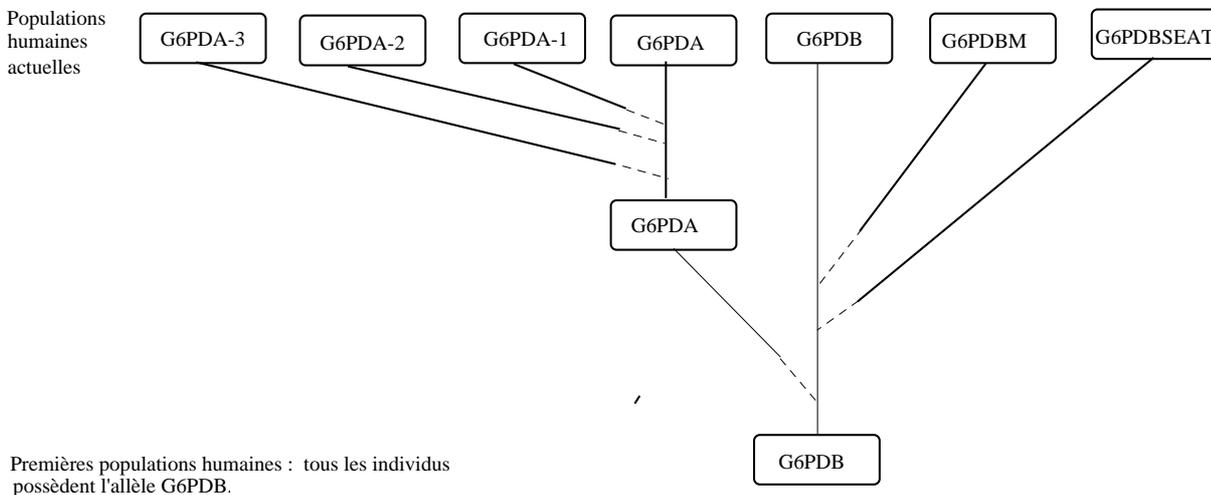
Fréquence de quelques allèles de la G6PD contenus dans la banque.

| Séquence nucléique | | | Polypeptide | |
|--------------------|--|-------------------------------------|-----------------------------------|------------------------|
| Noms des allèles | Nucléotides changés Nature - Position | Codons changés Nature - Position | A.A. changés Nature - Position | Type de mutation |
| G6pdb | (Référence) | (Référence) | (Référence) | |
| G6pda | A376 G | AAT126 GAT | Asn126Asp | Substitution faux sens |
| G6pda-1 | A376 G | AAT126 GAT | Asn126Asp | Substitution faux sens |
| | G202 A | GTG68 ATG | Val68Met | Substitution faux sens |
| G6pda-2 | A376 G | AAT GAT | Asn126Asp | Substitution faux sens |
| | G680 T | CGC227 CTC | Arg227Leu | Substitution faux sens |
| G6pda-3 | A376 G | AAT126 GAT | Asn126Asp | Substitution faux sens |
| | T968 C | CTG323 CCG | Leu323Pro | Substitution faux sens |
| G6pdm | C563 T | TCC188 TTC | Ser188Phe | Substitution faux sens |
| G6pdseat | G844 C | GAT282 CAT | Asp282His | Substitution faux sens |

Caractéristiques de quelques allèles du gène de la G6PD.

Les allèles de ce gène peuvent donc servir de support pour étudier la notion de polymorphisme génique. Certes, les allèles diffèrent uniquement par des substitutions faux sens et ne permettent pas d'illustrer les diverses modalités des différences alléliques (les allèles résultant de mutations non sens, d'insertions et délétions décalantes codant pour des protéines tronquées totalement non fonctionnelles sont sans doute rapidement éliminées par la sélection naturelle). Mais ce système allélique est probablement le meilleur outil pour faire saisir la micro évolution d'un gène et sa diversification au cours de l'évolution des populations humaines.

L'allèle G6pdb, le plus fréquent dans toutes les populations, est sans doute l'allèle ancestral (c'est aussi le plus proche de celui séquencé chez le Chimpanzé). L'allèle G6pda, répandu en Afrique, résulte d'une substitution au nucléotide 376 de la région codante (A376G) dans l'allèle G6pdb. Les trois allèles G6pda⁻ présentent cette même différence avec G6pdb plus une autre : tous les trois résultent de mutations survenues au cours de la gaméto-genèse d'individus G6pda et sont donc apparus postérieurement à cet allèle. Par contre, l'allèle G6pdm diffère de G6pdb par une substitution autre que celle trouvée dans G6pda : il provient d'une mutation intervenue chez un individu G6pdb mais on ne peut situer chronologiquement sa formation par rapport à G6pda.



Filiation possible entre les allèles de G6PD. Les pointillés indiquent seulement que le moment et l'ordre d'apparition des allèles par mutation au cours de l'histoire de l'humanité sont inconnus.

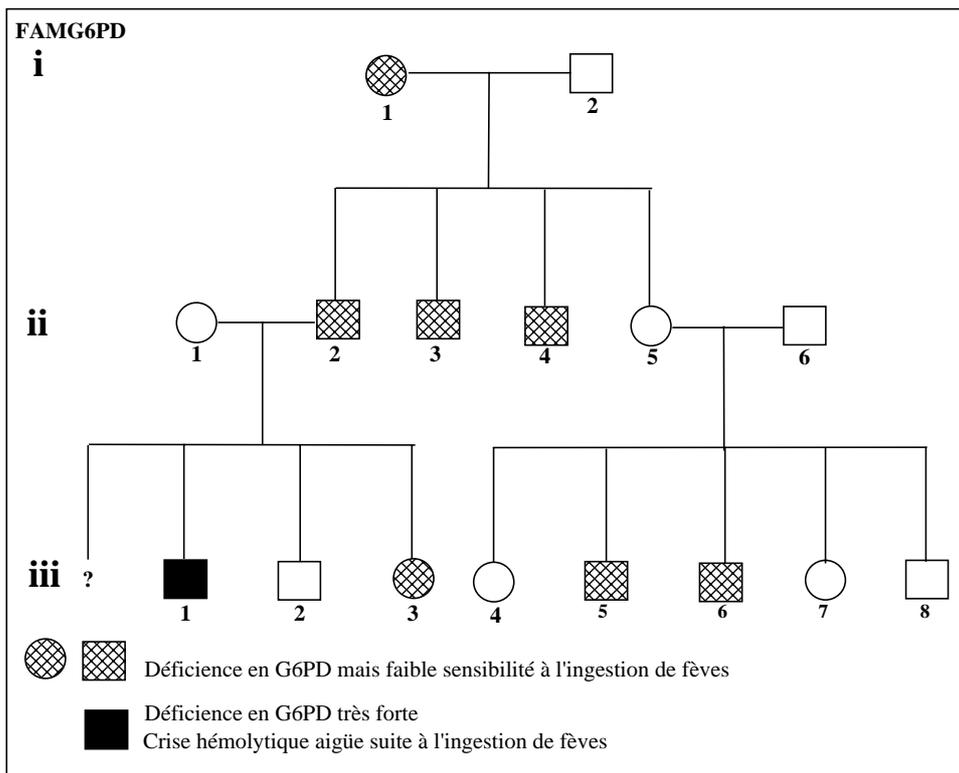
La déficience en G6PD

La déficience en G6PD est l'enzymopathie la plus répandue : elle affecterait 400 millions de personnes dans le monde. Les régions les plus touchées sont l'Afrique tropicale, le moyen Orient, l'Asie tropicale et subtropicale. Un certain nombre d'allèles codent pour une enzyme G6PD déficiente. La déficience n'est jamais totale : l'absence d'enzyme G6PD est sans doute incompatible avec la vie. Les manifestations cliniques sont la jaunisse néonatale, une anémie hémolytique, et dans des cas sévères des séquelles neurologiques. Des crises aiguës d'anémie hémolytique peuvent être déclenchées par des infections, des ingestions de fèves et divers médicaments (comme la primaquine). Heureusement, seule une faible proportion des malades déficients en G6PD présentent une anémie hémolytique chronique et pour les autres, en dehors des crises hémolytiques, il n'y a aucun symptôme particulier. L'action favorisante de l'ingestion de fèves sur le déclenchement des crises hémolytiques est surtout nette chez les personnes possédant l'allèle G6pdm.

Le phénotype des femmes hétérozygotes possédant un allèle « normal » et un allèle « déficient »

Il est classique de considérer comme récessif le phénotype G6PD déficient et c'est ce que traduit l'arbre généalogique soumis à l'analyse. En réalité, la situation est plus complexe et le phénotype des femmes hétérozygotes variable, certaines pouvant manifester des signes cliniques de déficience. Cela est en relation avec l'inactivation au hasard d'un des chromosomes X dans chacune des cellules de l'organisme, inactivation qui peut atteindre l'un ou l'autre des chromosomes X. La femme hétérozygote possède deux populations d'hématies, l'une G6PD déficiente, l'autre avec une enzyme G6PD efficace. L'importance relative de ces deux populations varie d'une femme à l'autre.

Utilisations pédagogiques en génétique humaine



Famille avec déficit en G6PD.

L'étude du polymorphisme génique à partir de G6PD peut être conduite en adoptant une démarche voisine de celle suivie pour l'alpha-antitrypsine.

Analyse classique

L'analyse de l'arbre généalogique conduit à la récessivité du phénotype G6PD déficient. En revanche, les deux interprétations, hérédité autosomale récessive ou hérédité récessive liée au sexe, sont plausibles même si la seconde apparaît plus probable.

L'interprétation suivant laquelle le gène serait situé sur un autosome implique que les individus non apparentés ii2, ii1 et ii6 soient hétérozygotes ; l'interprétation suivant laquelle l'hérédité serait liée au sexe nécessite seulement que la femme ii1 soit hétérozygote. En outre, la différence entre les deux phénotypes G6PD déficients des personnes iii1 et iii3 doit conduire à supposer que des allèles G6PD déficients différents existent dans cette famille.

Analyse génotypique

La détermination des génotypes des individus de l'arbre réalisée grâce à l'utilisation des enzymes de restriction (fichier **enzg6pd.zym**) permet de tester ces hypothèses.

| Enzymes du fichier enzg6pd.zym | |
|---------------------------------------|---------------------|
| Nom | Motif(s) reconnu(s) |
| Nla III | CATG |
| Fok I | GGATG,CATCC |
| Mbo II | GAAGA,TCTTC |

| Nla III | Fok I | Mbo II | Allèles reconnus |
|---------|-------|--------|------------------|
| 7 | 8 | 10 | G6pda |
| 8 | 8 | 10 | G6pda-1 |
| 7 | 7 | 10 | G6pdb |
| 7 | 7 | 11 | G6pdm |

Allèles de la G6PD reconnus par les enzymes de restriction.

| Allèles des individus | Nombre de sites reconnus par | | | Allèles du gène |
|-----------------------|------------------------------|-------|--------|-----------------|
| | Nla III | Fok I | Mbo II | |
| F1-i1all1 | 8 | 8 | 10 | G6pda-1 |
| F1-i1all2 | 8 | 8 | 10 | G6pda-1 |
| F1-i2all | 7 | 7 | 10 | G6pdb |
| F1-ii1all1 | 7 | 7 | 10 | G6pdb |
| F1-ii1all2 | 7 | 7 | 11 | G6pdm |
| F1-ii2all | 8 | 8 | 10 | G6pda-1 |
| F1-ii3all | 8 | 8 | 10 | G6pda-1 |
| F1-ii4all | 8 | 8 | 10 | G6pda-1 |
| F1-ii5all1 | 8 | 8 | 10 | G6pda-1 |
| F1-ii5all2 | 7 | 7 | 10 | G6pdb |
| F1-ii6all | 7 | 8 | 10 | G6pda |
| F1-iii1all | 7 | 7 | 11 | G6pdm |
| F1-iii2all | 7 | 7 | 10 | G6pdb |
| F1-iii3all1 | 8 | 8 | 10 | G6pda-1 |
| F1-iii3all2 | 7 | 7 | 11 | G6pdm |
| F1-iii4all1 | 7 | 7 | 10 | G6pdb |
| F1-iii4all2 | 7 | 8 | 10 | G6pda |
| F1-iii5all | 8 | 8 | 10 | G6pda-1 |
| F1-iii6all | 8 | 8 | 10 | G6pda-1 |
| F1-iii7all1 | 8 | 8 | 10 | G6pda-1 |
| F1-iii7all2 | 7 | 8 | 10 | G6pda |
| F1-iii8all | 7 | 7 | 10 | G6pdb |

Allèles du gène de la G6PD chez les individus de FAMG6PD.

L'existence d'un seul allèle G6PD dans la banque pour les hommes et de deux allèles pour les femmes doit conduire les élèves à bien s'approprier la notion d'hérédité liée au sexe. La détermination du génotype de iii1 met bien en valeur que dans un cas d'hérédité liée au sexe, le chromosome X d'un garçon et les allèles des gènes qu'il porte proviennent de sa mère.

BIBLIOGRAPHIE

- Alberts B. et al. : *Biologie Moléculaire de la Cellule*, Interéditions, 1988.
- Albin R., Tagle D. : *Genetics and molecular biology of huntington's disease*. Trends in Neurosciences, vol. 18, n° 1, 1995, p. 11-14.
- Antonarakis S.E., Kazazian H.H., Orkin S.H. : *DNA polymorphism and molecular pathology of the human globin gene cluster*, Human Genetics, n° 19, 1985, p. 1-14.
- Bishop M. J., Rawlings C. J. (eds) : *Nucleic Acid and Protein Sequence Analysis. A Practical Approach*, IRL Press, 1987.
- Bouchard B. : *Les protéines de régulation de la mélanogenèse. Génétique moléculaire des albinismes cutanés*. Médecine Science, vol. 9, 1993, p. 425-430.
- CELL, vol. 21, octobre 1980 :
- Barallel F.E. et al. : *The primary structure of the human ϵ globin gene*, p. 621-626.
 - Efstadiadis A. et al. : *The structure and evolution of the human β globin gene family*, p. 653-668.
 - Lawn R.M. et al. : *The nucleotide sequence of the human β globin gene*, p. 647-651.
 - Slightom J.L. et al. : *Human fetal $G\gamma$ et $A\gamma$ globin genes*, p. 627-638.
 - Spritz R.A. et al. : *Complete nucleotide sequence of the human δ globin gene*, p. 639-646.
- Colombani J. : *HLA - Fonctions immunitaires et applications médicales*. J. Libbey, Eurotext, 1992.
- Crystal R. et al. : *Molecular basis of alpha-1-Antitrypsin deficiency*. Am. J. of Med., vol 84 (suppl. 6A), juin 1988.
- Crystal R. et al. : *The alpha-1-antitrypsin gene and its mutations*. Chest, vol. 95, 1, 1989, p. 196-208.
- Doolittle R.F. (ed.) : *Molecular Evolution : Computer Analysis of Protein and Nucleic Acid Sequences*, Methods in Enzymology, Vol 183, Academic Press Inc., 1990.
- Doolittle R.F. (ed.) : *Of URFS and ORFS. A Primer on how to analyse derived Amino Acid Sequences*, Oxford University Press, 1986.
- Gribskon M., Devereux J. : *Sequence Analysis Primer*, Macmillan, Stockton Press, 1991.
- Hakomori S. : *Les rôles cachés des groupes sanguins*, La Recherche, mai 1993, p. 548-554.
- Heijne G. Von : *Sequence Analysis in Molecular Biology. Treasure Trove or Trivial Pursuit*, Academic Press Inc. 1987.
- Higgins D.G., Bleasby A.J., Fuchs R. : *CLUSTAL V: improved software for multiple sequence alignment*, Comput. Appl. Biosci. 8, p. 189-191, 1992.
- Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G., Gibson T.J. : *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*, Nucleic Acids Res. 22, p. 4673-4680, 1994.
- Humphries et al. : *On the molecular genetics of retinites pigmentosa*. Science, vol 256, mai 1992, p. 804-808.
- Kaplan J.-C., Dollfus H. : *Coup d'œil sur la complexité (à propos des rétinites)*. In *L'explosion de la génétique humaine*. Science et Vie hors-série, 1992.
- Kaplan J.-C., Delpech M. : *Biologie moléculaire et médecine*, Flammarion, 1992.

Kazazian H.H. et al : *Comparison of deficiency alleles of the β globin and factor VIII*, Cold Spring Harbor symposium, 1986.

Lesk A.M. (Ed.) : *Computational Molecular Biology. Sources and Methods for Sequence Analysis*, Oxford University Press, 1988.

Lesk A.M. : *Protein architecture. A practical approach*, IRL Press, 1991.

Liehaber S.A. et al. : *Cloning and complete nucleotide sequence of human 5' α globin gene*, Proc. Nat. Acad. Sci. USA, vol. 77, 12, 1980, p. 7054-7058.

Lopez de Castro J. : *HLA-B27 and HLA-A2 subtypes : structure, evolution and function*. Immunology Today, vol. 10, n° 7, 1989.

Maillet-Baron L., Soussi T. : *Séquençage des acides nucléiques*, Lavoisier, Tec./Doc, 1992.

Orkin S.H., Kazazian H.H. : *The human β globin gene*, Ann. Rev. Gen, 18, 1984, p. 131-171.

Rosa J. : *Un modèle en recherche clinique : la drépanocytose*, Médecine/Sciences, 2, 1986, p. 558-567.

Singer M., Berg P. : *Gènes et Génome*, Vigot, 1992.

Spritz R. : *Molecular genetics of oculocutaneous albinism*. Seminars in Dermatology, vol. 12, n° 3, 1993, p. 167-172.

The Huntington's Disease Collaborative Research Group : *A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes*. Cell, vol. 72, 1993, p. 971-983.

Thomas M.A. : *Les thalassémies*, Bull. de l'APBG, 4, 1989, p. 793-800.

Vulliamy T., Mason P., Luzzato L. : *The molecular basis of glucose-6-phosphate deshydrogenase deficiency*. Trends in Genetics, vol. 8, n°4, 1992.

Watson J. et al. : *Biologie moléculaire du gène*, Interéditions, 1990.

Yamamoto S. et al. : *Molecular genetic basis of the histo-blood group ABO system*, Nature, vol. 345, mai 1990, p. 229-233.

Yamamoto S., Hakomori S. : *Cloning and characterisation of DNA complementary to human UDP-GalNAcTransferase mRNA*, Biol. Chem., 265, n° 2, 1990, p. 1146-1151.

Publications disponibles à l'INRP

J.-C. Hervé, N. Salamé, B. Therrié : *Analyse de séquences de gènes et de protéines avec le logiciel SEQAID II*. INRP, 1993.

J.-C. Hervé, N. Salamé, B. Therrié : *SEQAID II : Thèmes d'étude en génétique humaine*. INRP, 1995.

Documents disponibles au CNDP

Cassette vidéo, *La drépanocytose, une maladie héréditaire*, CNDP, 1992, VHS 27 min. Référence 002 P6254.

Cédérom, *Infogène, ressources pour l'étude de l'information génétique*, CNDP, 2002. Référence 755 A0300

INDEX DE LA BANQUE DE SÉQUENCES

LES CHAÎNES DE L'HÉMOGLOBINE

Alpha

| | |
|--------------|---|
| alpha.adn | <i>Gène (complet) de la globine alpha humaine (brin non transcrit)</i> |
| alphadnc.adn | <i>ADNc (correspondant à l'ARNm complet) du gène codant pour la globine alpha humaine</i> |
| alphacod.adn | <i>Partie strictement codante du gène de la globine alpha humaine (brin non transcrit)</i> |
| alpha.pro | <i>Séquence protéique de la globine alpha humaine</i> |
| alphacod.arn | <i>ARNm correspondant à la partie strictement codante du gène de la globine alpha humaine</i> |
| alpharnm.arn | <i>ARNm complet du gène codant pour la globine alpha humaine</i> |

Bêta

Séquences normales

| | |
|-------------|---|
| beta.adn | <i>Gène (complet) de la globine bêta humaine (brin non transcrit)</i> |
| betadnc.adn | <i>ADNc (correspondant à l'ARNm complet) du gène codant pour la globine bêta humaine</i> |
| betacod.adn | <i>Partie strictement codante du gène codant pour la globine bêta humaine</i> |
| beta.pro | <i>Séquence protéique de la globine bêta humaine</i> |
| betacod.arn | <i>ARNm correspondant à la partie strictement codante du gène codant pour la globine bêta humaine</i> |
| betarnm.arn | <i>ARNm complet du gène codant pour la globine bêta humaine</i> |
| betavar.adn | <i>Variant normal (partie strictement codante) du gène codant pour la globine bêta humaine</i> |

Séquences mutées

Drépanocytose

| | |
|-------------|--|
| drepcod.adn | <i>Partie codante de l'allèle du gène bêta de la globine humaine responsable de la drépanocytose</i> |
| drepcod.arn | <i>ARNm correspondant à la partie strictement codante de l'allèle du gène qui code pour la globine bêta humaine responsable de de la drépanocytose</i> |
| drep.pro | <i>Séquence protéique résultant de l'expression de l'allèle du gène codant pour la globine bêta humaine responsable de la drépanocytose</i> |

Hémoglobinoses C

| | |
|-------------|---|
| hemccod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable de l'hémoglobinoses C</i> |
| hemccod.arn | <i>ARNm correspondant à la partie strictement codante d'un allèle du gène codant pour la globine bêta humaine de l'hémoglobinoses C</i> |
| hemc.pro | <i>Chaîne protéique résultant de l'expression de l'allèle du gène codant pour la globine bêta humaine responsable de l'hémoglobinoses C</i> |

Thalassémies

| | |
|-------------|--|
| tha1cod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable d'une thalassémie grave en Chine</i> |
| tha1cod.arn | <i>ARNm correspondant à tha1cod.adn</i> |
| tha1.pro | <i>Séquence protéique correspondant à tha1</i> |
| tha2cod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable d'une thalassémie majeure en Sardaigne</i> |
| tha2cod.arn | <i>ARNm correspondant à tha2cod.adn</i> |
| tha2.pro | <i>Séquence protéique correspondant à tha2</i> |
| tha3cod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable d'une thalassémie majeure en Inde</i> |
| tha3cod.arn | <i>ARNm correspondant à tha3cod.adn</i> |
| tha3.pro | <i>Séquence protéique correspondant à tha3</i> |
| tha4cod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable d'une thalassémie majeure dans les pays méditerranéens</i> |
| tha4cod.arn | <i>ARNm correspondant à tha4cod.adn</i> |
| tha4.pro | <i>Séquence protéique correspondant à tha4</i> |
| tha5cod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable d'une thalassémie</i> |
| tha5cod.arn | <i>ARNm correspondant à tha5cod.adn</i> |
| tha5.pro | <i>Séquence protéique correspondant à tha5</i> |
| tha6cod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable d'une thalassémie majeure en Inde</i> |
| tha6cod.arn | <i>ARNm correspondant à tha6cod.adn</i> |
| tha6.pro | <i>Séquence protéique correspondant à tha6</i> |
| tha7cod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable d'une thalassémie</i> |
| tha7cod.arn | <i>ARNm correspondant à tha7cod.adn</i> |
| tha7.pro | <i>Séquence protéique correspondant à tha7</i> |
| tha8cod.adn | <i>Partie codante d'un allèle du gène codant pour la globine bêta humaine responsable d'une thalassémie majeure en Inde</i> |
| tha8cod.arn | <i>ARNm correspondant à tha8cod.adn</i> |
| tha8.pro | <i>Séquence protéique correspondant à tha8</i> |

Gamma

| | |
|--------------|--|
| gamma.adn | <i>Gène (complet) de la globine gamma humaine (brin non transcrit)</i> |
| gammadnc.adn | <i>ADNc (correspondant à l'ARNm complet) du gène codant pour la globine gamma humaine</i> |
| gammacod.adn | <i>Partie codante du gène codant pour la globine gamma humaine</i> |
| gamma.pro | <i>Séquence protéique de la globine gamma humaine</i> |
| gammacod.arn | <i>ARNm correspondant à la partie strictement codante du gène codant pour la globine gamma humaine</i> |
| gammarnm.arn | <i>ARNm complet du gène codant pour la globine gamma humaine</i> |

Delta

| | |
|--------------|---|
| delta.adn | <i>Gène (complet) de la globine delta humaine (brin non transcrit)</i> |
| deltadnc.adn | <i>ADNc (correspondant à l'ARNm complet) du gène codant pour la globine delta humaine</i> |

| | |
|--------------|--|
| deltacod.adn | <i>Partie codante du gène codant pour la globine delta humaine</i> |
| delta.pro | <i>Séquence protéique de la globine delta humaine</i> |
| deltacod.arn | <i>ARNm correspondant à la partie strictement codante du gène codant pour la globine delta humaine</i> |
| deltarnm.arn | <i>ARNm complet du gène de la globine delta humaine</i> |

LE SYSTÈME ABO DES GROUPES SANGUINS

| | |
|----------|---|
| acod.adn | <i>Partie strictement codante de l'allèle A du gène responsable du système ABO des groupes sanguins</i> |
| bcod.adn | <i>Partie strictement codante de l'allèle B du gène responsable du système ABO des groupes sanguins</i> |
| ocod.adn | <i>Partie codante de l'allèle O du gène responsable du système ABO des groupes sanguins</i> |

LA PHÉNYLALANINE HYDROXYLASE (PHÉNYLCÉTONURIE)

| | |
|-------------|---|
| phe1.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe2.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe3.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe4.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe5.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe6.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe7.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe8.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe9.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe10.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe11.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe12.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe13.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe14.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe15.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe16.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phe17.cod | <i>Partie codante d'un allèle du gène codant pour la phényl-alanine hydroxylase</i> |
| phenorm.cod | <i>Partie strictement codante d'un allèle normal (de référence) du gène codant pour la phényl-alanine hydroxylase</i> |
| phenorm.pro | <i>Séquence protéique d'un allèle normal (de référence) du gène codant pour la phényl-alanine hydroxylase</i> |

L'ALPHA-ANTITRYPSINE

| | |
|-----------|--|
| atm'1.cod | <i>Allèle M'1 du gène codant pour l'alpha-antitrypsine (partie strictement codante ; brin non transcrit)</i> |
| atm1.cod | <i>Allèle M1 du gène codant pour l'alpha-antitrypsine (partie strictement codante ; brin non transcrit)</i> |
| atm2.cod | <i>Allèle M2 du gène codant pour l'alpha-antitrypsine (partie strictement codante ; brin non transcrit)</i> |
| atm3.cod | <i>Allèle M3 du gène codant pour l'alpha-antitrypsine (partie strictement codante ; brin non transcrit)</i> |

| | |
|-------------|--|
| atnull1.cod | <i>Allèle null1 du gène codant pour l'alpha-antitrypsine (partie strictement codante ; brin non transcrit)</i> |
| atnull2.cod | <i>Allèle null2 du gène codant pour l'alpha-antitrypsine (partie strictement codante ; brin non transcrit)</i> |
| ats.cod | <i>Allèle S du gène codant pour l'alpha-antitrypsine (partie strictement codante ; brin non transcrit)</i> |
| atz.cod | <i>Allèle Z du gène codant pour l'alpha-antitrypsine (partie strictement codante ; brin non transcrit)</i> |
| chat.cod | <i>Partie strictement codante du gène codant pour l'alpha-antitrypsine chez le Chimpanzé</i> |

LE GÈNE IT15 (CHORÉE DE HUNTINGTON)

| | |
|-------------|--|
| c11.adn | <i>Allèle du gène IT15</i> |
| c13.adn | <i>Allèle du gène IT15</i> |
| c16.adn | <i>Allèle du gène IT15</i> |
| c17.adn | <i>Allèle du gène IT15</i> |
| c20.adn | <i>Allèle du gène IT15</i> |
| c21.adn | <i>Allèle du gène IT15</i> |
| c22.adn | <i>Allèle du gène IT15</i> |
| c23.adn | <i>Allèle du gène IT15</i> |
| c33.adn | <i>Allèle du gène IT15</i> |
| c40.adn | <i>Allèle du gène IT15</i> |
| c43.adn | <i>Allèle du gène IT15</i> |
| c45.adn | <i>Allèle du gène IT15</i> |
| c48.adn | <i>Allèle du gène IT15</i> |
| c50.adn | <i>Allèle du gène IT15</i> |
| c75.adn | <i>Allèle du gène IT15</i> |
| it15cod.adn | <i>Région codante complète du gène IT15 (brin non transcrit)</i> |

LA GLUCOSE 6 PHOSPHATE DÉSHYDROGÉNASE (G6PD)

| | |
|--------------|--|
| g6pda-1.cod | <i>Allèle A-1 de la G6PD humaine</i> |
| g6pda-2.cod | <i>Allèle A-2 de la G6PD humaine</i> |
| g6pda-3.cod | <i>Allèle A-3 de la G6PD humaine</i> |
| g6pda.cod | <i>Allèle A de la G6PD humaine</i> |
| g6pdb.cod | <i>Allèle B de la G6PD humaine</i> |
| g6pdm.cod | <i>Allèle M de la G6PD humaine</i> |
| g6pdseat.cod | <i>Allèle Seattle de la G6PD humaine</i> |

LE SYSTÈME HLA

HLA A

| | |
|--------------|---|
| hlaa0101.cod | <i>Partie strictement codante de l'allèle HLAA0101</i> |
| hlaa0201.adn | <i>Gène complet de l'allèle HLAA0201 (brin non transcrit)</i> |
| hlaa0201.dnc | <i>ADNc de l'allèle HLAA0201</i> |
| hlaa0201.cod | <i>Partie strictement codante de l'allèle HLAA0201</i> |
| hlaa0201.pro | <i>Séquence protéique de HLAA0201</i> |
| hlaa0210.cod | <i>Partie strictement codante de l'allèle HLAA0210</i> |

| | |
|--------------|--|
| hlaa0211.cod | <i>Partie strictement codante de l'allèle HLAA0211</i> |
| hlaa0212.cod | <i>Partie strictement codante de l'allèle HLAA0201</i> |
| hlaa1101.cod | <i>Partie strictement codante de l'allèle HLAA1101</i> |
| hlaa1102.cod | <i>Partie strictement codante de l'allèle HLAA1102</i> |
| hlaa2401.cod | <i>Partie strictement codante de l'allèle HLAA2401</i> |
| hlaa2402.cod | <i>Partie strictement codante de l'allèle HLAA2402</i> |
| hlaa2403.cod | <i>Partie strictement codante de l'allèle HLAA2403</i> |

HLA B

| | |
|--------------|--|
| hlab2701.cod | <i>Partie strictement codante de l'allèle HLAB2701</i> |
| hlab2702.cod | <i>Partie strictement codante de l'allèle HLAB2702</i> |
| hlab2703.cod | <i>Partie strictement codante de l'allèle HLAB2703</i> |
| hlab2704.cod | <i>Partie strictement codante de l'allèle HLAB2704</i> |
| hlab2705.adn | <i>Gène complet de l'allèle HLAB2705</i> |
| hlab2705.dnc | <i>ADNc de l'allèle HLAB2705</i> |
| hlab2705.cod | <i>Partie strictement codante de l'allèle HLAB2705</i> |
| hlab2705.pro | <i>Séquence protéique de l'allèle HLAB2705</i> |
| hlab2706.cod | <i>Partie strictement codante de l'allèle HLAB2706</i> |

HLA C

| | |
|--------------|--|
| hlaaw101.cod | <i>Partie strictement codante de l'allèle HLACW101</i> |
| hlaaw102.cod | <i>Partie strictement codante de l'allèle HLACW102</i> |
| hlaaw201.cod | <i>Partie strictement codante de l'allèle HLACW201</i> |
| hlaaw202.cod | <i>Partie strictement codante de l'allèle HLACW202</i> |
| hlaaw301.adn | <i>Gène complet de l'allèle HLACW301</i> |
| hlaaw301.dnc | <i>ADNc de l'allèle HLACW301</i> |
| hlaaw301.cod | <i>Partie strictement codante de l'allèle HLACW301</i> |
| hlaaw301.pro | <i>Séquence protéique de l'allèle HLACW301</i> |

LE CMH DU CHIMPANZÉ

| | |
|--------------|-----------------------------------|
| chlaa108.cod | <i>Allèle du CMH du Chimpanzé</i> |
| chlaa126.cod | <i>Allèle du CMH du Chimpanzé</i> |
| chlab1.cod | <i>Allèle du CMH du Chimpanzé</i> |
| chlab2.cod | <i>Allèle du CMH du Chimpanzé</i> |

LES HORMONES HYPOPHYSAIRES ET PLACENTAIRES

| | |
|-------------|---|
| fsh.pro | <i>Séquence protéique de la FSH bêta humaine</i> |
| fshadnc.adn | <i>ADNc du gène codant pour la FSH bêta humaine</i> |
| fshb.adn | <i>Gène (complet) du gène codant pour la FSH bêta humaine (brin non transcrit)</i> |
| fshcod.adn | <i>Partie strictement codante du gène codant pour la FSH bêta humaine</i> |
| gh.adn | <i>Gène (complet) de l'hormone de croissance (Growth Hormone) humaine (brin non transcrit)</i> |
| gh.pro | <i>Séquence protéique de l'hormone de croissance humaine</i> |
| ghadnc.adn | <i>ADNc du gène codant pour l'hormone de croissance humaine</i> |
| ghcod.adn | <i>Partie strictement codante du gène codant pour l'hormone de croissance humaine</i> |
| hcgb.adn | <i>Gène (complet) de la HCG bêta (Hormone Chorionique Gonadotrope) humaine (brin non transcrit)</i> |

| | |
|--------------|---|
| hcgb.pro | <i>Séquence protéique de l'hormone HCG bêta humaine</i> |
| hcgbadnc.adn | <i>ADNC du gène codant pour l'hormone HCG bêta humaine</i> |
| hcgbcod.adn | <i>Partie strictement codante du gène codant pour l'hormone HCG bêta humaine</i> |
| hlp.adn | <i>Gène complet de l'hormone HLP (Hormone lactogène placentaire) humaine (brin non transcrit)</i> |
| hlp.pro | <i>Séquence protéique de HLP</i> |
| hlpadnc.adn | <i>ADNC du gène codant pour HLP</i> |
| hlpcod.adn | <i>Partie strictement codante du gène codant pour HLP</i> |
| hprl.adn | <i>Gène complet de l'hormone de la prolactine (HPRL) humaine (brin non transcrit)</i> |
| hprl.pro | <i>Séquence protéique de HPRL</i> |
| hprladnc.adn | <i>ADNC du gène codant pour HPRL</i> |
| hprlcod.adn | <i>Partie strictement codante du gène codant pour HPRL</i> |
| lh.pro | <i>Séquence protéique de l'hormone LH</i> |
| lhb.adn | <i>Gène complet de l'hormone LH humaine</i> |
| lhbadnc.adn | <i>ADNC du gène codant pour l'hormone LH</i> |
| lhbcod.adn | <i>Partie strictement codante du gène codant pour l'hormone LH</i> |
| tsh.pro | <i>Séquence protéique de l'hormone TSH bêta humaine</i> |
| tshb.adn | <i>Gène complet codant pour l'hormone TSH bêta humaine (brin non transcrit)</i> |
| tshbadnc.adn | <i>ADNC du gène codant pour l'hormone TSH bêta humaine</i> |
| tshbcod.adn | <i>Partie strictement codante du gène codant pour l'hormone TSH bêta humaine</i> |

LE GÈNE DE LA TYROSINASE

| | |
|--------------|--|
| tyralba1.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase, responsable d'un albinisme</i> |
| tyralba2.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase, responsable d'un albinisme</i> |
| tyralba3.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase, responsable d'un albinisme</i> |
| tyralba4.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase, responsable d'un albinisme</i> |
| tyralba5.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase, responsable d'un albinisme</i> |
| tyralbb1.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase, responsable d'un albinisme</i> |
| tyralbts.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase, responsable d'un albinisme</i> |
| tyrcod1.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase (référence 1)</i> |
| tyrcod2.cod | <i>Partie strictement codante d'un allèle du gène de la tyrosinase (référence 2)</i> |

LES GÈNES DES PIGMENTS RÉTINIENS

La rhodopsine

| | |
|-------------|---|
| rhonorm.cod | <i>Partie strictement codante d'un allèle normal (de référence) du gène codant pour la rhodopsine</i> |
| rhoret1.cod | <i>Partie codante d'un allèle du gène codant pour la rhodopsine</i> |
| rhoret2.cod | <i>Partie codante d'un allèle du gène codant pour la rhodopsine</i> |
| rhoret3.cod | <i>Partie codante d'un allèle du gène codant pour la rhodopsine</i> |
| rhoret4.cod | <i>Partie codante d'un allèle du gène codant pour la rhodopsine</i> |
| rhoret5.cod | <i>Partie codante d'un allèle du gène codant pour la rhodopsine</i> |

rhoret6.cod *Partie codante d'un allèle du gène codant pour la rhodopsine*
rhoret7.cod *Partie codante d'un allèle du gène codant pour la rhodopsine*
rhoret8.cod *Partie codante d'un allèle du gène codant pour la rhodopsine*
rhoret9.cod *Partie codante d'un allèle du gène codant pour la rhodopsine*

La périphérine

pernorm.cod *Partie strictement codante d'un allèle normal (de référence) du gène codant pour la périphérine (PER)*
perret1.cod *Partie strictement codante d'un allèle du gène codant pour la PER, responsable de rétinite*
perret2.cod *Partie strictement codante d'un allèle du gène codant pour la PER, responsable de rétinite*
perret3.cod *Partie strictement codante d'un allèle du gène codant pour la PER, responsable de rétinite*
perret4.cod *Partie strictement codante d'un allèle du gène codant pour la PER, responsable de rétinite*
perret5.cod *Partie strictement codante d'un allèle du gène codant pour la PER, responsable de rétinite*

La phosphodiesterase

pdenorm.cod *Partie strictement codante d'un allèle normal (de référence) du gène codant pour la phosphodiesterase (PDE)*
pderet1.cod *Partie strictement codante d'un allèle du gène codant pour la PDE, responsable de rétinite*
pderet2.cod *Partie strictement codante d'un allèle du gène codant pour la PDE, responsable de rétinite*
pderet3.cod *Partie strictement codante d'un allèle du gène codant pour la PDE, responsable de rétinite*
pderet4.cod *Partie strictement codante d'un allèle du gène codant pour la PDE, responsable de rétinite*

INDEX DES THÈMES D'ÉTUDE

EXPRESSION DE L'INFORMATION GÉNÉTIQUE

Globine alpha

- Gène et ARNm codant
- Séquence peptidique
- ARNm et ADNc

Globine bêta

- Gène et ARNm codant
- Séquence peptidique
- ARNm et ADNc

Globine gamma

- Gène et ARNm codant
- Séquence peptidique
- ARNm et ADNc

Globine delta

- Gène et ARNm codant
- Séquence peptidique
- ARNm et ADNc

RELATIONS GÉNOTYPE-PHÉNOTYPE

Phénotype drépanocytaire

Phénotypes thalassémiques

- Séquences nucléiques
- Séquences protéiques

Phénotype groupes sanguins ABO

Phénotype phénylcétonurique

POLYMORPHISME DES GÈNES

Polymorphisme de ABO

Polymorphisme de AT

- Allèles de AT
- Arbre avec sujets ayant un déficit en AT

Polymorphisme de IT15

- Allèles du gène IT15

Arbre avec sujets atteints de chorée de Huntington

Polymorphisme de HLA

Polymorphisme de HLA A

Polymorphisme de HLA B

Polymorphisme de HLA C

Polymorphisme du CMH du Chimpanzé

FAMILLES MULTIGÉNIQUES

Gènes des globines

Gènes des HLA

Gènes des hormones hypophysaires et placentaires

Gènes des hormones LH, FSH, TSH et HCG

Gènes des hormones GH, HPRL et HLP

GÉNOTYPES ET PRÉVISIONS EN GÉNÉTIQUE HUMAINE

Le déficit en Glucose 6 Phosphate Déshydrogénase [G6PD]

Allèles de G6PD

Arbre généalogique avec sujets ayant un déficit en G6PD

L'albinisme

Allèles du gène de la tyrosinase

Arbres généalogiques avec sujets albinos

Arbre 1 avec sujets atteints d'albinisme

Arbre 2 avec sujets atteints d'albinisme

Arbre 3 avec sujets atteints d'albinisme

Arbre 4 avec sujets atteints d'albinisme

Les rétinites pigmentaires

Les allèles de la rhodopsine

Les allèles de la PDE (Phospho Di Estérase)

Les allèles de la périphérine

Arbres généalogiques avec sujets atteints de rétinites

Arbre 1 avec sujets atteints de rétinites

Gène de la PDE
Gène de la rhodopsine
Gène de la périphérine

Arbre 2 avec sujets atteints de rétinites

Gène de la PDE
Gène de la rhodopsine
Gène de la périphérine

Arbre 3 avec sujets atteints de rétinites

Gène de la PDE
Gène de la rhodopsine
Gène de la périphérine

Arbre 4 avec sujets atteints de rétinites

Gène de la PDE
Gène de la rhodopsine
Gène de la périphérine

Famille 5 avec sujets atteints de rétinites

Gène de la PDE
Gène de la rhodopsine
Gène de la périphérine

INDEX DES PROGRAMMES ET DOCUMENTS

DOCUMENTS GÉNÉRAUX

Désignation des AA avec une lettre ou trois lettres tabpro.bmp¹

L'HÉMOGLOBINE HUMAINE NORMALE

Tétramère de l'Hb, squelette carboné hb4chbis.bmp

L'HÉMOGLOBINE HUMAINE DRÉPANOCYTAIRE²

Carte de répartition de la drépanocytose carte.bmp
Observation microscopique de sang drépanocytaire globules.bmp
Circulation capillaire, sang normal et drépanocytaire normdrep.avi
RX de la hanche normale et drépanocytaire hanche.bmp
Électrophorèse d'Hb normale et drépanocytaire electro.bmp
Électrophorèse de l'ADN normal et drépanocytaire adn.bmp
Chaîne β de l'Hb drépanocytaire hbbetask.bmp
Tétramères de l'Hb drépanocytaire hbbetadr.bmp

POLYMORPHISME DE L'ALPHA-ANTITRYPSINE

Tableau de fréquence des allèles de AT freqat.bmp
Arbre avec sujets atteints de déficit en AT arbreat.bmp
Filiation entre les allèles de l'alpha-antitrypsine filiat.bmp

POLYMORPHISME DU GÈNE IT15

Fréquence de quelques allèles du gène IT15 frq2it15.bmp
Fréquence d'allèles non morbides du gène IT15 freqit15.bmp
Fréquence d'allèles morbides du gène IT15 morbit15.bmp
Arbre avec sujets atteints de chorée de Huntington arbrit15.bmp

POLYMORPHISME DE HLA

Fréquence de quelques allèles de HLAA freqhla.bmp
Filiation possible entre des allèles de HLAB filihla.bmp
HLAA associé à un peptide de l'hépatite B hlatotal.bmp
HLAA, corbeille hlaorb1.bmp

¹ Les fichiers des images (extension bmp) et la séquence vidéo numérisée (extension avi) sont installés dans un sous-répertoire d'*Anagène* dénommé BANQUE.

² Les images et la séquence vidéo numérisée sont extraites de la vidéo-cassette *La drépanocytose* éditée par le CNDP.

| | |
|---|--------------|
| HLAA de même spécificité sérologique. AA modifiés | hlaorb2.bmp |
| HLAA de spécificité sérologique différente. AA modifiés | hlaorb4.bmp |
| HÉRÉDITÉ DU DÉFICIT EN G6PD | |
| Fréquence de quelques allèles du gène de la G6PD | frqg6pd.bmp |
| Filiation possible entre allèles de la G6PD | filig6pd.bmp |
| Enzymes pour les allèles du gène de la G6PD | enzg6pd.bmp |
| Arbre avec sujets ayant un déficit en G6PD | f1g6pd.bmp |
| HÉRÉDITÉ DE L'ALBINISME | |
| Nombre de sites reconnus par les enzymes | tabtyros.bmp |
| Arbre 1 avec sujets atteints d'albinisme | f1albino.bmp |
| Arbre 2 avec sujets atteints d'albinisme | f2albino.bmp |
| Arbre 3 avec sujets atteints d'albinisme | f3albino.bmp |
| Arbre 4 avec sujets atteints d'albinisme | f4albino.bmp |
| HÉRÉDITÉ DES RÉTINITES PIGMENTAIRES | |
| Enzymes pour les allèles du gène de la rhodopsine | enzrho.bmp |
| Arbre 1 avec sujets atteints de rétinites | f1retini.bmp |
| Arbre 2 avec sujets atteints de rétinites | f2retini.bmp |
| Arbre 3 avec sujets atteints de rétinites | f3retini.bmp |
| Arbre 4 avec sujets atteints de rétinites | f4retini.bmp |
| Arbre 5 avec sujets atteints de rétinites | f5retini.bmp |

تم نشر هكنا الملف بواسطة قرص **تجربتي** مع الباكالوريا

tajribatybac@gmail.com

facebook.com/tajribaty

jjel.tk/bac

تم نشر هكنا الملف بواسطة قرص **تجربتي** مع الباكالوريا

tajribatybac@gmail.com

facebook.com/tajribaty

jjel.tk/bac