

2 - Polymorphismes génétiques

L. SCHIBLER, D. VAIMAN, E.P. CRIBIU

INRA Laboratoire de Génétique Biochimique et de Cytogénétique, 78352 Jouy-en-Josas cedex.

e-mail : schibler@biotech.jouy.inra.fr

Origine du polymorphisme de l'ADN

Résumé. Le développement des techniques de biologie moléculaire a permis la mise en évidence d'une très importante variabilité de l'ADN tant au niveau chromosomique (remaniements) qu'au niveau nucléotidique (mutations ponctuelles). Les différentes sources de variabilité sont décrites. Les mécanismes potentiellement impliqués dans la genèse de ces polymorphismes sont passés en revue.

Le polymorphisme génétique a été défini comme l'existence, dans une même population, de deux ou plus de deux formes discontinues, dont la plus rare ne peut être maintenue simplement par une mutation récurrente. Le développement des techniques d'analyse moléculaire a permis de relier cette diversité au polymorphisme biochimique dans un premier temps, puis au polymorphisme de l'ADN, c'est-à-dire aux différences existant à un même locus entre individus d'une même espèce. Généralement, seules les différences observées avec une fréquence supérieure à 1 % sont considérées comme des polymorphismes, les autres étant assimilées à des anomalies. Il est possible d'identifier des polymorphismes de l'ADN à différentes échelles, allant du niveau chromosomique au niveau nucléotidique, certaines caractéristiques du génome comme les séquences répétées favorisant l'apparition d'un polymorphisme. Le présent document a pour objectif de présenter ces différentes sources de variabilité, en précisant, quand cela est possible, les mécanismes potentiellement impliqués dans la genèse de ce polymorphisme.

1 / Types de polymorphisme de l'ADN

1.1 / Translocations, inversions, délétions, fusions et insertions

Les études cytogénétiques de cas pathologiques humains ou animaux ont révélé l'existence d'altérations plus ou moins importantes des molécules d'ADN, conduisant à la formation de remaniements chromosomiques (Popescu 1989). Dans certaines populations naturelles, certains remaniements chromosomiques se maintiennent à des fréquences

élevées et, par conséquent, peuvent être considérés comme des polymorphismes : les fusions centriques chez les gazelles (*Gazella subgutturosa*) et l'oryx d'Arabie (*Oryx leucorix*) ou les inversions paracentriques chez *Drosophila melanogaster*, les passereaux et les pigeons. Chez les animaux d'élevage certaines translocations Robertsoniennes (1/29 chez les bovins, translocations Massey 5/26, 8/11 et 7/25 dans certaines races ovines) peuvent être considérées comme un polymorphisme chromosomique. Des polymorphismes consécutifs à des inversions ont également été décrits, dans la région q13 du chromosome 2 humain par exemple (inversion péricentrique). Les délétions/insertions, qui sont généralement la cause d'anomalies phénotypiques graves, peuvent cependant être des polymorphismes, comme dans le cas de la variation de taille du chromosome Y chez l'Homme, les bovins ou le cheval. A l'échelle moléculaire, ces remaniements chromosomiques peuvent provoquer des fusions de gènes, en orientation directe ou inversée, souvent associées à des délétions allant de quelques nucléotides à une centaine de bases. Dans le cas de la leucémie myéloïde chronique par exemple, une translocation apparemment réciproque entre les chromosomes 9 et 22 entraîne la fusion du gène bcr (Hsap22) et du proto-oncogène c-abl (Hsap9), avec perte variable d'une partie de la région 5' de c-abl.

L'analyse moléculaire a mis en évidence l'existence d'événements de délétions, fusions, inversions ou insertions à l'échelle de quelques bases à quelques kilobases. Toute insertion ou délétion d'un nombre de bases non multiple de trois entraîne une modification du cadre de lecture (mutation frameshift lorsqu'elle se produit dans un exon traduit). Ce type de mutation fait généralement apparaître en aval un codon non sens prématuré conduisant à la formation d'une protéine tronquée. Une variation d'un nombre de bases multiple de trois maintient le cadre de lecture, mais entraîne la perte ou l'addition d'acides aminés, comme dans le cas de la délétion $\Delta F508$ de la mucoviscidose.

1.2 / Mutations ponctuelles

Les mutations ponctuelles, qui n'intéressent qu'une région très limitée du génome (quelques nucléotides, au maximum), sont la source prépondérante de polymorphisme. En effet, les estimations réalisées chez l'Homme suggèrent qu'un individu non consanguin présente une base à l'état hétérozygote tous les 200 à 1000 nucléotides.

La modification la plus simple est le remplacement d'une base par une autre. La substitution d'une base purique par une base purique (A / G) ou d'une base pyrimidique par une base pyrimidique (C / T) constitue une transition. Les remplacements d'une base purique par une base pyrimidique et réciproquement (A,G / C,T) sont des transversions. Il existe donc quatre transitions et huit transversions possibles. Dans les régions codantes, ces substitutions peuvent modifier la nature de l'acide aminé codé, en fonction principalement de leurs positions dans le codon. En effet, compte tenu de la dégénérescence du code génétique et en ne considérant que les triplets codant pour des acides aminés, 5 % des mutations de la première base, 0 % des mutations de la seconde et 72 % des mutations de la troisième base n'entraînent pas de changement d'acide aminé (mutations silencieuses). Pour la troisième base, près de 90 % des transitions sont silencieuses, contre 30 % des transversions. Par exemple, le glutamate est codé par les deux codons GAA et GAG. Une transition de la troisième base du codon (A→G) est silencieuse, alors que les transversions (A→C ou A→T) entraînent le remplacement du glutamate par l'aspartate. Dans l'hypothèse d'une pression de sélection visant à conserver la fonction du peptide codé, les transitions sont donc probablement mieux tolérées que les transversions. C'est sans doute pourquoi les différences observées entre les séquences des gènes orthologues de différentes espèces de mammifères sont plus souvent causées par des transitions. Des bases peuvent également être délétées ou ajoutées, conduisant à une modification du cadre de lecture.

1.3 / Séquences répétées en tandem (VNTR : Variable Number of Tandem Repeats)

Les séquences répétées en tandem (ou VNTR pour Variable Number of Tandem Repeats) sont de taille variable, constituées de répétitions en tandem d'un motif unitaire de taille également variable. Selon la taille du motif et de la répétition, on distingue les satellites (Csink et Henikoff 1998), les minisatellites

(Jarman et Wells 1989) et les microsatellites (Karlin et Burge 1995).

a / Satellites

La centrifugation en gradient de densité d'ADN génomique a mis en évidence des bandes mineures (ou satellites) d'une densité différente de la bande majeure représentant la fraction principale de l'ADN. Classiquement, trois bandes peuvent être identifiées, chacune comportant différentes familles de séquences répétées appelées séquences satellites. Le motif répété en tandem comporte de 2 à 100 nucléotides. Il existe dans les unités de répétition une organisation hiérarchisée de séquences où l'on peut reconnaître des motifs de plus en plus courts. Chez les primates par exemple, les satellites II contiennent une répétition en tandem du motif AATTC, alors que l'ADN satellite alpha, constituant principal de l'hétérochromatine centromérique, est caractérisé par une répétition d'une unité de 171 paires de bases. La variation du nombre de répétitions des satellites entraîne un polymorphisme de quantité d'hétérochromatine constitutive en région péri-centromérique, visualisé par la technique de bandes C chez certaines espèces comme les bovins chez lesquels huit satellites ont été décrits.

Ces séquences ne sont généralement pas transcrites et sont considérées comme de l'ADN 'poubelle' ou 'camelote' (junk DNA). Le tableau 1 résume les principales caractéristiques des satellites.

b / Minisatellites

Les minisatellites hypervariables sont extrêmement polymorphes, tant du point de vue du nombre de répétitions que de la taille des unités répétées. Les unités partagent cependant un motif central (ou core) GGGCAGGANG. Ce motif présente une homologie avec la séquence *chi*, séquence signal de recombinaison chez *E. Coli* et serait de ce fait susceptible de favoriser les recombinaisons et donc le polymorphisme des minisatellites. Le tableau 2 en présente les principales caractéristiques.

L'ADN télomérique constitue une autre famille majeure de minisatellites. Le motif hexanucléotidique TTAGGG est répété sur environ 10 à 15 kb chez l'Homme et 30 kb chez la souris. De même que pour les satellites, le rôle biologique des minisatellites n'est pas élucidé, sauf dans le cas de l'ADN télomérique dont une fonction supposée est de protéger l'extrémité des chromosomes de la

Tableau 1. Caractéristiques des satellites.

Classe	Satellites II et III	Satellite I (AT riche)	ADN α
Taille de l'unité de base	5 pb	25-48 pb	171 pb
Nombre de répétitions		Jusqu'à plus de 1000	
Taille de la séquence		De 100 kb à plusieurs Mb	
Répartition chromosomique	La plupart des chromosomes	Hétérochromatine de la plupart des chromosomes	Hétérochromatine centromérique de tous les chromosomes

Tableau 2. Caractéristiques des minisatellites.

Classe	Famille hypervariable	Famille télomérique
Taille de l'unité de base	9-24 pb	6 pb
Nombre de répétitions	De 20 à 50	
Taille de la séquence	De 0,1 à 20 kb	
Répartition chromosomique	Tous les chromosomes, souvent en régions télomériques	Tous les télomères

dégradation et d'en permettre la réplication.

c / *Microsatellites* (ou *STR : Short Tandem Repeats*)

Les microsatellites comportent des séries de répétitions d'un motif de base court (1 à 4 pb) et sont répartis sur l'ensemble du génome. Le nombre de répétitions ne dépasse pas 25 et la taille de la séquence n'excède généralement pas 150 paires de bases. Le tableau 3 présente la fréquence des principales séries chez les mammifères.

Bien que les répétitions alternées purine-pyrimidine (comme dans le cas des répétitions CA) soient capables de modifier la conformation de l'ADN *in vitro* (ADN-Z), la fonction *in vivo* des microsatellites reste inconnue.

Les répétitions de trinuécléotides en tandem constituent un cas particulier, dans la mesure où un comportement anormal de microsatellites (CAG)_n ou (CCG)_n a pu être montré, en liaison avec certaines maladies humaines. En effet, au dessus d'une certaine longueur critique, les répétitions sont instables au cours de la mitose et de la méiose et ne sont pratiquement jamais transmises sans modifications à la descendance : des extensions ou des délétions peuvent apparaître, en fonction de la longueur de la répétition et du sexe du parent. Chez les patients atteints de la maladie de l'X fragile, le motif CGG, normalement répété de 6 à 54 fois, est présent de 200 à plus de 1000 fois. Par un mécanisme inconnu, ces extensions affectent la méthylation de l'ADN et la structure chromatidienne, conduisant à l'inactivation

des gènes adjacents. D'autres sites fragiles n'entraînant aucun effet phénotypique ont également été observés sur d'autres chromosomes humains, ovins et bovins (constrictions achromatiques).

2 / Mécanismes mis en œuvre dans la génération du polymorphisme

2.1 / Des erreurs du système de réparation expliquent les mutations ponctuelles

Le taux de mutation basal de l'ADN reflète les erreurs inévitables lors de la réplication ou de la réparation de l'ADN. En effet, la fidélité des ADN polymérasés n'est pas absolue et des erreurs de réplication surviennent avec une fréquence d'environ 10⁻⁹ par nucléotide incorporé, soit environ une mutation par division cellulaire sur l'ensemble du génome. Au cours d'une vie humaine, l'organisme effectuant de l'ordre de 10¹⁷ divisions cellulaires, le génome subit donc environ 10¹⁷ mutations somatiques. Pour la lignée germinale, on estime à 24 et (30+23n+5 où n = âge en années-15) le nombre de divisions nécessaires pour produire respectivement un ovule et un spermatozoïde. Chaque ovule est donc porteur d'environ 24 mutations ponctuelles et chaque spermatozoïde d'un homme de 25 ans comporte près de 265 mutations liées aux erreurs de réplication (Strachan et Read

Tableau 3. Fréquences des principaux types de microsatellites chez les mammifères.

Série (taille du motif)	Séquence du motif	Fréquence	Remarque
Mononucléotide	(A) _n	Un tous les 5-10 kb	0,3 % du génome
Dinucléotide	(CA) _n	Un tous les 25-100 kb	0,5% du génome
	(TC) _n	Un tous les 50 kb	0,2 % du génome
Trinucléotide	(TTA) _n (AGC) _n	Un tous les 300-500 kb	
Tétranucléotide	(AATC) _n , (AATG) _n , (ACAG) _n , (AAAT) _n , (AAAG) _n		Dérive des séquences poly(A) des familles Alu

1999).

L'ADN est, de plus, soumis à des modifications chimiques spontanées, principalement des réactions d'hydrolyse :

- des dépurinations spontanées, dues à la labilité de la liaison N-glycosidique reliant les bases puriques au squelette de polydésoxyribosephosphate. Cette liaison peut être hydrolysée et la purine (A ou G) remplacée par un groupement hydroxyle sur le carbone 1'. Il s'agit d'événements très fréquents, 5000 bases puriques étant perdues chaque jour par chaque cellule humaine ;

- des transitions tautomériques ($-NH_2 \rightleftharpoons =NH$ ou $C=O \rightleftharpoons =C-OH$), en particulier des désaminations avec transition de l'adénine en hypoxanthine, de la guanine en xanthine, de la cytosine en uracile. Ce dernier cas survient avec une fréquence d'environ 100 bases par jour et par cellule chez l'Homme ;

- des transversions qui consistent en une substitution d'une pyrimidine (C ou T) par une purine (A ou G) ou inversement.

L'ADN génomique peut également être altéré par exposition à de multiples mutagènes, endogènes (radicaux libres) ou exogènes (rayonnement UV) qui peuvent induire la formation de liaisons covalentes entre deux pyrimidines voisines (CC, CT ou TT) sur le même brin d'ADN.

La plupart de ces modifications (cassure d'un brin, création de liaisons covalentes entre bases complémentaires, modification chimique d'une base, excision d'une base...) sont généralement reconnues et corrigées par les systèmes enzymatiques de réparation des cellules, à l'exception des cas de désamination d'une cytosine méthylée. En effet, dans ce cas, le produit formé (5-méthyle uracile) correspond à la thymine, non reconnue comme élément étranger à l'ADN et de ce fait non réparée. La transition accidentelle C→T est donc généralement fixée. Les régions riches en cytosines méthylées, en particulier les doublets CG, sont donc des sites vulnérables et constituent des points chauds (Hot-spots) de mutation. Si la réparation est imparfaite, la mutation accidentelle peut être fixée. Les mutations qui surviennent dans la lignée germinale peu-

vent être transmises aux générations ultérieures, générant potentiellement un polymorphisme, tandis que les mutations somatiques ne concernent que l'individu chez lequel elles surviennent. Pour les cellules somatiques, si la mutation est sans effet sur le fonctionnement de la cellule, le clone cellulaire mutant se dilue dans la population et la mutation reste inaperçue. Si la mutation est délétère, le clone s'éteint spontanément. Si au contraire, la mutation procure un avantage sélectif, le clone muté peut proliférer, comme dans le cas des cancers par exemple.

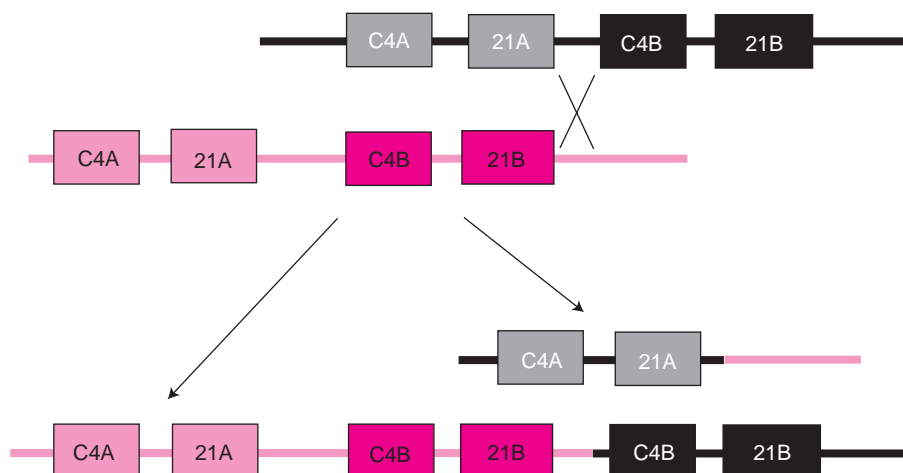
2.2 / Les recombinaisons inégales expliquent les délétions et duplications

Des délétions et duplications peuvent résulter de recombinaisons 'illégitimes' entre séquences très semblables mais non homologues. Ces séquences favorisent l'apparition d'erreurs d'alignement (mésappariement), conduisant à des crossing-over inégaux. Ce mécanisme aboutit à l'élimination d'une région sur l'un des chromosomes et à la duplication sur l'autre chromosome, comme le montre la figure 1 (Cooper *et al* 1995).

Ces recombinaisons sont favorisées par l'existence de séquences semblables contiguës, codantes ou non. Les séquences répétées dispersées comme les séquences Alu chez l'Homme, compte tenu de leur homologie et de leur fréquence (1 tous les 4 kb chez l'Homme), sont susceptibles de jouer un rôle important dans ce phénomène. De même, ces erreurs d'appariement sont fréquentes dans les régions de répétitions en tandem, compte tenu de la forte homologie de séquences entre les répétitions. Tout crossing-over intervenant lors de ces mésappariements produit une augmentation du nombre de répétitions sur l'une des chromatides et une délétion d'un nombre équivalent de répétitions sur l'autre.

D'un point de vue évolutif, la recombinaison inégale serait à l'origine de la création de familles de

Figure 1. Recombinaison inégale et délétion/fusion. La contiguïté, sur un même chromosome, de séquences identiques ou semblables favorise les crossing-over inégaux lors de la méiose. Dans le cas des gènes dupliqués en tandem de la stéroïde 21-hydroxylase (21A et 21B) et du complément (C4A et C4B) dans le domaine des classes III du complexe HLA, 21A et 21B ne diffèrent que par 88 nucléotides sur les 3400 que comporte la séquence. La recombinaison inégale conduit à la formation de gamètes porteurs soit d'une délétion, soit d'une duplication.



gènes à partir d'un gène ancestral.

2.3 / Conversion génique

La conversion génique est un mécanisme de recombinaison permettant un transfert d'information génétique qui, par son résultat, évoque un double crossing-over (Jeffreys *et al* 1994). La conversion, au sens moléculaire, produit le remplacement d'une séquence d'ADN par une autre, apparentée mais non allélique (conversion interlocus) ou allélique (conversion interallélique). En particulier, ce remplacement peut se produire entre séquences répétées en tandem ou dispersées. De plus, il peut être asymétrique et polarisé, certaines séquences se substituant plus fréquemment à une autre que la réciproque. A l'échelle de l'individu, ce mécanisme s'oppose au polymorphisme dans la mesure où il tend à homogénéiser les séquences. A l'échelle de la population en revanche, il est en mesure d'accélérer la fixation de nouveaux allèles (passage de l'état hétérozygote à l'état homozygote), générés par mutation ponctuelle par exemple, et contribue ainsi à la création d'un polymorphisme. Cette modification peut concerner quelques bases, mais peut aussi s'étendre sur quelques kb, sans modification de la quantité totale d'ADN. Les conversions intrachromosomiques sont environ dix fois plus fréquentes que les conversions interchromosomiques entre homologues.

Le mécanisme de conversion suppose la formation d'un hétéroduplex intermédiaire permettant la correction, éventuellement biaisée dans une direction, à partir d'un brin donneur, comme le schématise la figure 2.

2.4 / Insertions de séquences mobiles (transposons) ou virales

Certaines séquences d'ADN répétitif dispersé possèdent une structure de transposons, éléments d'ADN instables, capables de migrer vers différentes régions du génome par rétrotransposition (leur transcrits ARN peuvent être convertis dans la cellule en ADN complémentaire réinsérable au sein de l'ADN génomique). Les deux principaux groupes de telles séquences présents chez les mammifères

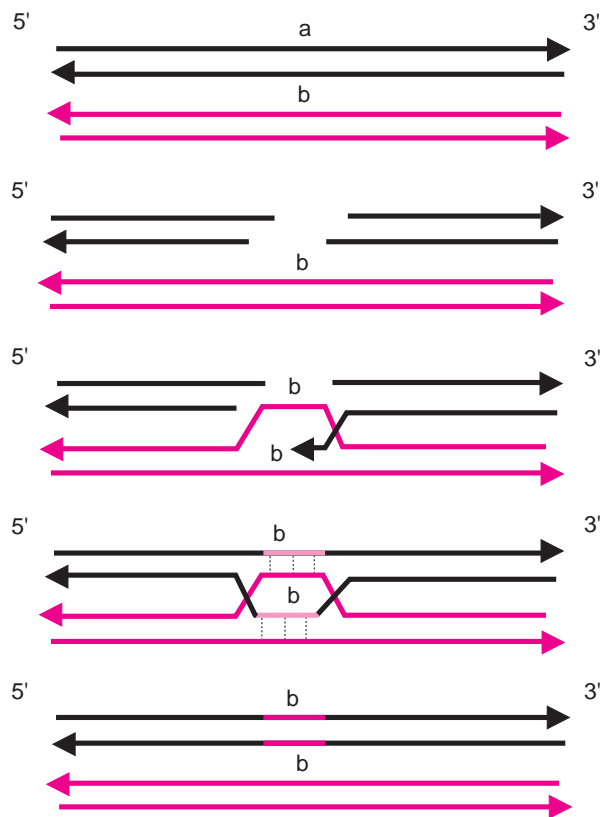


Figure 2. Mécanisme de la conversion génique. Exemple d'une conversion entre deux chromosomes homologues.

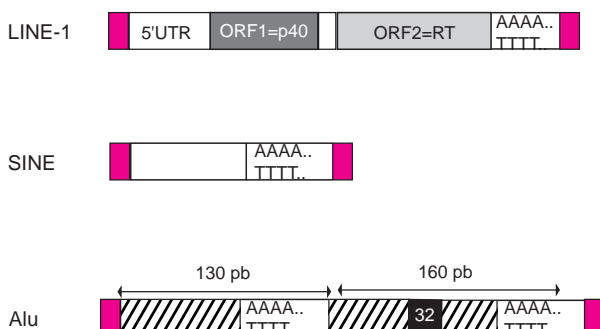
- 1- Les deux brins de chaque duplex sont orientés 5'→3' par les flèches, les régions homologues étant appariées.
- 2- Une brèche est ouverte et élargie par des endonucléases sur les brins accepteurs.
- 3- L'un des brins donneurs envahit le second duplex accepteur en formant une boucle.
- 4- Les brèches sont réparées en 3' par la séquence complémentaire des brins donneurs et ligation des extrémités.
- 5- Les séquences des brins du duplex accepteur ont été modifiées au profit des séquences donneuses.

Figure 3. Structure des LINES et SINES.

L'élément consensus des LINES mesure 6,1 kb et possède deux cadres ouverts de lecture (ORF). ORF1 code pour une protéine de fonction inconnue (p40) et ORF2 pour une transcriptase inverse (RT). Les éléments LINES sont encadrés par de courtes répétitions dupliquées (en rouge).

Les SINES sont également flanqués de courtes répétitions directes (6 à 18 pb), mais ne possèdent pas de cadre ouvert de lecture. Leur transposition résulte certainement de l'action en trans de transcriptases inverses, comme celles codées par les LINES.

Les répétitions Alu constituent un cas particulier de SINE, particulièrement abondant dans le génome des primates. La séquence Alu typique est un dimère répété en tandem d'une séquence de 120 pb (zones hachurées) suivie d'une zone riche en A/T. Les répétitions ne sont cependant pas symétriques car l'une des unités comporte une insertion de 32 pb (en noir).



constituent les LINES et les SINES dont la structure est présentée à la figure 3 (Smit 1996).

Les LINES (Long INterspersed Elements) correspondent à des séquences de 6 à 7 kb, alors que les SINES (Short INterspersed Elements) mesurent environ 300 pb. Les LINES L1 sont les plus courants chez l'Homme et sont répétés près de 50 000 fois, représentant ainsi près de 10 % du génome. Les SINES sont également très fréquents, comptant environ 500 000 copies par génome et correspondant à près de 5 % de l'ADN génomique total.

Bien que les phénomènes de transposition soient actuellement peu connus chez les eucaryotes supérieurs, plusieurs pathologies ont été associées à des événements de mutagenèse insertionnelle soit de LINE (dans le cas de l'hémophilie A, par insertion dans l'exon 14 du facteur VIII), soit de séquence Alu (dans un cas de neurofibromatose, par insertion dans un intron du gène NF1). De façon plus globale, il est probable que ces séquences aient un rôle

important dans la plasticité du génome au cours de l'évolution et du développement.

2.5 / L'excision de boucles chromatiniennes est responsable de délétions

Le mécanisme d'excision de boucles chromatiniennes, indépendant de toute homologie de séquence, a été mis en évidence dans le cas des gènes d' α et β globine. Il consisterait, lors de la réplication, en l'excision des boucles formées entre deux points d'ancrage de la matrice chromatidienne distants de 30 à 130 kb.

2.6 / Le glissement intra-chromatidien (slippage) modifie le nombre de répétitions des VNTR

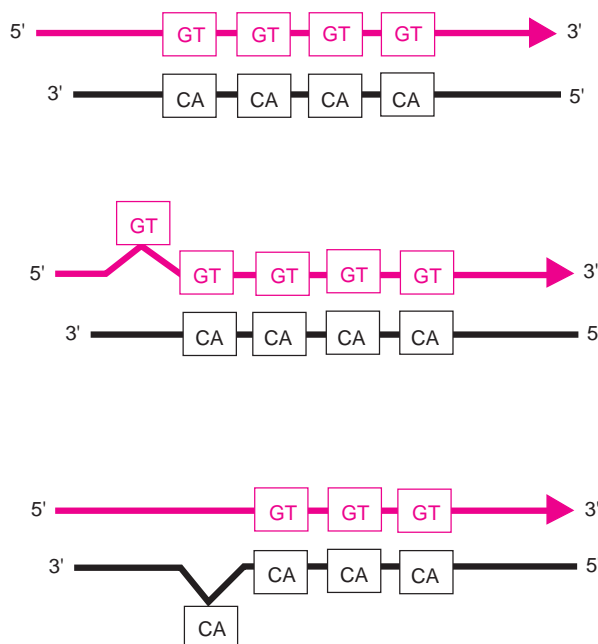
L'analyse d'haplotypes obtenus à l'aide de microsatellites a montré l'apparition d'allèles nouveaux de microsatellites (CA)_n (le taux d'évolution des microsatellites a été estimé à 10⁻³-10⁻⁵), sans modification des loci flanquants, suggérant que ces nouveaux allèles ne résultent pas d'un crossing-over inégal. Ces nouveaux allèles ne diffèrent des allèles d'origine que par une seule unité répétée, l'augmentation de taille semble résulter d'un glissement de la polymérase (Freimer et Slatkin 1996), consécutif à un mésappariement (figure 4). Ce phénomène serait ainsi la reproduction *in vivo* du phénomène observé *in vitro* lors de la PCR, au cours duquel la Taq polymérase s'avère incapable de reproduire fidèlement le nombre de répétitions d'origine et génère par glissement des bandes multiples, au lieu de la bande unique attendue.

Figure 4. Glissement intra-chromatidien.

1 - Réplication normale : un brin est néosynthétisé dans le sens 5'→3' (brin rouge) à partir d'un brin matrice (brin noir).

2 - Glissement en arrière : si le glissement du brin en croissance se fait dans le sens 3'→5', il en résulte une boucle sur ce brin et une addition d'un TG supplémentaire.

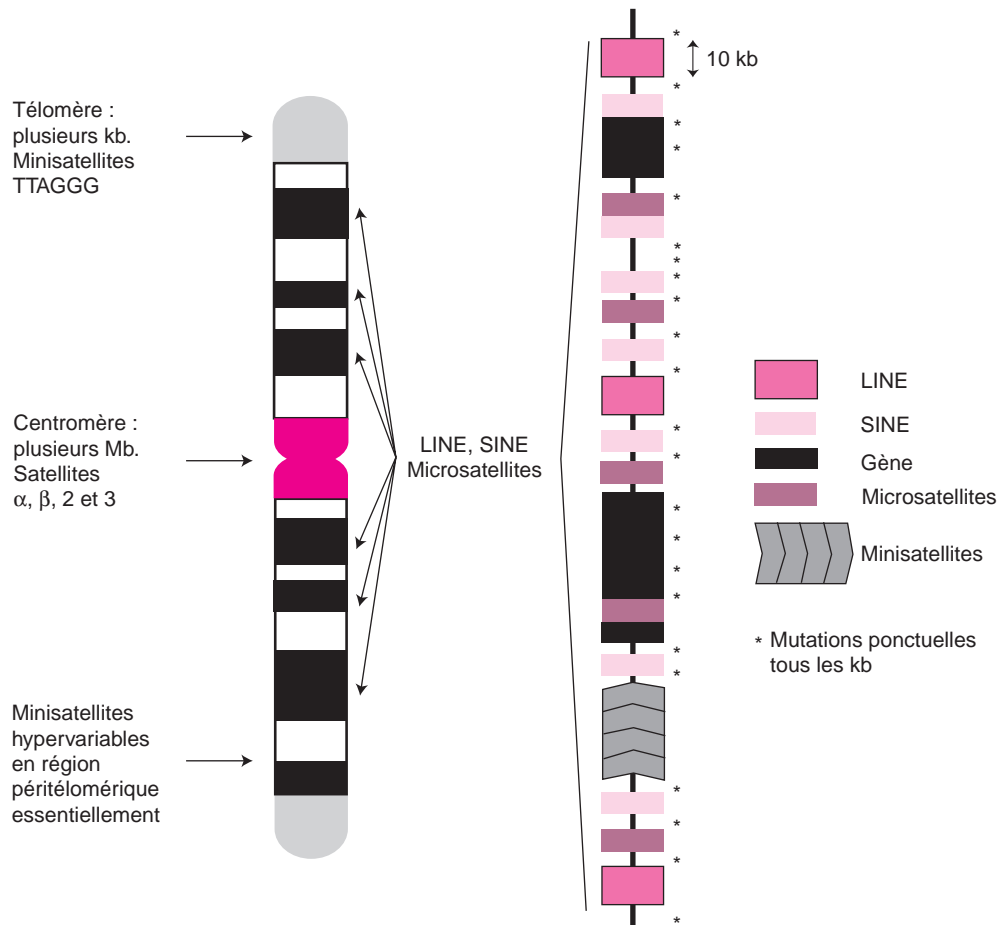
3 - Glissement en avant : si le glissement du brin en croissance se fait dans le sens 5'→3', il en résulte une boucle sur le brin matrice et une délétion d'un TG.



Conclusion

L'ADN recèle donc un important polymorphisme résultant soit d'accidents purement aléatoires, soit de mécanismes favorisés par certaines structures du génome (VNTR). La figure 5 résume ces informations en termes de localisation, de types de séquences et de polymorphismes ainsi que de fréquence.

Tous ces polymorphismes, moteur de l'évolution, peuvent être utilisés, soit dans des programmes de cartographie génétique et de clonage positionnel, soit dans des approches de diagnostic ou d'identification des individus par exemple. Cependant, en fonction de leurs caractéristiques ou des techniques permettant de les révéler, certains types sont plus adaptés à l'une ou l'autre des utilisations. C'est ainsi que les microsatellites ont supplanté les autres marqueurs dans les approches de cartographie génétique, pour des raisons à la fois méthodologiques

Figure 5. Structure du génome et polymorphisme.


Références

(automatisation) et génétiques (taux de polymorphisme, répartition). En revanche, les minisatellites restent les marqueurs de choix pour l'identification des individus (médecine légale).

Cooper D.N., Krawczak M., Antonorakis S.E., 1995. The nature and mechanisms of human gene mutation. In : C. Scriver, AL Beaudet, WS Sly, D. Valle (eds), *Metabolic and molecular bases of inherited disease*, 7th ed., 259-261. McGraw-Hill, New-York.

Csink A.K., Henikoff S., 1998. Something from nothing: the evolution and utility of satellite repeats. *Trends Genetics*, 14, 200-204.

Freimer N.B., Slatkin M., 1996. Microsatellites: evolution and mutational processes. *Ciba Found Symposium*, 197, 51-67.

Jarman A.P., Wells R.A., 1989. Hypervariable minisatellites: recombinators or innocent bystanders? *Trends Genetics*, 5, 367-371.

Jeffreys A., Tamaki K., MacLeod A., Monckton D.G., Neil D.L., Armour J.A.L., 1994. Complex gene conversion events in germline mutation at human microsatellites. *Nature Genetics*, 6, 136-145.

Karlin S., Burge C., 1995. Dinucleotide relative abundance extremes: a genomic signature. *Trends Genetics*, 11, 283-290.

Popescu C.P., 1989. *Cytogénétique des mammifères d'élevage*. INRA, Paris, 114 p.

Smit A.F., 1996. The origin of interspersed repeats in the human