

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).

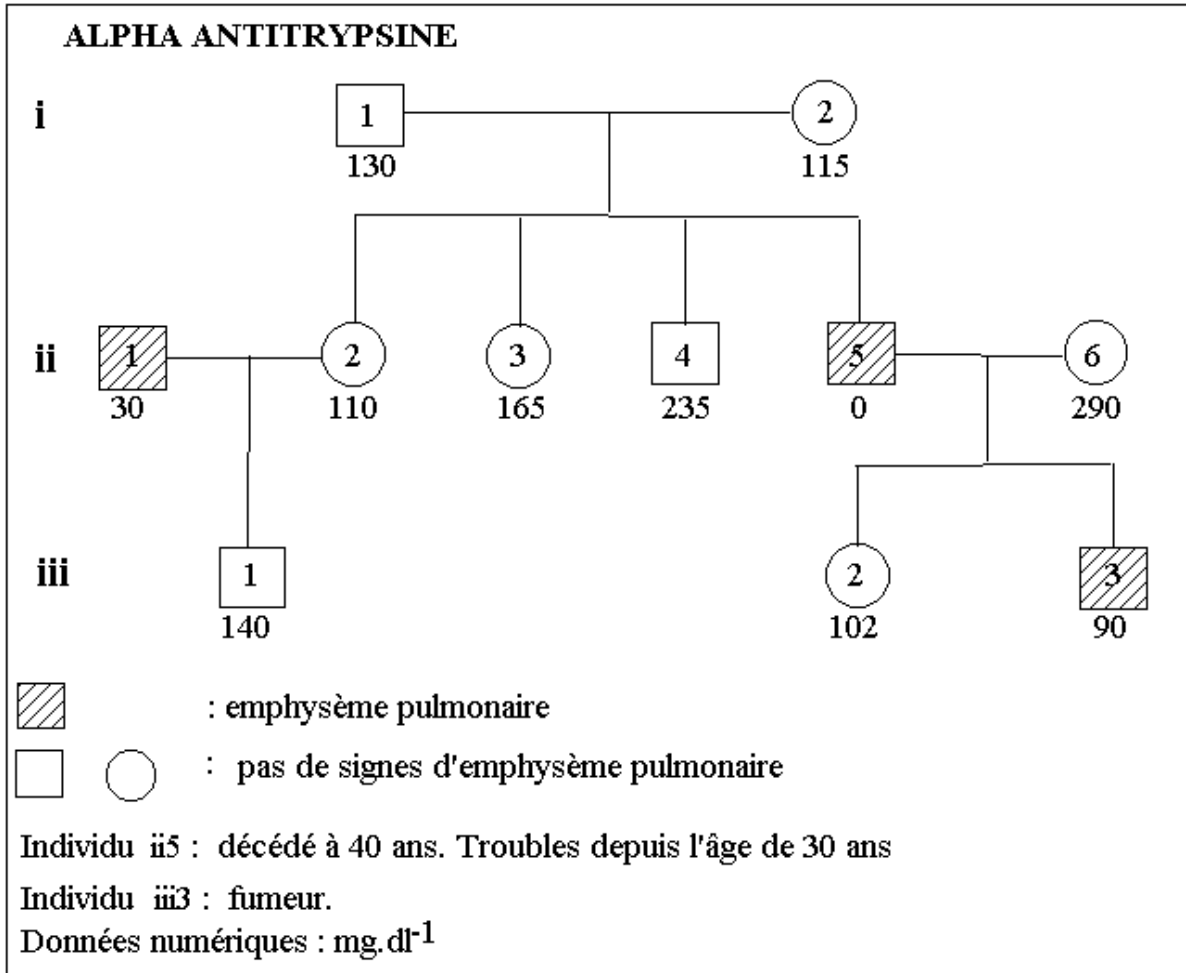
Remarque : il est indiqué de choisir comme allèle de référence non pas M1 qui est le plus fréquent, mais M'1 que les généticiens estiment l'allèle initialement présent dans les premières populations humaines.

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.

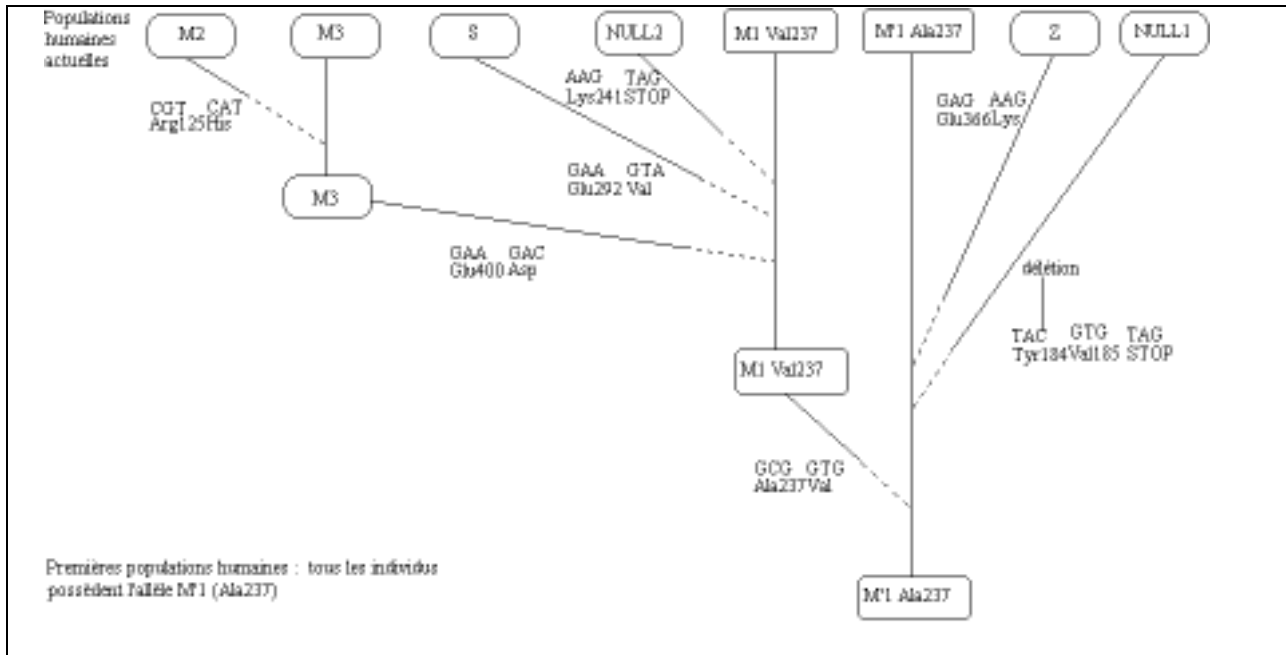
Allèles des individus	Nucléotides changés Nature - Position	Allèles du gène
F1-i1all1.at	Référence	M'1
F1-i1all2.at	-C552	NULL1
F1-i2all1.at	C710 → T	M1
F1-i2all2.at	-C552	NULL1
F1-ii1all1.at	G1096 → A	Z
F1-ii1all2.at	G1096 → A	Z
F1-ii2all1.at	Référence	M'1
F1-ii2all2.at	-C552	NULL1
F1-ii3all1.at	-C552	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.

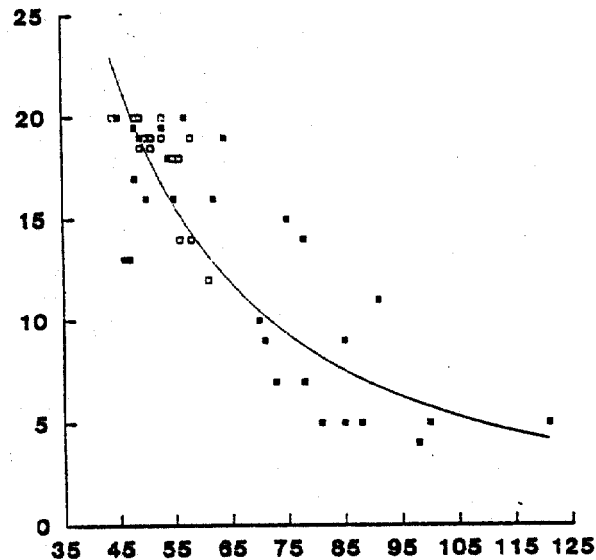
En ce qui concerne le phénotype macroscopique, les individus hétérozygotes possédant un seul allèle morbide présentent les mêmes symptômes que ceux ayant deux allèles morbides : mouvements involontaires désordonnés, perturbations du psychisme, perte de mémoire, troubles du langage et finalement démence. 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.

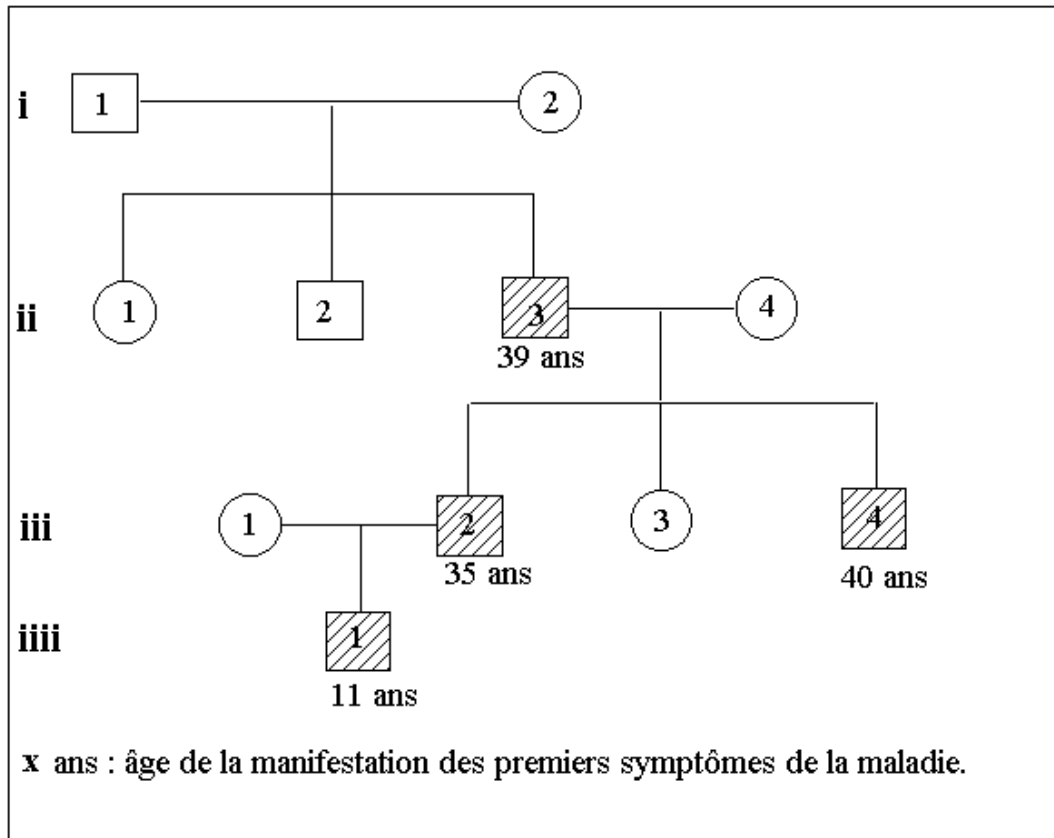
On demande aux élèves de rechercher les différences entre les allèles du gène et de dégager les caractéristiques de ceux qui sont à l'origine de la chorée. On peut prendre comme référence l'allèle dont la fréquence est la plus élevée dans les deux populations, à savoir C17. Les élèves utiliseront la 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 :

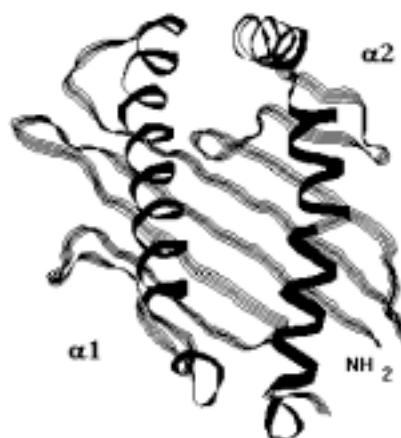
- une partie extracellulaire constituée par 3 domaines de 90 acides aminés chacun : les domaines alpha1, alpha2 et alpha3 ;
- une partie intramembranaire riche en acides aminés hydrophobes permettant l'ancrage solide de la molécule dans la membrane ;
- une partie intracytoplasmique constituant l'extrémité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

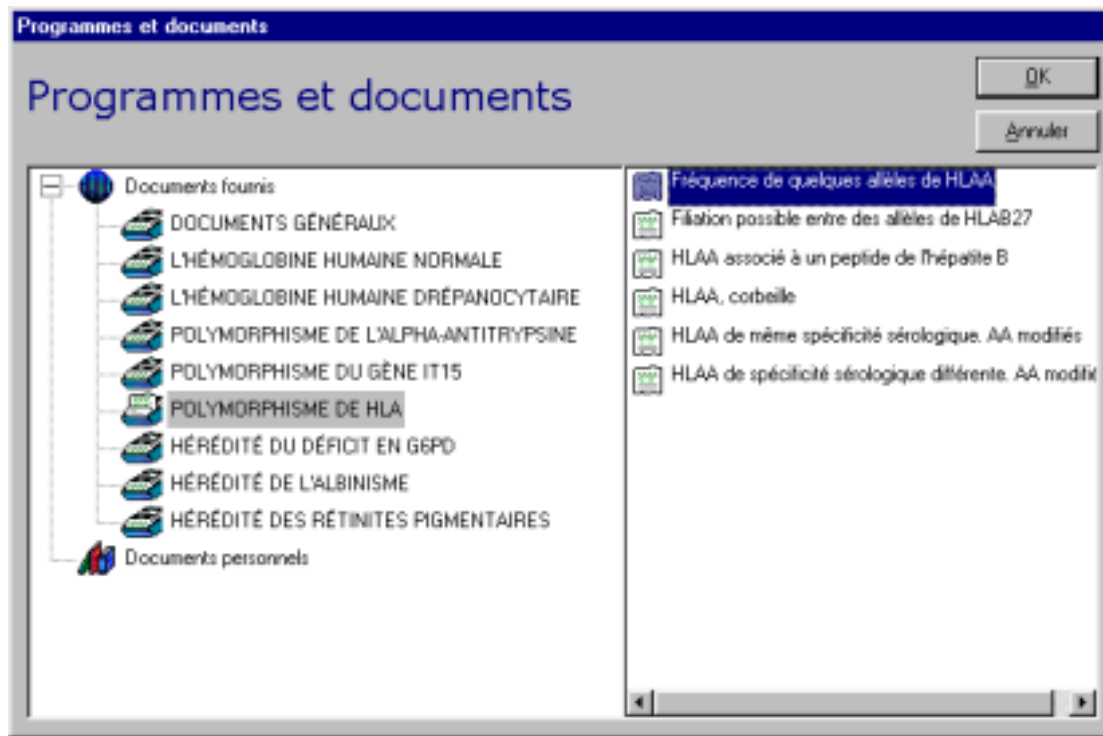
Récemment, l'analyse au niveau moléculaire du CMH du Chimpanzé a fourni des données corroborant cette hypothèse. Chez le Chimpanzé comme chez l'Homme, il y a trois locus polymorphiques pour les antigènes de classe I. On s'est aperçu que **le degré de similitude entre des allèles CHLAA du Chimpanzé (CH 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

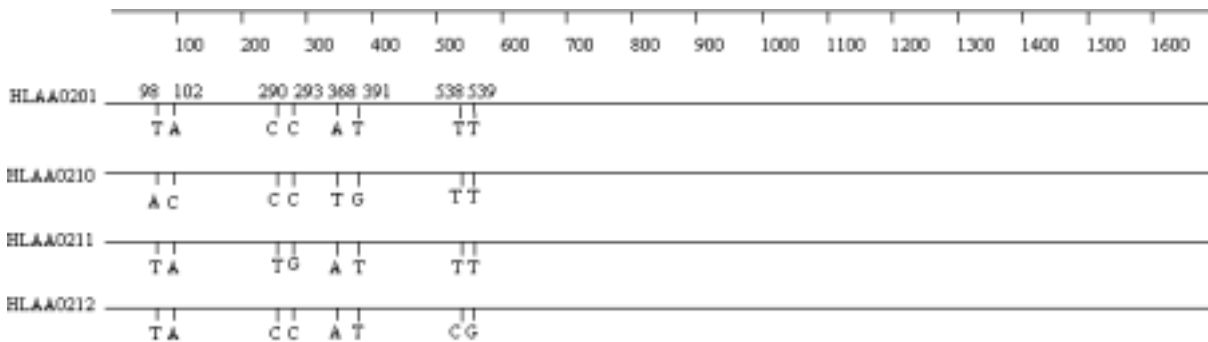
Les seules séquences sur les allèles HLAA suffisent pour dégager les idées essentielles. On peut proposer de rechercher séparément, en utilisant la modalité de « comparaison simple » :

- Les différences entre les allèles correspondant à une même spécificité sérologique : ceux du groupe HLAA2 ou ceux du groupe HLAA24.

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 HLAA02 et HLAA24.



Représentation des différences entre quelques allèles de HLAA02

- 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 HLAA0101, HLAA0201, HLAA1101 et HLAA2401.

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

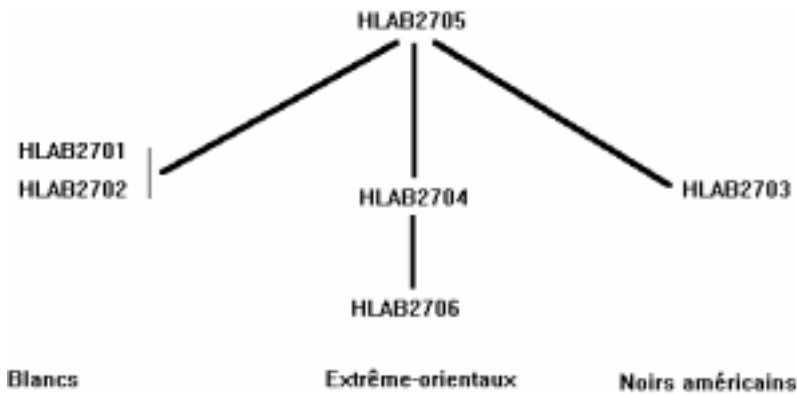
Comme pour le gène codant pour l'alpha-antitrypsine, l'objectif de cette étude est de faire prendre conscience que les allèles présents dans les populations humaines sont l'aboutissement actuel des mutations survenues dans le passé.

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----- alpha1 -----I----- alpha2 -----I----- alpha3 -----I							
	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.