

STABILITÉ ET VARIATION DES GÉNOMES ET ÉVOLUTION

Innovations génétiques – Mutations ponctuelles et filiations entre allèles

Allèles du gène de l'alpha-antitrypsine

Informations scientifiques

[Cf. page 83 \(classe de première\)](#)

Pistes d'exploitation pédagogique des données fournies

L'exploitation des données fournies (séquences et documents) dans le thème d'étude sur l'alphaAT permet de bâtir les notions relatives à l'apparition de nouveaux allèles par mutations, le gène de l'alpha AT étant très polymorphe. Les mutations à l'origine de ces allèles sont des substitutions et des délétions d'un nucléotide. Une filiation entre les allèles peut être établie, permettant ainsi de reconstituer une partie de l'histoire évolutive de ce gène. Des documents complémentaires permettent également de discuter de l'origine d'une mutation et de sa diffusion dans les zones géographiques voisines.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Mutations ponctuelles et filiations entre allèles/Allèles du gène de l'alpha-antitrypsine** permet d'atteindre **Allèles AT** qui charge le fichier `alleles-AT.edi` affichant les séquences nucléiques strictement codantes de quelques allèles du gène de l'alpha AT – allèles M'1, M1, M2, M3, Z, S, NULL 1 et NULL 2

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Mutations ponctuelles et filiations entre allèles/Allèles du gène de l'alpha-antitrypsine** permet de charger les fichiers :

- **InfoAlphaAT.bmp** permettant de définir le phénotype à différents niveaux (clinique, moléculaire et biochimique). Il précise également l'influence de la fumée de cigarette. Il peut donc être utilisé en introduction ou en complément de l'analyse de la filiation des gènes ;
- **AllelesAT.bmp** affichant les informations fournies dans ce tableau qui permettent de relier le phénotype biochimique (concentration plasmatique en alpha AT) et le phénotype clinique pour les différents allèles de l'alpha AT. Les données sur la fréquence des différents allèles permettent également de préciser la notion de polymorphisme ;
- **GenoPhenoAlphaAT.bmp** affichant un document pouvant servir de base à une discussion sur la notion de dominance/récessivité des allèles ;
- **filiationAT.bmp** présentant une activité pédagogique envisageable consistant à proposer cette filiation aux élèves, en leur demandant de la justifier à partir de l'exploitation des séquences fournies (ils doivent identifier et placer sur chaque branche la mutation permettant de la justifier) ;
- **repartitionSZ.bmp** affichant les cartes de répartition qui permettent de discuter de l'origine des allèles S et Z. On constate que la fréquence de l'allèle S est très nettement plus élevée en Espagne, et plus particulièrement en Galice ; de plus, cette fréquence décroît quand on s'éloigne de ces régions. On peut donc supposer que cet allèle est apparu dans cette région. La fréquence la plus élevée de l'allèle « Z » se trouve dans les populations du nord-ouest de l'Europe. On peut supposer que la mutation à l'origine de cet allèle est intervenue dans la lignée germinale d'un individu d'une de ces populations. Ensuite, non seulement l'allèle « Z » a diffusé dans la population nordique, mais a aussi réussi à se répandre dans les populations du sud-est européen.

L'allèle M'1 est considéré comme allèle de référence. La comparaison des autres allèles à cet allèle de référence permet d'aborder les notions relatives aux mutations et au polymorphisme génique. La traduction de ces séquences et la comparaison des séquences protéiques obtenues peuvent permettre de discuter des effets des mutations sur le phénotype moléculaire, biochimique et clinique.

La comparaison des autres allèles avec l'allèle M'1 permet de définir un certain nombre de mutations, substitutions faux-sens, substitutions non-sens, délétions. Cette série d'allèles permet d'introduire la notion de filiation entre allèles : un allèle « b » apparaît par mutation d'un allèle existant « a » dans la lignée germinale d'un individu. Il se répand par la suite plus ou moins dans la population. En mutant chez un autre individu, il

est à l'origine d'un nouvel allèle « c ». Bien entendu, les élèves doivent saisir que les allèles « a » et « b » persistent dans la population.

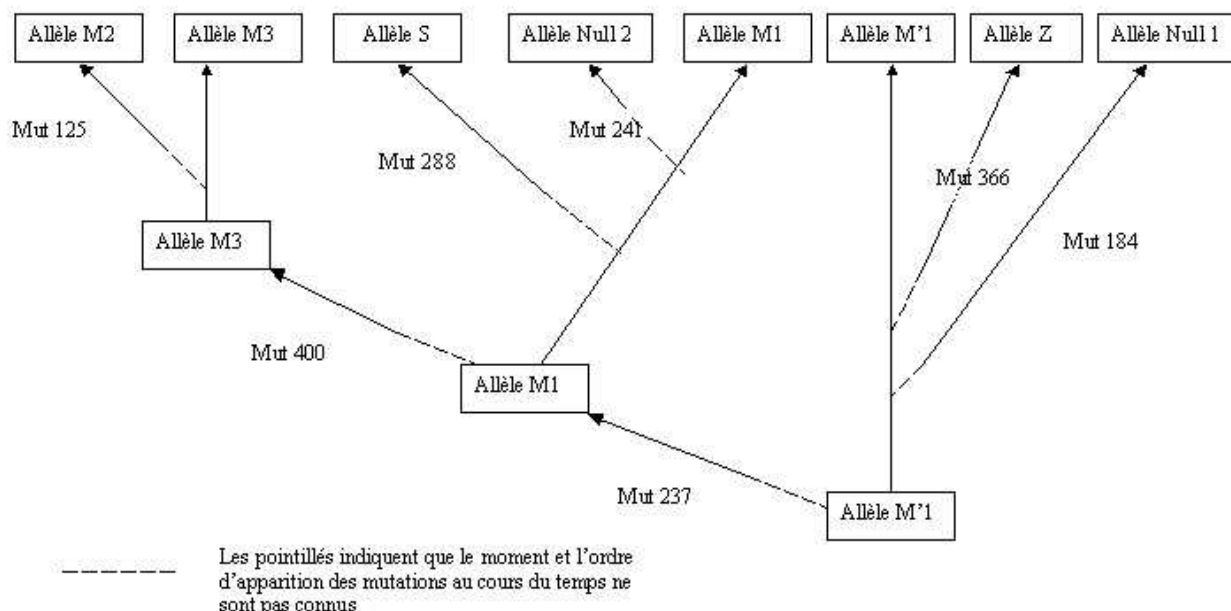
Les allèles « b » et « c » ont en commun la première mutation qui les différencie de l'allèle « a ». L'allèle « c » possède en plus la mutation qui lui est propre. Cela fournit la base de raisonnement que traduit le document sur la filiation entre les allèles de l'alpha-antitrypsine.

La filiation proposée suppose qu'il n'y avait qu'un seul allèle de ce gène dans les premières populations humaines. Si cela est semblé-t-il exact pour le gène de l'alpha-antitrypsine, ce n'est pas vrai pour tous les gènes. Ainsi, il est certain que les premières populations humaines étaient polymorphes pour les gènes du système HLA.

Résultats obtenus par l'exploitation des séquences nucléiques fournies

Allèle	Différence par rapport à l'allèle de référence (M'1)		Différence entre la protéine codée par M'1 et la protéine codée par cet allèle
	Nucléotide	Codon	
M1	710 : C → T	Codon 237 : GCG → GTG	A → V
M2	374 : G → A	Codon 125 : CGT → CAT	R → H
	710 : C → T 1200 : A → C	Codon 237 : GCG → GTG Codon 400 : GAA → GAC	A → V E → D
M3	710 : C → T	Codon 237 : GCG → GTG	A → V
	1200 : A → C	Codon 400 : GAA → GAC	E → D
S	710 : C → T	Codon 237 : GCG → GTG	A → V
	863 : A → T	Codon 288 : GAA → GTA	E → V
Z	1096 : G → A	Codon 366 : GAG → AAG	E → K
Null 1	552 : délétion de C	Codon 184 : TAC → TAG	183 acides aminés au lieu de 418
Null 2	710 : C → T	Codon 237 : GCG → GTG	A → V
	721 : A → T	Codon 241 : AAG → TAG	240 acides aminés au lieu de 418

Tableau de comparaison des allèles de l'alpha AT et des protéines correspondantes



Filiation des allèles de l'alpha AT

Les numéros des mutations se rapportent aux codons et non aux nucléotides

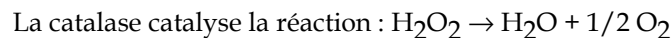
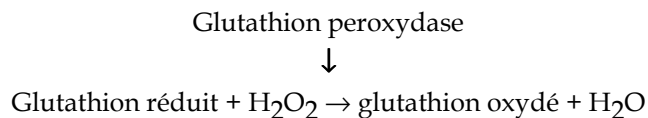
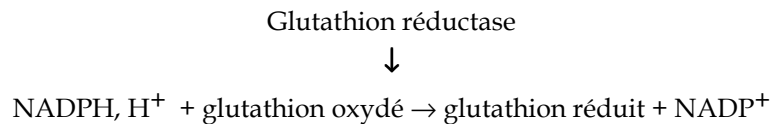
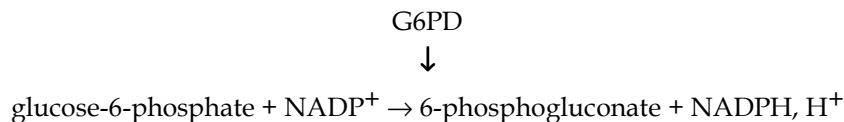
Innovations génétiques – Mutations ponctuelles et filiations entre allèles

- Allèles du gène de la G6PD

Informations scientifiques

Rôle de l'enzyme G6PD (glucose-6-phosphate déshydrogénase)

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 :



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 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 treize exons et mesure 18 kilo paires de bases environ. Toutefois, sa région codante ne comprend que 1 545 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

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. 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.

Pistes d'exploitation pédagogique du gène de la G6PD

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques - Mutations ponctuelles et filiations entre allèles/Allèles du gène de la G6PD** permet d'atteindre **Allèles G6PD** qui charge le fichier alleles-G6PD-HS.edi affichant les séquences nucléiques strictement codantes de quelques allèles du gène de la G6PD.

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Mutations ponctuelles et filiations entre allèles/Allèles du gène de la G6PD** permet de charger les fichiers :

- **roleG6PD.bmp** affichant les informations suffisantes sur les conséquences de la déficience de l'enzyme ;
- **frequenceG6PD.bmp** indiquant que les allèles G6PDA et G6PDA- ont une fréquence élevée en Afrique ; à vrai dire, on les trouve aussi dans les Amériques et dans les régions où il y a des populations d'origine africaine. L'allèle G6PDM est ainsi nommé car il est polymorphique dans les pays du bassin méditerranéen, y compris l'Afrique du Nord ;
- **filiationG6PD.bmp** présentant une activité pédagogique envisageable qui propose cette filiation aux élèves, en leur demandant de la justifier à partir de l'exploitation des séquences fournies (ils doivent identifier et placer sur chaque branche la mutation permettant de la justifier).

La comparaison des divers allèles avec l'allèle G6PDB pris comme référence ainsi que l'étude des conséquences des différences sur le polypeptide permet d'établir le tableau suivant.

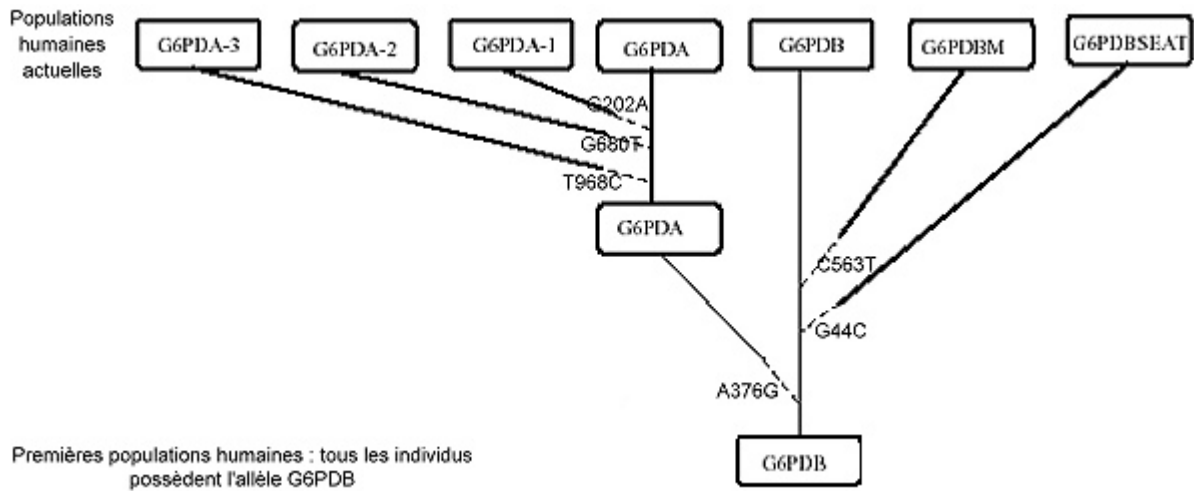
Séquence nucléique			Polypeptide	
Noms des allèles	Nucléotides changés Nature - Position	Codons changés Nature - Position	A.A. changés Nature - Position	Type de mutation
G6PDB	(Référence)	(Référence)	(Référence)	
G6PDA	A376 G	AAT126 GAT	Asn126Asp	Substitution faux-sens
G6 PDA -1	A376 G	AAT126 GAT	Asn126Asp	Substitution faux-sens
	G202 A	GTG68 ATG	Val68Met	Substitution faux-sens
G6 PDA -2	A376 G	AAT GAT	Asn126Asp	Substitution faux-sens
	G680 T	CGC227 CTC	Arg227Leu	Substitution faux-sens
G6 PDA -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 éliminés par la sélection naturelle).

Ce système allélique est aussi un outil pour faire saisir la filiation entre allèles d'un gène.

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

Innovations génétiques – Mutations ponctuelles et filiations entre allèles

- Allèles de la chaîne bêta de l'hémoglobine

Informations scientifiques

Les informations sur la structure de la globine, l'allèle HbC et le phénotype drépanocytaire ont été fournis précédemment. On trouvera ici quelques compléments sur le phénotype thalassémique et les allèles (tous récessifs) qui en sont à l'origine.

[Cf. page 51 \(classe de première\)](#)

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 qu'elles offrent 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.

Pistes d'exploitation pédagogique des données fournies

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Mutations ponctuelles et filiations entre allèles/Allèles du gène de la globine bêta** permet d'atteindre :

- **Séquences nucléiques** qui charge le fichier GlobineBetaADN.edi affichant les séquences des allèles du gène bêta globine correspondant à trois allèles codant pour une chaîne bêta fonctionnelle (l'allèle HbA étant le plus fréquent), à l'allèle HbS à l'origine du phénotype drépanocytaire et à celles de huit allèles entraînant le phénotype thalassémique ;
- **Séquences protéiques** qui charge le fichier GlobineBetaPRO.edi affichant les séquences protéiques correspondantes.

Le fichier contenant l'ensemble de ces séquences est sans doute le plus approprié pour envisager les divers types de mutations ponctuelles et leurs conséquences sur le phénotype moléculaire. Le principe de la démarche est de :

- comparer les séquences nucléiques des différents allèles du gène de la bêta globine par rapport à l'allèle HbA (allèle considéré comme référence) de façon à identifier les différents types de mutations ponctuelles ;
- comparer les séquences polypeptidiques codées par les divers allèles à la séquence de référence codée par la séquence HbA ;
- mettre les différences au niveau polypeptidique en relation avec celles du niveau nucléique pour faire le point sur les conséquences des mutations sur le phénotype moléculaire.

On aboutit ainsi aux conclusions suivantes :

- certaines mutations (substitution) sont muettes (variant 1) car elles n'ont aucune conséquence sur la séquence du polypeptide du fait de la redondance du code génétique ;
- d'autres substitutions sont neutres car entraînant un changement de la séquence du polypeptide sans en modifier les propriétés (variant 2) ;
- d'autres substitutions sont non-sens car entraînant la synthèse d'un polypeptide non fonctionnel à cause de l'apparition anticipé d'un codon stop ;
- d'autres mutations sont des délétions ou des insertions d'un ou de plusieurs nucléotides entraînant un décalage du cadre de lecture, l'apparition d'un codon stop anticipé et donc la synthèse d'un polypeptide raccourci et non fonctionnel.

Innovations génétiques – Mutations ponctuelles et filiations entre allèles

Allèles du gène IT15 (chorée de Huntington)

Informations scientifiques

La chorée de Huntington

Cette maladie se traduit par des mouvements involontaires et désordonnés, des perturbations du psychisme, des pertes de mémoire, des troubles du langage, et finalement de la démence. Ces troubles sont liés à une destruction progressive des neurones cérébraux, notamment de ceux des corps striés, impliqués dans la motricité.

Le gène de la huntingtine (IT 15)

Le gène IT15 est situé sur le bras court du chromosome 14. Il a été isolé et séquencé en 1995 et code pour une protéine, la huntingtine, dont le rôle est actuellement inconnu. Il s'exprime dans de nombreux tissus, notamment les neurones cérébraux des corps striés. Ce gène comprend 67 exons, et sa région codante 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.

On peut regrouper les allèles en deux catégories, en fonction du nombre de répétitions du triplet CAG et du phénotype clinique associé.

Les « variants normaux »

Ces allèles ont un nombre de répétitions du triplet CAG compris entre 9 et 39 ; les allèles les plus fréquents sont ceux qui ont un nombre de répétitions du triplet CAG compris entre 17 et 21. La fréquence de la majorité de ces allèles est supérieure à 1 % : le gène IT15 est donc très polymorphe.

Le tableau ci-dessous indique les fréquences de ces allèles dans une population allemande (d'après *Hum. Mol. Gen.*, 1993, vol. 2, n° 12, p. 2 063-2 067).

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

Les « variants morbides »

Ils sont à l'origine de la chorée de Huntington, et possèdent un triplet CAG répété entre 36 et 121 fois, avec une fréquence maximale comprise entre 42 et 46 répétitions.

La très grande majorité des personnes atteintes de chorée de Huntington ont un allèle où le triplet CAG est répété plus de quarante fois. Il reste une indétermination pour un triplet répété entre trente-six et trente-neuf fois car selon les cas la possession d'un tel allèle est associée aux signes de la maladie et dans d'autres cas non.

La présence d'une expansion anormalement importante de l'acide aminé glutamine dans la huntingtine conduit à la mort neuronale par l'activation de la machinerie apoptotique : c'est une mort par apoptose (activation d'un programme intrinsèque de mort cellulaire).

Le tableau ci-dessous présente la fréquence des allèles morbides dans une population de personnes atteintes de chorée de Huntington. Toutes les personnes atteintes étaient hétérozygotes (d'après *Hum. Mol. Gen.*, 1993, vol. 2, n° 12, p. 2 063-2 067)

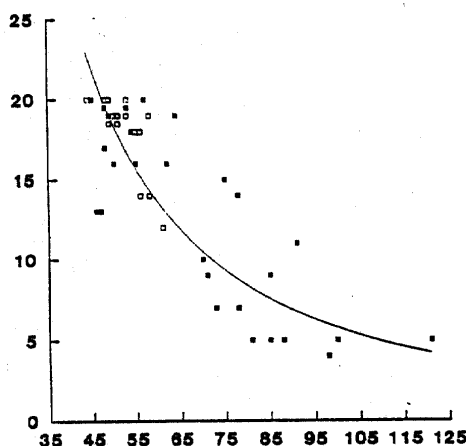
Nombre de répétitions du triplet CAG de l'allèle	Fréquence de l'allèle (en %)
40	6,69
41	11,15
42	11,46
43	10,83
44	12,42
45	12,10
46	8,91
47	7
48	5,10
49	2,23
50	1,91
51	1,59
52	2,87
53	0,96
54	0,64
55	1,27
56	0
57	0,64
58	0,96
61	0,32
63	0,32
73	0,32
75	0,32

On a constaté l'apparition de cas de chorée de Huntington dans des familles où la maladie n'avait pas été identifiée précédemment. Étant 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 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 mutation par expansion au cours de la spermatogenèse conduisant à un allèle où le triplet CAG était répété plus de quarante fois. En outre, dans les familles où le père est 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 trente répétitions, le gène IT15 ait tendance à muter avec expansion du triplet au cours de la spermatogenèse. Cette instabilité n'a pas été retrouvée pour les allèles dont le nombre de répétitions est inférieur à trente. Ainsi, les études au niveau moléculaire de ce gène semblent, pour un certain type d'allèles, révéler une fréquence de mutations assez élevée.

Les relations génotype/phénotype

Chez un individu hétérozygote les deux allèles s'expriment et on trouve donc les deux types de huntingtine dans les cellules. Le phénotype morbide et, donc, 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 n'est pas encore élucidé. On pense que la huntingtine avec expansion de la glutamine agit dans le noyau pour induire l'apoptose. 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 corrélations 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.



Pistes d'exploitation pédagogique des données fournies

Le gène IT15, dont certains allèles responsables de la chorée de Huntington, est particulièrement polyallélique et polymorphe. Bien que le rôle de la protéine codée par ce gène, la huntingtine, soit encore inconnu, les données fournies permettent d'aborder les notions de polymorphisme génique, de mutations par **expansion** d'un motif (triplet) et de **néomutations**, et les relations génotype/phénotype (relations de dominance/récessivité).

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques - Mutations ponctuelles et filiations entre allèles/Allèles du gène IT15 (Chorée de Huntington)** permet d'atteindre :

- **Allèles IT15** qui charge le fichier alleles-IT15.edi affichant les séquences strictement codantes de quinze allèles du gène IT15, neuf « variants normaux » et six « variants morbides ». Le gène IT15 étant long (9 435 nucléotides en moyenne), on s'est limité, pour chaque allèle, à une portion de la séquence codante, commençant au triplet d'initiation ATG et encadrant la partie répétée (le triplet répété commence à la position 52) ;
- **Famille (Chorée de Huntington)** qui charge le fichier allelesFamillechoree.edi affichant les séquences strictement codantes des allèles du gène IT 15 possédés par les membres de la famille dont l'arbre généalogique est fourni. Le gène IT15 étant long (9 435 nucléotides en moyenne), on s'est limité, pour chaque allèle, à une portion de la séquence codante, commençant au triplet d'initiation ATG et encadrant la partie répétée (le triplet répété commence à la position 52).

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Mutations ponctuelles et filiations entre allèles/Allèles du gène IT15 (chorée de Huntington)** permet d'atteindre :

- **Informations sur la chorée de Huntington** qui charge le fichier choree.bmp présentant des symptômes de la chorée de Huntington et du gène IT15. En introduction, ce document permet de motiver l'étude du polymorphisme de ce gène ;
- **Fréquence de quelques allèles du gène IT15** qui charge le fichier frequenceallelesIT15.bmp affichant les tableaux de fréquence de quelques allèles du gène IT15 dans une population de personnes non malades et dans une population de personnes atteintes de la chorée de Huntington ;
- **Arbre avec sujets atteints de chorée de Huntington** qui charge le fichier arbrechoree.jpg présentant l'arbre généalogique d'une famille touchée par la chorée de Huntington. L'âge d'apparition des premiers symptômes de la maladie des membres atteints de la chorée est indiqué ;
- **Répétition du triplet CAG en fonction de l'âge** qui charge le fichier courbeIT15.jpg

enregistrer dans la banque de séquences ou dans les thèmes d'étude personnels. Penser à reprotéger les données avant de poursuivre ! (Il suffit de retourner dans **Options** et de cliquer sur **Protéger les données**.)

Ne pas lancer d'alignement avec discontinuité pour un grand nombre de séquences en même temps si la capacité de votre ordinateur est limitée !

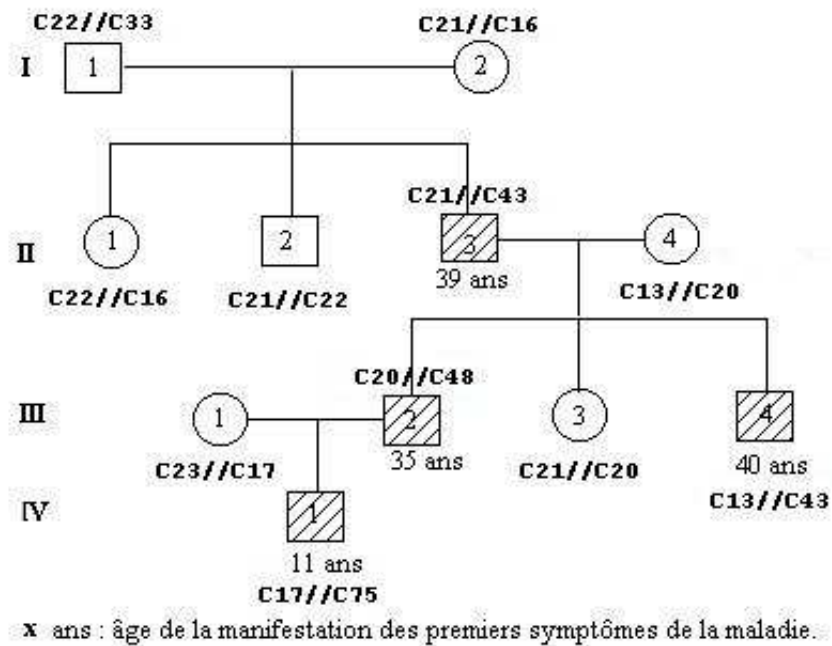
Mise en évidence de néomutations (origine de nouveaux allèles)

La lecture de l'arbre généalogique de la famille confronté avec le document **Répétition du triplet CAG en fonction de l'âge** (fichier courbeIT15.jpg) conduit à suspecter l'apparition de nouveaux allèles au sein de cette famille. La détermination des génotypes permet de tester l'idée.

La comparaison (alignement avec discontinuité) entre les allèles (fichier allelesFamillechoree.edi) d'un individu des générations I, II et III, et l'allèle C21 permet de déterminer le génotype des individus. Pour l'individu IV, 1, il faut utiliser comme références soit l'allèle 21 soit l'allèle 50.

NB : Le choix de ces allèles de référence n'a aucune signification biologique. Compte tenu de l'algorithme d'alignement, les allèles C21 et C50 sont ceux qui visualisent le mieux l'expansion de chacun des allèles des individus de la famille.

Allèles des individus	Nombre de répétitions du triplet CAG	Allèle du gène IT15
I1 allèle 1	22	C22
I1 allèle 2	33	C33
I2 allèle 1	21	C21
I2 allèle 2	16	C16
II1 allèle 1	22	C22
II1 allèle 2	16	C16
II2 allèle 1	21	C21
II2 allèle 2	22	C22
II3 allèle 1	21	C21
II3 allèle 2	43	C43
II4 allèle 1	13	C13
II4 allèle 2	20	C20
III1 allèle 1	23	C23
III1 allèle 2	17	C17
III2 allèle 1	20	C20
III2 allèle 2	48	C48
III3 allèle 1	21	C21
III3 allèle 2	20	C20
III4 allèle 1	13	C13
III4 allèle 2	43	C43
IV1 allèle 1	17	C17
IV1 allèle 2	75	C75



Arbre généalogique d'une famille avec individus atteints de chorée

Les génotypes des individus atteints de chorée indiquent que les allèles à l'origine de cette maladie sont dominants. En conséquence, il y a une contradiction entre les phénotypes des individus de la première génération et le génotype de l'individu II3. Celui-ci possède l'allèle C21 de sa mère, mais aucun allèle de son père. Cela conduit à l'idée que l'allèle C43 de II3 provient d'une mutation subie par un des allèles du père lors de la spermatogenèse (sans doute l'allèle C33). Cette idée de néomutation permet également l'interprétation du génotype de l'individu IV1.

Innovations génétiques – Duplications et familles multigéniques

Gènes des globines

Informations scientifiques

[Cf. page 124 \(classe terminale\)](#)

L'exploitation des données fournies permet de bâtir les notions relatives à l'apparition de nouveaux gènes par duplication génique suivie de mutations. Les gènes de la famille des globines codent pour des protéines ayant conservé globalement la même fonction (fixation de dioxygène grâce au groupement hème). La disposition des différents gènes de cette famille sur deux chromosomes différents permet d'aborder le phénomène de transposition (ou translocation) génique. Quelques données paléontologiques sont utilisables pour dater approximativement certaines duplications géniques.

Pistes d'utilisation pédagogique

Les gènes des globines forment une famille multigénique

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Duplications et familles multigéniques/Gènes des globines** permet d'atteindre :

- **Nucléiques** qui charge le fichier genes-glob-humaines.edi affichant les séquences nucléiques représentant la partie strictement codante (du codon d'initiation au codon stop) des gènes de globines humaines (alpha 1, alpha 2, bêta, gamma A, gamma G, zêta, delta, epsilon, thêta, myoglobine) ;
- **Protéiques** qui charge le fichier pro-glob-humaines.edi affichant les séquences protéiques des globines humaines (alpha 1, alpha 2, bêta, gamma A, gamma G, zêta, delta, epsilon, thêta, myoglobine).

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Duplications et familles multigéniques/Gènes des globines** permet d'atteindre :

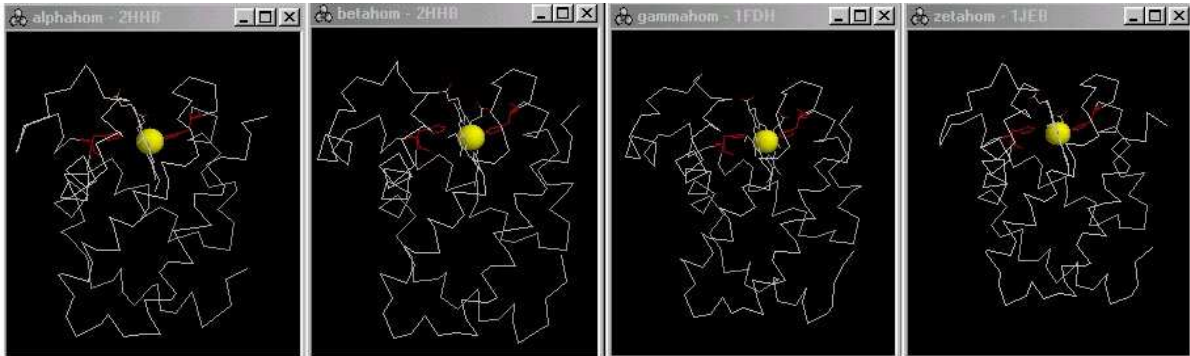
- **Présentation des globines humaines** qui charge le fichier globineshumaines.bmp affichant un petit texte précisant l'existence de plusieurs globines humaines, accompagné d'un graphique exprimant, en pourcentage, la synthèse de ces différentes globines en fonction de l'âge. Peut servir à poser le problème de l'existence de plusieurs globines humaines ayant le même rôle, et qui sont produites à différents moments de la vie de l'être humain ;
- **Localisation chromosomique** qui charge le fichier chrombetaglob.jpg affichant un schéma de localisation des gènes de globines humaines, y compris les pseudogènes sur les chromosomes 11 (gènes du groupe bêta) et 16 (gènes du groupe alpha). Ce document peut être important pour bien faire comprendre qu'il s'agit de gènes différents et non d'allèles d'un même gène. L'existence de deux familles de gènes localisées sur deux chromosomes différents peut aussi suggérer une parenté plus grande entre les gènes du groupe bêta ou entre les gènes du groupe alpha qu'entre les gènes de deux groupes différents. Utilisé après avoir posé la notion de duplication génique, ce document est utile pour introduire la notion de transposition d'un gène ;
- **Comparaison des globines en 3D** qui charge le fichier comp3Dglobines.jpg affichant les représentations en squelette carboné et en rubans de quatre globines humaines : les globines alpha, bêta, gamma et zêta. Ces structures ont été obtenues avec le logiciel *RasTop*. La similitude de structure des globines présentées peut suggérer l'idée d'une parenté entre ces molécules, donc entre les gènes qui les codent ;
- **Filiation globines** qui charge le fichier arbreglobines.jpg révélant la phylogénie qui traduit les relations de parenté entre l'ensemble des différents gènes de globines humaine (myoglobine, famille alpha, famille bêta). Des indications de dates sont portées sur cette figure ; elles situent approximativement dans le temps les principales duplications géniques ; elles ont été obtenues en combinant deux sortes d'information : les différentes globines présentes chez différents taxons et les dates d'apparition de ces taxons ;
- **Datation des duplications** qui charge le fichier datadupli.jpg rappelant sous forme d'un tableau les dates d'apparition des principales classes de Vertébrés, ainsi que des indications permettant de dater les duplications alpha/bêta et bêta/delta.

Fichiers des molécules en 3D

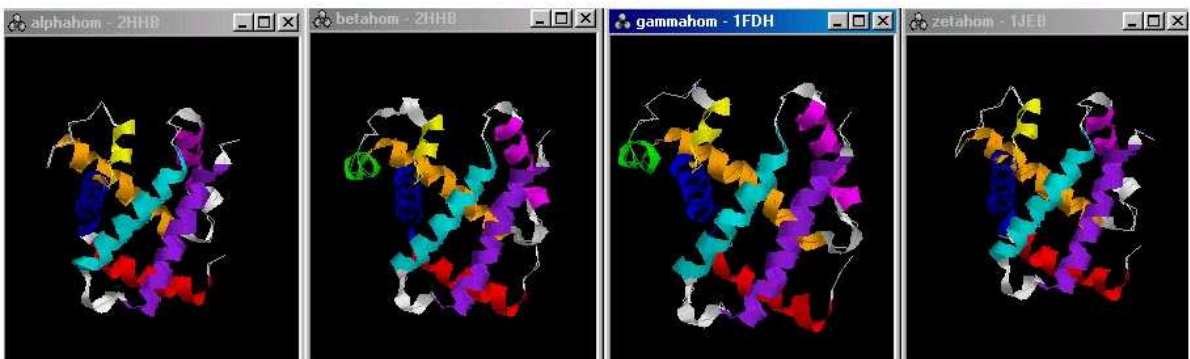
Dans le dossier 3D sous Anagene2, le répertoire **FamilleGlobines** contient les fichiers .pdb suivants : alphahom.pdb, betahom.pdb, gammahom.pdb et zetahom.pdb.

Des similitudes de structure et de fonction pour les globines humaines

Toutes les globines humaines, globines alpha, bêta, zêta, gamma G, gamma A, epsilon présentent des structures qui se ressemblent : de telles similitudes confortent l'idée d'une origine commune.



Représentation en squelette carboné de 4 globines humaines
(le groupement hème est mis en évidence)



Représentation en rubans de 4 globines humaines
(les secteurs en hélice sont mis en évidence de façon colorée)

Des similitudes de séquences protéiques et nucléiques

Les séquences nucléiques ou les séquences peptidiques des globines humaines présentent un degré de similitude qui ne peut être dû au hasard. Les globines sont donc des protéines homologues qui dérivent donc d'un même gène ancestral (de la même façon, on observe des similitudes dans l'organisation de ces différents gènes : ils possèdent tous trois exons et deux introns et ont des longueurs proches). Les gènes des globines forment donc une famille multigénique.

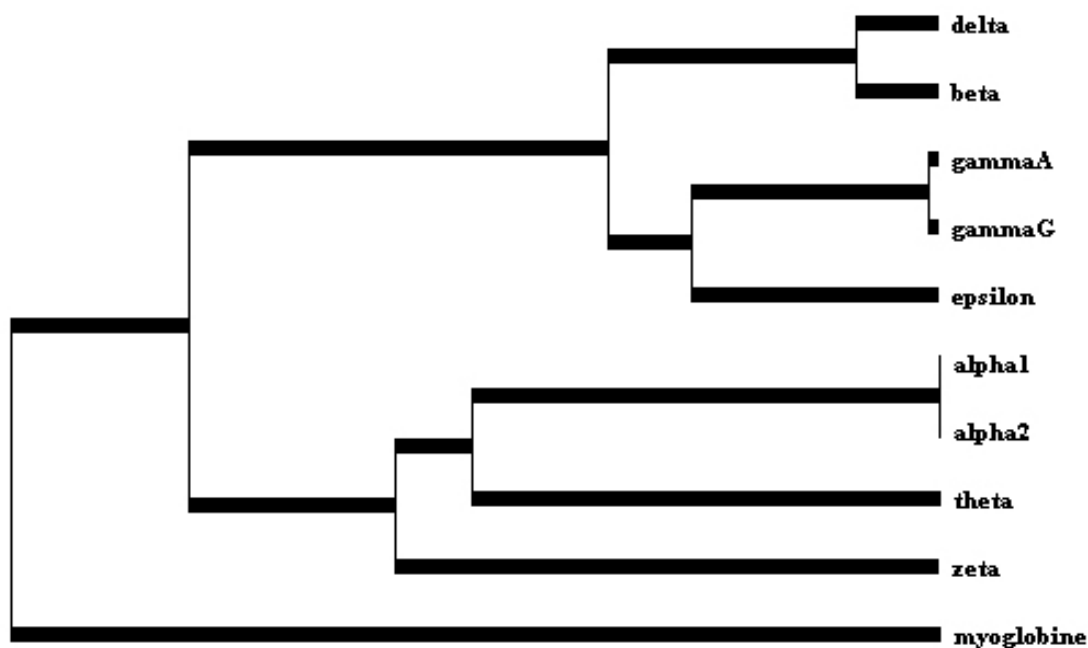
Afin d'établir l'arbre de filiation des gènes de cette famille de globines humaines, on relève les similitudes entre les séquences nucléiques de ces gènes, ce qui permet de construire une matrice :

	Alpha1	Alpha2	Bêta	GammaA	GammaG	Delta	Epsilon	Thêta	Zêta	Myoglobine
Alpha1	100									
Alpha2	99,5	100								
Bêta	57	56,8	100							
GammaA	54,5	54,1	75,9	100						
GammaG	54,5	54,1	76,1	99,8	100					
Delta	56,3	55,9	92,6	76,1	76,4	100				
Epsilon	54,5	54,3	78,8	83,6	83,8	78,6	100			
Thêta	73,2	73	52,7	53,8	53,8	52,4	53,1	100		
Zêta	68,8	68,5	55	55,9	55,9	54,5	56,9	64,8	100	
Myoglobine	40,9	40,6	40,2	42,2	41,9	40,2	41,9	41,1	40,9	100

Matrice des identités (en %) entre les globines humaines obtenue
à partir d'un alignement avec discontinuité des séquences codantes des gènes pour ces protéines

NB : Ces résultats ont été obtenus en choisissant une séquence de référence (la première) et en maintenant toutes les autres dans l'ordre alphabétique des noms. Du fait de l'algorithme de comparaison, ces résultats peuvent être légèrement différents de ceux obtenus par comparaison des séquences deux à deux.

On déduit de cette matrice l'arbre d'évolution probable des gènes des globines humaines :



Arbre de filiation des globines humaines obtenu avec le logiciel Phylogène

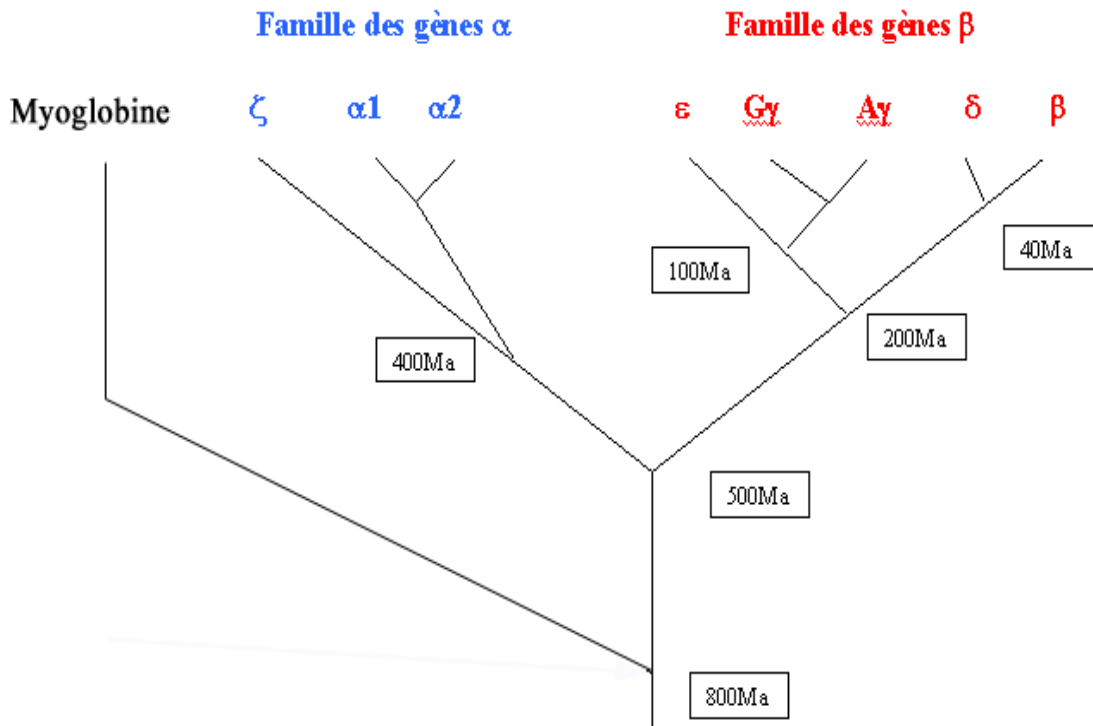
Les mécanismes à l'origine de cette famille multigénique

Chaque nœud de l'arbre obtenu précédemment correspond à une duplication génique ; sur chaque branche, des mutations différentes ont été fixées (les mutations apparaissant au hasard), responsables d'une évolution indépendante des deux duplicatas.

Un phénomène de translocation est à faire intervenir pour expliquer la localisation des gènes sur des chromosomes différents. Cette translocation (ou transposition) a dû accompagner la duplication à l'origine des deux familles de gènes (famille alpha et famille bêta).

Tous les Vertébrés, à l'exception des Agnathes, ayant deux gènes de globine (alpha et bêta) et les plus anciens poissons étant datés d'environ 450 mA alors que les premiers Vertébrés sont apparus il y a environ 550 mA, on en déduit que la première duplication a dû avoir lieu entre ces deux dates.

Arbre montrant l'évolution des gènes des globines, déduite des différences de séquences nucléiques.



La famille des globines est donc issue de duplications suivies de mutations qui amènent à une diversification des gènes.

Il est possible de proposer divers exercices utilisant les documents sur l'arbre d'évolution des gènes de globine et sur les dates d'apparition des principales classes de Vertébrés. Par exemple, l'arbre indique une date de 200 millions d'années pour la duplication gamma-bêta, bien antérieure à l'apparition des différents ordres de Mammifères. On peut demander aux élèves de tester cette donnée à l'aide des séquences des globines bêta et gamma humaines et de la globine bêta du Chien. La comparaison de ces séquences montre que les similitudes entre la globine bêta humaine et la globine bêta du Chien sont plus importantes qu'entre les globines bêta et gamma humaines. Cela signifie que la divergence Primates/Carnivores est postérieure à la duplication bêta-gamma. L'ancêtre commun à l'Homme et au Chien possédait déjà deux gènes bêta et gamma. C'est un caractère ancestral, la duplication ayant eu lieu bien avant, au tout début de l'histoire des Mammifères placentaires.

Innovations génétiques – Duplications et familles multigéniques

Gènes des opsines

Informations scientifiques

[Cf. page 129 \(classe terminale\)](#)

Pistes d'exploitation pédagogique

L'exploitation des données fournies permet d'aborder la notion d'apparition de nouveaux gènes par duplication génique suivie de mutations différentes. Les nouveaux gènes codent ici pour des protéines qui, tout en ayant conservé la même fonction globale (absorption de radiations lumineuses) présentent des propriétés légèrement différentes (les longueurs d'ondes absorbées par les différents pigments ne sont pas les mêmes).

Plusieurs arguments suggèrent l'idée d'une origine commune pour les gènes des opsines :

- les opsines présentent de nombreuses similitudes : leurs propriétés et leurs structures se ressemblent. Elles sont toutes impliquées dans la conversion de la lumière en mouvement ionique puis en signal nerveux. Chaque molécule photosensible contient sept hélices alpha transmembranaires ;
- les séquences d'acides aminés des opsines absorbant le vert et des opsines absorbant le rouge sont particulièrement semblables. Trois résidus localisés près du rétinol déterminent la différence spectrale entre les récepteurs pour le vert et pour le rouge. Le remplacement d'un résidu polaire par un résidu non polaire (par exemple, la sérine par l'alanine) à chacune de ces positions déplace λ_{\max} vers le rouge d'environ 10 nm. Trois acides aminés expliquent la majeure partie de la différence de 30 nm entre les maxima d'absorption ;
- les séquences nucléotidiques ou les séquences peptidiques des opsines humaines présentent des similitudes importantes.

La famille multigénique des opsines

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques – Duplications et familles multigéniques/Gènes des opsines** permet d'atteindre :

- **Nucléiques** qui charge le fichier genes-Opsines.edi affichant les séquences nucléiques strictement codantes des gènes des opsines humaines rouge, verte et bleue ;
- **Protéiques** qui charge le fichier pro-opsines.edi affichant les séquences protéiques des opsines rouge, verte et bleue humaines.

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Duplications et familles multigéniques/Gènes des opsines** permet d'atteindre :

- **Informations opsines** qui charge le fichier presopsines.jpg affichant un texte de présentation des opsines humaines (incluant le spectre d'absorption) et précisions sur les opsines présentes chez les autres Primates. Ce document peut être utilisé pour poser le problème (existence, chez un même organisme, de plusieurs molécules aux rôles très voisins). Il est aussi utilisable pour dater la dernière duplication génique à partir des informations concernant les opsines présentes chez les autres Primates ;
- **Spectres d'absorption des opsines** qui charge le fichier spectropsines.jpg présentant le spectre d'absorption des opsines humaines ;
- **Comparaison structure 3D opsines** qui charge le fichier comparopsines.jpg révélant la comparaison des structures 3D (représentation en rubans) des trois opsines humaines (document obtenu avec le logiciel *RasTop*). Les similitudes constatées sont un argument en faveur d'une origine commune de ces protéines, donc des gènes qui les codent ;
- **Localisation chromosomique** qui charge le fichier chromopsines.jpg montrant la localisation chromosomique des gènes des opsines humaines. Une fois vue leur origine commune et les mécanismes à l'origine de leur apparition, ce document permet de discuter de la plasticité importante du génome, en abordant notamment le phénomène de transposition de gènes ;
- **Phylogénie des opsines** qui charge le fichier phyloopsines.jpg montrant la phylogénie des gènes des opsines humaines (sans légende). Ce document peut être complété de façon à y placer les mécanismes intervenant dans l'histoire de ces gènes (duplications et fixations de mutations indépendamment dans les deux branches).

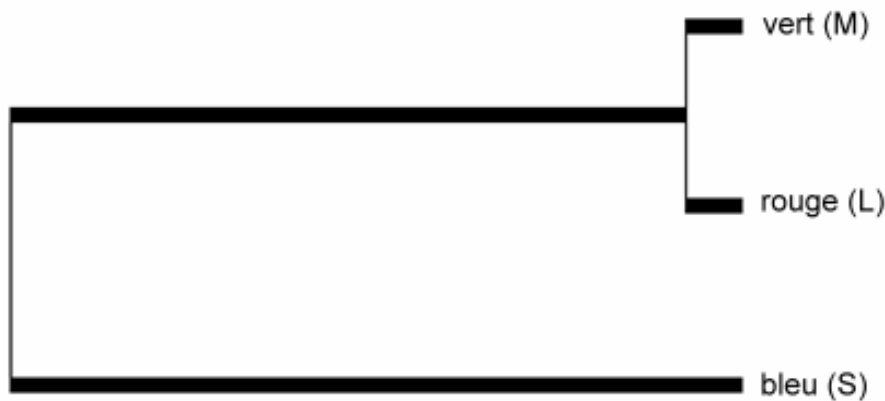
La comparaison de séquences protéiques et/ou nucléiques permet de dégager la notion de molécules homologues (protéines homologues ou gènes homologues) grâce à l'importance des similitudes constatées. Le relevé des différences permet de préciser le degré de parenté entre les protéines et/ou les gènes, et donc d'établir une phylogénie pour ces gènes.

	Gène de l'opsine rouge (L)	Gène de l'opsine verte (M)	Gène de l'opsine bleue (S)
Gène de l'opsine rouge	100 %		
Gène de l'opsine verte	97,7 %	100 %	
Gène de l'opsine bleue	57,1 %	57,6 %	100 %

Matrice des identités créée à partir d'un alignement avec discontinuité des gènes des opsines

	Opsine « rouge » (L)	Opsine « verte » (M)	Opsine « bleue » (S)
Opsine rouge	100 %		
Opsine verte	95,9 %	100 %	
Opsine bleue	42 %	43,1 %	100 %

Matrice des identités créée à partir d'un alignement avec discontinuité des opsines (séquences protéiques)



Arbre phylogénétique obtenu à partir de l'exploitation des matrices des identités

La phylogénie obtenue permet de discuter des mécanismes à l'origine de l'apparition de gènes différents à partir d'un seul.

Datation des duplications

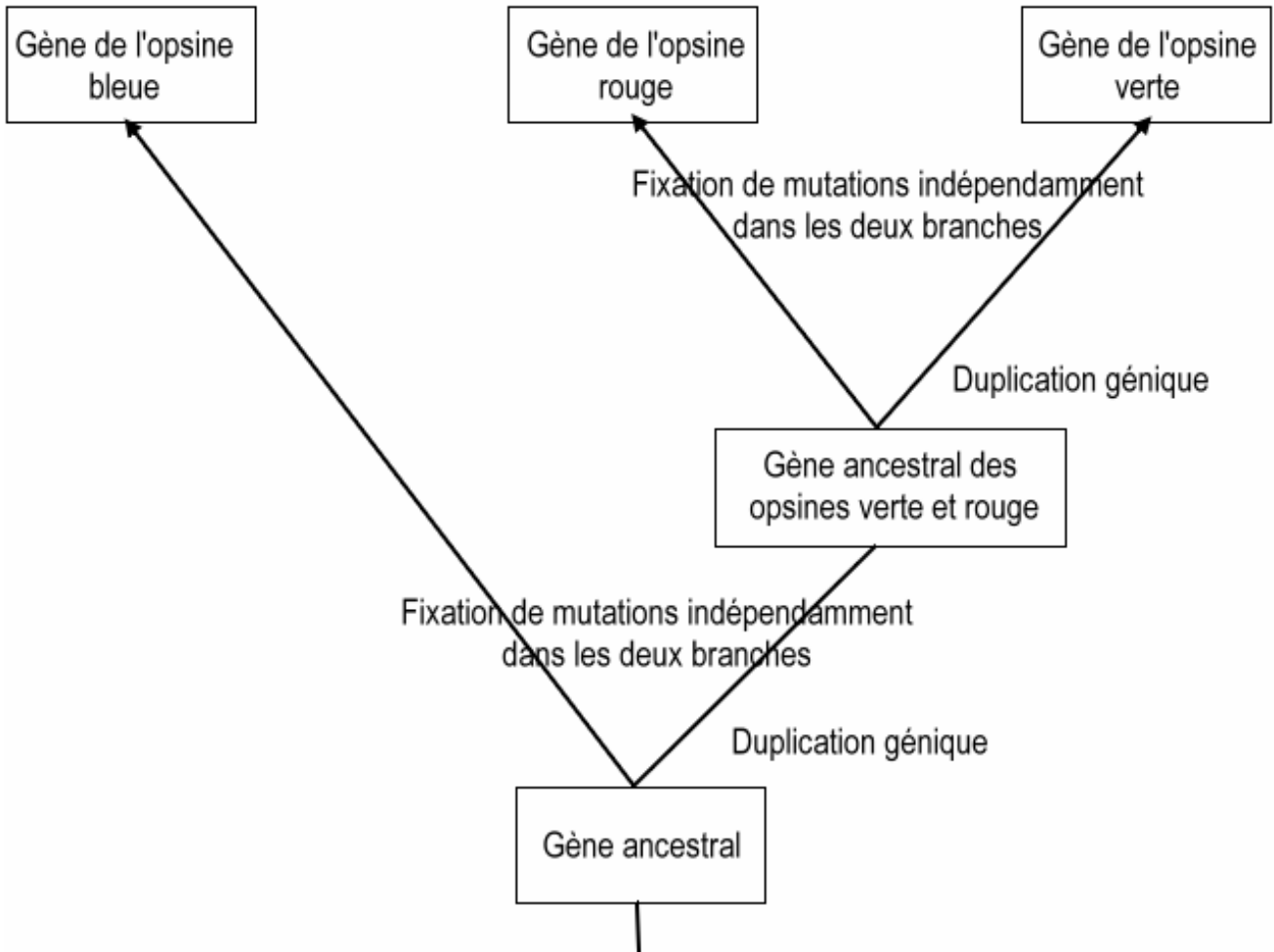
La prise en compte d'observations réalisées chez d'autres Primates permet de dater approximativement la dernière duplication génique : seuls les Singes de l'Ancien Monde possèdent trois gènes d'opsine et notamment un gène d'opsine rouge et un gène d'opsine verte. Les Singes du Nouveau Monde ne possédant que deux gènes, on peut dire que la duplication à l'origine des gènes des opsines verte et rouge à partir d'un même gène ancestral a dû avoir lieu dans la lignée menant aux Singes de l'Ancien Monde, et a donc dû se produire après la séparation de cette lignée de celle des Singes du Nouveau Monde, soit il y a environ 40 à 23 Ma.

Translocation de gènes

Les gènes des opsines provenant tous d'un même gène ancestral, mais étant situés sur deux chromosomes différents, il faut imaginer un phénomène de translocation d'un des deux exemplaires de gènes lors de la duplication du gène ancestral.

L'histoire évolutive de la famille des opsines

On peut donc demander aux élèves de compléter l'arbre phylogénétique à l'aide des informations fournies par la comparaison des gènes des opsines comme ci-dessous :



Reconstitution de l'histoire de la famille des opsines

Innovations génétiques – Duplications et familles multigéniques

Gènes des hormones hypophysaires et placentaires

Informations scientifiques

Les hormones antéhypophysaires et placentaires forment deux familles multigéniques : la famille des gènes LH, FSH, TSH et HCG et la famille des gènes GH, HPRL et HLP.

La famille des gènes LH, FSH, TSH et HCG

Les hormones antéhypophysaires LH, FSH et TSH et l'hormone HCG sont des glycoprotéines formées par l'assemblage de deux chaînes polypeptidiques, les chaînes alpha et bêta. Ces quatre hormones diffèrent uniquement par leur chaîne bêta qui est la chaîne responsable de la spécificité fonctionnelle de ces hormones.

Ces hormones jouent des rôles différents dans l'organisme : les hormones LH et FSH participent au contrôle hormonal du fonctionnement de l'appareil reproducteur chez l'homme et chez la femme, l'hormone TSH stimule le fonctionnement thyroïdien et l'hormone HCG stimule le corps jaune lors d'un début de grossesse.

Ces chaînes bêta sont codées par des gènes qui sont situés sur les chromosomes différents (chromosome 1 pour le gène de TSH bêta, chromosome 11 pour le gène de FSH bêta et chromosome 19 pour les gènes de LH bêta et de HCG bêta).

Plusieurs arguments permettent de dire que ces gènes forment une famille multigénique :

- ces quatre gènes présentent la même structure globale (trois exons et deux introns) ;
- ces quatre gènes et les protéines pour lesquelles ils codent présentent des similitudes de séquences assez importantes ;
- au niveau protéique, on peut constater la présence de douze résidus cystéine à des sites identiques pour les quatre hormones. Or, ces résidus, en permettant la formation des ponts disulfures, jouent un rôle essentiel dans la stabilisation de la structure tridimensionnelle de ces chaînes et donc de leurs propriétés.

On peut donc considérer que les similitudes constatées au niveau protéique et nucléique traduisent une origine commune pour toutes ces hormones et les gènes qui les codent. Ces gènes sont issus d'un même gène ancestral, et ils sont apparus par duplications géniques suivies de mutations indépendantes. La duplication la plus récente est celle qui est à l'origine des gènes LH bêta et HCG bêta. Ce qu'il est important de noter ici, c'est que la divergence **des copies dupliquées a conduit à des gènes codant pour des polypeptides ayant des fonctions différentes.**

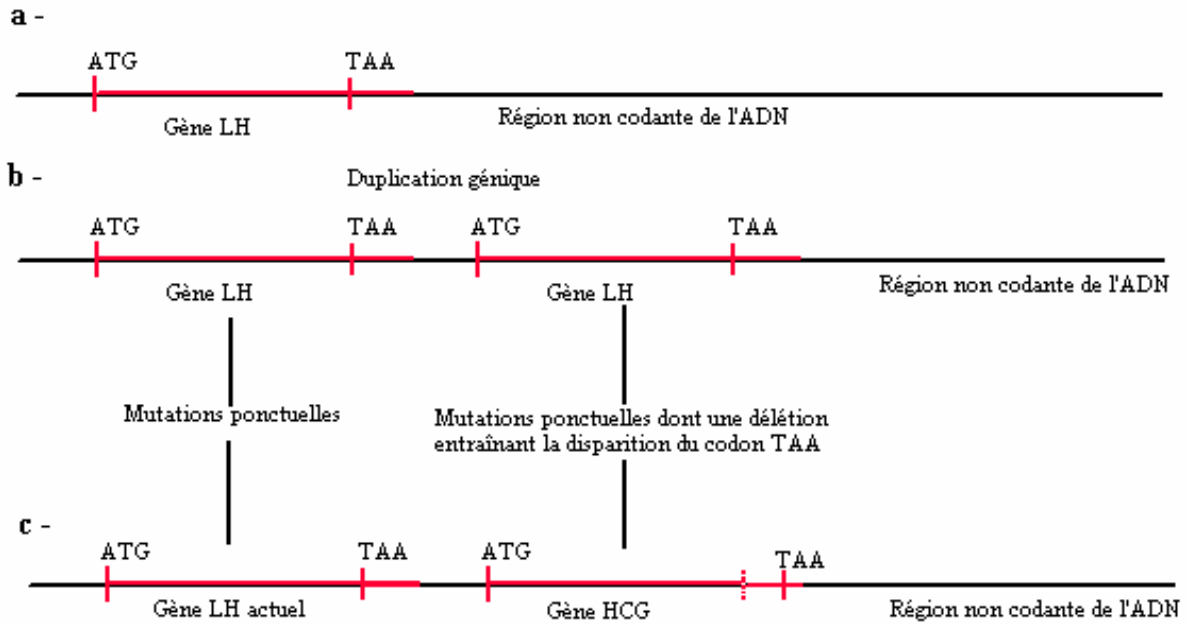
Tous les Vertébrés possèdent le gène de LH bêta, mais seuls les Mammifères placentaires (notamment les Primates et les Équidés) possèdent le gène de HCG bêta. On peut donc penser que le gène HCG bêta provient d'un gène LH bêta ancestral, la duplication étant postérieure à l'apparition des Mammifères placentaires.

La comparaison des protéines LH bêta et HCG bêta montre que ces deux protéines n'ont pas la même longueur : 165 acides aminés pour HCG bêta et 141 acides aminés pour LH bêta (méthionine comprise). Or, la comparaison des séquences nucléiques des gènes de ces deux hormones montre une délétion d'un nucléotide (en position 402) ; cette délétion entraîne un décalage du cadre de lecture lors de la traduction, décalage qui fait disparaître le codon stop présent dans le gène de LH (codon n° 142). C'est cette mutation par délétion qui est donc responsable de l'allongement de la protéine HCG bêta par rapport à la protéine LH bêta.

Remarque

La duplication du gène ne concerne pas uniquement la séquence strictement codante mais aussi celle qui se trouve en amont et en aval de celle-ci. Pour faciliter la compréhension de l'allongement de la séquence codante de HCG (et donc de sa protéine) par rapport à LH, nous avons fourni une séquence appelée LHADNC constituée par la séquence codante de LH, suivie d'une séquence de nucléotides homologue de la fin de celle de HCG mais non transcrite dans LH. La comparaison de LHADNC et de HCG permet d'expliquer que la protéine HCG a 24 acides aminés de plus que LH : par suite de la délétion, le gène HCG incorpore dans sa séquence codante la région en aval de LH non transcrite.

On peut conduire les élèves à la représentation suivante :



a - Situation initiale il y a plus de 60 Ma dans une espèce ancestrale des Primates.

b - Duplication génique de LH bêta.

c - Situation actuelle : la délétion dans le duplicata, en entraînant un décalage du cadre de lecture, a fait disparaître le codon stop. La région codante du gène incorpore une partie de l'ADN non codant situé en aval, jusqu'à un nouveau codon stop. En toute rigueur, pour bien comprendre tout cela, il faut savoir que l'ARNm comporte une région normalement non codante après le codon stop UAA. Si ce codon stop disparaît, l'ARNm est traduit jusqu'à la rencontre d'un nouveau codon stop.

La famille des gènes GH, HPRL et HLP

Les hormones GH (hormone de croissance produite par l'antéhypophyse), HPRL (hormone prolactine, produite par l'antéhypophyse) et HLP (hormone lactogène placentaire produite par le placenta) ont des actions différentes mais apparentées. La HLP agit sur la glande mammaire en stimulant la synthèse protéique, comme la HPRL (toutefois plus efficace), et la GH stimule également la synthèse protéique dans de nombreux tissus.

Ces hormones sont codées par des gènes situés sur des chromosomes différents : chromosome 6 pour le gène de la HLP, chromosome 17 pour les gènes de la GH et la HPRL.

Les similitudes de séquences nucléiques et protéiques constatées permettent de dire que ces gènes forment une famille multigénique :

Ces trois gènes ont donc une origine commune et sont issus d'un même gène ancestral par duplications géniques suivies de mutations différentes.

La duplication la plus récente est celle qui est à l'origine des gènes GH et HLP.

Tous les Vertébrés tétrapodes possèdent les gènes GH et HPRL, mais seuls les Mammifères placentaires possèdent le gène HLP.

On peut donc considérer qu'une première duplication (à l'origine des gènes GH et HPRL) a eu lieu chez les premiers tétrapodes, avant que n'apparaissent les Amphibiens, et qu'une deuxième duplication a eu lieu beaucoup plus récemment, il y a environ 100 Ma.

Pistes d'exploitation pédagogique des données fournies

L'exploitation des données fournies sur les hormones antéhypophysaires et placentaires permet de dégager les notions liées aux mécanismes d'évolution du génome, et notamment à l'apparition de nouveaux gènes par duplication génique suivie de mutations indépendantes.

Ici, **les mutations qui sont apparues ont amené à des fonctionnalités nouvelles** pour les protéines codées puisque toutes ces hormones ont des rôles différents. Enfin, les gènes codant pour ces hormones étant situés sur des chromosomes différents, on illustre bien avec l'étude de cet exemple la souplesse du génome au cours de l'évolution.

Les gènes des hormones étudiées ici constituent deux familles multigéniques : la famille des gènes LH, FSH, TSH et HCG, et la famille des gènes de la GH, HPRL et HLP.

La comparaison des séquences protéiques et nucléiques au sein de chacun des groupes proposés (famille des gènes LH, FSH, TSH, HCG et famille des gènes GH, HPRL et HLP) permet de mettre en évidence des similitudes moléculaires et de dégager la notion de familles multigéniques. La notion de duplication génique suivie de mutations indépendantes permet alors d'expliquer l'apparition de nouveaux gènes à partir d'un même gène ancestral.

La prise en compte du taux d'identités permet de préciser les relations de parenté au sein de chacune de ces familles. Les données paléontologiques fournies permettent quant à elles de comprendre comment on peut dater approximativement certaines duplications géniques.

Enfin, les données fournies peuvent servir de support à une réflexion sur les effets des mutations et sur leur conservation :

- l'étude des gènes et protéines LH et HCG permet de montrer comment une mutation peut, en supprimant un codon stop, entraîner un allongement de la protéine synthétisée, ce qui a contribué, avec l'apparition d'autres mutations, à l'obtention d'une protéine aux fonctionnalités nouvelles ;
- l'alignement multiple des séquences protéiques des hormones LH, FSH, HCG et TSH permet de repérer l'existence de sites identiques remarquables, en particulier les cystéines. Or, ces résidus, en permettant la formation de ponts disulfures, jouent un rôle essentiel dans la stabilisation de la structure tridimensionnelle de la protéine et donc dans ses propriétés. On peut alors discuter du fait que ces sites soient très conservés d'une hormone à l'autre : une mutation à leur niveau doit arriver (aussi fréquemment qu'ailleurs dans la molécule), mais ne se fixe pas car elle entraîne une non-fonctionnalité de la protéine, donc nuit à la survie de l'individu porteur qui ne peut alors pas la transmettre.

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques - Duplications et familles multigéniques** permet d'atteindre :

- **Gènes GH, HPRL et HLP** pour accéder à :

- **Nucléiques** qui charge le fichier genes-familleGH.edi affichant les séquences nucléiques strictement codantes (du codon d'initiation au codon stop) des gènes de GH (hormone de croissance produite par l'antéhypophyse), HPRL (hormone prolactine produite par l'antéhypophyse), HLP (hormone lactogène placentaire produite par le placenta) ;

- **Protéiques** qui charge le fichier hormones-familleGH.edi affichant les séquences protéiques des hormones GH (hormone de croissance produite par l'antéhypophyse), HPRL (hormone prolactine produite par l'antéhypophyse), HLP (hormone lactogène placentaire produite par le placenta).

- **Gènes des hormones FSH, LH, TSH et HCG** pour accéder à :

- **Nucléiques** qui charge le fichier genes-familleLH.edi affichant les séquences nucléiques strictement codantes (du codon d'initiation au codon stop) des gènes de LH (hormone lutéinique produite par l'antéhypophyse), FSH (*Folliculate Stimulating Hormon* produite par l'antéhypophyse), TSH (*Thyroid Stimulating Hormon*, produite par l'antéhypophyse) et HCG (hormone gonadotrophine chorionique, produite par les villosités choriales) ;

- **Protéiques** qui charge le fichier hormones-familleLH.edi affichant les séquences protéiques des hormones LH (hormone lutéinique produite par l'antéhypophyse), FSH (*Folliculate Stimulating Hormon* produite par l'antéhypophyse), TSH (*Thyroid Stimulating Hormon*, produite par l'antéhypophyse) et HCG (hormone gonadotrophine chorionique, produite par les villosités choriales).

Documents fournis

Dans la banque de documents, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques/Duplications et familles multigéniques** permet d'atteindre **Gènes des hormones FSH, LH, TSH et HCG** et **Gènes GH, HPRL et HLP** pour accéder à :

- **Information hormones** qui charge le fichier infoshormones.bmp présentant, pour chacune des hormones considérées ici, la localisation chromosomique du gène, le rôle, les cellules productrices et les groupes de Vertébrés qui la synthétisent. Ce document peut être utilisé pour discuter des phénomènes de transposition de gène lors de la duplication, de l'acquisition de nouvelles fonctionnalités pour les protéines codées par les nouveaux gènes produits et de la datation approximative de certaines duplications géniques ;

- **Duplication LH HCG** qui charge le fichier dupliLH_HCG.jpg et **Duplication GH HLP** qui charge le fichier dupliGH_HLP.jpg : ces documents peuvent servir de base de discussion pour poser la notion de duplication génique ; ils peuvent être à compléter par les élèves ou à justifier à partir des analyses de séquences effectuées ;

- **Phylogénie famille GH** qui charge le fichier phylofamilleGH.jpg résumant l'histoire évolutive de la famille des gènes GH/HPRL/HLP et qui peut être complété par les élèves ou à justifier à partir des analyses de séquences effectuées. Une autre utilisation possible de ces documents consisterait à les considérer comme des modèles de référence de l'histoire d'une famille multigénique donnée, et à demander aux élèves en quoi ils sont confortés par l'analyse des séquences à leur disposition.

Comparaisons obtenues pour la famille des gènes LH, FSH, TSH et HCG

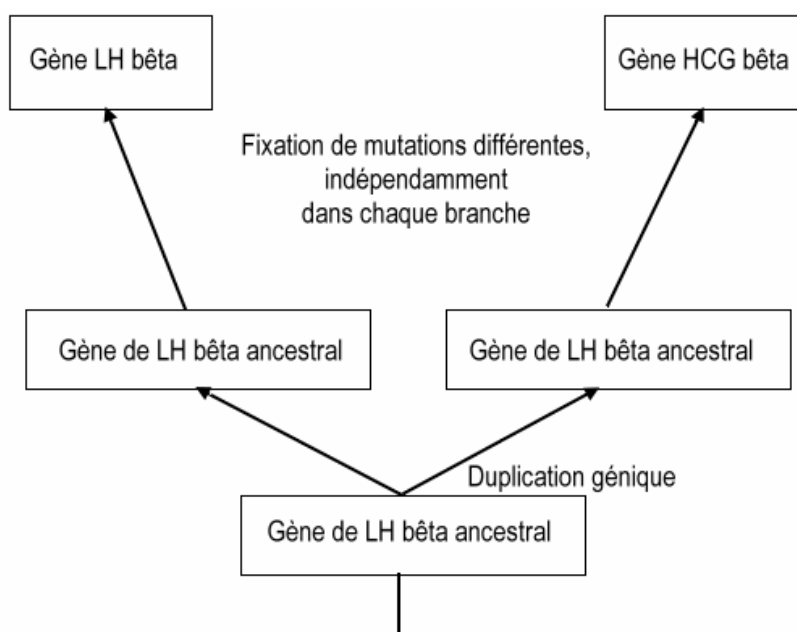
	LH bêta	HCG bêta	FSH bêta	TSH bêta
LH bêta	100 %			
HCG bêta	79,1 %	100 %		
FSH bêta	54,4 %	54,4 %	100 %	
TSH bêta	45,5 %	46,5 %	45,8 %	100 %

Matrice des identités obtenue à partir d'un alignement avec discontinuité des gènes de ces hormones

	LH bêta	HCG bêta	FSH bêta	TSH bêta
LH bêta	100 %			
HCG bêta	71,5 %	100 %		
FSH bêta	34,1 %	33,3 %	100 %	
TSH bêta	39,1 %	34,8 %	35,5 %	100 %

Matrice des identités obtenue à partir d'un alignement avec discontinuité des séquences protéiques de ces hormones

Le plus fort taux d'identité est entre les gènes LH bêta et HCG bêta (ou entre les hormones elles-mêmes). On en déduit que ces deux gènes dérivent d'un gène ancestral récent, par un phénomène de duplication génique ; les différences qu'ils présentent aujourd'hui s'expliquent par la fixation de mutations différentes dans les deux cas :



La LH bêta étant présente chez tous les Vertébrés alors que la HCG bêta n'étant présente que chez quelques ordres de Mammifères placentaires (Primates, Équidés), on peut dater approximativement la duplication à l'origine des gènes de ces deux hormones : elle est postérieure à l'apparition des Mammifères placentaires. Les Équidés n'étant pas les Mammifères les plus étroitement apparentés aux Primates, on peut supposer que l'apparition de HCG résulte d'une évolution convergente dans les lignées des Primates et des Équidés. D'autre part, cela permet de dire que le gène ancestral devait coder pour une molécule de type LH bêta.

La mise en relation des mutations dans les gènes LH bêta et HCG bêta avec les différences observées au niveau protéique permet de discuter de l'effet de certaines mutations : ici, une mutation par délétion en position 402 apparue dans le gène HCG bêta provoque la disparition d'un codon stop et donc l'allongement de la protéine synthétisée (la HCG bêta possède 165 acides aminés, la LH bêta n'en possède que 141).

Les taux d'identité entre les gènes FSH bêta, TSH bêta et LH bêta/HCG bêta sont suffisants pour dire que ces gènes sont homologues (et que ces protéines sont homologues), mais les différences ne sont pas assez significatives pour pouvoir réaliser une phylogénie pertinente (entre 54,4 % et 45,5 %, la différence n'est pas assez nette pour pouvoir vraiment distinguer sur une phylogénie les duplications à l'origine de ces gènes, d'un point de vue chronologique). Comme on obtient quatre gènes à partir d'un gène ancestral unique, on peut dire qu'il y a eu trois duplications dans cette famille de gènes.

Les gènes des hormones FSH et TSH n'étant pas sur le même chromosome que les gènes LH et HCG, il faut également faire intervenir un phénomène de transposition (ou translocation) de gènes lors de chacune des deux duplications les plus anciennes. Les quatre hormones n'ayant pas les mêmes rôles, on peut discuter de l'effet des mutations apparues et fixées différemment selon les branches ; ces mutations sont à l'origine des fonctionnalités nouvelles acquises par ces hormones.

Comparaisons obtenues pour la famille des gènes GH, HPRL et HLP

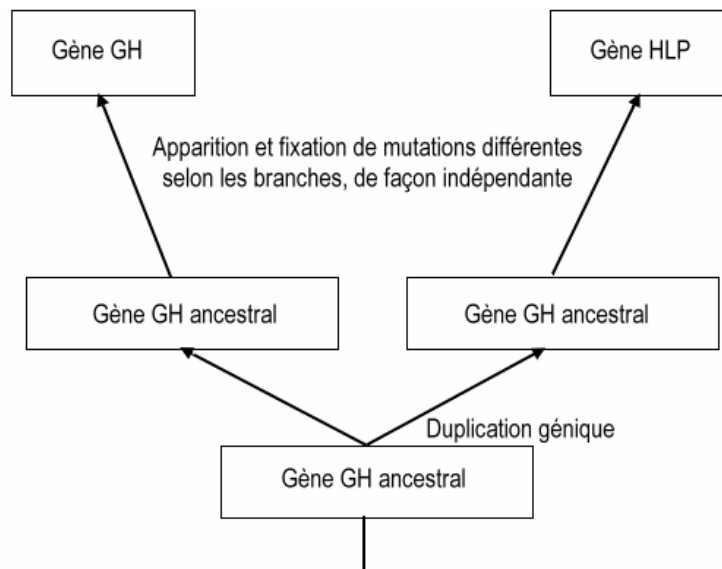
	GH	HPRL	HLP
GH	100 %		
HPRL	44,3 %	100 %	
HLP	92,7 %	47,1 %	100 %

Matrice des identités obtenue à partir d'un alignement avec discontinuité des gènes de ces hormones

	GH	HPRL	HLP
GH	100 %		
HPRL	22,9 %	100 %	
HLP	85,3 %	24,4 %	100 %

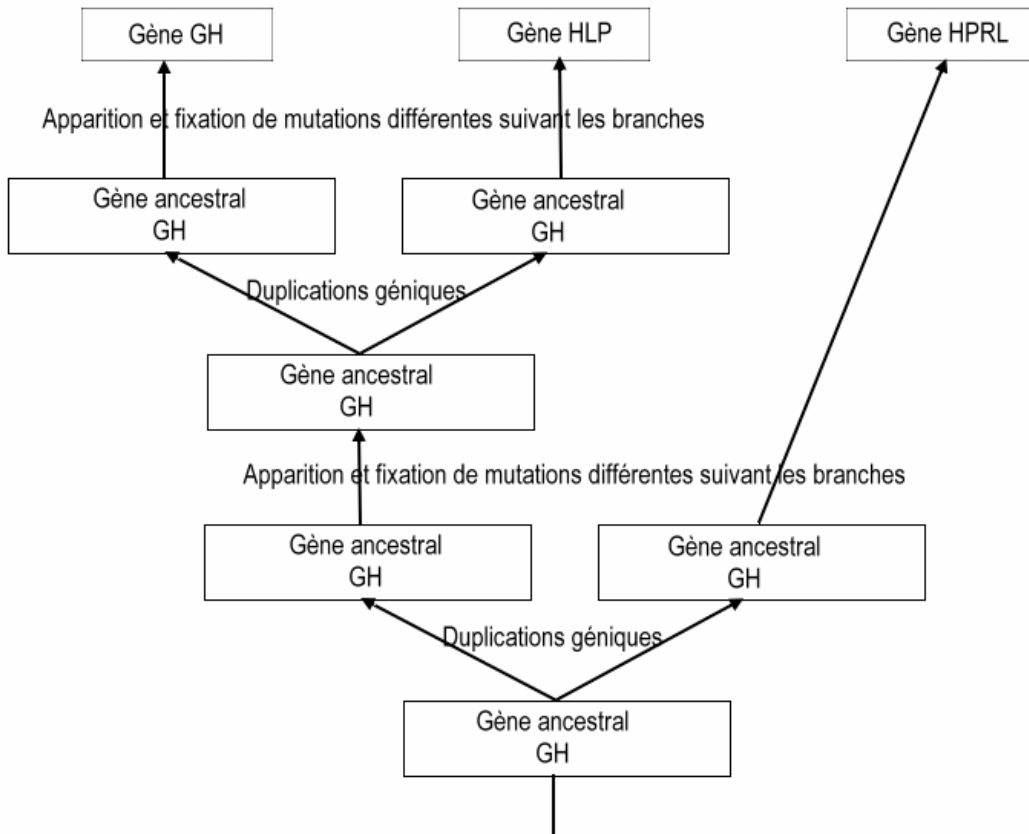
Matrice des identités obtenue à partir d'un alignement avec discontinuité des séquences protéiques de ces hormones

Le plus fort taux d'identité est entre les gènes GH et HLP (ou entre les hormones elles-mêmes). On en déduit que ces deux gènes dérivent d'un gène ancestral récent, par un phénomène de duplication génique ; les différences qu'ils présentent aujourd'hui s'expliquent par la fixation de mutations différentes dans les deux cas :



La GH étant présente chez tous les Vertébrés alors que la HLP n'est présente que chez les Mammifères placentaires, on peut dater approximativement la duplication à l'origine des gènes de ces deux hormones chez l'Homme : elle a eu lieu chez une espèce à l'origine de tous les placentaires actuels, soit environ entre 85 et 100 Ma. D'autre part, cela permet de dire que le gène ancestral devait coder pour une molécule de type GH.

Les taux d'identité entre les gènes HLP et GH/HPRL sont beaucoup plus faibles, mais suffisants pour dire que ces gènes sont homologues (et que ces protéines sont homologues), surtout si l'on considère la conservation de plusieurs acides aminés à des positions précises dans les protéines codées et la structure globale identique de ces gènes. Comme on obtient trois gènes à partir d'un gène ancestral unique, on peut dire qu'il y a eu deux duplications dans cette famille de gènes.



Les gènes des hormones GH et HLP n'étant pas sur le même chromosome que le gène HPRL, il faut également faire intervenir un phénomène de transposition (ou translocation) de gènes lors de la duplication la plus ancienne (tous les Vertébrés possédant les gènes GH et HPRL, on peut dire que la duplication à l'origine de ces deux gènes est plus ancienne que l'apparition du groupe des Amphibiens, soit il y a environ 400 Ma).

Les trois hormones n'ayant pas les mêmes rôles, on peut discuter de l'effet des mutations apparues et fixées différemment selon les branches ; ces mutations sont à l'origine des fonctionnalités nouvelles acquises par ces hormones.

Innovations génétiques – Duplications et familles multigéniques

– Gènes homéotiques

Informations scientifiques

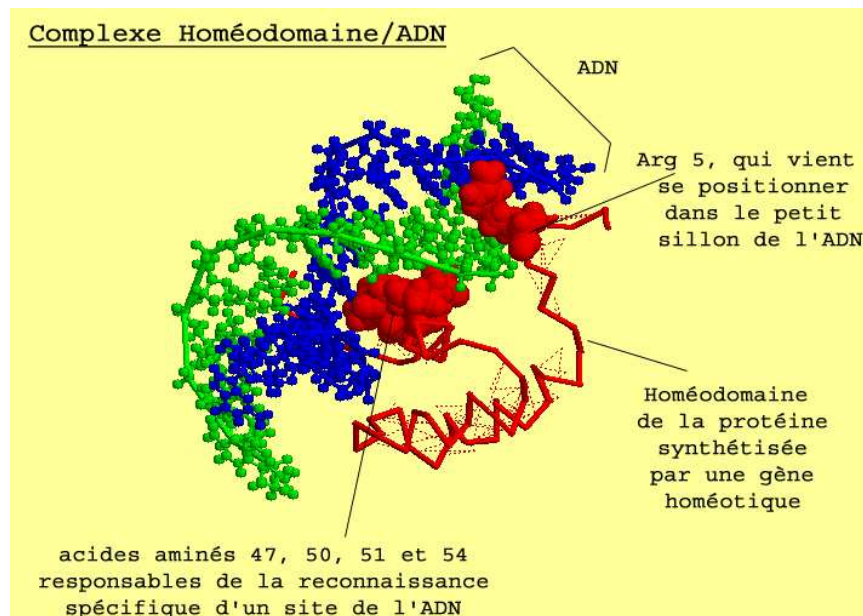
[Cf. page 45 \(classe de seconde\)](#)

Les gènes homéotiques interviennent au cours du développement embryonnaire, notamment pour diriger l'expression d'autres gènes et permettre ainsi la mise en place des différents organes ; on les qualifie de gènes « maîtres » ou gènes « architectes ». Ils déterminent la disposition des organes le long de l'axe antéro-postérieur.

Les gènes homéotiques sont des gènes régulateurs (activation ou inhibition) de l'expression d'autres gènes.

Ce sont des gènes à homéoboîte (ou homeobox), c'est-à-dire qu'ils codent pour une protéine dont un domaine est très conservé au cours de l'évolution. Ce domaine protéique, appelé homéodomaine, a été découvert en 1983 ; il comporte soixante acides aminés et sa structure secondaire comporte trois hélices formant le motif HLH (Helix Loop Helix). Ce domaine se fixe à l'ADN, ce qui contribue à ouvrir localement l'ADN pour permettre la transcription.

La figure ci-dessous a été obtenue avec le logiciel *RasTop*.



Les gènes homéotiques sont organisés en complexes homéotiques chez les animaux.

Les complexes homéotiques

La disposition et l'organisation des gènes homéotiques sont bien connues chez deux organismes : la Drosophile et la Souris.

Chez la Drosophile, les gènes homéotiques forment le complexe Hom-C porté par le chromosome 3 ; ce complexe est composé de deux groupes de gènes : le groupe Antennapedia et le groupe Bithorax, chacun de ces groupes comprenant plusieurs gènes :

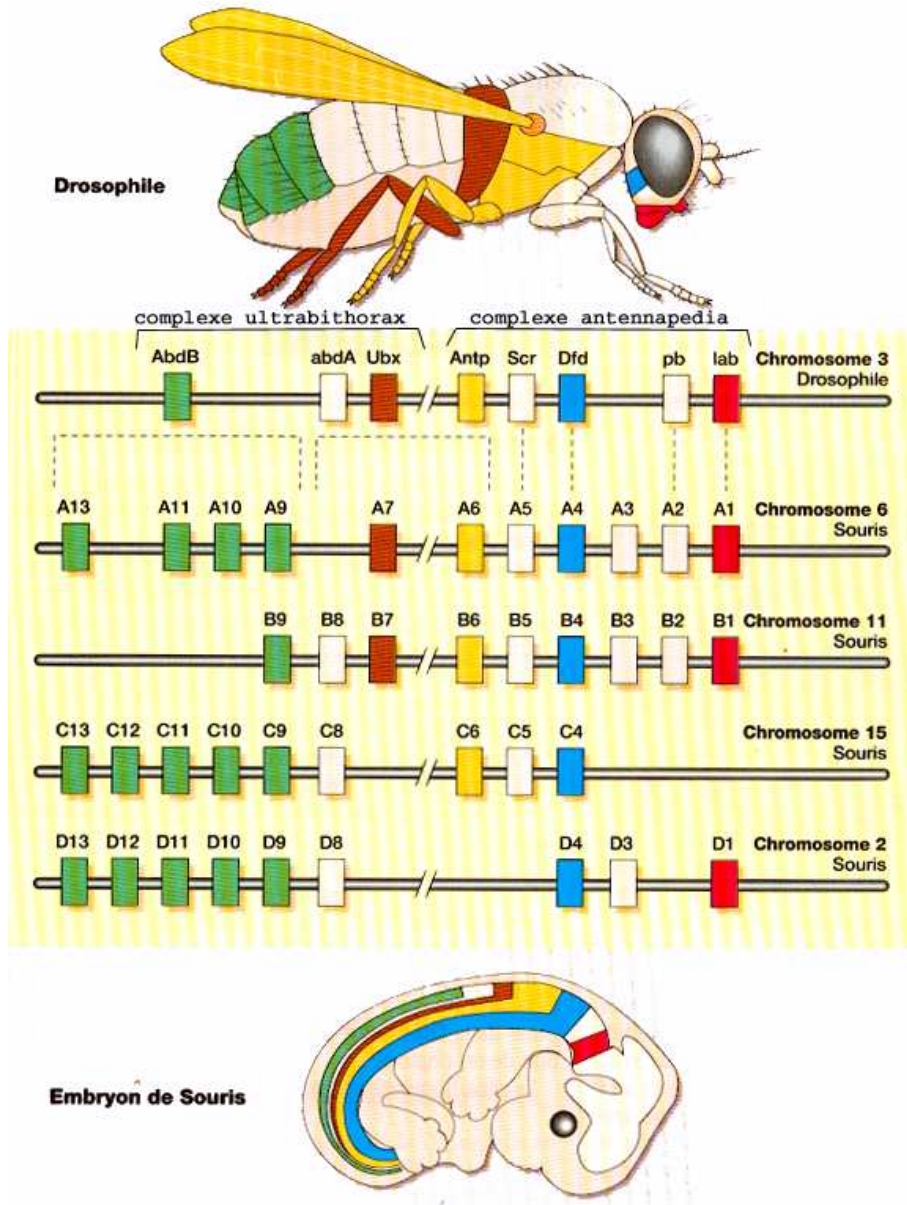
- Antennapedia : gènes lab (labial), Antennapedia (antp), Sex comb reduced (Scr), Deformed (dfd), proboscipedia (pb) ;
- Bithorax : gènes ultrabithorax (ubx), abdominal A (AbdA) et abdominal B (AbdB).

Chez la Souris, les gènes homéotiques forment quatre complexes Hox répartis sur quatre chromosomes :

- Complexe des gènes Hox A sur le chromosome 6 ;
- Complexe des gènes Hox B sur le chromosome 11 ;
- Complexe des gènes Hox C sur le chromosome 15 ;
- Complexe des gènes Hox D sur le chromosome 2.

Il y a colinéarité entre la disposition des gènes homéotiques sur les chromosomes et les régions du corps dans lesquelles ils s'expriment : en parcourant l'ADN de 3' vers 5', on trouve des gènes dont les lieux d'action s'échelonnent de l'avant vers l'arrière de l'animal.

Disposition des gènes homéotiques et secteurs d'expression correspondants



Lors de l'embryogenèse, la combinaison des produits de plusieurs gènes Hox donne une identité relative aux cellules embryonnaires le long de l'axe antéro-postérieur. Chaque cellule embryonnaire est affectée d'une valeur positionnelle, qui résulte d'une combinaison de plusieurs gènes Hox.

Pistes d'exploitation pédagogique

Séquences et documents

Fichiers des séquences

Dans la banque de thèmes d'étude, le chemin **Stabilité et variation des génomes et évolution/Innovations génétiques - Duplications et familles multigéniques/Gènes homéotiques** permet d'atteindre :

- **Nucléiques** qui charge le fichier BoxFamillesHomme.edi affichant les séquences de gènes chez l'Homme et la Drosophile ;
- **Protéiques** qui charge le fichier DomFamillesHomme.edi affichant les séquences protéiques correspondantes.

Les gènes homéotiques des Arthropodes et des Vertébrés forment une famille multigénique

Au sein d'un même organisme, par exemple la Souris, tous les gènes homéotiques présentent une similitude importante de séquence au niveau de l'homéoboîte : ce sont donc des gènes homologues, cette homologie témoignant de leur origine commune. De plus, l'ordre des gènes sur chaque chromosome (donc au sein d'un même complexe) est similaire d'un complexe à l'autre. On peut donc dire que l'on a des gènes paralogues (exemple : HoxA4, HoxB4, HoxC4 et HoxD4 sont paralogues).

Comparaisons entre les gènes homéotiques des Vertébrés et des Arthropodes :

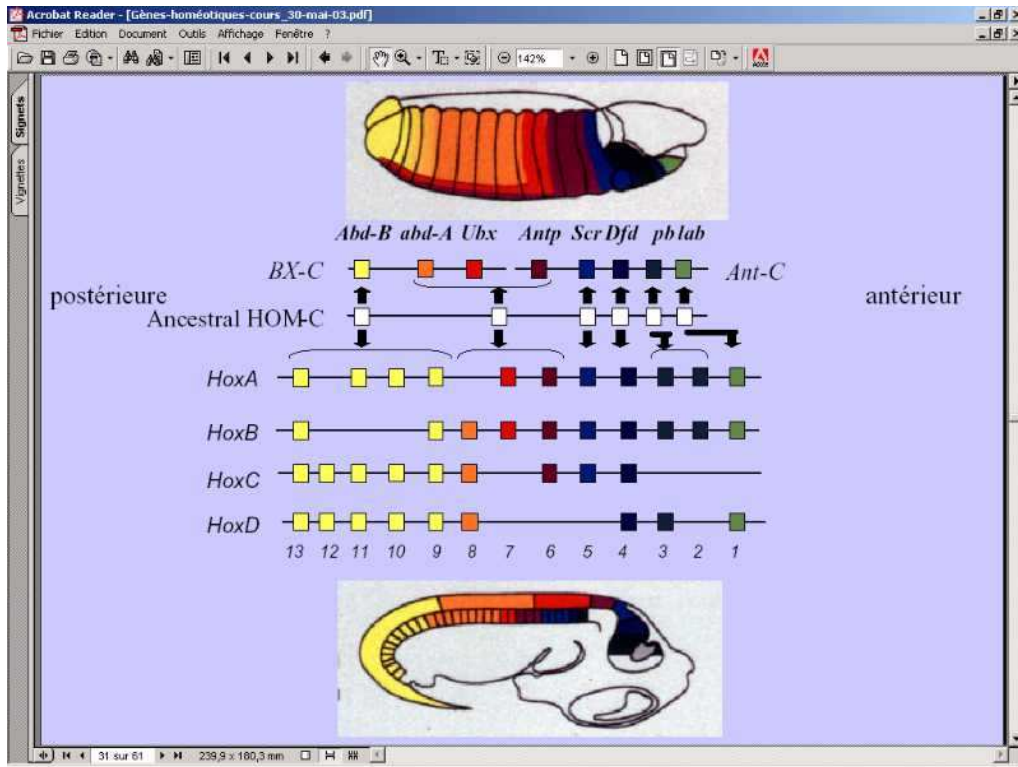
- les gènes des complexes Hox de Souris et Hom-C de Drosophile présentent des similitudes dans leur ordre de disposition sur les chromosomes et, dans les deux cas, ils sont impliqués dans le positionnement et l'identité cellulaire le long de l'axe antéro-postérieur ;
- des expériences de transgénèse interspécifique ont été réalisées avec succès :
 - la substitution du gène lab muté chez la Drosophile (mutation qui entraîne la formation d'une tête très petite) par le gène Hox B1 de Poulet restaure la morphogénèse normale de la tête de Mouche ;
 - l'insertion du gène Hox B6 de Souris chez la Drosophile entraîne la transformation des antennes en pattes thoraciques ;
 - d'une manière générale, un gène homéotique de Vertébré possède son homologue chez les Insectes ; inséré par transgénèse interspécifique dans le génome de la Drosophile, le gène homéotique de la Souris ou de l'Homme induit des transformations homéotiques semblables à celles produites par son homologue ;
- la comparaison de tous les gènes homéotiques présents chez les Arthropodes et les Vertébrés met en évidence une grande similitude au niveau de l'homéoboîte. Cette similitude de séquence permet de dire que tous ces gènes sont homologues, c'est-à-dire qu'ils ont une origine commune ; ils forment donc une famille multigénique. Tous ces gènes se sont formés à partir d'un même gène ancestral par duplications géniques suivies de mutations.

	Domaine-B4 Homme	Domaine-B7 Homme	Domaine-Dfd Drosophile	Domaine-Ubx Drosophile
Domaine-B4 Homme	100			
Domaine-B7 Homme	83,3	100		
Domaine-Dfd Drosophile	88,3	83,3	100	
Domaine-Ubx Drosophile	75	90	73,3	100

Tableaux des identités (en %) entre plusieurs homéodomains de gènes homéotiques de Vertébrés et d'Arthropodes (valeurs obtenues à l'aide d'un alignement réalisé avec le logiciel Anagène)

Le fait qu'il y ait plus de similitudes entre les gènes Hoxb4 de l'Homme et Dfd de la Drosophile qu'entre les gènes HoxB4 et HoxB7 indique que la duplication à l'origine de B4 et B7 est antérieure à la séparation des lignées des Insectes et des Vertébrés. Autrement dit, l'ancêtre commun à l'Homme et à la Drosophile avait un gène à l'origine du Dfd de la Drosophile et du B4 de l'Homme. En outre, il possédait un gène à l'origine du gène HoxB7 et du gène Ubx (plus forte similitude entre Ubx et B7 qu'entre HoxB4 et HoxB7). Le même raisonnement peut être tenu pour les gènes HoxB1, HoxB9, AbdB et Lab.

Des comparaisons du même type entre les gènes homéotiques des Vertébrés et de la *Drosophile* ont conduit à proposer le schéma évolutif suivant :



Ce schéma indique que l'ancêtre commun aux Vertébrés et aux Insectes possédait un complexe homéotique formé de six gènes, eux-mêmes résultant de duplications antérieures. Ces six gènes sont à l'origine des complexes Hom de la *Drosophile* et Hox des Vertébrés. L'*Amphioxus* est un Cordé qui ne possède qu'un seul complexe Hox. Les deux duplications à l'origine des quatre complexes existant chez tous les Vertébrés ont donc eu lieu dans la lignée des Cordés aboutissant à l'ancêtre commun à tous les Vertébrés.

Il est possible de faire travailler les élèves sur le schéma évolutif ci-dessus, les amener à extraire les informations essentielles puis à tester la validité de ce schéma à partir des séquences des gènes homéotiques de la *Drosophile* et de la Souris.