

EXERCICES DE GENETIQUE ET DE GENETIQUE MOLECULAIRE

PARTIE I

Ces exercices d'applications de cours ont pour but de présenter les principales notions de la génétique fondamentale. Ils sont proposés avec un corrigé. Quelques mots clés indiquent les points abordés par chaque sujet.

Polymorphisme et Phénylcétonurie Conséquences de diverses mutations sur le phénotype, interactions entre gènes et avec l'environnement	2
Maladies de la coagulation Etudes de pedigrees, calculs de probabilité	14
Dystrophie des ceintures Pedigree, test de complémentation fonctionnelle, diagnostic prénatal	33
Facteur IX Traduction, les mutations et leurs conséquences au niveau des protéines	38
Diagnostic prénatal de l'atrophie gyrée Test de complémentation fonctionnelle, cellules eucaryotes, diagnostic prénatal	43

Monique MASSELOT
Génétique
Université P. et M. Curie, Paris.

Polymorphisme et phénylcétonurie

Les symptômes cliniques de la phénylcétonurie sont une grave arriération mentale et des troubles du caractère. La plupart des sujets atteints de phénylcétonurie et non traités, sont victimes d'un retard mental extrêmement sévère. Le poids du cerveau de ces individus est au-dessous de la moyenne, la myélinisation de leurs nerfs est défectueuse et leurs réflexes sont hyperactifs. L'espérance de vie des phénylcétonuriques non traités est considérablement raccourcie. 50% d'entre eux sont morts à l'âge de vingt ans, et les trois quarts à l'âge de trente ans. Cette atteinte du cerveau est due à l'effet toxique d'un excès de phénylalanine (Phé).

Cet excès de phénylalanine est dû à un déficit en phénylalanine hydroxylase qui catalyse la transformation de Phé en Tyrosine (figure 1). La phénylalanine non utilisée s'accumule.

La phénylalanine hydroxylase active qui catalyse la transformation est un homo tétramère qui interagit avec la tétrahydrobioptérine comme cofacteur. Ce cofacteur est recyclé cycliquement.

1°) Formes du gène

Le gène qui code la phénylalanine hydroxylase (PAH) est situé sur le bras long du chromosome 12 (bande 12q23.2). Il est long de 90kbp et contient 12 exons. L'ARNm mature correspondant fait 2,4kbp. On connaît actuellement 463 formes différentes (allèles) de ce gène, qui entraînent ou non des phénotypes modifiés.

Statistiques par type mutationnel







Type de mutation	No.	Représentation graphique	%
Faux sens	291		62.85%
Délétion	60		12.96%
Epissage	52		11.23%
Silencieuse	26		5.62%
Non sens	24		5.18%
Insertion	6		1.30%

Tableau 1. Répartition de la plupart des formes alléliques du gène PAH.

Expliquez en quoi consiste chacun de ces types mutationnels. Se définissent-ils au niveau de l'ADN ou dans les conséquences au niveau protéique ? En conclusion faites un tableau des modifications de l'ADN et des conséquences qu'elles entraînent sur une protéine

En observant la structure du code génétique donnez une méthode pour déterminer ce que la modification d'une base pourrait entraîner au niveau protéique. Prenez comme exemple l'arginine en considérant les 6 codons possibles.

Parmi les allèles du gène PAH on trouve de nombreux cas où Arg (R) est remplacé par un autre acide aminé (ou un stop=X). Pour une même position, on peut trouver différents AA, d'une position à l'autre les remplaçants ne sont pas les mêmes (voir tableau 2). Quelles sont les causes de ces différences. Déterminez pour chaque cas, le triplet d'origine du brin non transcrit de l'ADN.

	Position où Arg (R) est trouvé dans la séquence de l'allèle de référence					
	176	241	243	252	261	413
AA trouvé à cette position chez différents allèles	Stop (X) Phé (P) Leu (L)	Leu (L) Cys (C) His (H)	Stop (X) Gln (Q) Leu (L)	Trp (W) Gln (Q) Gly (G)	Stop (X) Gln (Q) Phé (P)	Ser (S) Phé (P) Cys (C)

Tableau 2. Quelques formes alléliques du gène PAH.

2°) Phénotype : Influence du milieu

Un couple a déjà un enfant de sexe mâle atteint de PKU et âgé de 9 ans. La phénylalanine plasmatique atteint un taux supérieur à 1000 µm/l (lorsque l'enfant ne suit aucun régime). Chez le second enfant, la PKU est diagnostiquée à la naissance (par le test de Guthrie). Ce second enfant est immédiatement soumis à un régime strict qui maintient son taux sérique à un niveau normal (120 µm/l). Actuellement le second enfant a 2 ans et demi et présente un développement normal tant physique que mental alors que son frère de 11 ans ½ présente un retard mental sévère (QI de 30). L'étude génétique des deux enfants a permis de montrer qu'ils étaient tous deux porteurs des deux mêmes formes mutées pour le gène PAH. (un allèle R408W et l'autre R252W).

Quelle conclusion s'impose ? Donnez une définition générale du phénotype.

3°) Phénotype des mutants de PAH

Parmi les 463 formes différentes du gène PAH on a pu en caractériser quelques unes : 31 modifications de la séquence protéique de l'enzyme n'entraînent aucune modification de l'individu ni de son taux de phénylalanine sérique.

Parmi les autres formes on distingue 4 cas .

- un taux de Phé sérique supérieur à la normale mais inférieur à 600 µm/l. Les sujets qui portent de tels allèles ne présentent aucun des signes cliniques de PKU ils ont juste une hyperphénylalaninémie (HPA non PKU) avec une alimentation normale.
- un taux sérique qui atteint ou dépasse 1000 µm/l avec la nécessité d'un régime pour éviter les atteintes cérébrales
 - si la dose journalière tolérable de phé apportée dans les aliments est comprise entre 400 et 600 mg/jour la forme de PKU est légère
 - si la dose journalière tolérable de phé apportée dans les aliments est comprise entre 350 et 400 mg/jour la forme de PKU est modérée
 - si la dose journalière tolérable de phé apportée dans les aliments doit être inférieure à 250 mg/jour la forme de PKU est classique.

Parmi les allèles vus précédemment certains ont pu être classés car on disposait de sujets porteurs de la même mutation sur les deux chromosomes homologues (homozygotes). Le tableau 3 montre les résultats.

forme mutés	phénotype	phé sérique (sans régime)	tolérance pour phe
I65T	modéré à fort	608 μ m/l	400 mg/j
D129G	léger	130 μ m/l	>500 mg/j
R176L	HPA non PKU	119 μ m/l	alimentation normale
R252W	classique	>1500 μ m/l	<300mg/j
R261Q	modéré à fort	1200 μ m/l	<600mg/j
R261X	classique	>1500 μ m/l	<300mg/j
R408W	classique	>1200 μ m/l	<370mg/j
R408Q	HPA non PKU	600 μ m/l	alimentation normale
K363fsdelG*	classique	>1500 μ m/l	<300 mg/j
Y377fsdelT	classique	nd	nd
L348V	modéré à fort	nd	>330 mg/j
V388M	modéré à fort	900-1200 μ m/l	500-600 mg/j
Y414C	HPA non PKU	<600 μ m/l	alimentation normale
R408W	classique	>1200 μ m/l	<370mg/j
R408Q	HPA non PKU	600 μ m/l	alimentation normale
Y417H	léger	240 μ m/l	>450 mg/j
jonction introns 12-13 G→A	classique	>1500 μ m/l	<370 mg/j
jonction introns 4-5 G→T	classique	nd	nd

*K363fsdelG signifie qu'une délétion d'un G au niveau du codon 363 a provoqué l'apparition d'un faux sens à partir de ce point

Tableau 3. Phénotype associé à quelques allèles du gène HPA chez les homozygotes.

Comment pouvez-vous expliquer cette variabilité phénotypique parmi les différents allèles?

Dans le tableau 3, certaines relations génotype/phénotype vous semblent-elles prévisibles ?

Ces observations modifient-elles votre définition du phénotype ?

4°) Etude in vitro

L'ADNc de différentes formes mutées du gène PAH a été cloné dans un vecteur d'expression qui a ensuite été utilisé pour transformer des cellules en culture de mammifères. On a ainsi pu étudier le comportement de différentes formes alléliques dans un même contexte (la cellule hôte). Les résultats sont homogènes pour un allèle donné (tableau 4).

Allèle	Quantité d'enzyme décelable*	ARNm décelable*	Activité spécifique de l'enzyme*
I65T	26%	100%	nd
R176L	42%	nd	nd
R241C	25%	100%	nd
R241H	23%	nd	nd
R243Q	<10%	100%	nd
R243X	<1%	nd	0,2%
R252Q	<3%	100%	3%
R252G	<3%	100%	<5%
R252W	<1%	100%	0,5%
R261Q	20%	100%	20%
L348V	25%	nd	55%
V388M	20%	100%	43%
R408W	1%	100%	17%
R408Q	55%	93%	60%
jonction introns 12-13 G→A	<1%	100%	nd

*valeurs données en % de la valeur correspondante obtenue avec l'allèle de référence fonctionnel.

nd = non déterminé

Tableau 4. Quelques caractéristiques d'allèles clonés du gène PAH.

Ces résultats confirment-ils vos conclusions de la question précédente. Quelle est la différence entre les deux types d'analyses.

5°) Autres influences

L'observation de nombreux individus atteints a montré que dans certains cas, pour le même génotype concernant le gène HPA, le phénotype pouvait être différent (tableau 5).

Génotype du sujet atteint	phénotype associé	taux sérique de phénylalanine	dose tolérable de phé
R408W/P281L	PKU classique	>1200µm/l	<20mg/kg/j
R408W/P281L	PKU classique	<700	<24
R408W/R252W	PKU classique	>1200	<20
R408W/R252W	modéré	nd	normal
R408W/R261W	PKU classique	>122	<20
R408W/R261W	modéré	nd	normal
R408W/R261W	normal	nd	normal

Tableau 5. Phénotype et génotype de quelques sujets.

La figure 1 montre qu'en même temps que la phénylalanine est utilisée, une molécule de tétrahydrobioptérine (bh4) l'est aussi. Parallèlement à la dégradation de Phé il existe un cycle de restauration de bh4.

Quels sont les gènes qui agissent dans le recyclage de bh4 ? Quelle peut être l'action d'une mutation sur un de ces gènes ?

Proposez une hypothèse pour expliquer les différences phénotypiques entre sujets de même génotype PAH ?

Corrigez votre définition du phénotype en fonction de ces nouvelles observations.

6°) Fréquence des hétérozygotes

L'incidence de la PKU est d'environ 1/16 000 naissances, ce qui correspond à environ 1 hétérozygote sur 60 dans la population. On se demande pourquoi des allèles mutés qui invalident autant les sujets atteints persistent à une fréquence aussi élevée dans la population au lieu d'être progressivement éliminés.

Une hypothèse est que les hétérozygotes présenteraient un avantage sélectif tel que la résistance à un produit toxique (dans le cas présent l'ochratoxine A qui est produite par des champignons). Dans les siècles précédents, la population qui se nourrissait de pain à base de farine produite avec des grains infestés par divers champignons était intoxiquée par cette mycotoxine et mourait (un certain pourcentage) alors que les hétérozygotes porteurs d'un allèle sauvage et d'un allèle muté pour PAH résistaient tous à cette intoxication.

Donnez le phénotype d'un sujet homozygote pour l'allèle sauvage, ou pour l'allèle R408W ainsi que de l'hétérozygote (un allèle sauvage/R408W) en fonction

- de la phénylcétonurie
- de la résistance à l'ochratoxine

Dans chaque cas dites ce qui est dominant (cette notion ayant déjà étudiée dans le secondaire).

Dominance / récessivité vous semble-t-elle une propriété caractérisant l'allèle ou le phénotype que cet allèle confère ? Justifiez votre point de vue.

Bibliographie

Base de données à

<http://www.pahdb.mcgill.ca>

E. Kayaalp, E. Treacy, P. J. Waters, S. Byck, P. Nowacki, C. R. Scriver.(1997) Human Phenylalanine Hydroxylase Mutations and Hyperphenylalaninemia Phenotypes: A Metanalysis of Genotype-Phenotype Correlations. *Am. J. Hum. Genet.* **61**:1309–1317

P. J. Waters, M. A. Parniak, P. Nowacki, C. R. Scriver(1998) In Vitro Expression Analysis of Mutations in Phenylalanine Hydroxylase: Linking Genotype to Phenotype and Structure to Function. *Human Mutation* **11**:pp4- 17

M. Krawczak, J. Zschocke (2003) A Role for Overdominant Selection in Phenylketonuria? Evidence From Molecular Data. *Human Mutation* **21**:pp394-397

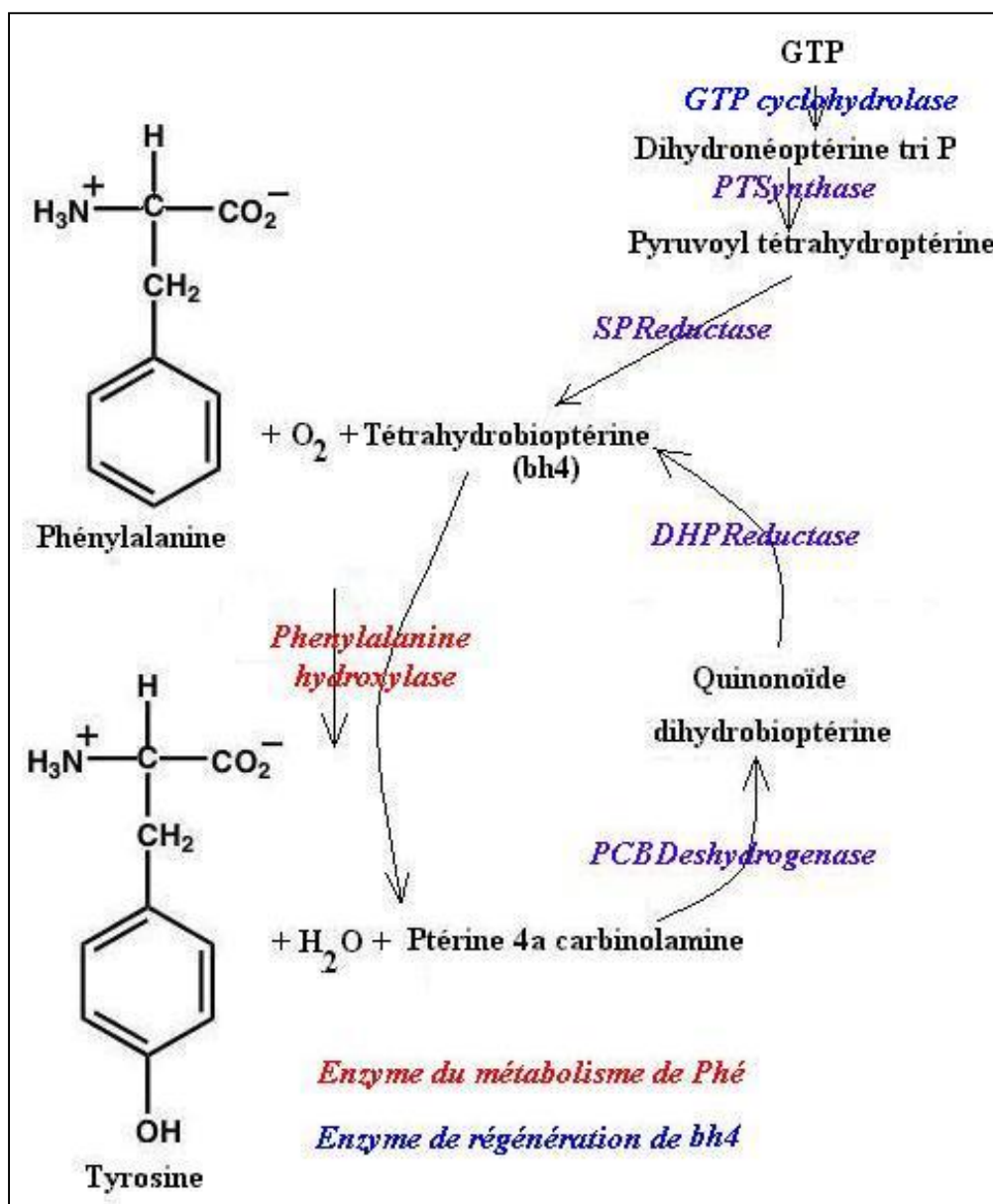


Figure 1. Utilisation de la phénylalanine et recyclage du cofacteur de la Phe hydroxylase.

Le code génétique								
		Seconde position du codon						
		U	C	A	G			
Première position du codon (ext 5')	U	UUU Phe(F)	UCU Ser(S)	UAU Tyr(Y)	UGU Cys(C)	U	Troisième position du codon (ext 3')	
		UUC Phe(F)	UCC Ser(S)	UAC Tyr (Y)	UGC Cys(C)			C
		UUA Leu L)	UCA Ser(S)	UAA Stop(X)	UGA Stop(X)			A
		UUG Leu(L)	UCG Ser(S)	UAG Stop(X)	UGG Trp(W)			G
	C	CUU Leu(L)	CCU Pro(P)	CAU His (H)	CGU Arg (R)	U		
		CUC Leu(L)	CCC Pro(P)	CAC His (H)	CGC Arg (R)	C		
		CUA Leu(L)	CCA Pro(P)	CAA Gln (Q)	CGA Arg (R)	A		
		CUG Leu(L)	CCG Pro(P)	CAG Gln (Q)	CGG Arg (R)	G		
	A	AUU Ile (I)	ACU Thr(T)	AAU Asn (N)	AGU Ser (S)	U		
		AUC Ile (I)	ACC Thr(T)	AAC Asn (N)	AGC Ser (S)	C		
		AUA Ile (I)	ACA Thr(T)	AAA Lys (K)	AGA Arg (R)	A		
		AUG Met(M)	ACG Thr(T)	AAG Lys (K)	AGG Arg (R)	G		
	G	GUU Val(V)	GCU Ala(A)	GAU Asp (D)	GGU Gly (G)	U		
		GUC Val(V)	GCC Ala(A)	GAC Asp (D)	GGC Gly (G)	C		
		GUA Val(V)	GCA Ala(A)	GAA Glu (E)	GGA Gly (G)	A		
		GUG Val(V)	GCG Ala(A)	GAG Glu (E)	GGG Gly (G)	G		

Corrigé : Polymorphisme et phénylcétonurie

Si la phénylalanine n'est plus métabolisée il y a accumulation de ce produit à des doses toxiques pour le développement du système nerveux central.

1°) Formes du gène

On peut remarquer la très grande différence de taille entre le gène génomique (90kbp) et l'ARNm mature (2,4kbp) qui montre que la majeure partie du génome humain n'est pas codante pour des protéines.

Définition des différents types de mutations.

Toutes ces mutations ne modifient que très localement la séquence du gène. Il faut les distinguer des remaniements plus importants qui pourraient entraîner la disparition complète du gène.

faux sens : un changement de base qui provoque un changement d'acide aminé dans la protéine.

délétion : une ou plusieurs bases perdues. Deux cas sont à distinguer

1. le nombre de bases perdu est multiple de 3, la protéine perd un ou plusieurs acides aminés mais la séquence de part et d'autre de la mutation reste inchangée.
2. le nombre de bases n'est pas multiple de 3, à partir du point de la mutation la lecture se fait de façon décalée (décalage du cadre de lecture). Très rapidement, on va rencontrer un codon non sens. La protéine est raccourcie et son extrémité COOH terminale est totalement modifiée

épissage : Lors de la maturation de l'ARNm pour permettre sa migration du noyau vers le cytoplasme, un site d'épissage est modifié, tout ou partie de l'ARNm est alors épissé anormalement provoquant une diminution ou une disparition de la forme active de la protéine.

silencieuse : le code génétique étant dégénéré (plusieurs codons codent un même acide aminé dans la plupart des cas) certaines substitutions de bases (en particulier sur la troisième position du codon) ne modifient pas la protéine ni son niveau d'expression. Dans ce cas les porteurs sont sains.

non sens : une substitution de base peut transformer un codon qui spécifie un acide aminé en un autre auquel ne correspond aucun ARNt adaptateur d'AA (codon Stop). Dans ce cas il y a arrêt prématuré de la chaîne protéique.

insertion : une ou plusieurs bases ajoutées. Deux cas sont à distinguer

1. le nombre de bases ajoutées est multiple de 3, la protéine gagne un ou plusieurs acides aminés mais la séquence de part et d'autre de la mutation reste inchangée.
2. le nombre de bases n'est pas multiple de 3, à partir du point de la mutation la lecture se fait de façon décalée (décalage du cadre de lecture). Très rapidement, on va rencontrer un codon non sens. La protéine est raccourcie et son extrémité COOH terminale est totalement modifiée.

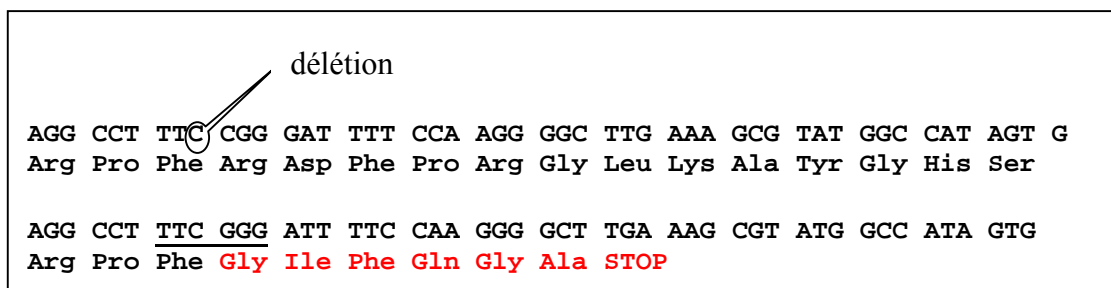
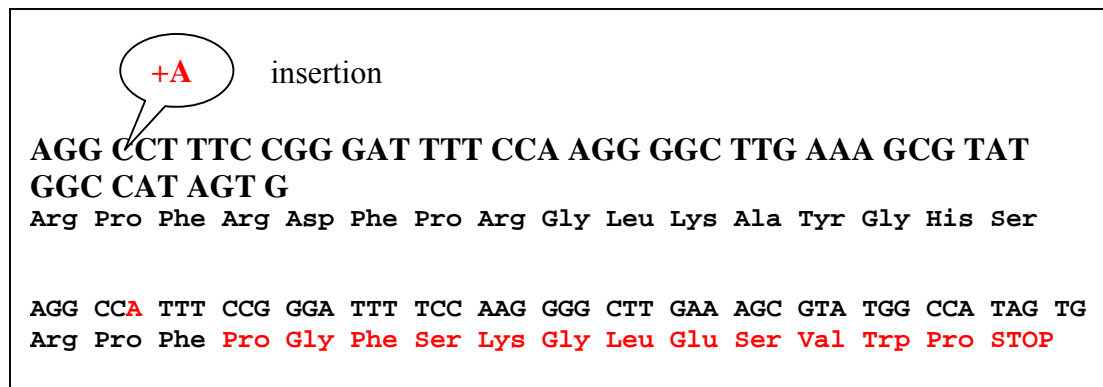


Figure 2. Insertion ou délétion d'un nombre de bases non multiple de 3 produisent des décalages du cadre de lecture.

La rencontre d'un codon stop dans un cadre de lecture décalé revient à une lecture au hasard (aucune pression de sélection n'a sélectionné cette phase de lecture comme codant une protéine nécessaire). On peut donc grossièrement admettre que les bases se succèdent au hasard. Si la fréquence des 4 bases est égale (donc 25%), la probabilité d'un codon stop est de 3/60 soit de l'ordre d'un sur 20 codons lus.

étendue de la mutation	modification au niveau de l'ADN	conséquence au niveau protéique	activité protéique
mutations locales	substitution	silencieuse :le même AA	L'activité protéique peut rester inchangée, être diminuée ou totalement perdue
		faux sens :AA différent	
		non sens Stop	
		modification d'un site d'épissage	
	insertion /délétion M3=0	gain ou perte d'un ou plusieurs AA	
insertion /délétion M3=1 ou 2	décalage du cadre de lecture avec perte de l'activité protéique		
remaniements importants	délétions	perte de la protéine active	
	translocations	modification possible du niveau d'expression de la protéine	
	insertion d'un élément mobile	perte de la protéine active	

Tableau 6. Conséquences des mutations de l'ADN au niveau protéique.

Sur le tableau du code génétique, si la troisième base est modifiée, les AA possibles sont ceux qui occupent la même case, si c'est la base 1, les AA possibles sont ceux rencontrés dans la même colonne et si c'est la seconde base, on peut avoir comme AAA de remplacement ceux de la même rangée.

Arg est un acide aminé codé par 6 codons répartis en 2 groupes.

	AGA	AGG	CGU	CGC	CGA	CGG
3ème base	Ser	Ser	-	-	-	-
2ème base	Lys Thr Ile	Lys Thr Met	His Pro Leu	His Pro Leu	Gln Pro Leu	Gln Pro Leu
1ère base	Stop Gly	Trp Gly	Cys Ser Gly	Cys Ser Gly	Stop Gly	Trp Gly

Tableau 7. Substitutions possibles d'une Arg par substitution d'une base.

L'examen de ce tableau permet de conclure que la position 176 avait comme triplet sur le brin d'ADN non transcrit (qui est identique à l'ARNm en remplaçant U par T) CGA comme les positions 243 et 261 ; alors qu'en position 252 on a CGG et en 241 et 413 soit CGC soit CGT.

Dans le cas d'une substitution sur une des bases d'un codon Arg on pourra obtenir 5 cas de mutations silencieuses (5 autres codons qui donnent aussi Arg)

1 ou 0 cas de non sens

4 à 6 cas de faux sens (suivant le codon envisagé, voir tableau 1)

Ces proportions varient pour différents aminoacides (en fonction du degré de dégénérescence). Elles reflètent à peu près les proportions globales trouvées de mutations silencieuses (5,62%), non sens (5,18%) et faux sens (62,85%). Il faut remarquer que si un AA n'est codé que par 2 codons, il y aura proportionnellement plus de possibilités de faux sens.

2°) Phénotype : Influence du milieu

Les deux enfants ont le même génotype pour le gène PAH, par contre la composante alimentaire de leur environnement est radicalement différente. Cette différence est suffisante à rendre compte de la différence de leur état. Un régime très strict en doses journalières de phé est connu pour son efficacité thérapeutique dans les cas de PKU.

Le phénotype est le résultat de l'interaction du génotype d'un individu pour un gène donné et de l'environnement dans lequel il est placé.

3°) Phénotype des mutants de PAH

Les différentes mutations ont des conséquences différentes au niveau de la protéine qui peut rester presque aussi active que la forme non mutée ou disparaître complètement. Ces différences d'intensité doivent être corrélées avec la gravité de la maladie.

Les décalages de phase de lecture (ici des délétions non multiples de 3), les erreurs d'épissage et les stop prématurés doivent être les plus graves au niveau du dysfonctionnement protéique. Ces types de mutation se rencontrent dans des formes alléliques qui confèrent la forme classique la plus sévère de PKU.

La définition du phénotype n'est pas modifiée par ces observations. Il faut cependant noter qu'un phénotype donné n'est pas attaché à un gène dès l'instant où il est muté mais dépend aussi de la mutation, autrement dit de la forme allélique.

4°) Etude in vitro

Il n'y a que certaines formes pour lesquelles on dispose à la fois des données in vivo et in vitro.

allèle HPA	quantité de protéine	quantité d'ARNm	phénotype
R176L	42%%	nd	HPA non PKU
R408Q	55%	100%	HPA non PKU
I65T	26%	100%	modéré à fort
R261Q	30%	100%	modéré à fort
L348V	25%	100%	modéré à fort
V388M	15%	100%	modéré à fort
R408W	<1%	100%	classique
jonction introns 12-13 G→A	<1%	100%	classique

Tableau 8. Comparaison des propriétés in vitro de certains allèles du gène PAH et des phénotypes qu'ils confèrent in vivo.

Dans ce tableau on voit nettement que lorsque le phénotype s'aggrave, la quantité d'enzyme active diminue voire même n'est plus décelable.

Dans le cas précédent on examinait l'effet d'une mutation dans différents organismes qui n'étaient pas identiques. Sur les 30 000 gènes humains il y en a de nombreux qui présentent diverses formes alléliques et leurs combinaisons sont en si grand nombre (2^{30000} si chacun n'avait que 2 formes alléliques soit 10^{100} génotypes différents, pour une population humaine de quelques milliards $12 \cdot 10^9$) que le génotype de chacun est unique.

In vitro, on a construit le plasmide (toujours le même vecteur de départ) recombiné avec la forme mutante. Chacun de ces vecteurs recombiné a servi à transformer des cellules en culture (cellules de fibroblastes de singe). Dans ce dernier cas, toutes les formes mutées sont exactement dans le même contexte.

5°) Autres influences

Plusieurs gènes interviennent dans le recyclage de bh4 (un par enzyme, gènes PTS, PCBD, DHPR, SPR, GCH1). Si l'un de ces gènes mute et code une protéine moins active, le recyclage sera plus lent et la conversion de Phé en Tyr peut en être ralentie d'où une accumulation de Phé. Chez les sujets n'ayant pas de mutation du gène PAH, des mutations dans chacun de ces 5 gènes ne provoque pas de phénotype [PKU] mais [HPA non PKU] ou pour le gène SPR, une déficience en neurotransmetteur non accompagnée d'HPA. Par contre, chez des sujets porteurs de deux allèles mutés pour le gène PAH, une mutation sur un de ces gènes est susceptible de modifier le phénotype [PKU].

Il faut donc reprendre la définition du phénotype :

« Le phénotype est le résultat de l'interaction du génotype d'un individu pour un gène donné et de l'environnement dans lequel il est placé. »
et la modifier :

Le phénotype est le résultat de l'interaction de tous les éléments du génotype d'un individu et de l'environnement dans lequel il est placé. C'est-à-dire que plus d'un gène peut intervenir pour déterminer un phénotype.

6°) Fréquence des hétérozygotes

génotype		phénotype	
		phénylcétonurique ?	résistant à l'ochratoxine ?
PAH+/PAH+		non malade	sensible
PAH+/R408X		non malade	résistant
R408W/R408W		malade	résistant
conclusion	génotype	PAH+ > R408W	PAH+ < R408W
	phénotype	[sain] > PKU]	[sensible] < [résistant]

On voit que la formulation en utilisant les génotypes conduit à une contradiction qui n'existe pas si l'on définit dominance et récessivité en terme de phénotype.

Dominance et récessivité sont des propriétés exclusives du phénotype.

Pour interpréter un résultat expérimental il faut formuler de la façon suivante : le phénotype [xxx] conféré par l'allèle n est récessif (ou dominant) devant le phénotype [sauvage] conféré par l'allèle sauvage. Avec cette formulation, si une forme allélique confère plusieurs phénotypes qui n'ont pas les mêmes propriétés, on n'aboutira pas à une contradiction puisqu'on traite chaque phénotype séparément.

Le Pénotype peut donc être considéré comme la résultante des interactions entre le génome de l'individu et l'environnement dans lequel il se développe. La formule générale s'écrit :

$$P = G * E$$

L'étoile symbolise l'ensemble des interactions, le phénotype ne devant pas être considéré comme la simple somme des effets du génome et de l'environnement.

On remarquera que certains caractères n'ont pas de déterminisme génétique (par exemple le fait de parler telle ou telle langue maternelle). Cependant, lorsqu'une composante génétique existe, l'effet des gènes est très variable selon le type de caractère, allant d'un déterminisme génétique quasi absolu (pour les groupes sanguins par exemple) à une participation plus modérée (déterminisme de la taille ou de la masse corporelle). Ces derniers cas sont l'objet des études de la Génétique Quantitative qui utilise les méthodes de l'analyse statistique, alliées à celles de la génétique moléculaire.

Maladies de la coagulation

La coagulation sanguine provient d'une séquence de réactions enzymatiques complexes, très bien contrôlées. L'événement précis qui déclenche la coagulation reste obscur.

1°) De nombreux cas de défaut de coagulation ont été décrits pour lesquels le plasma est déficient dans un ou plusieurs des facteurs impliqués dans l'hémostase (à l'exception des ions Ca^{++} et des phospholipides) voir tableau 1. Leur tableau clinique manque de spécificité. De telles affections sont héréditaires et rares dans la population générale : de 10^{-5} à 10^{-6} . Pour chaque cas on admettra que le défaut envisagé, lorsqu'il est héréditaire est monogénique.

Dans tous ces pedigrees, l'un ou les deux parents ne sont pas figurés lorsque l'on n'a aucune indication à leur sujet. On les suppose alors sains, non apparentés à la famille et donc issus de la population générale.

Pour les pedigrees A, B, ... F, répondre aux questions suivantes :

- L'affection est-elle héréditaire?
- Récéssivité ou dominance du phénotype muté par rapport au phénotype sauvage?
- Le gène concerné est-il porté par le chromosome X ou autosomique? (Lorsque le croisement démonstratif n'existe pas quel est le mode de transmission le plus probable ?).
- Quel est le génotype des individus atteints?

2°) Dès 1947 Pavlovsky montra que le mélange des plasmas obtenus de deux malades "hémophiles" pouvait se comporter comme un plasma normal ou anormal suivant l'origine des deux plasmas. Interprétation de ce résultat. Proposer une méthode de classification des différents "hémophiles".

On a ainsi élucidé, au moins en partie, le processus de la coagulation qui est schématisé sur la figure 1.

De plus on a mis au point le dosage spécifique de certains des facteurs de coagulation. Ceci a permis d'étudier des cas de déficience d'un facteur donné.

3°) Déterminisme de l'expression du facteur VIII

Le pedigree C représente une famille dont les membres malades ont une déficience en Facteur VIII (hémophilie A, la plus fréquente : 10^{-4}). Ce facteur fait aussi défaut chez les individus malades des pedigrees E, F, G et H. L'individu GI-1 est également le EIII-5 et sa femme GI-2 est la descendante FIII-1. Que peut-on conclure de l'ensemble de ces pedigrees (E+F+G).

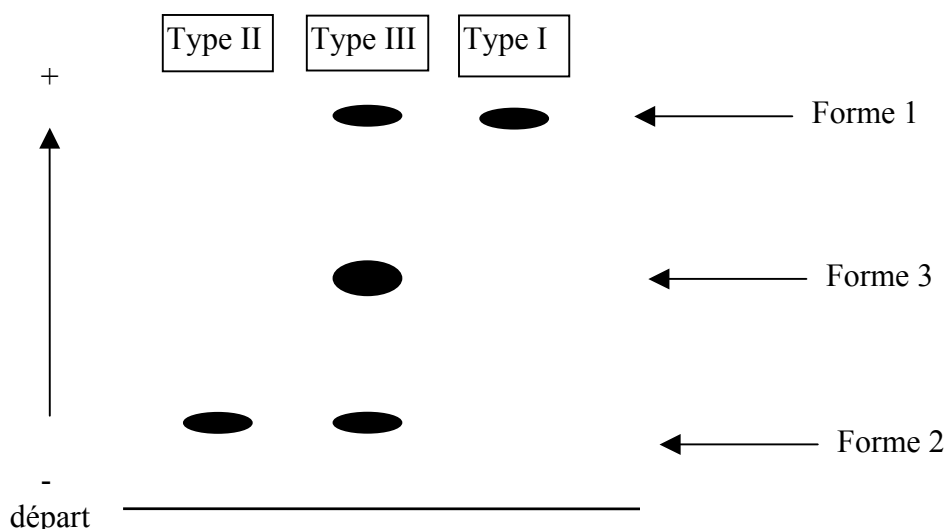
Quel est le mode de transmission de l'hémophilie dans le pedigree H ?

Que doit-on penser de l'ensemble de ces résultats concernant le facteur VIII ?

4°) Déterminisme génétique du facteur XIII

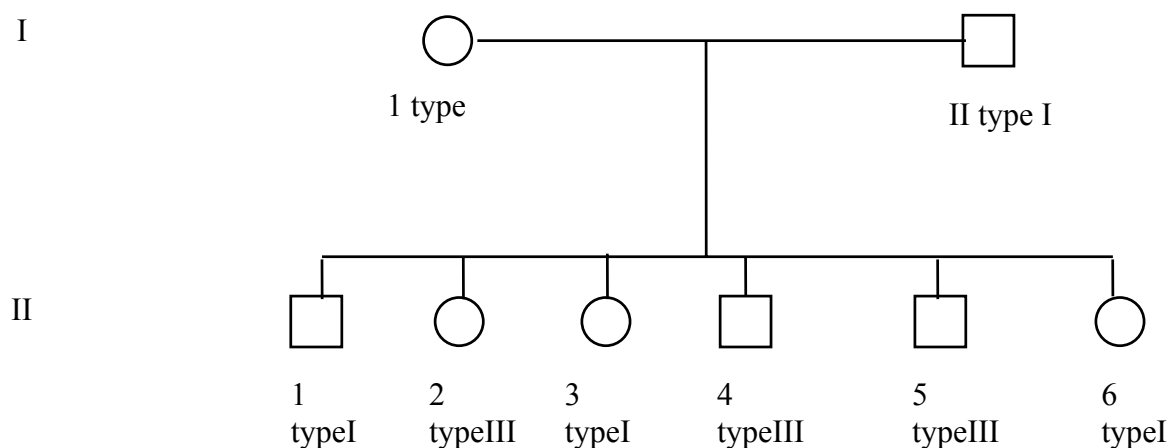
Le facteur XIII est composé de deux sous unités A et de deux sous unités B (A₂ B₂). On se demande si la déficience en facteur XIII est portée par le chromosome X ou par un autosome.

Dans la population australienne on a découvert un polymorphisme de la sous unité A du facteur XIII. Le plasma de différents individus est soumis à une électrophorèse suivie d'une réaction colorée spécifique du facteur XIII. On obtient la figure suivante :



Quelle interprétation peut-on proposer pour chaque type d'individu en admettant que la sous unité A soit codée par un gène unique.

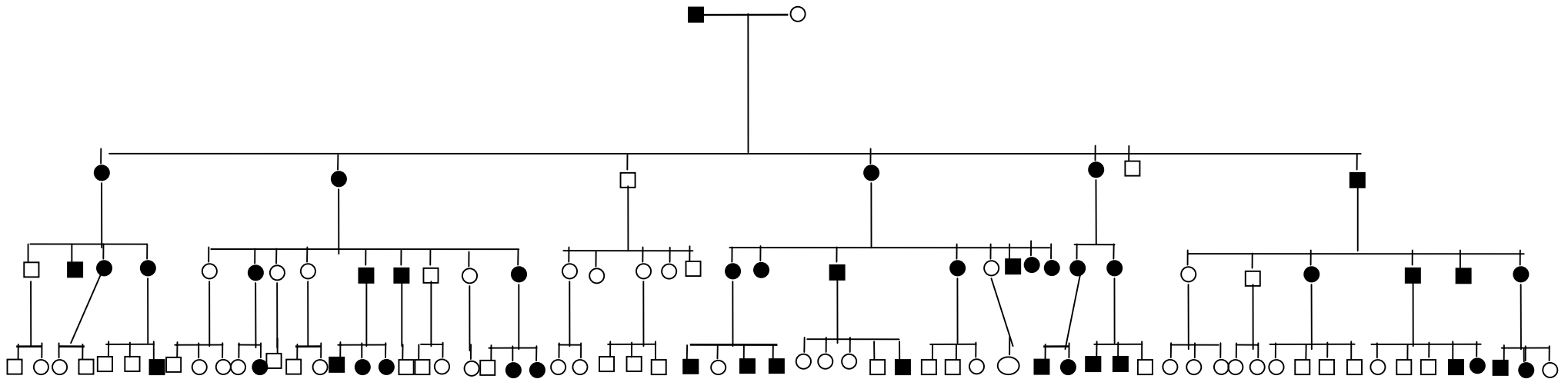
Dans une famille on a étudié la distribution de ces différents types (tous les membres de la famille sont en bonne santé) :



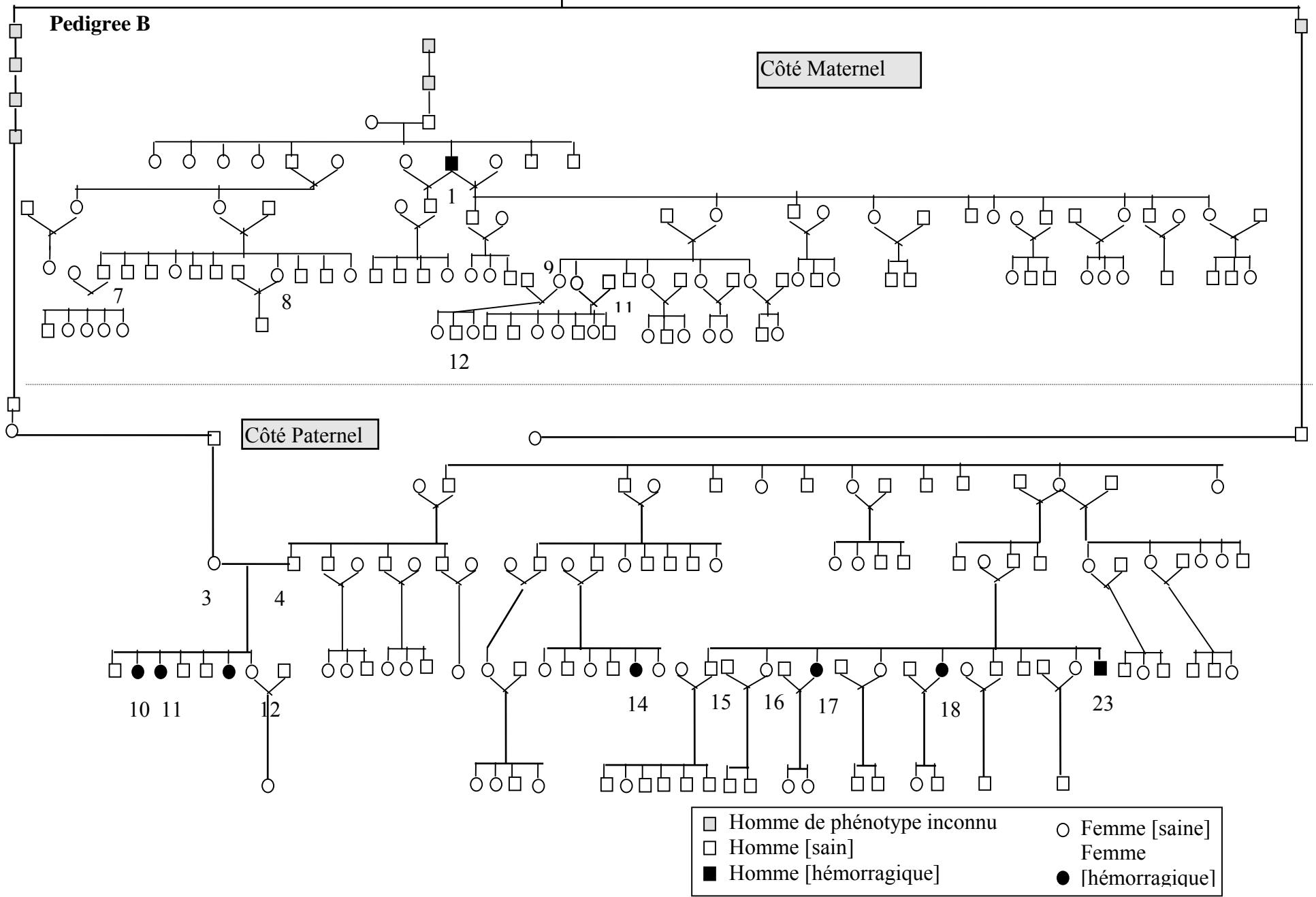
Que démontre ce pedigree?

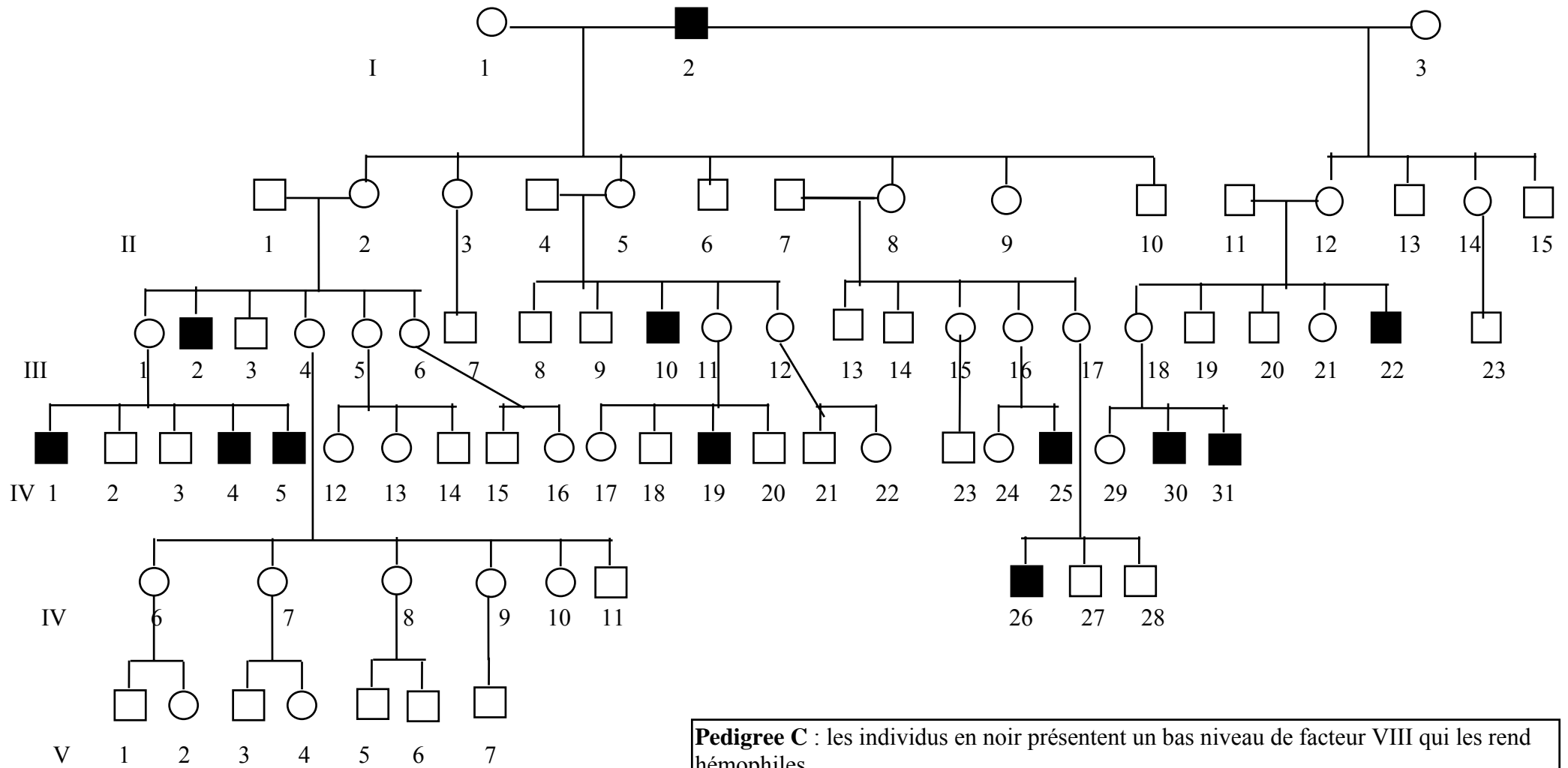
Bibliographie

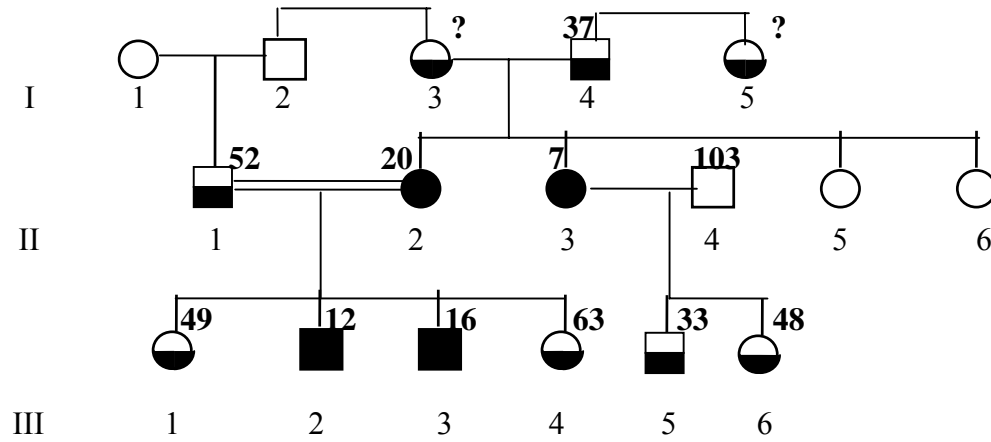
- Board : Am.J.Hum.Genet. **31** 116-124 (1979) FXIII question 4
 Hensen , Mattern, Loeliger : Thromb. Diath. Haemorrhag. **14** pp 341-345 (1965) FVIII dominant pedigree H
 Holmberg , Nilsson : Brit.Med.J. **3** pp 317-320 (1972) vWd pedigree A
 Israels , Lempert , Gilbertson :Lancet **1** pp 1375-1380(1951)pedigreeG
 Kingsley : Q.J.Med.Oxford **23** pp323-329 (1954) F V pedigree B
 Ratnoff , Bennett : Sci. **179** pp1291-1298 (1973) revue
 Rapaport et al : Blood **18** pp149-165 (1961) F XI pedigree D
 Whittaker , Copeland , Graham : Am.J.Hum.Genet. **14** pp149-158 (1962): pedigree C F VII



Pedigree de la famille A:maladie de von Willebrand. Les sujets en noir sont atteints

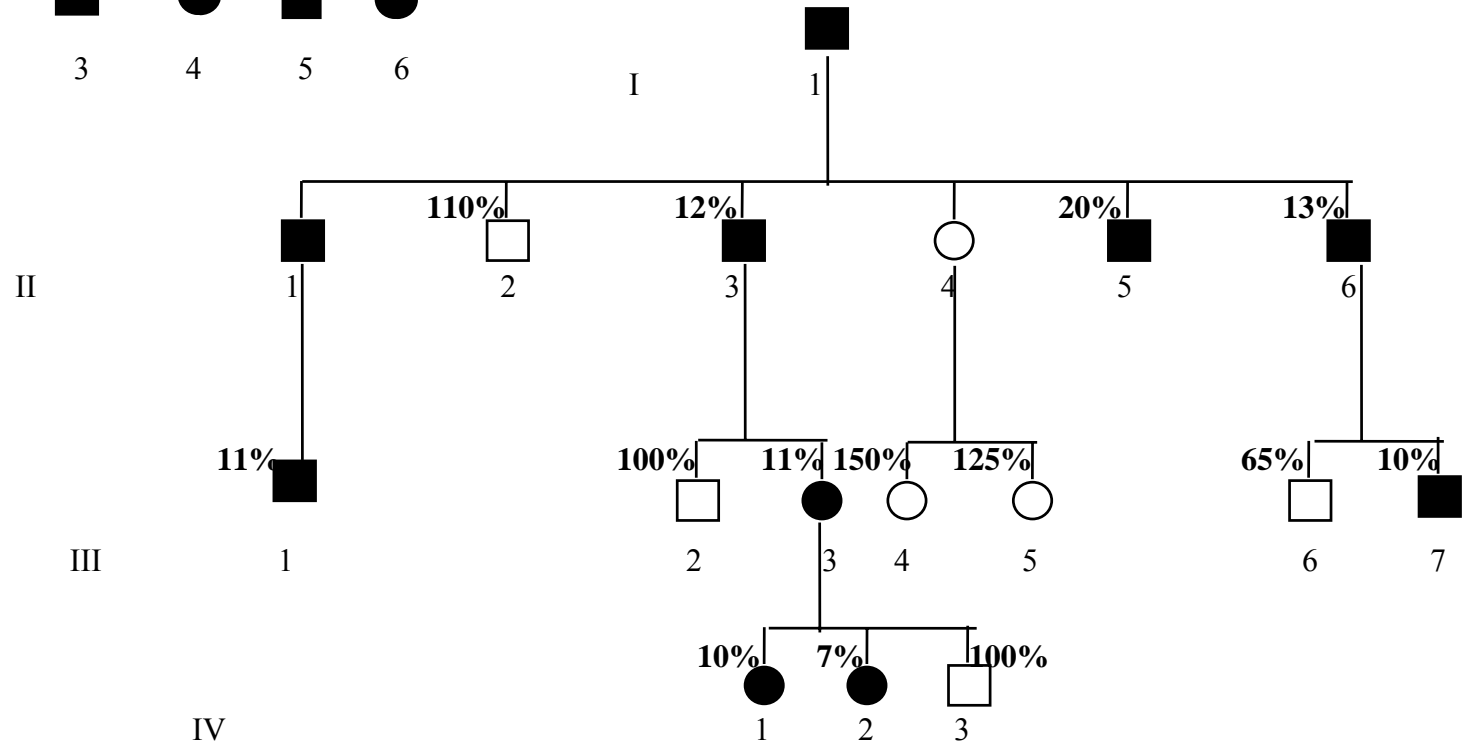


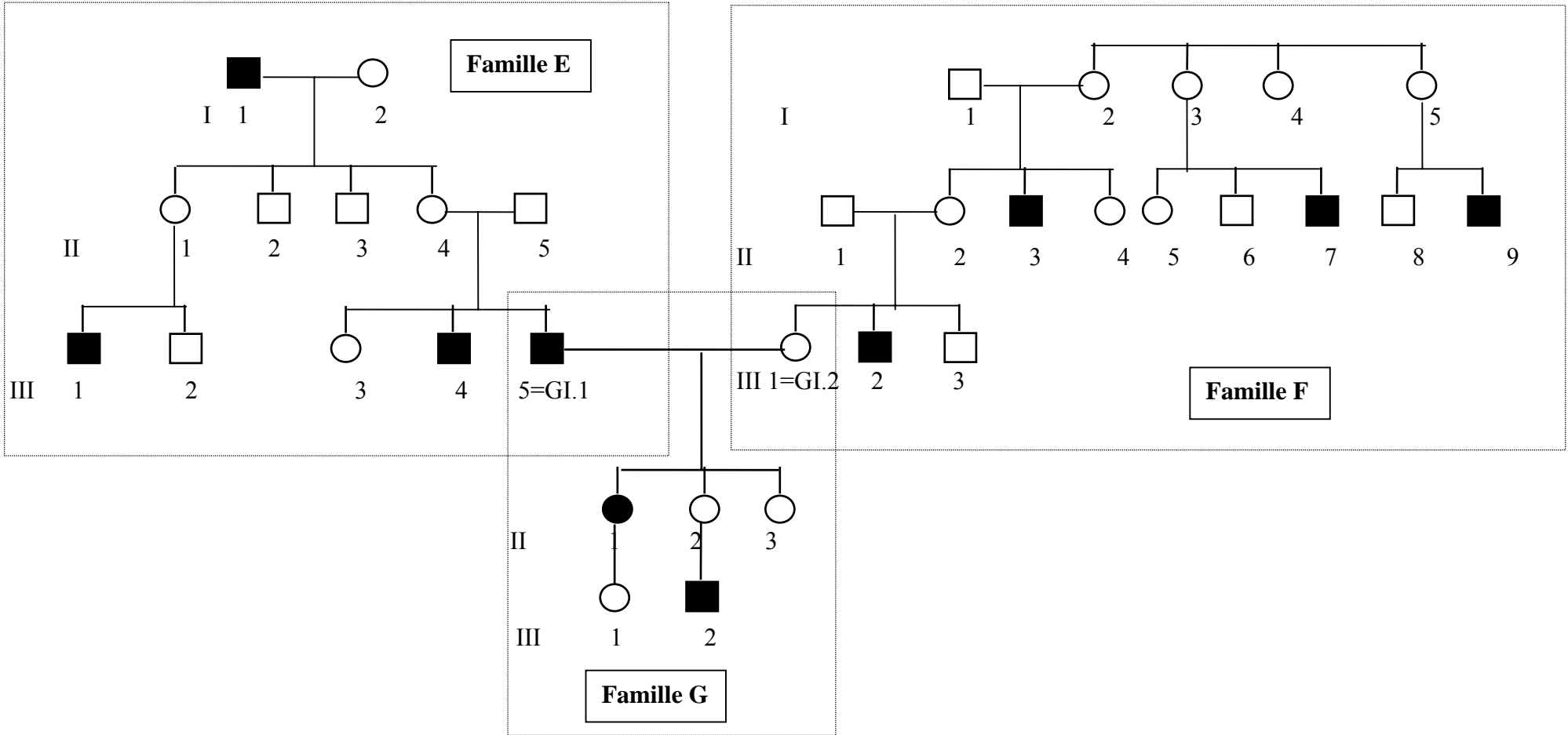




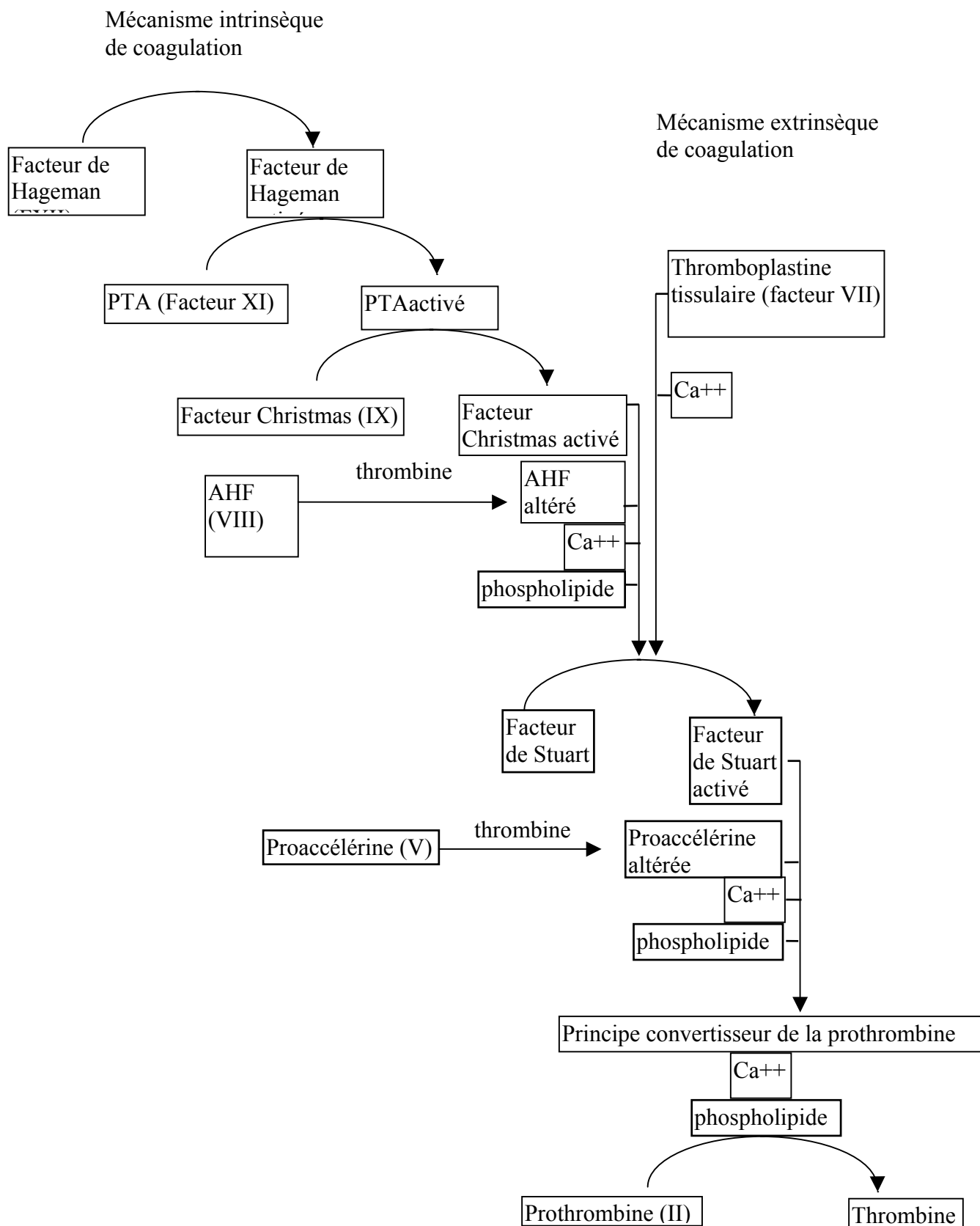
Pedigree D : la ligne double indique un mariage consanguin. Les nombres en **gras** indiquent le % de PTA. Les individus noirs sont ceux qui présentent des symptômes majeurs hémorragiques, ceux en noir et blanc n'ont qu'une déficience mineure en PTA.

Pedigree H : les % indiquent le niveau de facteur VIII (100= valeur moyenne d'un individu sain).





Mécanisme de la coagulation sanguine chez un sujet normal



Les facteurs de coagulation sanguine

Partie 1

Substance	concentration plasmatique normale	Hérédité de la déficience	fréquence estimée /10 ⁶	Test de laboratoire
Fibrinogène Hypofibrinogénie Afibrinogénie Dysfibrinogénie	1500-4000	Autosomique récessif Autosomique récessif Autosomique récessif	<0,5	Fibrinogène<100mg/dl Fibrinogène absent(immuno) Libération retardée de fibrinopeptide
Prothrombine	150	Autosomique récessif	<0,5	essai spécifique
thromboplastine tissulaire VII	-	pas de déficience connue	-	essai spécifique
Proaccélélerine	10	Autosomique récessif	<0,5	essai spécifique
Proconvertine	<1	Autosomique récessif	<0,5	essai spécifique, aggrégation normale des plaquettes à la ristocétine
Facteur antihémophilique VIII	<0,5	sur l'X récessif	60-80	essai spécifique
Facteur Christmas (PTC) IX	5	sur l'X récessif	15-20	essai spécifique
Facteur de Stuart	8	Autosomique récessif	<0,5	essai spécifique
PTA (plasma thromboplastine antécédant) XI	5	Autosomique récessif	1	essai spécifique
Facteur de Hageman	35	Autosomique récessif	1	essai spécifique

.../...

Les facteurs de coagulation sanguine

Suite

Facteur de Hageman	35	Autosomique récessif	1	essai spécifique
Facteur de stabilisation de la fibrine	20	Autosomique récessif	<0,5	essai spécifique
Prékallitréine (facteur de Fletcher)	30	Autosomique récessif	<0,5	caillots solubles dans de l'urée 5M ou C ₂ H ₃ ClO ₂ 1%
kininogène (facteur de Fitzgerald)	80	Autosomique récessif	<0,5	essai spécifique
Facteur de von Willebrand V	16	Autosomique semidominant	5-10	protéine vW diminuée (immuno)
Protéine C	5	pas de déficience connue	-	-
Protéine S	5	pas de déficience connue	-	-
Antithrombine III	240	Autosomique dominant	<0,5	essai spécifique
α ₂ -antiplasmine	70	Autosomique récessif	1 cas identifié	temps de lyse du caillot écourté
inhibiteur de protéine C	?	Autosomique récessif	<0,5	tests combinés des facteurs V et VIII
Plasminogène	150	Autosomique récessif	<0,5	temps de lyse du caillot allongé

Corrigé Maladies de la coagulation

1°) Le défaut de coagulation est un phénotype rare dans la population générale 10^{-5} . Si dans une famille donnée on trouve nettement plus d'individus hémorragiques que dans la population générale, on peut en déduire qu'un ancêtre a apporté l'allèle muté et qu'il l'a transmis à sa descendance donc que dans cette famille le défaut de coagulation est héréditaire.

Pour chaque pedigree, il faut d'abord déterminer si le phénotype muté est dominant ou récessif devant le phénotype sauvage. Ensuite on cherche à déterminer la position du gène en question : sur l'X ou autosomique.

Si le phénotype muté est dominant, chaque enfant hémorragique aura au moins un parent hémorragique. Dès l'instant où l'un des descendants sera sauvage, il ne possèdera pas l'allèle muté et toute sa descendance sera sauvage. Un allèle dominant se « voit » de génération en génération.

Au contraire si le phénotype muté est récessif devant le phénotype sauvage, les individus hétérozygotes de phénotype sauvage transmettent l'allèle muté sans qu'il soit visible, jusqu'au croisement qui produira, à partir de deux parents de phénotype sauvage, un ou des enfants hémorragique(s). Un tel croisement a souvent lieu entre individus apparentés et produit dans ce cas des individus dits consanguins.

On peut, dans certains cas, calculer les probabilités d'apparition des différents cas et montrer, par exemple, que la probabilité est en faveur de la dominance du phénotype muté face au phénotype sain est la plus élevée, ce qui permet de préférer l'une des différentes hypothèses.

Considérations statistiques

Pour un gène autosomique dans la population générale, avec le phénotype mutant récessif devant le phénotype sauvage, soient deux formes alléliques d'un gène A qui ont respectivement les fréquences p et q.

$$\begin{aligned} \text{fr } a^- &= q \\ \text{fr } a^+ &= p \\ \text{avec } p + q &= 1 \end{aligned}$$

Dans la population générale où l'on admet que les gamètes se rencontrent au hasard, on peut calculer la fréquence de chaque classe génotypique.

gamètes	a+ fréquence p	a- fréquence q
a+ fréquence p	$\frac{a^+}{a^+}$ [sain] $\frac{a^+}{a^-}$ fréquence p^2	$\frac{a^+}{a^-}$ [sain] $\frac{a^-}{a^-}$ fréquence pq
a- fréquence q	$\frac{a^+}{a^-}$ [sain] $\frac{a^-}{a^-}$ fréquence pq	$\frac{a^-}{a^-}$ [atteint] $\frac{a^-}{a^-}$ fréquence q^2

Les homozygotes sauvages de génotype $\frac{a^+}{a^+}$ ont la fréquence p^2

Les hétérozygotes sauvages également de génotype $\frac{a^+}{a^-}$ ont la fréquence $2pq$

Les homozygotes mutés de génotype $\frac{a^-}{a^-}$ ont la fréquence q^2

donc parmi les individus sains, les hétérozygotes ont la fréquence $\frac{2pq}{p^2 + 2pq} = \frac{2pq}{1 - q^2}$,

ce qui est peu différent de $2pq$ si l'allèle muté est rare.

Ici si la fréquence des hémorragiques est 10^{-5} , celle de l'allèle muté est donc de 0,0032 et celle des hétérozygotes 0,0064.

Pour un gène sur l'X il faut faire le calcul soit chez les femelles (identique à ce qui précède)

soit chez les mâles. Les mâles atteints de génotype $\frac{a^-}{\square}$ ont la fréquence de l'allèle a-

c'est-à-dire q et les mâles sains de génotype $\frac{a^+}{\square}$ ont la fréquence p

Cas du pedigree A. Quel que soit le phénotype dominant, dans le couple fondateur il y a au moins un des parents qui est porteur d'un allèle muté du gène en cause.

On peut constater qu'il n'y a aucun couple de parents sains qui ait des enfants atteints (pas de saut de génération), ce qui est l'indice d'un phénotype [atteint] dominant sur le phénotype [sain]. Peut-on en faire une démonstration rigoureuse ?

Faisons l'hypothèse que le phénotype sain est dominant sur le phénotype muté. Dans ce cas, tous les individus atteints doivent avoir leurs 2 parents porteurs d'un allèle muté. Or il y a 16 couples dans ce cas. Pour chacun des couples, un des parents appartient à la famille A et peut donc être raisonnablement porteur, mais l'autre est issu de la population générale et sa fréquence est de l'ordre de 0,006 (voir le calcul ci-dessous). Si tel est bien le cas, le pedigree observé doit avoir une fréquence de $0.006^{16} = 2,8 \times 10^{-36}$ ce qui revient à dire qu'en pratique on n'avait aucune probabilité de l'observer (la population totale de la planète est de l'ordre de 7×10^9 habitants)

Si au contraire le phénotype atteint est dominant sur le phénotype sain, tout individu sain est homozygote pour l'allèle sauvage (Probabilité 1). La probabilité de ce pedigree à partir de la génération 1 ne met plus en jeu celle de tirer un hétérozygote de la population générale.

C'est donc ce cas qui est hautement probable par rapport à l'autre. On peut conclure [malade] dominant sur [sain] pour le phénotype rencontré dans la famille A.

Le gène en jeu est-il sur l'X ou autosomique ? Le sujet dont le génotype s'écrira différemment et se transmettra aussi de façon différente est un homme atteint. Comparons ce que l'on attend dans chaque hypothèse.

mâle [m] x femelle [+]																																					
<p style="text-align: center;">hypothèse gène AUTOSOMIQUE</p> <table style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">$a-$</td> <td style="padding: 2px 10px;">x</td> <td style="padding: 2px 10px;">$a+$</td> </tr> <tr> <td style="padding: 2px 10px;">$a+$</td> <td></td> <td style="padding: 2px 10px;">$a+$</td> </tr> <tr> <td style="text-align: center; padding: 2px 10px;">↓</td> <td></td> <td style="text-align: center; padding: 2px 10px;">↓</td> </tr> <tr> <td style="padding: 2px 10px;">2t.g.</td> <td></td> <td style="padding: 2px 10px;">1 type de gamètes</td> </tr> </table> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <tr> <td style="width: 20px;"></td> <td style="width: 20px;">$a+$</td> </tr> <tr> <td style="width: 20px;">$a-$</td> <td>$a-/a+$ [atteint]</td> </tr> <tr> <td style="width: 20px;">$a+$</td> <td>$a+/a+$ [sain]</td> </tr> </table> <p style="margin-top: 10px;">mâles et femelles [+] et [m]</p> <p style="margin-top: 10px;">La présence, dans la descendance de mâles atteints ou de femelles atteintes élimine l'hypothèse de la localisation sur l'X.</p>	$a-$	x	$a+$	$a+$		$a+$	↓		↓	2t.g.		1 type de gamètes		$a+$	$a-$	$a-/a+$ [atteint]	$a+$	$a+/a+$ [sain]	<p style="text-align: center;">hypothèse gène SUR l'X</p> <table style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tr> <td style="padding: 2px 10px;">$a-$</td> <td style="padding: 2px 10px;">x</td> <td style="padding: 2px 10px;">$a+$</td> </tr> <tr> <td style="padding: 2px 10px;">\neg</td> <td></td> <td style="padding: 2px 10px;">$a+$</td> </tr> <tr> <td style="text-align: center; padding: 2px 10px;">↓</td> <td></td> <td style="text-align: center; padding: 2px 10px;">↓</td> </tr> <tr> <td style="padding: 2px 10px;">2 types de gamètes</td> <td></td> <td style="padding: 2px 10px;">1 t.g.</td> </tr> </table> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <tr> <td style="width: 20px;"></td> <td style="width: 20px;">$a+$</td> </tr> <tr> <td style="width: 20px;">\neg</td> <td>$a+/\neg$ mâle [sain]</td> </tr> <tr> <td style="width: 20px;">$a-$</td> <td>$a-/a+$ femelle [atteinte]</td> </tr> </table> <p style="margin-top: 10px;">mâles [+] et femelles [atteinte]</p> <p style="margin-top: 10px;">Seules des considérations statistiques peuvent faire pencher pour cette hypothèse en l'absence de filles [+] et de mâles [m].</p>	$a-$	x	$a+$	\neg		$a+$	↓		↓	2 types de gamètes		1 t.g.		$a+$	\neg	$a+/\neg$ mâle [sain]	$a-$	$a-/a+$ femelle [atteinte]
$a-$	x	$a+$																																			
$a+$		$a+$																																			
↓		↓																																			
2t.g.		1 type de gamètes																																			
	$a+$																																				
$a-$	$a-/a+$ [atteint]																																				
$a+$	$a+/a+$ [sain]																																				
$a-$	x	$a+$																																			
\neg		$a+$																																			
↓		↓																																			
2 types de gamètes		1 t.g.																																			
	$a+$																																				
\neg	$a+/\neg$ mâle [sain]																																				
$a-$	$a-/a+$ femelle [atteinte]																																				

Dans le cas de la famille A, on constate que des pères atteints ont à la fois filles et garçons sains et malades. Cela exclut la possibilité d'un gène localisé sur l'X.

Famille A : la maladie est héréditaire, dominante et autosomique

Pedigree B : il y a des sauts de génération. On voit plusieurs enfants malades ayant des parents sains. La mutation est un événement suffisamment rare pour ne pas être observé plus d'une fois dans une fratrie. Si plusieurs enfants sont atteints, ce doit être parcequ'ils ont reçu un allèle atteint d'un de leurs parents (ou des 2). Or les deux parents dont au moins un est porteur sont tous deux sains donc le phénotype hémorragique est récessif devant le phénotype sauvage.

Ces couples n'auront pas le même génotype si le gène en cause est autosomique ou sur l'X. Comme précédemment, comparons leur descendance dans chaque hypothèse.

Parmi les descendants de 2 parents sains mais porteurs on trouve aussi bien des mâles que des femelles atteintes : gène autosomique.

mâle [+] x femelle [+]																			
<p style="text-align: center;">hypothèse gène AUTOSOMIQUE</p> <p>S'ils ont des enfants atteints c'est qu'ils sont porteurs.</p> $\begin{array}{cc} \underline{a-} & \times & \underline{a-} \\ a+ & & a+ \\ \downarrow & & \downarrow \end{array}$ <p style="text-align: center;">2 types de gamètes</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th>a-</th> <th>a+</th> </tr> </thead> <tbody> <tr> <th>a-</th> <td>a-/a-[atteint]</td> <td>a-/a+[sain]</td> </tr> <tr> <th>a+</th> <td>a-/a+[sain]</td> <td>a-/a+[sain]</td> </tr> </tbody> </table> <p>mâles et femelles [+] et [m] la présence d'une femelle [m] élimine l'hypothèse sur l'X</p>		a-	a+	a-	a-/a-[atteint]	a-/a+[sain]	a+	a-/a+[sain]	a-/a+[sain]	<p style="text-align: center;">hypothèse gène SUR l'X</p> <p>S'ils ont des enfants atteints c'est que la femelle est porteuse.</p> $\begin{array}{cc} \underline{a+} & \times & \underline{a-} \\ \neg & & a+ \\ \downarrow & & \downarrow \end{array}$ <p style="text-align: center;">2 types de gamètes</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th>a-</th> <th>a+</th> </tr> </thead> <tbody> <tr> <th>\neg</th> <td>a-/\neg ♂[atteint]</td> <td>a+/\neg ♂[sain]</td> </tr> <tr> <th>a+</th> <td>a-/a+ ♀[saine]</td> <td>a-/a+ ♀[saine]</td> </tr> </tbody> </table> <p>mâles [+] et [m] femelles [+] seules des considérations statistiques peuvent faire pencher pour cette hypothèse en l'absence de filles [m]</p>		a-	a+	\neg	a-/ \neg ♂[atteint]	a+/ \neg ♂[sain]	a+	a-/a+ ♀[saine]	a-/a+ ♀[saine]
	a-	a+																	
a-	a-/a-[atteint]	a-/a+[sain]																	
a+	a-/a+[sain]	a-/a+[sain]																	
	a-	a+																	
\neg	a-/ \neg ♂[atteint]	a+/ \neg ♂[sain]																	
a+	a-/a+ ♀[saine]	a-/a+ ♀[saine]																	

Dans cette famille des couples sains ont à la fois filles et garçons atteints, ce qui exclut la localisation sur l'X du gène en cause.

Famille B : la maladie est héréditaire, récessive et autosomique.

Pedigree C : il y a des sauts de génération , phénotype hémorragique récessif devant le phénotype sauvage. On constate que dans cette famille il n'apparaît que des garçons atteints, de plus, les parents qui transmettent la maladie dans la famille sont tous des femmes. Ce qui fait suspecter une localisation du gène sur l'X . Pour démontrer cette localisation on peut dans tous les croisements qui donnent des enfants atteints noter la répartition de ceux-ci:

phénotype	cas observés	cas attendus dans l'hypothèse d'un gène autosomique
mâles [sains]	10	12.375
mâles [hémorragiques]	12	4.125
femelles [saines]	11	12.375
femelles [hémorragiques]	0	4.125
total	33	33

Il faut que les nombres soient suffisants (la plus petite classe théorique ≥ 5) pour que le calcul du χ^2 soit valide, ce qui n'est pas le cas ici.

On peut compter le nombre de couples d'individus sains qui ont eu des enfants atteints (ici 8).

Dans l'hypothèse autosomique le mâle qui est puisé dans la population générale doit être hétérozygote. Répété 8 fois ce tirage a une probabilité de $(0,0064)^2 = 2,8 \times 10^{-18}$

Dans l'hypothèse sur l'X le mâle sain ne peut être que porteur de l'allèle a + et sa fréquence est 1.

Les deux types d'arguments permettent souvent de conclure comme c'est le cas ici.

Famille C maladie héréditaire, récessive, sur l'X.

Pedigree D : il y a des sauts de génération, phénotype hémorragique récessif sur le phénotype sauvage. Mais si on regarde le dosage de PTA on peut distinguer les hétérozygotes dont le phénotype est semi dominant [taux moyen de PTA], des homozygotes +/+ de phénotype sauvage [taux normal]. Parmi les hétérozygotes, on trouve aussi bien des mâles que des femelles : gène autosomique.

Famille D maladie héréditaire, récessive ou semi dominant selon le phénotype examiné, la maladie ou le taux de PTA , autosomique.

Pedigrees E et F : il y a des sauts de génération, phénotype hémorragique récessif sur le phénotype sauvage. **Gène sur l'X** (même type d'arguments que pour le pedigree C avec un nombre de couples moins grand dans chaque cas).

2°) Les plasmas contiennent non pas les gènes eux mêmes mais leur produit, la protéine. Si le mélange de deux plasmas d'individus hémorragiques se comporte normalement c'est que la protéine qui faisait défaut à l'un est apportée par l'autre. C'est une sorte de complémentation au niveau protéique. Et si les plasmas complémentent, les individus qui ont fourni ces plasmas sont mutés dans des gènes différents. Sinon ils peuvent être mutés dans le même gène, mais dans ce cas il peut également y avoir des phénomènes secondaires d'assemblage des sous-unités. C'est tout de même un moyen d'effectuer un premier classement.

3°) Parmi les descendants du couple G-1 x G-2 on trouve une fille atteinte donc homozygote mutée. Elle a reçu un allèle muté de son père qui est hémorragique et un de sa mère qui doit donc être hétérozygote. Ses deux parents n'appartiennent pas à la même famille néanmoins le phénotype de cette fille montre que le gène muté dans les familles E et F est le même (cas de non complémentation).

Dans la famille H le gène en cause est autosomique et donc obligatoirement différent. Dans la famille E et dans la famille C les gènes sont sur l'X. Rien ne permet d'affirmer que c'est le même ou qu'ils sont différents.

La production du facteur VIII est au moins sous la dépendance de 2 gènes (depuis on a identifié le gène de la famille E comme étant le gène de structure du Facteur VIII).

4°) Les types I et II sont produits par des individus ne portant qu'une forme allélique pour la sous unité A : les homozygotes. Chez l'hétérozygote les deux formes alléliques coexistent et produisent autant de sous unités A1 que A2. Ce qui donne 25% de dimères (A1,A1) forme 1, 25% de dimères (A2,A2) forme 2 et 50% de dimères (A1,A2) forme 3.

Deux hypothèses (autosomique ou sur l'X) : génotype possible des représentants du pedigree :

typeI	typeII	typeIII
$\frac{a1}{a1} \cdot \text{ou} \cdot \frac{a1}{\neg}$	$\frac{a2}{a2} \cdot \text{ou} \cdot \frac{a2}{\neg}$	$\frac{a1}{a2}$
I-2mâle II-1mâle II-3femelle II-6femelle		I-1femelle II-2femelle III-4mâle III-5mâle

Cependant l'existence de deux mâles de type III montre que les mâles portent, comme les femelles, deux allèles. **Ce gène est donc autosomique.**

Dystrophies des ceintures

Dans tout l'exercice on admettra que les maladies étudiées, lorsqu'elles sont héréditaires, sont monogéniques.

I) Dans une famille A on a recensé plusieurs individus présentant une dystrophie musculaire des ceintures pelvienne et scapulaire, qui est une affection que l'on rencontre rarement dans la population générale (1/40 000). Le pedigree ci-joint a permis d'établir que, dans la famille A, cette affection était :

- due à un facteur génétique
- récessive
- liée au sexe

Faites la démonstration de ces affirmations.

II) Une seconde famille B présente également quelques cas de dystrophie musculaire analogues à ceux de la famille précédente.

1°) Cette affection peut-elle être attribuée ici à des causes génétiques ? Justifiez. Si oui, quel en est le déterminisme ?

2°) Les individus III-2 de la famille A et IV-1 de la famille B se marient; ils ont sept enfants tous normaux. Quelle conclusion s'impose?

3°) Auriez-vous pu prédire ce résultat compte tenu des pedigrees des deux familles? Justifiez.

III) Ecrivez les génotypes des individus suivants lorsque cela est possible :

famille A : I-1, I-2

II-2, II-4, II-6, II-8, II-10

III-1, III-2, III-3, III-4, III-5

famille B : I-1, I-2

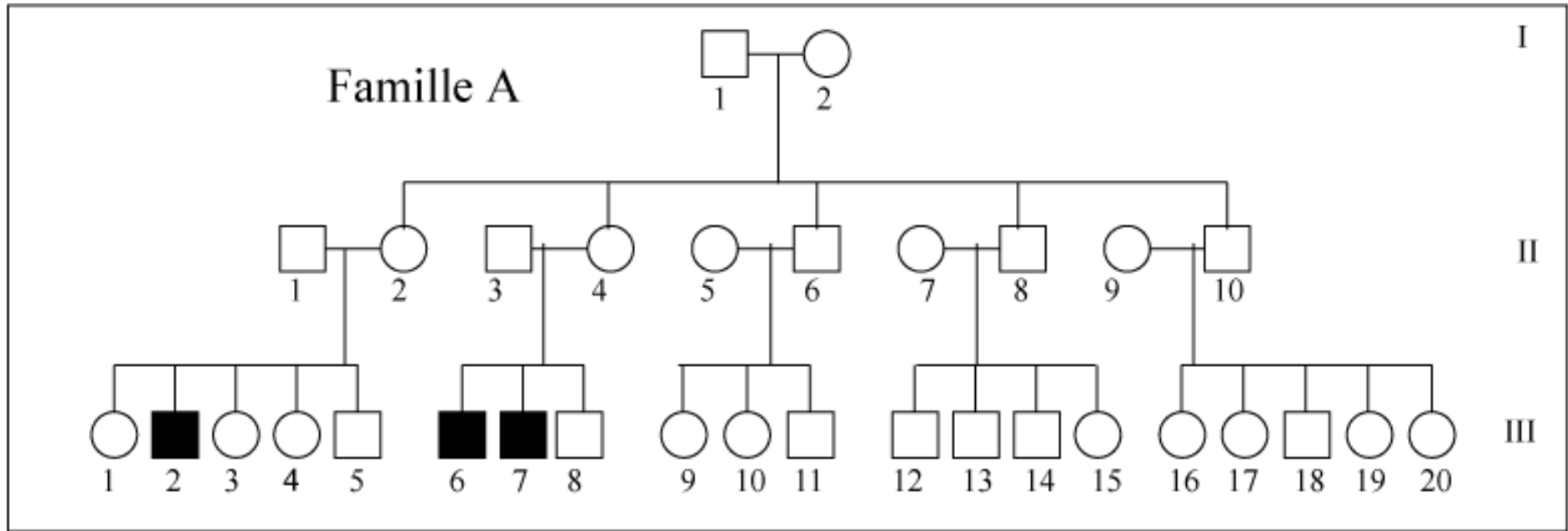
II-1, II-2,

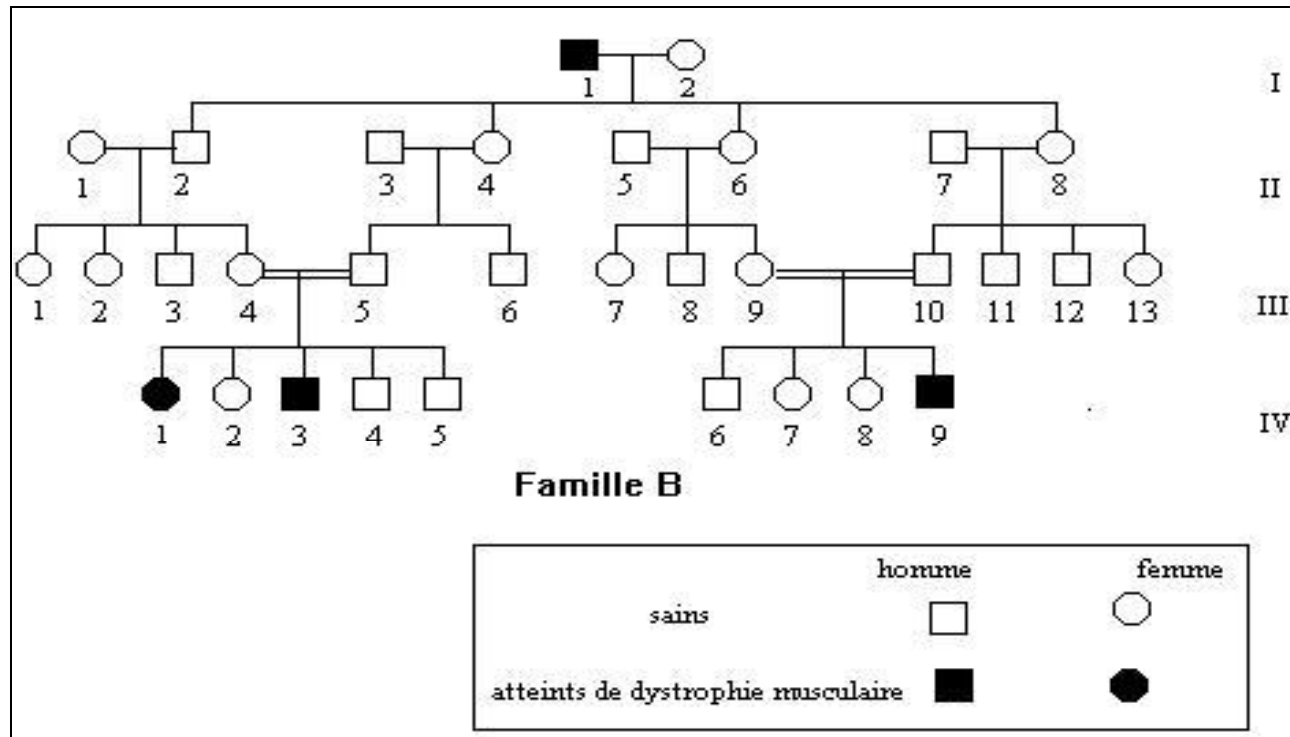
III-2, III-3, III-4

IV-1, IV-3, IV-9

IV) Les individus IV-2 et IV-6 de la famille B envisagent d'avoir un enfant. Quelle est la probabilité que celui-ci soit atteint ?

Même question si le couple était IV-2 et IV-9 .





Corrigé : Dystrophie des ceintures

Rappel

La dystrophie des ceintures est un phénotype rare dans la population générale $2,5 \times 10^{-5}$. Si dans une famille donnée on trouve nettement plus d'individus atteints, on peut en déduire qu'un ancêtre a apporté l'allèle muté et qu'il l'a transmis à sa descendance donc que dans cette famille la dystrophie des ceintures est héréditaire.

Pour ce pedigree il faut d'abord déterminer si le phénotype muté est dominant ou récessif devant le phénotype sauvage. Ensuite on cherche à déterminer la position sur l'X ou autosomique du gène en question.

Si le phénotype muté est récessif devant le phénotype sauvage les individus hétérozygotes de phénotype sauvage transmettent l'allèle muté sans qu'il soit visible jusqu'au croisement qui produira, à partir de deux parents sauvages, un ou des enfants dystrophique. Un tel croisement est souvent consanguin.

Au contraire si le phénotype muté est dominant chaque enfant dystrophique aura au moins un parent dystrophique. Dès l'instant où un descendant sera sauvage il aura perdu l'allèle muté et toute sa descendance sera sauvage. On suit l'allèle muté de génération en génération. Mais cela pourrait-il aussi être observé dans l'hypothèse de la récessivité? Si tel était le cas, chaque parent sain ayant un ou plusieurs enfants atteint devrait être hétérozygote (ou du moins chaque femme si le gène est sur l'X). On compte combien de parents de ce type sont issus de la population générale. On peut estimer la fréquence p de l'allèle muté dans la population générale à partir de la fréquence des individus atteints $(a1/a1) p^2$ et en déduire la fréquence dans cette population générale des hétérozygotes porteurs sains $(2(1-p^2)p)$: de l'ordre de 1 à 2 %. Si l'hypothèse de la récessivité nécessite n parents sains porteurs issus de la population générale, la probabilité du pedigree observé est de $(1 \text{ à } 2\%)^n$.

Par contre, dans l'hypothèse de la dominance, il suffit d'un seul individu atteint ($a1/a+$) (dit fondateur) issu de la population générale pour observer ce pedigree. Il suffit de comparer les probabilités des deux hypothèses, on choisit la plus probable.

I) Dans cette famille, sur un nombre limité de descendants, on voit plusieurs individus atteints de dystrophie donc la fréquence est significativement plus élevée que dans la population générale ($3/32$) et on a le droit de penser que la dystrophie est héréditaire dans cette famille.

Le couple II-3 x II-4 a deux enfants atteints. Compte tenu de la rareté des mutations, la seule interprétation plausible est que les enfants ont hérité un allèle muté de leur(s) parent(s) qui sont sains. Le phénotype [dystrophique] est donc récessif sur le phénotype sauvage.

On trouve uniquement des enfants mâles atteints dans la descendance de couples sains. C'est l'indice d'une localisation possible sur le chromosome X. Pour établir cette localisation on peut dans tous les croisements entre deux parents sains qui ont des enfants atteints voir la répartition de ceux-ci :

mâles [sains]	7
mâles [atteints]	3
femelles [saines]	10
femelles [atteints]	0

	femelles observées	si localisation sur le X	si localisation autosomique
sains	10	10	$\frac{3}{4}$ du total soit $10 \cdot \frac{3}{4} = 7,5$
atteints	0	0	$\frac{1}{4}$ du total soit $10 \cdot \frac{1}{4} = 2,5$
	total 10	$\chi^2 = 0$	test du χ^2 non valable : une catégorie théorique est < 5

Aucune des deux hypothèses ne peut être rejetée de cette façon.

On peut évaluer la probabilité d'une telle famille sous chacune des deux hypothèses.

Dans l'hypothèse autosomique, le mâle qui est puisé dans la population générale doit être hétérozygote.

Dans la population générale on admet en première approximation que les gamètes se rencontrent au hasard (hypothèse de Hardy-Weinberg). Ici, pour la dystrophie on est en présence de deux formes alléliques

a⁺ sauvage, soit p la fréquence des gamètes qui le portent

a1 mutant, soit q la fréquence des gamètes qui le portent (dans cette catégorie on regroupe tous les allèles responsables de la dystrophie).

On a p+q=1 (ce sont les deux seules formes possibles).

Si les gamètes se rencontrent au hasard on attend la composition suivante de la population générale:

gamète mâle gamète femelle	a ⁺ fréquence p	a1 fréquence q(1-p)
a ⁺ fréquence p	$\frac{a^+}{a^+}$ [sain] p ²	$\frac{a^+}{a1}$ [sain] pq
a1 fréquence q(1-p)	$\frac{a^+}{a1}$ [sain] pq	$\frac{a1}{a1}$ [atteint] q ²

Pour connaître les valeurs de p et q on sait que la fréquence des dystrophiques dans la population générale est

$$1/40\ 000 \text{ donc } q^2 = 1/40\ 000 \quad q = \sqrt{1/40\ 000} = 1/2 \times 10^{-2} = 0,005$$

$$2pq \text{ fréquence des hétérozygotes} = 2 \times 0,005 \times (1-0,005) = 0,0095$$

fréquence des individus sains non porteurs = $(0,995)^2 = 0,989025$
 fréquence de tous les individus sains $0,00995 + 0,989025 = 0,998975$

Dans un calcul rigoureux il faudrait corriger la fréquence des hétérozygotes, en effet on sait qu'ils sont sains:

fréquence des hétérozygotes parmi les sains = $0,00995/0,998975 = 0,0099521$ (soit approximativement différent de 10^{-2})

Pour que la famille A existe il faut

- que l'un des deux parents de la génération I soit hétérozygote, afin que les enfants du couple puissent également l'être : Probabilité = 10^{-2}
- que chaque parent puisé dans la population générale et qui a eu un ou plusieurs enfants atteints soit hétérozygote. Il y a 2 sujets II-1 et II-3 : Probabilité = $(10^{-2})^2 = 10^{-4}$

La probabilité de cette famille, sous l'hypothèse autosomique, est donc $10^{-2} * 10^{-4} = 10^{-6}$.

Dans l'hypothèse « gène sur l'X », la femelle porteuse saine I-2 a pour génotype (a1/a+). Dans ce cas la plupart des malades sont des mâles (a1/Y) fréquence $q=1/40000$. Une femelle hétérozygote de la population générale a une fréquence de $2(1/40000)(1-1/40000) = 10^{-4}$. Cette dernière hypothèse plus probable est retenue.

Le calcul des probabilités permet souvent de conclure, comme c'est ici le cas.

II) Dans cette famille, sur un nombre limité de descendants, on voit plusieurs individus atteints de dystrophie donc la fréquence est significativement plus élevée que dans la population générale et l'on a le droit de penser que la dystrophie est héréditaire dans cette famille.

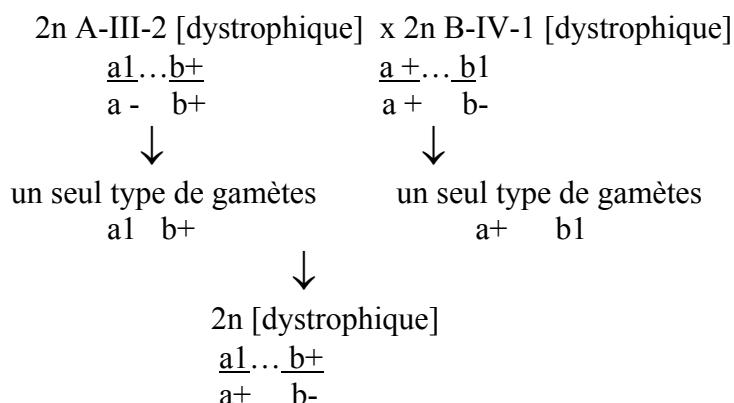
1°) Le couple III-4 x III-5 a deux enfants atteints. L'événement mutation est trop rare pour rendre compte de ces deux descendants malades. Il faut donc supposer que si 2 enfants sont atteints, c'est que les parents sains sont cependant porteurs. Le phénotype [dystrophique] est récessif sur le phénotype sauvage. On trouve des enfants des deux sexes atteints parmi les descendants d'un couple sain. Le gène en cause est donc autosomique. Or, dans l'hypothèse sur l'X si une fille est atteinte, son père doit être hémizyote pour l'allèle muté et il ne peut être sain.

Les deux couples qui ont donné naissance à des enfants atteints constituent des unions consanguines. La consanguinité augmente la probabilité de recevoir de chaque parent une même copie d'un allèle ancestral. Cette remarque n'a de valeur que dans le cas d'une maladie à transmission récessive autosomique.

2°) Il s'agit d'un test de complémentation destiné à déterminer si deux mutants présentant le même phénotype sont touchés dans le même gène ou dans des gènes différents.

Puisque le phénotype [dystrophique] est récessif dans chacune des deux familles, les individus atteints sont homozygotes. Chaque parent fournit un seul type de gamètes. Les gamètes de A-III-2 sont porteurs d'une mutation (au moins) qui inactive un gène qu'on appellera A. Cet allèle différent de celui de la souche sauvage sera noté a1. Les enfants (diploïdes) résultant du croisement ont reçu l'allèle a1 de ce parent. Puisque leur phénotype est [sain] c'est que l'allèle du gène A fourni par l'autre parent est a+. Ce second gamète fourni par B-IV-1 provient d'un individu atteint et il est donc muté dans un autre gène: B il porte l'allèle b1 que les enfants ont reçu. Pour ce second gène, l'allèle fourni par le parent de la

famille A est forcément b^+ puisque les enfants diploïdes sont sains. Donc le premier gamète haploïde a pour génotype $a^1 b^+$ et le second gamète haploïde $a^+ b^1$. Il y a COMPLEMENTATION: les deux haploïdes sont mutés sur DES GENES DIFFERENTS. Le croisement s'écrit:



Les deux familles sont mutées dans des gènes différents.

3°) Ce résultat était prévisible puisque les deux gènes avaient des localisations différentes, l'un autosomique et l'autre sur l'X.

III) Dans la famille A :

* les mâles sains, II-1, 3, 5, 7 et 9 ont pour génotype a^+/Y . (ainsi que III-5, 8, 11, 12, 13, 14, et 18)

* les mâles dystrophiques, III-12, 6 et 7 sont a^1/Y

* les femelles saines qui ont eu des enfants atteints sont hétérozygotes a^+/a^1 . Il s'agit de II-2 et 4

* les filles dont le père était sain et les mères issues de la population générale sont très probablement a^+/a^+ . Ce sont les femmes III-9, 10, 15, 16, 17, 19 et 20.

* les filles saines issues de parents sains dont un est porteurs peuvent être soit a^+/a^+ soit a^+/a^1 . Ce sont les femmes III-1, 3 et 4.

Dans la famille B, que les sujets soient mâles ou femelle ne joue pas puisque le gène est porté par un autosome.

* Les individus sains issus de la population générale et qui n'ont eu que des descendants sains ont une probabilité de 0,99 d'être b^+/b^+ . Ce sont les sujets I-2 et II-1. (ainsi que II-3, 5 et 7).

* Les individus sains qui ont eu dans leur descendance des sujets atteints (F1 ou F2) sont porteurs b^+/b^1 . Ce sont les sujets II-2 et III-4. (ainsi que II-4, 6 et 8 et III-5, 9 et 10).

* Les individus sains qui n'ont pas eu de descendance mais qui ont des parents porteurs peuvent être b^+/b^+ ou b^+/b^1 . Il s'agit de III-2 et 3. (ainsi que III-1, 6, 7, 8, 11, 12 et 13 et IV-2, 4, 5, 6, 7 et 8).

* Les sujets atteints portent deux allèles mutés. Ce sont les sujets I-1 et IV-1, 3 et 9. Pour I-1 aucune consanguinité n'est signalée, les deux allèles peuvent être différents : b1/b2. Pour les sujets de la génération IV les trois possibilités coexistent : b1/b1, b1/b2, b2/b2.

IV) Les sujets IV-2 et IV-6 sont tous deux issus de parents hétérozygotes ils ont une chance sur deux que chaque parent leur ait donné un gamète a1 , soit une probabilité de 0,25 d'être homozygote sain et 0,5 d'être hétérozygote porteur. On sait qu'ils sont sains (proba = 1) donc :

$$\text{probabilité d'être homozygote} = \frac{0,25}{0,25 + 0,75} = 0,33$$

$$\text{probabilité d'être hétérozygote} = \frac{0,25}{0,25 + 0,75} = 0,67$$

		sujet sain		
		homozygote Prob 0,33	hétérozygote Prob 0,67	
		gamètes b+ Pb 1	gamètes b+ Pb 0,5	gamètes b1 Pb 0,5
homozygote Prob 0,33	gamètes b+ Pb 1	$\frac{b+}{b+}$ [sain] Pb 0,33 ²	$\frac{b+}{b+}$ [sain] Pb 0,33x0,67x0,5	$\frac{b+}{b1}$ [sain] Pb 0,33x0,67x0,5
hétérozygote Prob 0,67	gamètes b+ Pb 0,5	$\frac{b+}{b+}$ [sain] Pb 0,33x0,67x0,5	$\frac{b+}{b+}$ [sain] Pb (0,67x0,5) ²	$\frac{b+}{b1}$ [sain] Pb (0,67x0,5) ²
	gamètes b1 Pb 0,5	$\frac{b+}{b1}$ [sain] Pb 0,33x0,67x0,5	$\frac{b+}{b1}$ [sain] Pb (0,67x0,5) ²	$\frac{b1}{b1}$ [atteint] Pb (0,67x0,5) ²

Ce qui donne comme probabilité d'avoir un enfant atteint:
 $((0,67 \times 0,50) \times (0,67 \times 0,50)) = 0,1089$

Si le couple était IV-2 et IV-9, ce dernier individu est homozygote b1. Il ne fournit qu'un seul type de gamètes et le tableau de gamètes devient donc:

		sujet
		Homozygote b1/b1 Prob 1
		gamètes b1 Pb 1
homozygote Prob 0,33	gamètes b+ Pb 1	$\frac{b+}{b1}$ [sain] Pb 0,33 b1
hétérozygote Prob 0,67	gamètes b+ Pb 0,5	$\frac{b+}{b1}$ [sain] Pb 0,67x0,5
	gamètes b1 Pb 0,5	$\frac{b1}{b1}$ [atteint] Pb 0,67x0,5

Et la probabilité d'avoir un enfant atteint :
 $0,67 \times 0,50 = 0,33$.

Facteur IX

Le facteur IX humain est une protéine de coagulation qui est synthétisée en une seule chaîne polypeptidique de 461 aminoacides (numérotés de -46 à +415). L'absence ou la dysfonction du facteur IX résulte en une déficience hémorragique: l'hémophilie B. Des patients hémophiles non apparentés portent des mutations diverses. Les individus atteints d'hémophilie B peuvent être divisés en deux classes :

- les CRM+ qui possèdent des niveaux normaux de facteur IX dosé par test immunologique mais qui ont une activité coagulante du facteur IX diminuée ou nulle.
- les CRM- qui ont des niveaux très réduits ou indétectables de facteur IX aussi bien par dosage d'activité coagulante que par test immunologique.

Différents patients hémophiles déficients en facteur IX ont été étudiés (détermination de l'activité coagulante et de l'immunogénicité) et on a déterminé la séquence du facteur IX.

patient	activité coagulante	dosage immunologique	facteur IX
CRM+			
Cambridge	2	50	Arg-1→Ser
London 3	1	48	Arg-4→Gln
Malm 6	<1	30	Arg4→Trp
md 1	3	100	Asp47→Lys
Alabama	1	20	Asp47→Gly
md 3	8	100	Asp49→Glu
London 7	10	49	Pro55→Ala
md 4	2	100	Asp64→Lys
md 5	2	100	Asp64→Val
md 6	52	20	Glu78→Asp
md 7	<1	30	Glu78→Lys
ChapelHill	3	80	Arg145→His
Unnamed	<1	?	Arg252→Leu
London 5	1	75	Arg333→Gln
Lake Elsinore	2	83	Ala390 →Val
Vancouver	1	69	Thr397 →Ileu
London 10	<1	12	Arg37 manque
CRM-			
London 8	2	2	Cys336→Arg
London 9	<1	0.4	Asn120→Tyr
London 11	<1	<2	voir séquence
London 12	<1	<0.1	voir séquence
Malm 1	<1	<0.1	voir séquence
X 1	<1	<0.1	comme + du début à 85 et et de 195 à la fin. AA 86 à 194 manquent
Malm 3	<1	<0.1	fin en Arg248
Malm 4	<1	<0.1	fin en Arg 29
Malm 5	<1	<0.1	fin en Trp194

On a également mis en évidence des cas où tout le gène codant pour le facteur IX est délété.

Interpréter les différentes mutations.

1°) Substitutions d'acides aminés :

A quelles mutations sont-elles dues,

Quelles propriétés du code génétique mettent-elles en évidence.

2°) Protéines raccourcies :

Pourquoi ce raccourcissement se fait-il toujours du même côté?

Quelles sont les mutations portées par les variants Malmö 3,4 et 5?

Quelles sont les mutations portées par les variants London 12,11 et Malmö 1?

Quelles indications peut-on tirer de la comparaison des séquences mutante et sauvage?

Quelles propriétés du code génétique mettent-elles en évidence?

Bibliographie

GREEN et al : EMBO J.8 1067-1072 (1989)

YOSHITAKE et al : Biochemistry 24 3736-3750 (1985)

Séquences de différentes formes alléliques du facteur IX

Séquence nucléotidique de 1 à 143 du facteur IX allèle sauvage et allèle London 12

Abréviations représentant les ambiguïtés de l'ADN, (le code est dit « dégénéré » *)

R=A ou G, Y=C ou T, H=A, C ou T, N=A, C, G ou T

FIX +	{	1	GlyLysLeuGluGluPheValGlnGlyAsnLeuGluArgArgGluCysMETGluGluLys
		1	GGNAARTTRGARGARTTYGTNCARGGNAAYTTRGARAGRARGARTGYATGGARGARAAR
			CTN CTN CGNCGN
			TCNCTN CTNCGN TCN
L12	{	1	GGNAARTGGAARAGYTRRTTYAARGGNACNTTRAGRGARAAAYGTNTGGAARAARAGYGTN
		1	GlyLysTrpLysSerLeuPheLysGlyThrLeuArgGluAsnValTrpLysLysSerVal
FIX +	{	21	CysSerPheGluGluArgGluValPheGluAsnThrGluArgThrThrGluPheTrpLys
		61	TGYAGYTTYGARGARAGRARGTNTTYGARAAYACNGARAGRACNACNGARTTYTGGAAR
			TCN CGN CGN
			CTN CTN CTN CTN
L12	{	61	GTNTTRAARAAYGARAARCCNTTRAARACNTTRAARGARCARTTRAAYCCNGGNAGYAGY
		21	ValLeuLysAsnGluLysProLeuLysThrLeuLysGluGlnLeuAsnProGlySerSer
FIX +	{	41	GlnTyrValAspGlyAspGlnCysGluSerAsnProCysLeu
		121	CARTAYGTNGAYGGNGAYCARTGYGARAGYAAAYCCNTGYTTR
			TCN CTN
			CTN TCN TCN
L12	{	121	ATGTTRATGGARATHAGYGTNAGYCCNATHCAY
		41	METLeuMETGluIleSerValSerProIleHis

- * Les travaux sur l'évolution du code génétique ont montré que la synonymie des codons était un caractère de résistance du code aux mutations ; le code actuel est ainsi particulièrement résistant aux perturbations qui viendraient modifier son information.
-
- C'est donc en fait un **code « évolué »** et non « dégénéré ».

Séquence nucléotidique de 310 à 340 du facteur IX allèle sauvage et allèle London 11.

Abréviations représentant les ambiguïtés de l'ADN, (puisque le code est dégénéré)

R=A ou G, Y=C ou T, H=A, C ou T, N=A, C, G ou T

FIX +	$\left\{ \begin{array}{l} 310 \\ 310 \end{array} \right.$	GlyTrpGlyArgValPheHisLysGlyArgSerAlaLeuValLeuGlnTyrLeuArgVal
		GGNTGGGGNAGRGTNTTYCAYAARGGNAGRAGYGCNTTTRGTNTTRCARTAYTTRAGRGTN
		CGN CGN CGNCTN CTN CTN CTNCTNCTNCTNCTN
L11	$\left\{ \begin{array}{l} 310 \\ 310 \end{array} \right.$	GGNTGGGGNAGRGTNCCNCARAGRGARATHAGYTTTRAGYAGYAGYGTN
		GlyTrpGlyArgValProGlnArgGluIleSerLeuSerSerSerVal

Séquence nucléotidique de 275 à 290 du facteur IX allèle sauvage et allèle London 11.

Abréviations représentant les ambiguïtés de l'ADN, (puisque le code est dégénéré)

R=A or G, Y=C or T, H=A, C or T, N=A, C, G or T

FIX +	$\left\{ \begin{array}{l} 273 \\ 273 \end{array} \right.$	LeuGluLeuAspGluProLeuValLeuAsnSerTyrValThrProIleCysIleAlaAsp
		TTRGARTTRGAYGARCCNTTTRGTNTTAAAYAGYTAYGTNACNCCNATHGTGYATHGCNGAY
		CTN CTN CTN CTN CTN TCN CTN CTN CTNCTNCTNCTNCTN
M1	$\left\{ \begin{array}{l} 273 \\ 273 \end{array} \right.$	TTRGARTTRGAYAGYGCNAARCARTTTRAGRTAYACNTAYTTTR
		LeuGluLeuAspSerAlaLysGlnLeuArgTyrThrTyrLeu

Corrigé : Facteur IX

1°) Substitutions d'acides aminés

mutant	n° de la base modifiée dans le codon	brin non transcrit	
		triplet sauvage	triplet mutant
Cambridge	3	AGA ou AGG	AGC ou AGT
London 3	2	CGx	CAA ou CAG
Malmö 6	1	CGG	TGG
md 1	1 et 3	GAC ou GAT	AAA ou AAG
Alabama	2	GAC ou GAT	GGC ou GGT
md 3	2	GAC ou GAT	GAA ou GAG
md 5	2	GAC ou GAT	GTC ou GTT
md 6	3	GAA ou GAG	GAC ou GAT
md 7	1	GAA ou GAG	AAA ou AAG
Chapel Hill	2	CGC ou CGT	CAC ou CAT
Unnamed	2	CGx	CTx
London 5	2	CGA ou CGG	CAA ou CAG
Lake Elsinor	2	GCx	GTx
Vancouver	2	ACA, ACC ou ACT	ATA, ATC ou ATT
London 10			délétion de 3 bases
London 8	1	TGC ou TGT	CGC ou CGT
London 9	1	AAC ou AAT	TAC ou TAT

mutant	n° de la base modifiée dans le codon	brin non transcrit	
		triplet sauvage	triplet mutant
London 11			délétion des 2 premières bases du triplet 314
London 12			délétion de la première base du triplet 3
Malmö 1			délétion de 8 bases à partir de la 1e du triplet 277
X			délétion de 327 bases (109 triplets)
Malmö 3	1	CGA	TGA (non sens)
Malmö 4	1	CGA	TGA(non sens)
Malmö 5	3	TGG	TGAou TAG (non sens)

Ces mutations montrent que le code génétique est non-chevauchant : une seule base changée ne modifie qu'un acide aminé et qu'il est sans virgule : la délétion d'un nombre de bases non-multiple de 3 décale le cadre de lecture.

2°) Le raccourcissement se fait toujours du côté COOH parce que la traduction se fait de 5' vers 3' et est arrêtée en cours par un codon non sens qui empêche d'aller plus loin en 3'. Donc les protéines sont synthétisées de NH₂ vers COOH (ADN brin non transcrit de 5' vers 3').

Les mutants Malmö 3, 4 et 5 sont des mutants non sens créés par la substitution d'une base. London 12, 11 et Malmö 1 portent des mutations dites de décalage de cadre de lecture (ou frameshift) résultant d'une délétion (ou une addition) d'un nombre de bases non multiple de 3.

La comparaison des séquences mutante et sauvage permet de déterminer la base mutée et d'établir sans séquençage de l'ADN sa séquence exacte en excluant les triplets qui ne rendent pas compte à la fois de la séquence protéique sauvage et de celle du mutant.

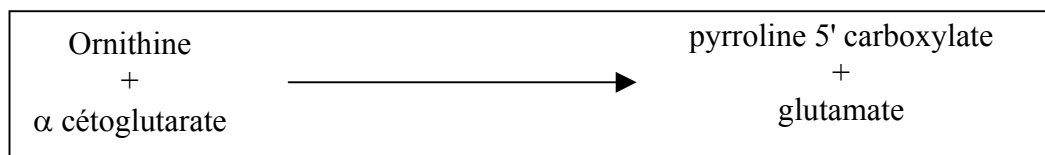
Le code est sans virgule puisqu'il peut y avoir des décalages du cadre de lecture du seul fait d'ajouter ou d'ôter un nombre de bases non-multiple de 3

Les cas L11, L12 et M1 ont un non-sens qui est lu respectivement 10, 49 et 10 AA après la mutation.

En admettant que la composition de l'ADN soit A=C=G=T=25% (ce qui n'est que rarement le cas) chaque codon parmi les 64 du code génétique a la même probabilité. On attend donc dans une séquence d'ADN au hasard (sans pression de sélection) 3 codons non sens pour 61 codons sens, soit une fréquence de l'ordre de 1/20 : 1 non-sens tous les 20 codons sens. Tel n'est pas le cas dans les phases ouvertes de lecture puisque la sélection naturelle a maintenu au cours des générations le gène fonctionnel. Par contre la séquence telle qu'elle est lue après une insertion (mutation) peut être assimilée à une séquence au hasard et les valeurs observées (10,49,10 avec une moyenne = 23) en sont le reflet.

Diagnostic prénatal de l'atrophie gyrée.

L'Ornithine Amino Transférase (OAT) est une enzyme de la matrice mitochondriale qui est codée par un gène nucléaire. Elle catalyse la réaction :



Chez l'être humain une déficience en OAT résulte en des niveaux augmentés en ornithine dans le plasma et est associée à une perte de vision due à une forme rare de dégénérescence chorioretinale progressive : l'atrophie gyrée. L'AG a une fréquence de 10^{-5} dans la population générale. Le défaut enzymatique est exprimé, entre autre, dans les fibroblastes de la peau lorsqu'ils sont mis en culture.

1°) Alors que cette affection est rare dans la population générale, on a pu établir le pedigree d'une famille dont plusieurs membres sont atteints (figure 1).

Montrer que l'AG est :

1. -héréditaire
2. -récessive
3. -autosomique.

(On admettra dans tout l'exercice que les défauts génétiques envisagés ont un déterminisme monogénique.)

2°) On a mis en culture des fibroblastes prélevés sur des individus non apparentés atteints d'AG. Ces fibroblastes diploïdes ont ensuite fusionné et les cellules résultantes tétraploïdes sont étudiées. Dans ces cellules on a dosé le niveau d'OAT :

diploïde	mutants					sauvage
mutants	A	B	C	D	E	+
A	0	0	0	0	0	101
B		0	0	0	0	99
C			0	0	0	103
D				0	0	105
E					0	97

Dans le tableau sont portés les niveaux d'activité d'OAT dans les cellules obtenues par fusion des deux parents correspondants.

Que peut-on dire du défaut enzymatique de chaque mutant par rapport au sauvage?
Comparer les mutants entre eux (expliquer le principe de cette expérience).

3°) L'étude du gène impliqué dans l'AG a permis :

- de cloner l'ADNc de ce gène (en rappeler le principe).
- de cloner l'ADN génomique en deux fragments (R5 et R10).
- de mettre en évidence que les différents mutants ne portaient pas de gène ayant des modifications repérables (délétion ou modification de carte de restriction).

On a donc cherché à repérer de façon indirecte un éventuel gène muté à l'aide du polymorphisme des fragments de restriction. Pour cela on a recherché un polymorphisme au voisinage du gène impliqué en utilisant comme sonde R5, R10 et deux morceaux de l'ADNc : l'extrémité 3' et l'extrémité 5'. Des lignées cellulaires établies à partir d'un certain nombre de sujets sains non apparentés issus de la population générale (34 à 48) ont été utilisées comme source d'ADN. Cet ADN total a été soumis à différentes digestions par un enzyme de restriction puis les fragments obtenus ont été séparés en fonction de leur taille par électrophorèse sur gel d'agarose, transférés sur membrane de nylon et hybridés avec une des 4 sondes marquées au ^{32}P définies ci-dessus. La digestion par 20 enzymes de restriction donne des bandes d'hybridation identiques chez tous les sujets. Par contre dans 6 cas on a trouvé une hétérogénéité :

enzyme de restriction	sonde	fragment	taille (kb)	fréquence allélique	nb de sujets examinés
Sau96I	R5	A1	0.66	0.24	44
		A2	0.56	0.76	
5RsaI	R5	B1	1.30	0.13	34
		B2	1.10	0.87	
BanI	R5	C1	5.40	0.95	37
		C2	3.30	0.05	
HincIOI	R10	D1	3.10	0.82	34
		D2	2.30	0.18	
MspI	extrémité 5' de l'ADNc	E1	4.10	0.78	48
		E2	3.30	0.22	
MspI	extrémité 3' de l'ADNc	F1	4.10	0.69	48
		F2	3.40	0.31	

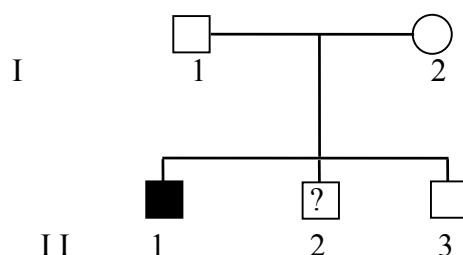
Interpréter cette hétérogénéité au niveau de l'ADN.

Quel est le génotype probable des individus testés, du point de vue de l'AG ?

Classer les polymorphismes par ordre d'intérêt décroissant (justifiez votre classement).

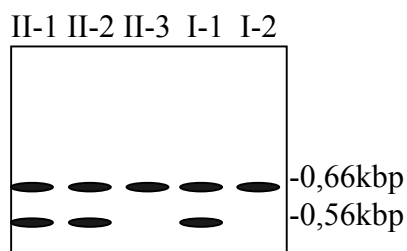
On a étudié l'ADN de tous les individus disponibles d'une famille dans les conditions précédentes permettant de révéler les fragments B1 et B2 (figure 2). Ecrire le génotype de tous les individus (même II-1)

4°) Dans une famille est né un enfant II-1 qui, à partir de la 7^{ème} année a présenté des signes d'AG :

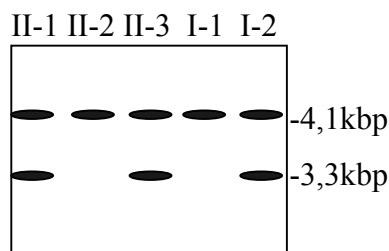


- Qu'est-ce que cela implique pour les individus de la génération I?
- Les individus II-2 et II-3 sont encore trop jeunes pour manifester un éventuel phénotype atteint. Quelle est la probabilité qu'ils le soient cependant?

- On a pu compléter l'analyse par l'étude de la ségrégation des bandes A1 et A2 (polymorphisme Sau96I) et E1 et E2 (polymorphisme Msp I).



Digestion par Sau96I



Digestion par MspI

- A l'aide de ces résultats, préciser le génotype des 5 individus de la famille.
- Les deux polymorphismes apportent-ils autant d'information? Pourquoi?
- Quelle précision apporte cette analyse des RFLP pour les individus II-2 et II-3?
- Quelle est la limite de la méthode?
- On a mis en évidence 5 polymorphismes différents dans la population générale (voir question 3). Expliquer l'intérêt d'avoir plusieurs polymorphismes liés à la même région du génôme (ici liaison au gène codant pour l'OAT).

Bibliographie

- PNAS **74** 5159-5161 (1977)
 Am J Hum Genet **30** 174-179 (1978)
 DNA **5** 493-501 (1986)
 Am J Hum Genet **42** 365-372 (1988)

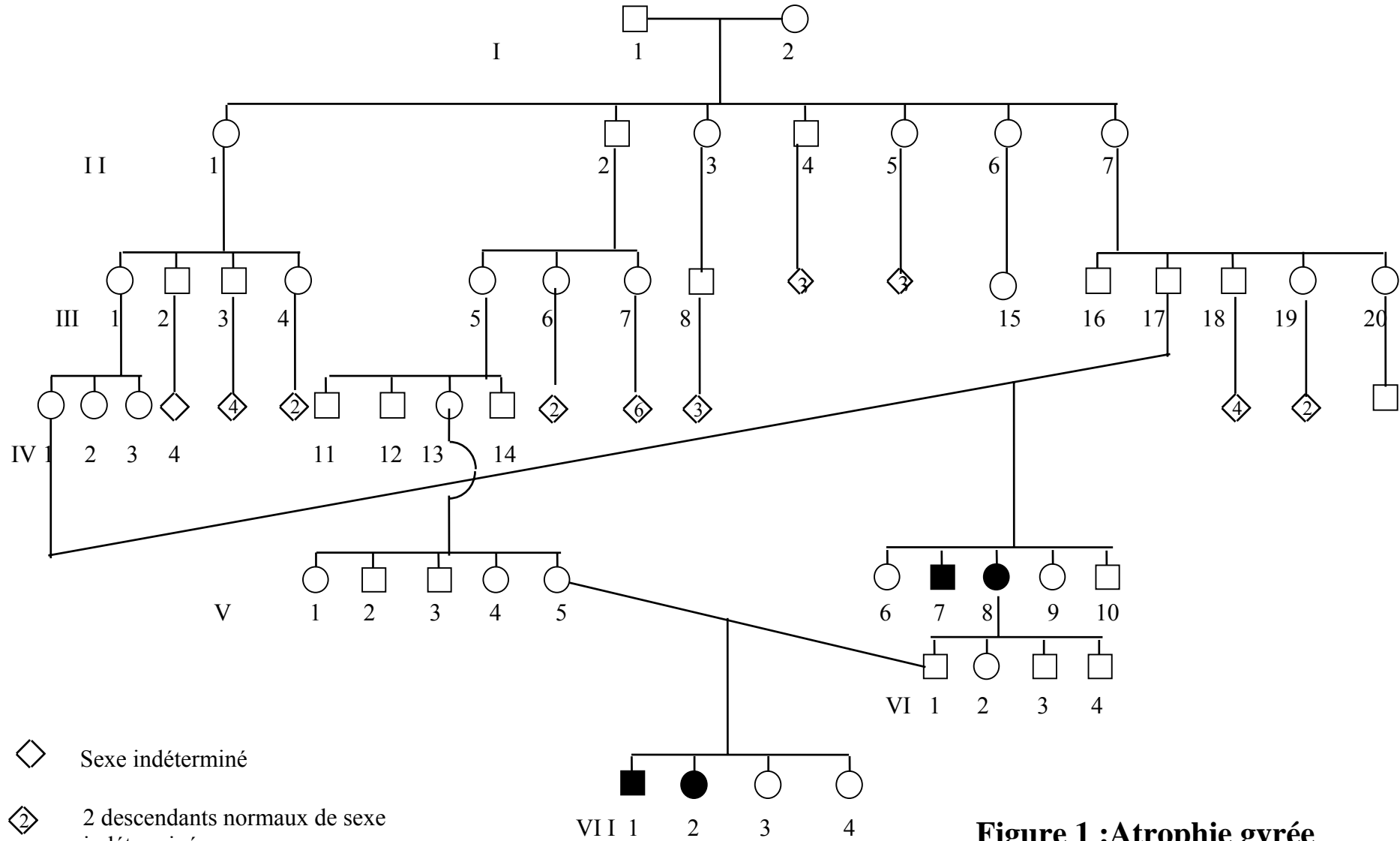
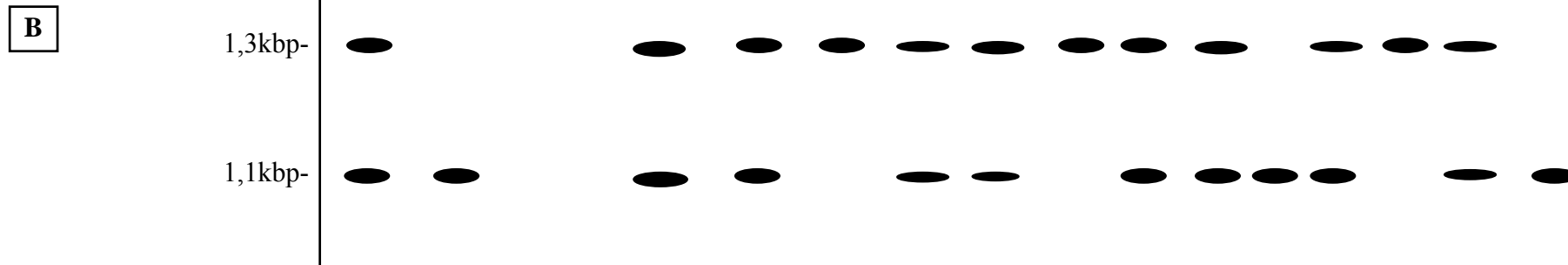
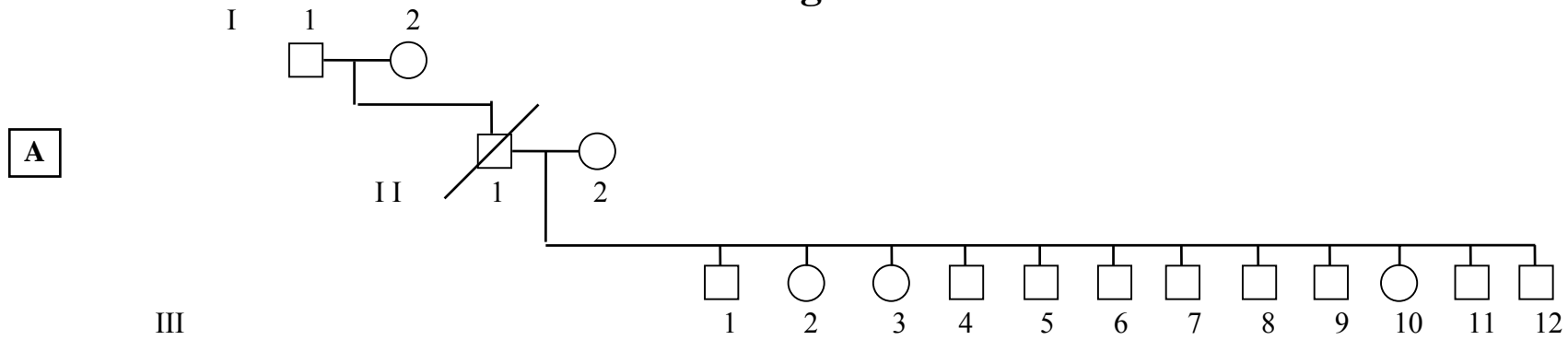


Figure 1 : Atrophie gyrée

Figure 2



A: généalogie d'une famille dont tous les membres sont sains. Le sujet II-1 est mort.

B: électrophorégramme des l'ADN total de chaque membre vivant de la famille considérée.

Corrigé: Diagnostic prénatal de l'atrophie gyrée.

1°) Dans cette famille on voit sur un nombre limité de descendants plusieurs individus atteints d'atrophie gyrée, la fréquence de cette maladie est donc significativement plus élevée dans la famille que dans la population générale, ce qui suggère fortement que l'atrophie gyrée dans cette famille est héréditaire. On pourrait également penser qu'un environnement commun particulier est à l'origine de cette maladie, mais la transmission du caractère selon le modèle mendélien va confirmer le déterminisme génétique de la maladie.

En effet on observe des sauts de génération entre individus atteints, des parents sains ont plusieurs enfants atteints, ce qui ne peut être expliqué par une mutation sur un des gamètes entre les deux générations. Donc le phénotype [AG] est récessif sur le phénotype sauvage.

Hypothèse gène autosomique

S'ils ont des enfants atteints
c'est qu'ils sont porteurs.

$$\begin{array}{c} \frac{a-}{a+} \quad \times \quad \frac{a-}{a+} \\ \downarrow \qquad \qquad \downarrow \\ \text{2 types de gamètes} \end{array}$$

	a-	a+
a-	a-/a-[atteint]	a-/a+[sain]
a+	a-/a+[sain]	a-/a+[sain]

mâles et femelles

[+] et [m]

la présence d'une femelle [m]

élimine l'hypothèse sur l'X

Hypothèse gène sur l'X

S'ils ont des enfants atteints
c'est que la femelle est porteuse.

$$\begin{array}{c} \frac{a+}{\neg} \quad \times \quad \frac{a-}{a+} \\ \downarrow \qquad \qquad \downarrow \\ \text{2 types de gamètes} \end{array}$$

	a-	a+
\neg	a-/ \neg [atteint]	a+/ \neg [sain]
a+	a-/a+[sain]	a-/a+[sain]

mâles [+] et [m]

femelles [+]

seules des considérations statistiques

peuvent faire pencher pour cette hypothèse

en l'absence de filles [m].

La présence de filles atteintes dans la descendance de parents sains démontre que le gène en jeu est autosomique.

Les deux couples qui ont donné naissance à des enfants atteints constituent des unions consanguines. Or la consanguinité n'a d'importance que si les deux parents transmettent la maladie, donc si le gène en cause est autosomique.

2°) Il s'agit d'un test de complémentation destiné à déterminer si deux mutants sont touchés dans le même gène ou dans des gènes différents.

Puisque le phénotype [OAT -] est récessif les cellules mises en culture, provenant d'individus atteints sont homozygotes. Ici le TCF s'effectue entre des diploïdes homozygotes et on observe le tétraploïde hétérozygote. Le mode de raisonnement est identique au cas classique.

L'observation du phénotype des tétraploïdes (mutant x sauvage) permet de dire que dans tous les cas le phénotype [OAT-] est récessif devant le phénotype sauvage dans les cellules en culture. Les conditions sont différentes de celles illustrées dans le pedigree, ce test n'est donc pas superflu.

Dans les croisements du type (mutant x mutant) le diploïde A est homozygote porteur d'un allèle muté (au moins) du gène A. Cet allèle différent de celui de la souche sauvage sera noté (a-). Le tétraploïde résultant du croisement a reçu les 2 allèles (a-) du mutant A. Puisque son phénotype est [OAT-] c'est que les 2 autres allèles sont aussi (a-). L'autre diploïde B est donc homozygote (a-) ce qui est suffisant pour rendre compte de son phénotype [OAT-].

Donc le diploïde A a pour génotype $\frac{a-}{a-}$ et le diploïde B $\frac{a-}{a-}$. Il n'y a **pas** de complémentation, les deux diploïdes sont mutés sur le **même** gène. Le croisement s'écrit:

$$\begin{array}{ccc}
 2n \text{ A [OAT-]} & \times & 2n \text{ C [OAT -]} \\
 \frac{a_1^-}{a_1^-} & & \frac{a_2^-}{a_2^-} \\
 & \downarrow & \\
 & & \frac{a_1^-}{a_2^-} \\
 4n \text{ [OAT -]} & & \frac{a_2^-}{a_2^-}
 \end{array}$$

On nomme les deux allèles a1 et a2 afin de préciser qu'ils n'ont aucune raison d'être identiques, puisqu'ils sont le résultats d'événements mutationnels indépendants.

L'examen de tout le tableau montre que tous les mutants sont mutés dans le même gène.

3°) Pour cloner l'ADN copie il faut extraire l'ARN total puis isoler l'ARNm caractérisé par une queue polyA en 3'. L'ARN total est centrifugé en gradient de sucrose et l'on ne garde que ce qui est entre 5 et 16S. On élimine ainsi la majeure partie des ARNt et des ARNr 18 et 28S. Cette fraction est mise sur colonne d'oligo dT. Les ARN polyA sont retenus et on les élue ensuite en augmentant la concentration en sels. Avec une amorce oligo dT comme primer

on fait copier l'ARNm par la réverse transcriptase. Le second brin de l'ADNc est complété par la polymérase Klenow par exemple. La boucle de réplication est éliminée par traitement à la nucléase S1 qui ne digère que l'ADN simple brin.

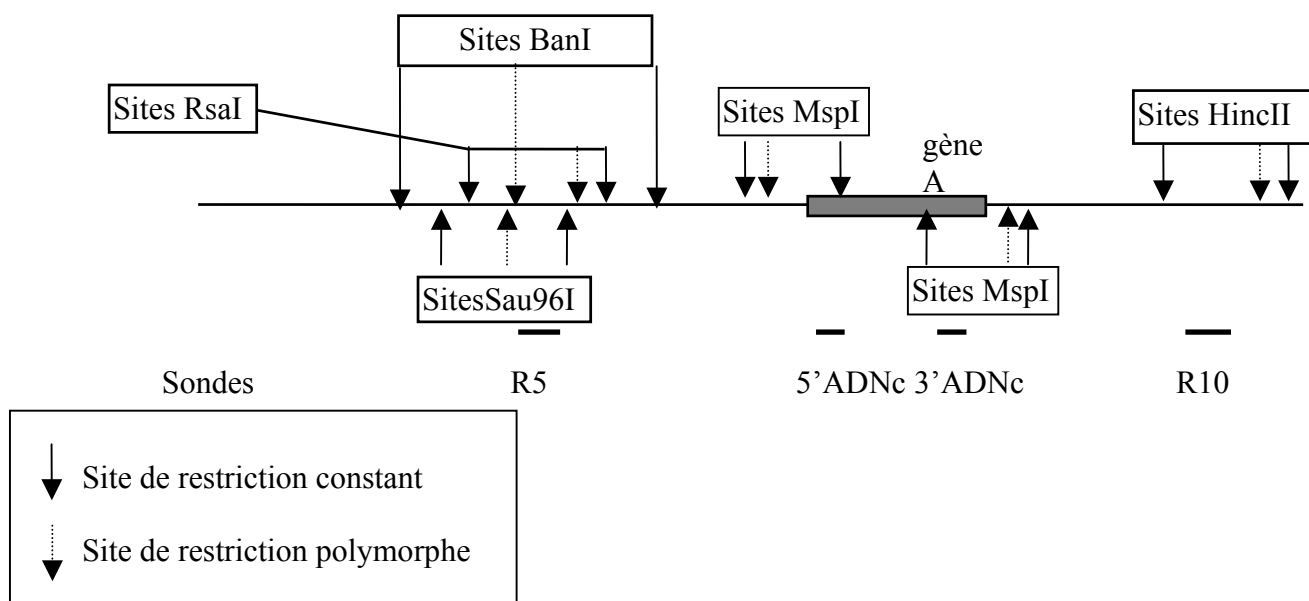
Ensuite il faut mettre des bouts collants aux fragments ainsi obtenus : Après traitement à l'EcoRI méthylase qui rend inaccessibles les sites EcoRI en les méthylant, on fait agir une ligase en présence de linkers EcoRI puis on digère par EcoRI.

Par ailleurs on a préparé un plasmide digéré par EcoRI. Les produits des deux digestions sont mélangés et reliqués. Le tout est utilisé pour transformer des bactéries.

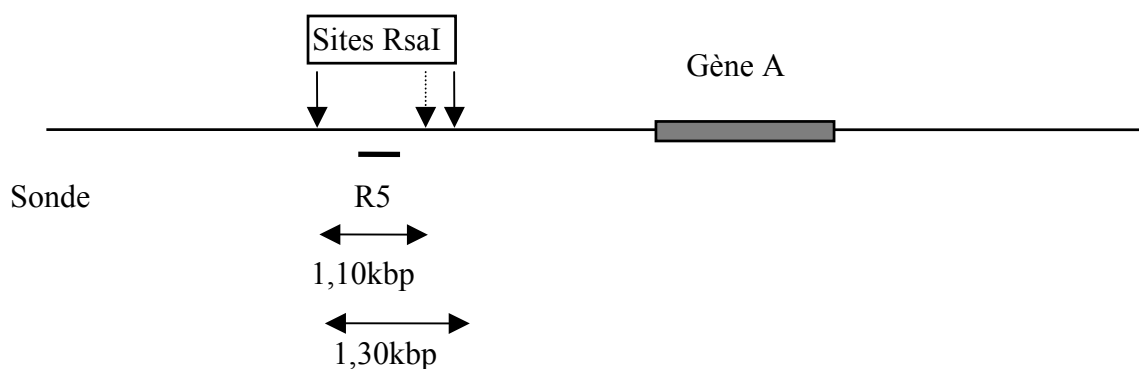
Les plasmides qui portent tous le gène de résistance à l'ampicilline servent pour transformer une souche bactérienne sensible à cet antibiotique. Seules les cellules qui auront reçu un plasmide pousseront sur un milieu contenant de l'ampicilline.

Connaissant la séquence de l'OAT on peut préparer une sonde synthétique qui une fois marquée radioactivement servira à hybrider les clones de la banque. Seuls les clones donnant une réponse positive seront étudiés plus avant.

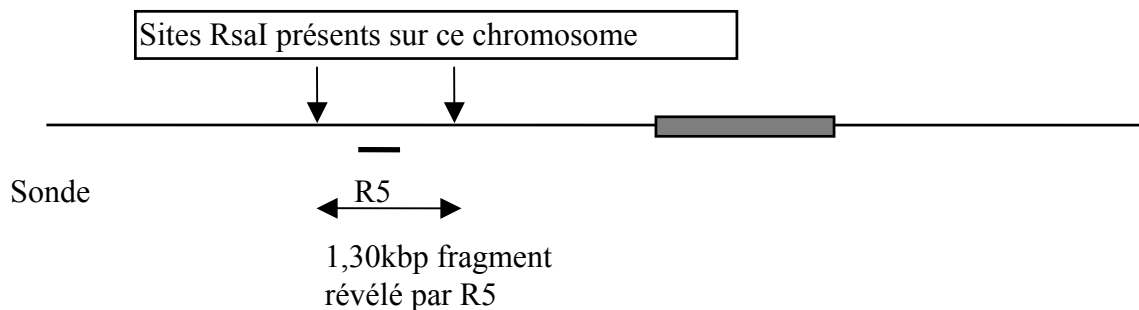
Carte chromosomique de la région portant le gène A et les 5 RFLP.



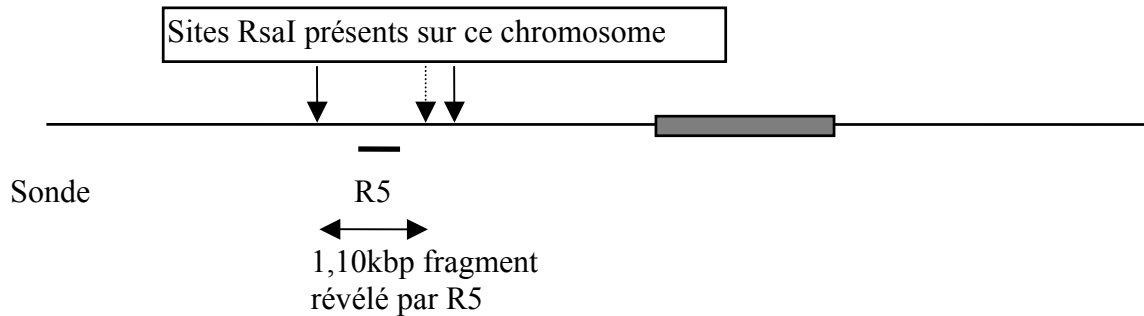
Détail dans le cas de la sonde R5 et du site polymorphe RsaI :



Cas d'un chromosome donnant l'haplotype B1 (chromosome de type 1)



Cas d'un chromosome donnant l'haplotype B2 (chromosome de type 2)



Le polymorphisme le plus intéressant au niveau de la population est celui que l'on trouvera le plus souvent. Pour que l'hétérozygote soit le plus fréquent possible il faut que les fréquences des deux haplotypes soient les plus proches possible de 0,5. C'est donc le polymorphisme associé à l'extrémité 3' de l'ADNc avec le site MspI qui est le plus intéressant au niveau de la population générale (dans laquelle ont été calculées les fréquences alléliques). Néanmoins pour un couple donné, si celui-ci n'est pas utilisable, le fait d'en avoir plusieurs donne d'autant plus de chances de trouver un marqueur polymorphe pour lequel les parents soient hétérozygotes et dans la bonne phase pour pouvoir faire un diagnostic.

Analyse de la figure 2.

Les individus I-2, III-7, III-9 ,III-12 ne présentent que l'haplotype B2. Ils sont donc homozygotes pour le chromosome de type 2.

et on écrit leur génotype : $\frac{r+, a+}{r+, a+}$

Les individus III-2, III-5 , III-10 ne présentent que l'haplotype B1. Ils sont donc homozygotes pour le chromosome de type 1

et on écrit leur génotype $\frac{r-, a+}{r-, a+}$

Les individus I-1, II-2, III-1, III-3, III-4, III-6, III-8, III-11 présentent à la fois l'haplotype B1 et l'haplotype B2. Ils sont donc hétérozygotes pour le chromosome et portent un exemplaire de type 1 et un exemplaire de type 2 :

et on écrit leur génotype $\frac{r-, a+}{r+, a+}$

Puisque parmi les descendants III il y a à la fois des homozygotes B1 et des homozygotes B2, c'est que chaque parent peut fournir les deux types. Donc tous deux sont hétérozygotes B1 et B2.

4°) Chacun des deux parents doit être porteur d'un allèle muté et d'un allèle sauvage pour le gène de l'OAT (sauf le cas peu probable où il y aurait eu une mutation, dans ce cas un seul des 2 parents serait hétérozygote porteur).

Chacun produit autant de gamètes mutants que de gamètes sauvages et la probabilité pour un de leur descendant d'être atteint est de 0,25.

Polymorphisme Sau96I-R5 $\begin{cases} A1 & 0,66 \text{ pas de site facultatif : } s- \\ A2 & 0,56 \text{ site facultatif présent : } s+ \end{cases}$

II-1 est homozygote pour l'allèle a- et hétérozygote pour le site polymorphe Sau96I puisqu'il présente les bandes A1 et A2.

Son génotype s'écrit:

$$\frac{s-, a-}{s+, a-}$$

La mère I-2 est homozygote A1 son génotype est donc:

$$\frac{s-, a-}{s-, a+}$$

Le père a donc sur un chromosome : A2 et a - qu'il a légué à II-1. Son génotype s'écrit :

$$\frac{s-, a+}{s+, a-}$$

En supposant qu'il n'y a pas de crossing-over entre le site polymorphe et le gène OAT on peut écrire le tableau de gamètes suivant :

gamètes femelles	gamètes mâles	
	s+, a-	s-, a+
s-, a-	$\frac{s-, a-}{s+, a-}$ [A1, A2, AG]	$\frac{s-, a-}{s-, a+}$ [A1, sain]
s-, a+	$\frac{s-, a+}{s+, a-}$ [A1, A2, sain]	$\frac{s-, a+}{s-, a+}$ [A1, sain]

D'après ce tableau on voit que II-2 est soit homozygote atteint soit hétérozygote sain alors que II-3 est soit hétérozygote sain mais porteur soit homozygote sain pour le gène A.

Polymorphisme MspI-5'ADNc $\begin{cases} E1 & 4,10 \text{ pas de site facultatif : } m- \\ E2 & 3,40 \text{ site facultatif présent : } m+ \end{cases}$

II-1 est homozygote pour l'allèle a- et hétérozygote pour le site polymorphe MspI puisqu'il présente les bandes E1 et E2.

Son génotype s'écrit:

$$\frac{m-, a-}{m+, a-}$$

Le père I-1 est homozygote E1, son génotype est donc:

$$\frac{m-, a-}{m-, a+}$$

La mère a donc sur un chromosome E2 et a- qu'elle a légué à II-1. Son génotype s'écrit :

$$\frac{m-, a+}{m+, a-}$$

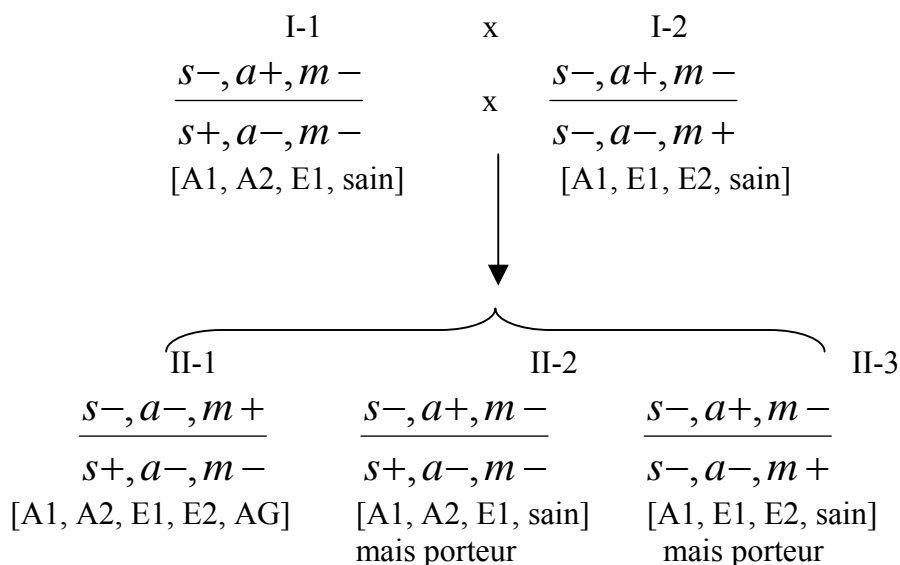
En supposant qu'il n'y a pas de crossing-over entre le site polymorphe et le gène OAT on peut écrire le tableau de gamètes suivant :

gamètes femelles	gamètes mâles	
	m+, a-	m-, a+
m-, a-	$\frac{m-, a-}{m+, a-}$ [E1, E2, AG]	$\frac{m-, a-}{m-, a+}$ [E1, sain]
m-, a+	$\frac{m-, a+}{m+, a-}$ [E1, E2, sain]	$\frac{m-, a+}{m-, a+}$ [E1, sain]

D'après ce tableau on voit que II-3 peut aussi bien être homozygote sain qu'hétérozygote atteint alors que II-2 est hétérozygote sain mais porteur.

Chaque polymorphisme a permis de marquer l'allèle muté d'un parent. Mais pour ce même polymorphisme l'autre parent est homozygote. Ce n'est ici que la combinaison des deux qui permet un diagnostic complet : les deux enfants II-2 et II-3 sont de phénotype [sain].

On peut établir la phase des 3 marqueurs. II-1 a reçu de son père : (s+ a- m-) et de sa mère (s- a- m+), ce qui permet d'écrire le pedigree avec les génotypes complets et de voir que les deux enfants II-2 et 3 sont porteurs



Cette méthode n'est valable que

1. si les sites polymorphes sont assez proches pour que le risque de CO entre le site et le gène soit infime
2. si les parents sont hétérozygotes et dans la même phase lorsque l'on n'a qu'un polymorphisme.