

2m11.2966.9

Université de Montréal

Analyse de l'évolution des éléments répétitifs de type LINE-1
chez l'humain

par
Jacinte Gauthier

Programme de biologie moléculaire
Faculté des études supérieures

Mémoire présenté à la faculté des études supérieures
en vue de l'obtention du grade de
Maître ès science (M.Sc.)
en biologie moléculaire

Novembre, 2001

©Jacinte Gauthier, 2001



QH
506
U54
2002
v.011

11/10

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé:

Analyse de l'évolution des éléments répétitifs de types LINE-1
chez l'humain

présenté par:

Jacinte Gauthier

a été évalué par un jury composé des personnes suivantes:

Dr Damian Labuda

Dr Edward Bradley

Dr Pierre Chartrand

RÉSUMÉ

Le génome humain serait composé de 45% de séquences répétitives dérivant d'éléments transposables. Les LINEs sont les éléments transposables les plus anciens, s'étant établis avec succès chez les mammifères. La famille de LINE 1 (L1) est la plus abondante et on retrouve 8×10^5 copies de LINEs de cette famille dispersées à travers tout le génome. Leur mécanisme de dispersion, soit la rétrotransposition, leur a permis de coloniser les génomes de mammifère avec succès. Cependant, la majorité des copies que l'on retrouve sont non fonctionnelles. Ces copies peuvent diverger avec le temps en accumulant des mutations aléatoires. Toutefois, il existe une plus grande similarité entre ces séquences à l'intérieur d'une même lignée qu'entre différentes lignées. Les éléments L1 évoluent donc de façon concertée. Ceci suggère qu'il puisse exister des interactions entre ces éléments permettant l'homogénéisation des séquences. L'un des mécanisme proposé est la conversion génique. Le but de notre travail était de déterminer si nous pouvions retrouver des évidences de conversion génique entre des éléments L1 dans le génome humain. Pour réaliser ce projet, nous avons sélectionné trois éléments L1 rapportés dans la littérature comme étant impliqués dans événements de recombinaisons génétiques et nous avons analysé leur séquence chez 10 individus. D'une part nous avons observé plusieurs formes d'haplotypes d'éléments L1 dues à l'accumulation de mutations ponctuelles au cours de l'évolution, ce qui pourrait être attribuable à l'instabilité de la région génomique étudiée soit la région fragile FRA3B du gène FHIT rapportée dans la littérature comme étant impliquée dans des événements de recombinaisons homologues. D'autre part, nous avons identifié une forme d'haplotype qui présentait, selon nos analyses, une zone de conversion génique suggérant l'implication de ce mécanisme dans l'évolution des éléments L1.

MOTS CLÉS

Conversion génique, éléments transposables, évolution concertée, recombinaison
homologue, génome humain.

ABSTRACT

The human genome is constituted for more than 45% by repetitive sequences derived from transposable elements. Amongst these, the oldest group of repetitive elements are the LINEs that are present in all mammals. The family of LINE-1 (L1) elements is the most abundant with over 8×10^5 copies that are dispersed all through the genome. Thus, the dispersion mode by retrotransposition of the L1 elements has permitted them to efficiently colonize the genome of mammals. However, the vast majority of L1 elements present in the genome are truncated and rearranged and thus believed to be non-functional. It would be expected that such genetic material would accumulate divergences over time in a random fashion. This has been observed to be the case. However L1 elements have been found to be more homologous one to the other within a specie than between species, which is indicative of a concerted evolution process. This is usually mediated by a mechanism of recombination designated as gene conversion that permits the homogenisation of diverging sequences. The objective of our research has been to determine if we could find evidences of gene conversion between L1 elements in the human genome. We have chosen three L1 elements that are located close one to the other and have determined their allelic variation in a population of 10 individuals. Our results have been that there are indeed numerous allelic variations for given L1 elements, thus indicating the accumulation of divergences over time. The extent of variations has been found to be up to 10 times higher than the average degree of allelic variation in the human genome. One of the allelic variations of one of the L1 elements is constituted of many modifications over a very short distance. Furthermore all these modifications could have originated from one of the other L1 element near by. This strongly suggests a gene conversion process between these two L1 elements and thus substantiates the implication of this mechanism in the evolution of L1 elements.

KEY WORDS

Gene conversion, transposable elements, concerted evolution, homologous recombination, human genome.

TABLE DES MATIÈRES

RÉSUMÉ ET MOTS CLÉS (FRANÇAIS)	iii
RÉSUMÉ ET MOTS CLÉS (ANGLAIS)	v
TABLE DES MATIÈRES	vii
LISTE DES FIGURES	ix
LISTE DES TABLEAUX	xi
LISTE DES ABRÉVIATIONS	xii
REMERCIEMENTS	xiv
I. INTRODUCTION	1
I.1 Les séquences répétitives de type LINE	1
I.2 Les LINEs et l'évolution dynamique du génome	3
I.3 Les LINEs et l'organisation structurale du génome	4
I.4 La recombinaison homologue	7
I.5 Les séquences répétitives et les réarrangements chromosomiques	9
I.6 L'évolution concertée des séquences répétitives	13
I.7 But du travail	16
I.8 Approche expérimentale	17
II. MATÉRIEL ET MÉTHODES	19
II.1 Sélection des éléments L1 et de la séquence intronique	19
II.2 Sélection de la région à cloner et à séquencer	19
II.3 Amplification des éléments L1 et de la séquence intronique	20
II.4 Clonage des fragments L1 A, B et C et de la séquence intronique	21
II.5 Extraction de l'ADN plasmidique	23
II.6 Digestions caractéristiques des clones obtenus	24
II.7 Séquençage de la région de 300 pb	24
III. RÉSULTATS	34
III.1 Résultats de l'approche expérimentale utilisée	34
III.2 Analyse de l'évolution d'un élément LINE-1 donné chez l'humain	61
III.3 Processus d'évolution d'un élément LINE-1	71
III.4 Analyse de la variabilité de la région étudiée.....	74
III.5 Interactions des éléments LINEs dans un contexte d'instabilité génomique	78
IV. DISCUSSION	81

V. BIBLIOGRAPHIE97

LISTE DES FIGURES

Figure 1 :	Représentation schématique d'un élément LINE-1 actif en rétrotransposition.	2
Figure 2 :	Implication des séquences répétitives dans l'organisation structurale de la chromatine dans le noyau.	6
Figure 3 :	Représentation schématique des principaux produits d'un événement de recombinaison entre deux séquences homologues non alléliques. ...	8
Figure 4 :	Représentation schématique de la recombinaison homologue avec crossing-over intrachromosomique et interchromosomique.	10
Figure 5 :	Représentation schématique de l'homogénéisation des membres d'une famille de séquences répétitives.	14
Figure 6 :	Schéma de la procédure de clonage du L1 A.	26
Figure 7 :	Schéma de la procédure de clonage du L1 B.	28
Figure 8 :	Schéma de la procédure de clonage du L1 C.	30
Figure 9 :	Schéma de la procédure de clonage de la séquence intronique.	32
Figure 10 :	Schéma de l'organisation des éléments L1 A, B et C sélectionnés dans la région fragile FRA3B du gène FHIT situé sur le chromosome 3 (3p14.2).	35
Figure 11 :	Alignement de la région de 300 pb sélectionnée des trois séquences L1 A, B et C retrouvées dans Gene Bank.	38
Figure 12 :	Illustration de la première méthode de clonage testée pour faire l'analyse des séquences L1.	41
Figure 13 :	Test de ligation pour vérifier l'étape de ligation de la première méthode de clonage.	44
Figure 14 :	Test pour vérifier l'étape de digestion aux extrémités de la première méthode de clonage.	46
Figure 15 :	Schéma des résultats attendus et présentation des résultats obtenus aux différentes étapes de clonage du L1 A chez l'individu 244.	50

- Figure 16 :** Schéma des résultats attendus et présentation des résultats obtenus aux différentes étapes de clonage du L1 B chez l'individu 244.53
- Figure 17 :** Schéma des résultats attendus et présentation des résultats obtenus aux différentes étapes de clonage du L1 C chez l'individu 244.56
- Figure 18 :** Lecture d'une séquence de la région de 300 pb d'un L1.59
- Figure 19 :** Présentation de la zone de conversion génique du L1 C allèle 4.69

LISTE DES TABLEAUX

Tableau I :	Analyse comparative des différents types d'haplotypes du L1 A.	62
Tableau II :	Distribution des différents types d'haplotypes du L1 A retrouvés pour chacun des 10 individus.	62
Tableau III :	Analyse comparative des différents types d'haplotypes du L1 B.	64
Tableau IV :	Distribution des différents types d'haplotypes du L1 B retrouvés pour chacun des 10 individus.	64
Tableau V :	Analyse comparative des différents types d'haplotypes du L1 C.	66
Tableau VI :	Distribution des différents types d'haplotypes du L1 C retrouvés pour chacun des 10 individus.	66
Tableau VII :	Analyse comparative des différents types d'haplotypes de la séquence intronique.	72
Tableau VIII :	Distribution des différents types d'haplotypes de la séquence intronique retrouvés pour chacun des 10 individus.	72
Tableau IX :	Présentation des différents haplotypes combinés retrouvés chez les dix individus.	76
Tableau X :	Variations de bases observées en comparant la séquence totale de 1200 pb de deux individus à la fois.	77

LISTE DES ABRÉVIATIONS

%	pourcentage
°C	degré(s) Celsius
µg	microgramme(s)
µL	microlitre(s)
ADN	acide désoxyribonucléique
ARN	acide ribonucléique
ATP	adénosine triphosphate
BSA	bovine serum albumin
FHIT	fragile histidine triad
kb	kilopaire(s) de base
LINE	long interspersed nuclear element
L1	LINE 1
min	minute(s)
M	molaire(s)
mL	millilitre(s)
mM	millimolaire(s)
ng	nanogramme(s)
PCR	polymerase chain reaction
pb	paire(s) de base
rpm	rotation par minute
sec	seconde(s)
SINE	short interspersed nuclear element
TE	tris-EDTA (10:1)
U	unité(s)

À ma famille,
mes parents,
mes amis
et à Sylvain

REMERCIEMENTS

Ce mémoire est le résultat de deux belles années de travail. J'ai eu beaucoup de plaisir à travailler pendant ces deux années grâce, en partie, au Dr Pierre Chartrand et à son équipe de laboratoire.

Je voudrais tout particulièrement remercier Dr. Pierre Chartrand pour m'avoir accueillie dans son laboratoire. Sa supervision, ses encouragements, ses précieux conseils ainsi que de nombreuses discussions enrichissantes m'ont permis de développer ma passion pour la recherche.

Je tiens aussi à remercier les membres du laboratoire, soit Nadine, Roxane, Hélène, Catherine, Hugo et Kevin pour leur support, leurs conseils, leur amitié et surtout pour tous les bons moments que nous avons passés ensemble. Merci aux gens de l'Institut du cancer pour m'avoir permis d'évoluer en recherche dans un milieu chaleureux.

Merci à tous et à chacun et au plaisir de se revoir

I. INTRODUCTION

I.1 Les séquences répétitives de type LINE

Selon l'analyse du génome humain, suite au dévoilement de la séquence complète, le génome serait composé de 45% de séquences répétitives dérivant d'éléments transposables (Int. Hum. Gen. Seq. Cons. 2001). Les deux types majeurs d'éléments transposables sont les éléments *LINEs* (*long interspersed nuclear elements*) et les éléments *SINEs* (*short interspersed nuclear elements*), représentés chez l'humain par la famille Alu (Singer 1982 ; Li *et al.* 2001). Les *LINEs* sont les éléments les plus anciens s'étant établis avec succès chez les mammifères. Leur mécanisme de dispersion, soit la rétrotransposition, leur a permis de coloniser les génomes de mammifère (Hutchison III *et al.* 1989 ; Fanning et Singer 1987). Ce mécanisme leur permet de se répliquer et de s'intégrer ailleurs dans le génome de façon autonome, puisqu'ils codent pour les machineries de transcription et d'intégration requises à leur déplacement (Furano 2000 ; voir figure 1, p.2). Cependant, la réplication d'éléments *LINEs* produit quelques copies compétentes et plusieurs copies non-compétentes tronquées en 5' et réarrangées (Fanning 1983 ; Scott *et al.* 1987 ; Voliva *et al.* 1983 ; Rogers 1985). En effet, la plupart des séquences *LINEs* que l'on retrouve ont une longueur d'environ 1 kb, alors que la forme active possède une longueur d'environ 6 kb (Fanning et Singer 1987 ; Voliva *et al.* 1983 ; Grimaldi *et al.* 1984). On retrouve trois familles de *LINEs* chez l'humain; soit *LINE 1*, *LINE 2* et *LINE 3*. Seulement quelques membres (30 à 60) de

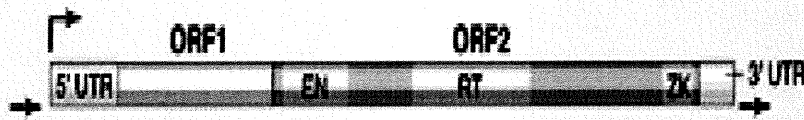


Figure 1 : Représentation schématique d'un élément LINE-1 actif en rétrotransposition.

L'élément L1 complet contient un promoteur interne en 5' (☐▶) et deux cadres de lecture ouverts; l'ORF1 et l'ORF2 (Open Reading Frame). L'ORF1 code pour une protéine se liant au brin d'ARN intermédiaire et l'ORF2 code pour une protéine possédant une activité endonucléase (EN) et reverse transcriptase (RT). L'élément L1 contient aussi une courte région non-transcrite (UTR: Untranslated region) en 5' et en 3', et un signal de polyadénylation (Kazazian 2000).

la famille de LINE 1 (L1) seraient encore actifs en rétrotransposition chez l'humain (Sassaman *et al.* 1997 ; Smit 1999). Selon les dernières données sur le génome humain, la famille L1 serait la plus abondante (Gu *et al.* 2000), constituant 20 % de notre génome et représentant 868 000 copies (Int. Hum. Gen. Seq. Cons. 2001).

I.2 Les LINEs et l'évolution dynamique du génome

La découverte des éléments transposables a fait naître le concept du génome dynamique. Ces séquences d'ADN, qui semblaient autrefois être perçues comme des parasites sans histoire ou sans fonction importante (Hickey 1982), semblent maintenant être considérées comme une force majeure dans le remodelage du génome eucaryote (Makalowski 2000 ; Moffat 2000). Leur activité de rétrotransposition permet leur insertion dans des régions codantes, la dérégulation de la transcription de gènes et la formation de sites de recombinaison homologue permettant le remodelage génomique (Smit 1999). Par exemple, une étude portant sur les éléments rétrotransposables BARE-1 du génome de l'orge suggère que ces éléments pourraient être responsables d'une forme de restructuration du génome via leur mécanisme de rétrotransposition, en réponse aux conditions de croissance de la plante (Wendel et Wessler 2000). Chez l'humain, l'intron 6 du gène EWS (*ewing's*), riche en séquences Alu, montre une très grande plasticité. Les séquences Alu seraient responsables du remodelage de cette région puisque leur transposition aurait permis l'expansion de cette région au cours de l'évolution, et plus récemment (10^5 dernières

années) cette région aurait diminué de plus de 50% par la recombinaison de séquences Alu, créant une délétion de 2480 pb (Zucman-Rossi *et al.* 1997). De plus, des séquences L1 seraient associées à l'évolution du chromosome Y par une inversion causée par la recombinaison entre des éléments L1 (Schwartz *et al.* 1998). Enfin, la rétrotransposition, qui permet aussi la mobilisation de séquences non-L1 en aval (Boeke et Pickeral 1999 ; Moran *et al.* 1999 ; Goodier *et al.* 2000), aurait donné de nouveaux domaines fonctionnels à des protéines et aurait contribué à l'augmentation de la complexité et de la longueur du génome chez les mammifères (Kazazian Jr. 2000 ; Pickeral *et al.* 2000 ; Smit 1996).

I.3 Les LINEs et l'organisation structurale du génome

La majorité des séquences L1 que nous retrouvons dans notre génome sont non-codantes, ne semblant pas à première vue posséder de fonction importante pour le génome. Cependant, elles sont dispersées à travers tout le génome, elles sont présentes chez tous les mammifères et elles conservent des caractéristiques importantes telles que l'abondance et l'homologie. Ceci nous amène à nous demander: quel est leur rôle dans le génome? Nous avons donc émis l'hypothèse que ces séquences pourraient avoir un rôle structural. Par exemple, la présence de ces séquences répétées pourrait être importante dans l'organisation structurale du génome. Nous savons que la fibre de chromatine est compactée dans le noyau d'une façon hautement organisée (Lamond 1998), permettant aux séquences homologues

non-alléliques telles que les séquences répétitives, de se retrouver rapprochées l'une de l'autre. Lorsque des séquences homologues se retrouvent proches, elles ont tendance à former des structures quadruplexes stables (Gaillard et Strauss 1994). Ces structures sont reconnues par des groupes de protéines non-histones de haute mobilité nucléaire (HMG1 et HMG2) (Gaillard et Strauss 1994). Ces protéines omniprésentes dans le noyau permettent de stabiliser ces structures (Lilley 1992). Cependant, la formation de ces structures semble dépendre de la composition de l'ADN (région riche en A-T, poly(CA)-poly(TG)) (Gaillard et Strauss 1994). Or, l'extrémité 3' des éléments L1 est composée de motifs (CA)_x(T)_y (Hutchison *et al.* 1989). On pourrait donc spéculer que l'interaction entre des éléments L1, via leur homologie, permettrait de stabiliser les domaines chromosomiques, c'est-à-dire des boucles chromatine rattachées à la matrice nucléaire (figure 2, p.6). D'autre part, une étude sur l'apoptose menée par Khodarev et ses collaborateurs rapporte que les éléments L1 seraient ciblés durant les étapes initiales de la fragmentation de l'ADN (Khodarev *et al.* 2000). Parmi leurs explications, ils proposent que les éléments L1 puissent être impliqués dans la formation de structures en boucle de l'ADN, puisqu'ils ont retrouvé des séquences L1 aux sites d'attache de l'ADN à la matrice, à l'endroit où le clivage apoptotique a lieu durant l'étape de fragmentation de l'ADN. Les séquences L1 pourraient donc être importantes dans la structure de l'ADN et la stabilisation des domaines chromosomiques. Cependant, ce rôle de stabilisateur implique l'association intime de ces séquences, représentant ainsi des substrats possibles pour la recombinaison homologe spontanée entre des séquences répétitives non alléliques.

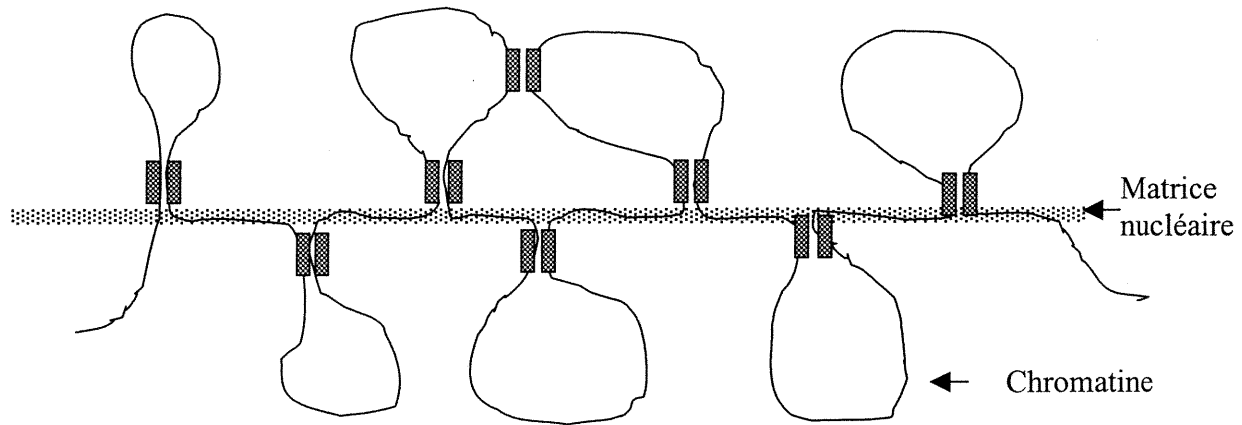


Figure 2 : Implication des séquences répétitives dans l'organisation structurale de la chromatine dans le noyau.

La fibre de chromatine forme des boucles rattachées à la matrice nucléaire, qui constituent les domaines chromosomiques. Les séquences répétitives (rectangles hachurés) permettraient la stabilisation de ces domaines chromosomiques.

I.4 La recombinaison homologue

La recombinaison homologue est un processus universel dans la vie et l'évolution des organismes (Kucherlapati et Smith 1998). Elle joue un rôle important dans le réassortiment des gènes et la ségrégation des chromosomes durant la méiose et c'est un mécanisme important dans la réparation de l'ADN au cours de la mitose d'une cellule. Les séquences répétitives sont distribuées à travers le génome des organismes complexes. L'homologie entre ces séquences leur permet de s'apparier et de promouvoir la recombinaison non allélique entre des fragments de chromatine non reliés. La recombinaison entre deux séquences homologues peut donner lieu principalement à deux types de produits, soit la conversion génique avec ou sans crossing-over. Dans la littérature on nomme simplement crossing-over le premier produit et conversion génique le deuxième (voir figure 3, p.8). Le crossing-over résulte de l'échange réciproque d'information génétique entre deux séquences d'ADN pouvant donner lieu à des réarrangements majeurs puisque ce ne sont plus les mêmes séquences flanquantes qui sont associées, tandis que la conversion génique résulte de l'échange non réciproque d'information génétique entre deux séquences d'ADN puisque l'échange s'effectue d'un chromosome vers l'autre en laissant le brin d'ADN donneur intact. Ce dernier produit est donc préférable pour conserver l'intégrité du génome puisque la modification est localisée et que les régions génomiques flanquantes des deux séquences homologues demeurent inchangées (Haber 2000).

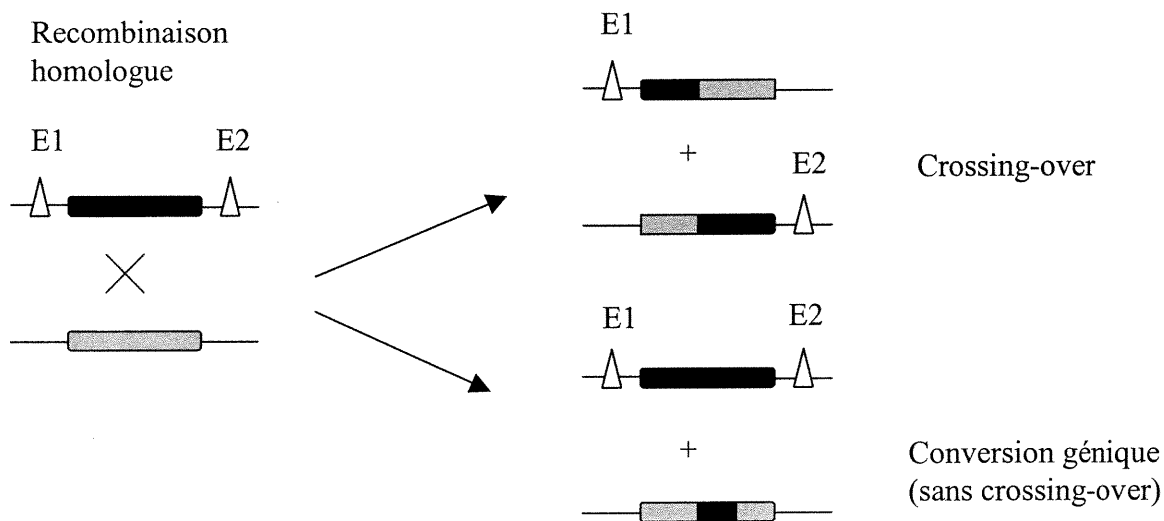


Figure 3 : Représentation schématique des principaux produits d'un événement de recombinaison entre deux séquences homologues non alléliques.

Les deux séquences homologues (rectangle noir et rectangle noir et blanc) s'alignent et, selon la résolution de l'événement de recombinaison (X), deux types de résultats sont possibles, soit la conversion génique ou le crossing-over. Le crossing-over résulte en l'échange réciproque d'information génétique entre les deux séquences homologues incluant l'échange des séquences génomiques flanquantes, comme par exemple des exons (E1 et E2). La conversion génique est quant à elle un échange non réciproque d'information génétique, c'est-à-dire que la modification est localisée et l'élément donneur ainsi que les séquences génomiques flanquantes des deux éléments ne sont pas modifiés.

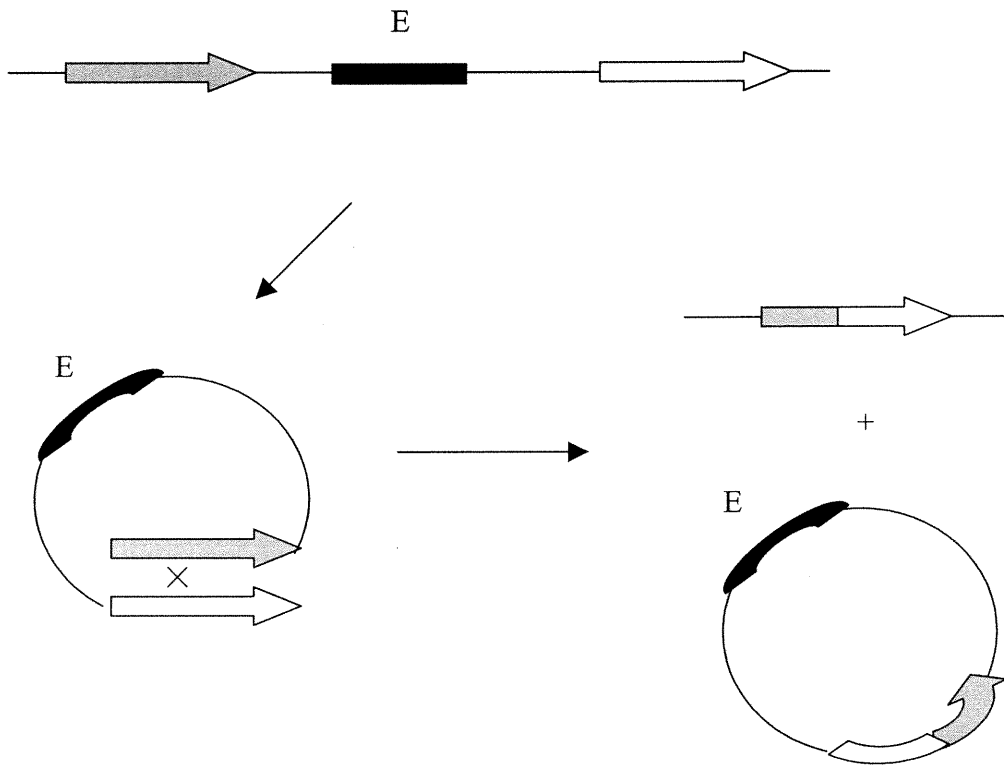
I.5 Les séquences répétitives et les réarrangements chromosomiques

L'abondance des séquences répétitives représente un défi pour la cellule qui doit maintenir la stabilité génomique puisque chacune des séquences représente autant de sites potentiels de recombinaison homologue. Si l'échange d'information génétique entre les deux séquences se résout par un échange réciproque, ceci peut mettre en jeu l'intégrité du génome. Ce type d'événement pourrait avoir lieu entre deux séquences homologues situées sur le même chromosome (intrachromosomiques) résultant en la délétion (illustré à la figure 4 A), p.10) ou l'inversion de séquences importantes comme des exons de gènes (Shaffer et Lupski 2000). Cet événement pourrait aussi se produire entre deux séquences provenant de deux chromosomes différents, interchromosomiques, donnant lieu à des réarrangements chromosomiques complexes comme des translocations chromosomiques (Shaffer et Lupski 2000; voir figure 4 B), p.10). De fait, il y a plusieurs évidences de réarrangements chromosomiques liés à la recombinaison entre des éléments répétitifs chez l'humain. La recombinaison entre des séquences répétitives de type Alu est connue pour être responsable d'une variété de désordres génétiques. Par exemple, la recombinaison homologue entre des séquences Alu du gène BRCA1, qui présente une forte densité en séquences Alu, contribuerait à une proportion significative des mutations de BRCA1 retrouvées dans les cancers héréditaires de l'ovaire ou du sein (Rohlf's *et al.* 2000). Les séquences Alu seraient aussi responsables de réarrangements associés à l'activation du gène MDR-1 (Multidrug Resistance-1) identifiés dans des cellules malignes humaines (Harada *et al.* 2000). De plus, la recombinaison entre des

Figure 4: Représentation schématique de la recombinaison homologue avec crossing-over intrachromosomique et interchromosomique.

A) La recombinaison (X) avec crossing-over entre deux séquences homologues situées sur le même chromosome impose la formation d'une boucle d'ADN lors de l'alignement des séquences. La résolution de cet événement comprend l'échange réciproque d'information génétique entre les deux séquences homologues et la délétion des séquences génomiques contenues entre ces deux éléments. B) La recombinaison avec crossing-over entre deux séquences homologues situées sur deux chromosomes différents se résout par une translocation chromosomique, c'est-à-dire l'échange réciproque d'information génétique entre les deux chromosomes. Les séquences homologues sont représentées par une flèche blanche et une flèche grise. La ligne mince représente les séquences génomiques et le rectangle noir un exon (E).

A) Recombinaison homologue
intrachromosomique



B) Recombinaison homologue
interchromosomique

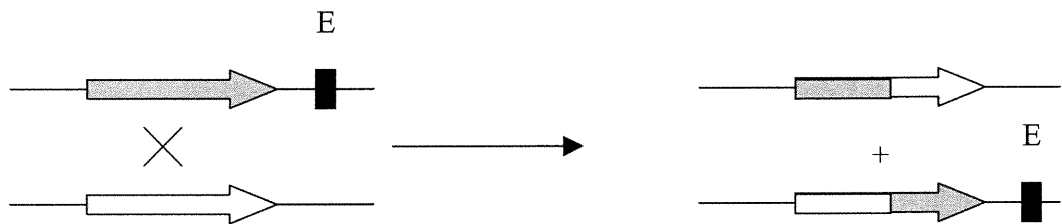


Figure 4

séquences Alu serait un mécanisme actif de formation de délétions dans le syndrome Lesch-Nyhan (Tvdik *et al.* 1998). Enfin, la recombinaison entre des séquences Alu pourrait être associée à l'événement de translocation constitutionnel le plus fréquent chez l'humain soit t(11;22) (Hill *et al.* 2000). Des études rapportent aussi l'implication de séquences L1 dans des réarrangements moléculaires associés à des maladies génétiques chez l'humain. Par exemple la recombinaison entre des éléments L1 serait responsable de délétions d'ADN génomique observées dans le syndrome Alport (Segal *et al.* 1999) et dans le cas d'une déficience de la phosphorylase kinase PHKB associée à des problèmes d'emmagasinement du glycogène (Burwinkel et Kilimann 1998). La recombinaison entre des éléments L1 serait aussi responsable de délétions au niveau du bras court du chromosome 3 associées à plusieurs types de cancers chez l'humain (Inoue *et al.* 1997 ; Mimori *et al.* 1999). Ces exemples de maladies génétiques ou de cancers causés par des événements de recombinaison homologue entre des séquences L1 présentent des évidences que les éléments L1 interagissent ensemble, puisque ces interactions mènent à des phénotypes que l'on peut caractériser. Mais qu'en est-il des échanges génétiques non réciproques (conversion génique) entre ces éléments qui ne présentent pas en soi un danger pour l'intégrité du génome et qui n'interviennent pas dans l'évolution générale du génome mais plutôt dans l'évolution précise de ces séquences?

I.6 L'évolution concertée des séquences répétitives

Nous avons vu que les éléments L1 évoluent par leur mécanisme de dispersion en générant de nouvelles copies colonisant le génome. Cependant, la majorité des copies de LINEs que l'on retrouve dans le génome humain sont très vieilles, c'est-à-dire que 80% de ces copies se seraient rétrotransposées il y a plus de 30 millions d'années (Smit *et al.* 1995). Mais l'observation que la majorité des copies présentes dans le génome sont défectives soulève la possibilité que d'autres mécanismes d'évolution existent. Les séquences L1 défectives peuvent diverger avec le temps par rapport à la séquence d'origine, en accumulant des mutations aléatoires à un taux semblable aux pseudogènes lors d'erreurs durant la réplication de l'ADN de l'hôte, puisqu'elles ne subissent pas la pression sélective (Voliva *et al.* 1983 ; Hardies *et al.* 1986 ; Scott *et al.* 1987 ; Pascale *et al.* 1993 ; DeBerardinis et Kazazian 1998). Cependant, les éléments L1 présentent un patron lignée-spécifique, c'est-à-dire qu'il existe une plus grande homologie entre ces séquences à l'intérieur d'une même lignée qu'entre différentes lignées. Les éléments L1 évoluent donc de façon concertée (Scott *et al.* 1987). La majorité des séquences répétitives démontrent un patron d'évolution concertée (Elder et Turner 1995 ; Liao 1999), suggérant que les membres d'une même famille répétitive n'évoluent pas indépendamment les uns des autres, et qu'il doit exister des interactions entre ceux-ci permettant l'élimination ou la propagation d'une variation dans les autres copies (Dover *et al.* 1982 ; Zimmer *et al.* 1980 ; voir figure 5, p.14). Ce processus moléculaire menant à l'homogénéisation des séquences d'ADN d'une même famille existe chez tous les organismes, des

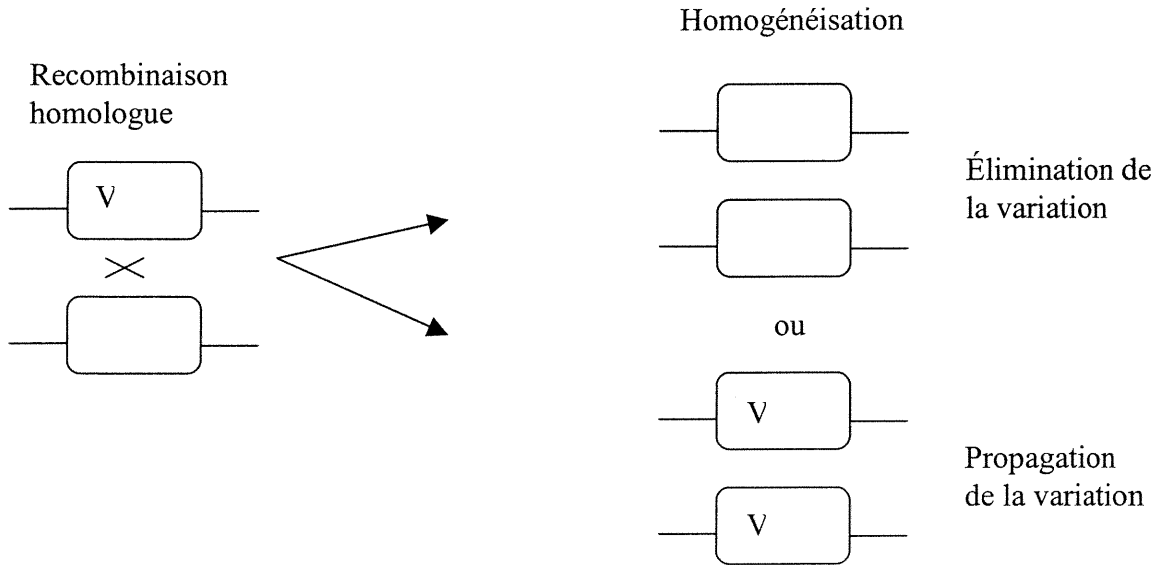


Figure 5: Représentation schématique de l'homogénéisation des membres d'une famille de séquences répétitives.

La recombinaison (X) entre deux séquences homologues (rectangles) permet l'élimination ou la propagation d'une variation (V) dans une famille de séquences répétitives selon quelle séquence (avec ou sans variation) sert de donneur dans l'événement de conversion. La ligne mince représente les séquences génomiques flanquantes.

bactéries jusqu'aux mammifères (Liao 1999). Par exemple, chez les primates, l'évolution concertée du locus RNU2 comprenant un alignement en tandem de gènes U2 fut très bien documentée (Pavelitz 1999). Nous observons aussi ce phénomène au niveau de familles multigéniques telles que celles codant pour les histones (Coen *et al.* 1982). Selon certaines évidences, la conversion génique, qui nécessite une interaction non réciproque entre deux séquences homologues, pourrait maintenir l'homogénéisation des séquences répétitives. En effet, une étude menée par Kass et ses collaborateurs suggère qu'un événement de conversion génique serait responsable de la conversion d'un membre d'une sous-famille d'élément Alu préexistant en une jeune sous-famille (Maeda *et al.* 1988; Kass *et al.* 1995). Des études portant sur l'évolution des éléments L1 chez la souris proposent, à partir d'études phylogénétiques, que des événements de recombinaison homologue entre des éléments L1 pourraient être responsables d'une part de leur évolution concertée (Jubier-Maurin *et al.* 1985) et d'autre part de la naissance d'une nouvelle famille L1 active (Saxton et Martin 1998 ; Mears et Hutchison 2001). De plus, notre laboratoire a démontré qu'une séquence L1 exogène pouvait interagir par conversion génique avec des éléments L1 endogènes dans des cellules de souris en culture (Belmaaza *et al.* 1990 ; 1994) et dans des lignées de cellules humaines (Richard *et al.* 1994 ; 1997). Nous avons aussi observé des évidences de conversion génique entre des éléments L1 chromosomiques chez la souris (Tremblay *et al.* 2000 ; 2001).

I.7 But du travail

Nous savons maintenant que les éléments L1 ont contribué à l'évolution dynamique du génome chez les mammifères. Leur abondance, leur dispersion et leur similarité font d'eux des sites privilégiés pour la recombinaison homologue. De plus, plusieurs études ont démontré des évidences d'interactions entre des éléments L1. Le but précis de ce travail est de déterminer si nous pouvons retrouver des évidences de conversion génique entre des éléments L1 dans le génome humain.

I.8 Approche expérimentale

Nous avons vu que les éléments L1 constituent une grande part de notre génome, soit environ le 1/5. La majorité de ces séquences sont non-codantes. Elles semblent, à priori, n'avoir aucune fonction biologique. Cependant elles conservent des propriétés importantes telles que l'homologie, l'abondance et la dispersion. Ces caractéristiques, en plus du fait qu'elles existent chez tous les mammifères, soulèvent la possibilité que ces éléments L1 puissent avoir un rôle à jouer au sein de l'organisation structurale du génome. Ce rôle de structure pourrait, par exemple, impliquer des interactions homologues entre ces séquences L1. Toutefois, ces interactions pourraient d'une part servir de substrat à la recombinaison homologue, tel que nous l'avons vu par des évidences indirectes dans des cas de maladies génétiques et de cancers, et d'autre part favoriser l'évolution des séquences L1 elles-mêmes et du génome.

Dans notre projet, nous nous sommes intéressés au comportement évolutif de ces séquences L1 chez l'humain. Nous nous sommes demandé si ces séquences interagissaient de façon spontanée au cours de l'évolution de l'espèce humaine

Notre démarche consistait dans un premier temps à choisir un élément LINE particulier et à comparer sa séquence chez plusieurs individus, afin de déterminer s'il existait plusieurs variations de cet élément dans la population. Nous avons choisi trois éléments L1 situés dans la région fragile FRA3B du gène FHIT pour lesquels nous connaissions la séquence. Ces trois éléments L1 furent sélectionnés puisque des études

rapportaient des interactions entre eux, caractérisées au niveau de plusieurs lignées de cellules tumorales (Inoue *et al.* 1997). Dans un deuxième temps, nous avons tenté de déterminer si les variations que l'on observait pour chacun des L1 était dues à l'accumulation de mutations ponctuelles au cours de l'évolution, ou si elles étaient le résultat d'interactions entre un élément L1 et d'autres éléments du génome. Pour ce faire, nous avons comparé les variations obtenues pour chacun des éléments L1 aux variations observées pour une séquence intronique, laquelle ne devrait pas subir la pression sélective au même titre qu'un élément L1. Enfin, nous avons vérifié si dans un contexte d'instabilité génomique, il y aurait augmentation des interactions entre les éléments L1. Nous avons ainsi comparé la séquence d'un élément donné à un endroit donné au niveau du tissu normal et tumoral chez des individus atteints du cancer du colon. Nous avons choisi de faire nos analyses sur ce type de cancer puisque nous avons accès à des biopsies de tumeur dans notre laboratoire par l'entremise du Dr. Mark Basik. De plus des études ont rapporté l'implication du gène FHIT dans ce type de cancer (tel que mentionné plus loin).

II. MATÉRIEL ET MÉTHODES

II.1 Sélection des éléments L1 et de la séquence intronique

Pour notre étude portant sur l'évolution des L1, nous avons sélectionné trois éléments L1 situés dans la région fragile FRA3B du gène FHIT localisé sur le chromosome 3. Les trois éléments L1 choisis, illustrés à la figure 10A), p.35 (section des résultats) , sont tronqués en 5' comme la majorité des L1 que l'on retrouve dans notre génome, ils sont plus grand que 1kb et sont situés sur le même chromosome. De plus ces trois éléments L1 que nous avons identifiés A, B et C sont répartis de part et d'autre de l'exon 5 du gène FHIT. Pour sa part, la séquence intronique choisie est située entre le L1 C et l'exon 5 et est une séquence unique dans le génome (figure 10A)).

II.2 Sélection de la région à cloner et à séquencer

L'alignement des séquences de chacun des trois L1 sélectionnés, publiées dans Gene Bank (#accession AF020503), nous a permis de déterminer une région de 300pb qui comportait le plus de bases différentes (31 bases) entre les trois éléments (figure 11, p. 38, section des résultats). Cette région de 300pb est située dans la portion 3' des L1. C'est la région la plus fréquente puisque la majorité des éléments

sont tronqués en 5'. Ces bases distinctes entre les trois éléments L1 permettaient d'identifier les partenaires de recombinaison homologue si l'événement avait lieu entre deux de ces éléments. Par contre, la région de 300pb sélectionnée pour la séquence intronique fut choisie au hasard parmi les séquences uniques de cette région.

II.3 Amplification des éléments L1 et de la séquence intronique

Nous avons amplifié les trois éléments L1 par PCR à l'aide du système Expand High Fidelity (Boehringer Mannheim) , au niveau du tissu sain périphérique et du tissu tumoral provenant de biopsies de tumeurs chez dix patients atteints du cancer du colon. Les échantillons d'ADN normal et tumoral nous ont été fournis par le Dr. Mark Basik, CHUM, Hôpital Hotel Dieu. Les amorces PCR sont situées de part et d'autre des éléments L1 dans des régions uniques du génome, tel qu'illustré aux figures 6, 7 et 8 (p.26 à 31), afin de favoriser une amplification spécifique. Les séquences des amorces utilisées pour l'amplification de chacun des éléments L1 sont les suivantes: L1 A amorce #1, 5'-TGTACAGTTCTGGCTATGTGGGTGTGTCCA-3' et amorce #2, 5'-AGAACAAGTACGAACACTGTCTGACCTTGC-3' ; L1 B amorce #3, 5'-AGGTATGGCTGTACCCCTACGAGGCATGGA-3' et amorce #4, 5'-GTAGGCATATGCTTGGAGTAGTAGTGATGG-3' ; L1 C amorce #5, 5'-CTTCAGCATTAACTCGACTTCCCCATAGT-3' et amorce #6, 5'-ACATTTAGAGTTGGCGGTCAAGGTGTGTTG-3'. Puis, les amorces utilisées

pour l'amplification de la séquence intronique (figure 9, p.32) sont : amorce #7, 5'-GCCGTTTGTA ACTAATACAGACGCAC-3' et amorce #8, 5'-GTATACTCCTCCCACGATGTAGGAGAAGAG-3'. L'amplification PCR de chacun des L1 pour chacun des individus a été effectuée sur 100ng d'ADN génomique, selon les recommandations du manufacturier (Boehringer Mannheim), dans l'appareil PCR DNA thermal cycler (Perkin Elmer Cetus) selon les conditions suivantes : 3 min à 95°C ; 10 cycles pour 1 min à 95°C, 1 min à 70°C et 4 min à 68°C ; et 20 cycles pour 1 min à 95°C, 1 min à 62°C et 4 min à 68°C (extension de 20 sec/cycle).

II.4 Clonage des fragments L1 A, B et C et de la séquence intronique

II.4.1 Méthode de clonage des L1 B et C (voir schéma de la procédure de clonage aux figures 7 et 8, p.28 à 31)

a) Digestion des produits PCR

Tout d'abord, nous avons précipité l'ADN amplifié en ajoutant 5 μ l d'acétate de sodium 3M et 100 μ l d'éthanol 100% au 50 μ l de produit PCR et en déposant les tubes à -80°C pendant 15 minutes. Par la suite le tout a été centrifugé à 13 000 rpm pendant 15 minutes et le culot a été resuspendu dans 38 μ l de TE 1X. Tous les enzymes de restriction utilisés provenaient de New England Biolabs. Les enzymes de restriction KpnI et NdeI, dont les sites de digestion sont situés à l'intérieur des L1 B

et C, ont été choisies pour faire le clonage de ces fragments. La digestion du produit PCR total s'effectuait à 37°C pendant environ 18 heures dans un volume total de 50µl en présence de 5µl de tampon NEB1 10x, 5µl de BSA 10x, 1µl d'enzyme KpnI (10U/µl) et 1µl d'enzyme NdeI (20U/µl).

b) Extraction de la bande d'ADN à cloner

La bande d'ADN à cloner fut isolée à partir d'un gel d'agarose 0,7%, selon le protocole fournit avec le kit d'extraction d'ADN QIAEXII (Qiagen). Ensuite, la concentration d'ADN obtenue a été évaluée sur gel d'agarose à l'aide d'un standard de concentration d'ADN, en vue du clonage.

c) Ligation des fragments de L1, B et C dans le plasmide p3'SS

La réaction de ligation était composée de 100ng d'insert (ADN L1) pour 25 ng de vecteur p3'SS (préalablement digéré aux sites de clonage) en présence de tampon de ligation et de l'enzyme T4 DNA ligase. La réaction d'un volume total de 20µl a été incubée à 16°C durant environ 18 heures.

d) Transformation bactérienne des produits de ligation

Les produits de ligation ont été introduits par électroporation (appareil *gene zapper*, Biorad) dans une souche de bactéries compétentes DH10B (préparée au laboratoire). Nous avons électroporé 2µl du produit de ligation dans 50µl d'une suspension de bactéries compétentes. Par la suite, les bactéries ont été diluées dans

1ml de milieu S.O.C et incubées à 37°C pendant 1 heure avec agitation avant d'être étalées sur des Pétri contenant du milieu LB agar avec ampicilline.

II.4.2 Clonage du L1 A et de la séquence intronique (voir schéma de la procédure de clonage aux figures 6 et 9, p.26 et p.32)

a) Ajout de nucléotides "A" aux extrémités 3' des produits d'amplification

Nous avons incubé le produit d'amplification du L1 A et de la séquence intronique à 72°C pendant 8 minutes en présence de 0,4 µl d'enzyme polymérase Amplitaq 5U/µl (Perkin Elmer).

b) Ligation de chacun des produits d'amplification L1 A et la séquence intronique dans le vecteur topo pCR2.1

Les produits de ligation incubés à la température ambiante pendant 10 minutes contenaient 2 µl de produit PCR, 1µl de solution de sel diluée à 300mM, 2.5 µl d'eau stérile et 0.5 µl (10 ng/ul) de vecteur topo pCR2.1 (Invitrogen).

c) Transformation bactérienne des produits de ligation

La transformation de ces produits de ligation a été effectuée selon la procédure décrite au point 4.1d

II.5 Extraction de l'ADN plasmidique

Pour extraire l'ADN plasmidique nous avons utilisé un protocole d'extraction d'ADN sans phénol/chloroforme. Nous avons repiqué 5 colonies bactériennes par individu pour chacun des trois éléments L1 choisis (ADN normal et tumoral) et pour la séquence intronique (ADN normal). L'extraction d'ADN plasmidique a été effectuée comme suit : le culot de bactéries récolté à partir de 1,5ml de culture bactérienne a été mis en présence de 300µl d'une solution préparée de TENS (concentrations finales : TE 1X ; 0,5% SDS ; 0,1N NaOH) et de 150µl d'acétate de sodium 3M pH5,2. Le tout a été centrifugé à 15 000 rpm pendant 2 minutes. Ensuite l'ADN était précipité avec 900µl d'éthanol 100% (-20°C) et centrifugé à 15 000 rpm. Enfin le culot a été rincé à l'éthanol 70% et dilué dans 100µl de TE 1X contenant 40µg/ml d'ARNase.

II.6 Digestions caractéristiques des clones obtenus

Pour s'assurer que les clones obtenus (5 clones/individu) possédaient les bons fragments d'ADN, nous avons effectué des digestions caractéristiques présentées dans la section des résultats. Les enzymes ont été utilisées selon les indications du manufacturier (New England Biolabs)

II.7 Séquençage de la région de 300pb

L'ADN (450ng) a été séquençé à l'aide de la trousse de séquence *thermo sequenase radiolabeled terminator cycle* (Amersham Pharmacia) en utilisant un oligo orienté dans le sens 5'-3' de l'ADN et en complétant la lecture de la séquence avec un oligo dans le sens inverse 3'-5'. Les résultats (ADN normal et tumoral) ont été analysés par l'alignement des 50 séquences obtenues (5 clones/individus) pour chacun des L1 (A,B et C) et pour la séquence d'ADN intronique.

Figure 6 : Schéma de la procédure de clonage du L1 A.

L'élément L1 A (flèche pleine) ainsi que ses régions flanquantes (lignes minces) furent amplifiés à l'aide des amorces 1 et 2 (flèches minces) donnant lieu à un produit d'amplification total de 3342 pb qui a été cloné directement dans le plasmide PCR2.1. Nous avons représenté le plasmide avec ses différentes caractéristiques ainsi que son site de clonage situé dans le gène lacZ permettant une sélection bleue/blanche des colonies portant l'insert.

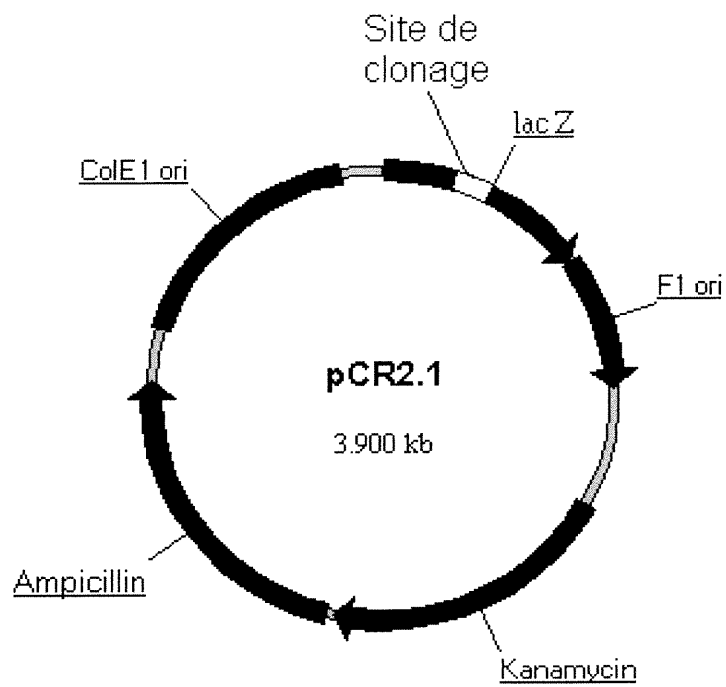
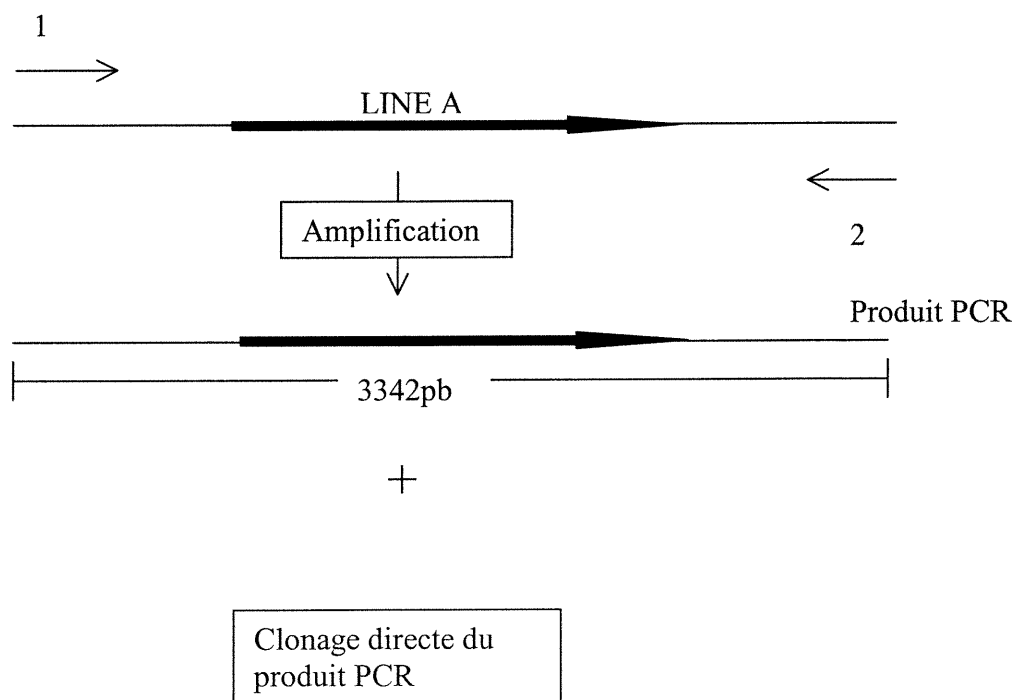
**Figure 6**

Figure 7 : Schéma de la procédure de clonage du L1 B

L'élément L1 B (flèche pleine) ainsi que ses régions flanquantes (lignes minces) furent amplifiés à l'aide des amorces 3 et 4 (flèche minces), donnant lieu à un produit d'amplification total de 4173 pb produisant un insert de 1301 pb suite à une digestion par les enzymes de clonage KpnI et NdeI (flèches verticales). Ensuite l'insert était cloné dans le vecteur p3'SS préalablement digéré par les enzymes KpnI et NdeI. Nous avons présenté le plasmide avec ses différentes caractéristiques ainsi que les sites de digestion (KpnI et NdeI) ayant servi au clonage. La portion du vecteur contenant le gène lacI a été éliminée afin de pouvoir utiliser les sites de restriction KpnI et NdeI pour le clonage. Cette délétion n'a pas affecté le bon fonctionnement du vecteur.

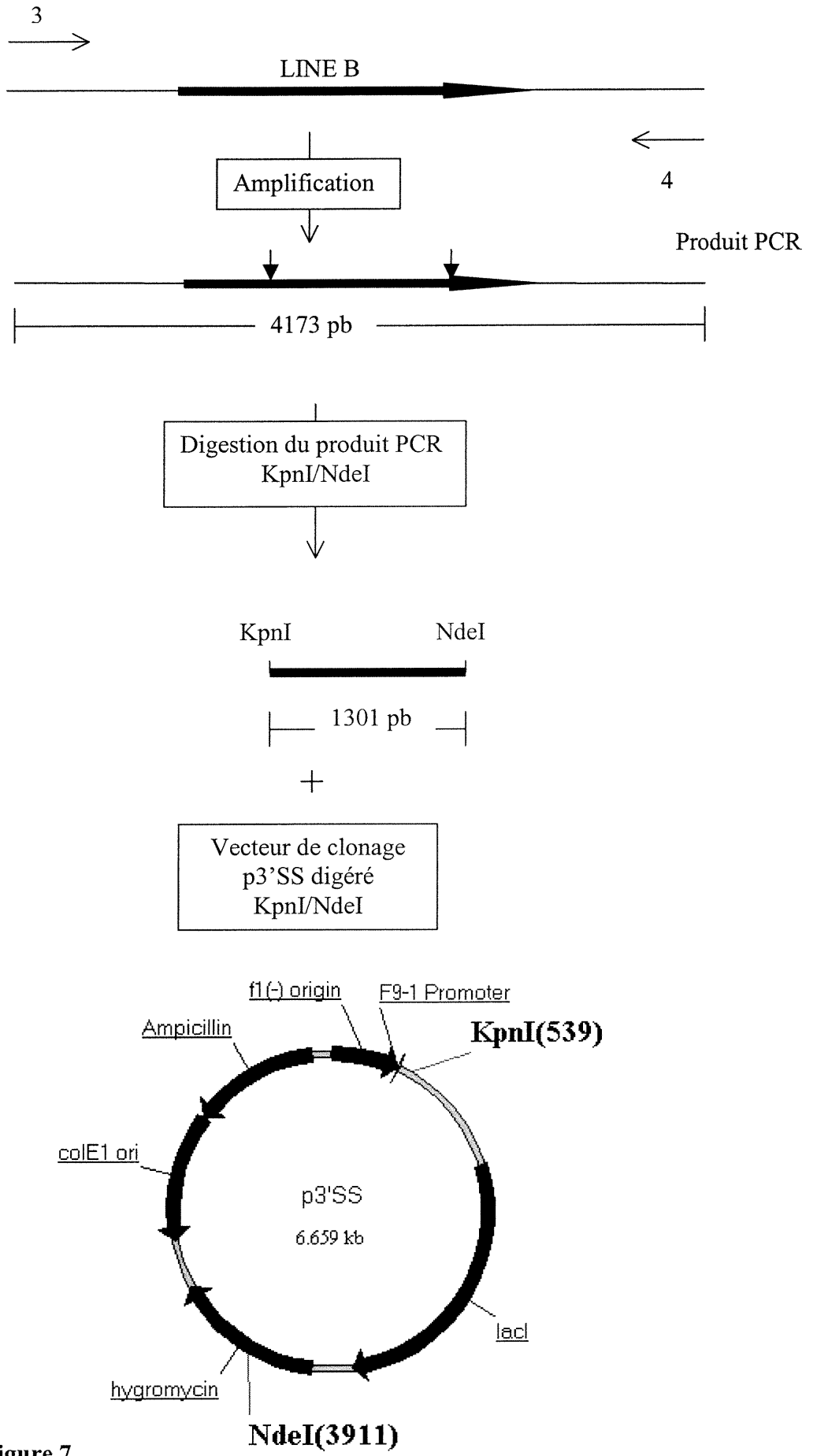


Figure 7

Figure 8 : Schéma de la procédure de clonage de L1 C

L'élément L1 C (flèche pleine) ainsi que ses régions flanquantes (lignes minces) furent amplifiés à l'aide des amorces 5 et 6 (flèche minces) donnant lieu à un produit d'amplification total de 2011 pb produisant un insert de 1313 pb suite à une digestion par les enzymes de clonage KpnI et NdeI (flèches verticales). Ensuite, l'insert était cloné dans le vecteur p3'SS qui était préalablement digéré par les enzymes KpnI et NdeI. Nous avons présenté le plasmide avec ses différentes caractéristiques ainsi que les sites de digestion (KpnI et NdeI) ayant servi au clonage. La portion du vecteur contenant le gène lacI a été éliminée afin de pouvoir utiliser les sites de restriction KpnI et NdeI pour le clonage. Cette délétion n'a pas affecté le bon fonctionnement du vecteur.

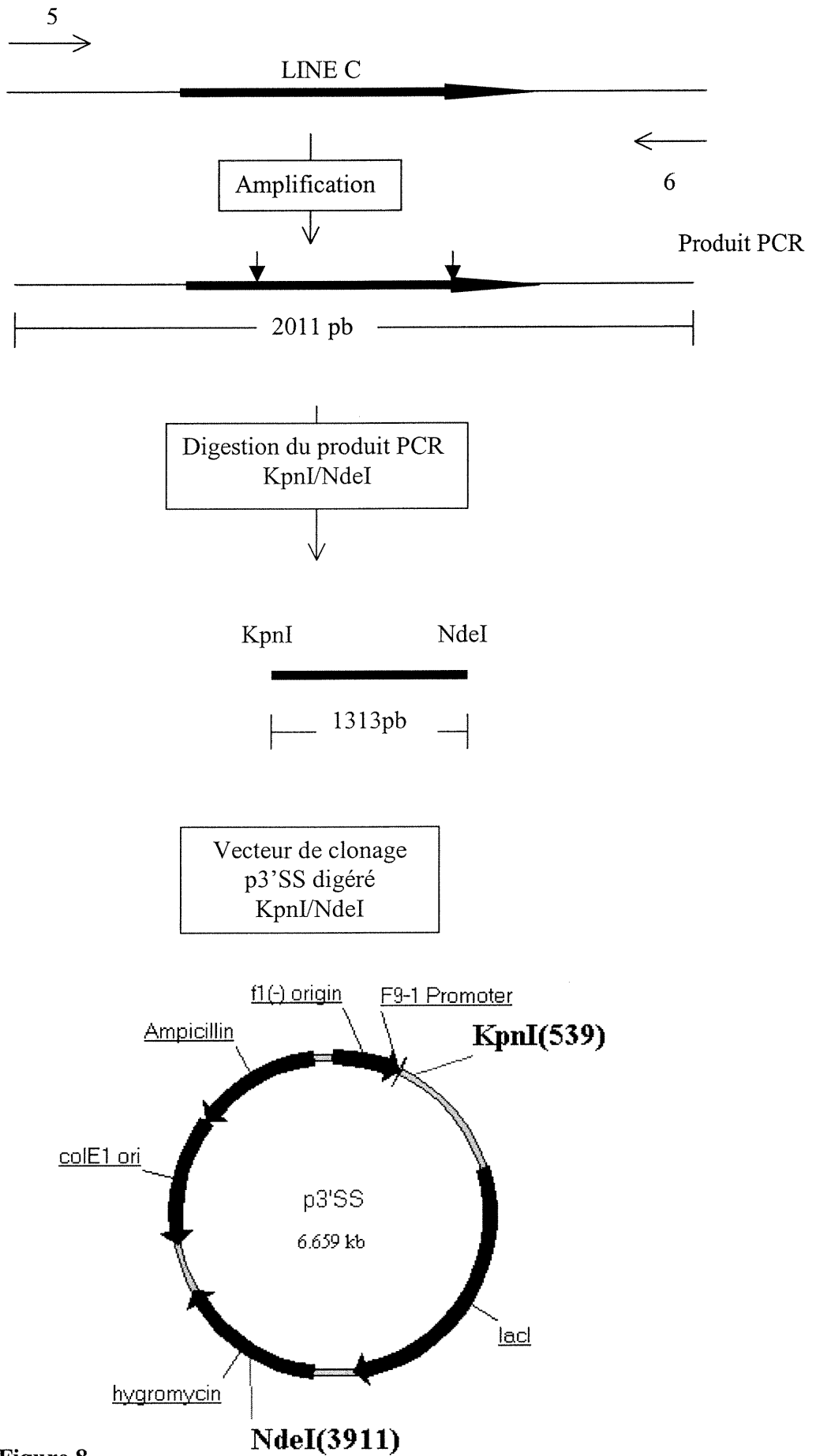
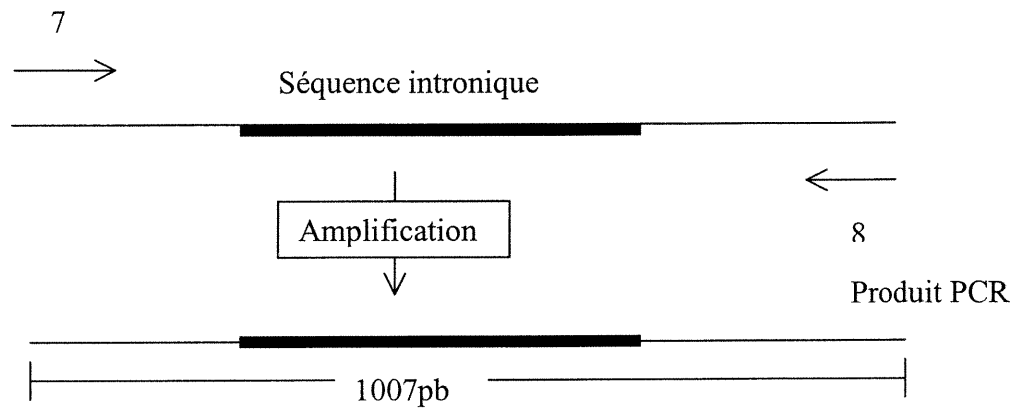


Figure 8

Figure 9 : Schéma de la procédure de clonage de la séquence intronique.

La séquence intronique (rectangle noir) ainsi que ses régions flanquantes (lignes minces) furent amplifiées à l'aide des amorces 7 et 8 (flèches minces), donnant lieu à un produit d'amplification total de 1007 pb, cloné directement dans le plasmide PCR2.1. Nous avons représenté le plasmide avec ses différentes caractéristiques ainsi que son site de clonage situé dans le gène lacZ permettant une sélection bleue/blanche des colonies portant l'insert.



+

Clonage directe du
produit PCR

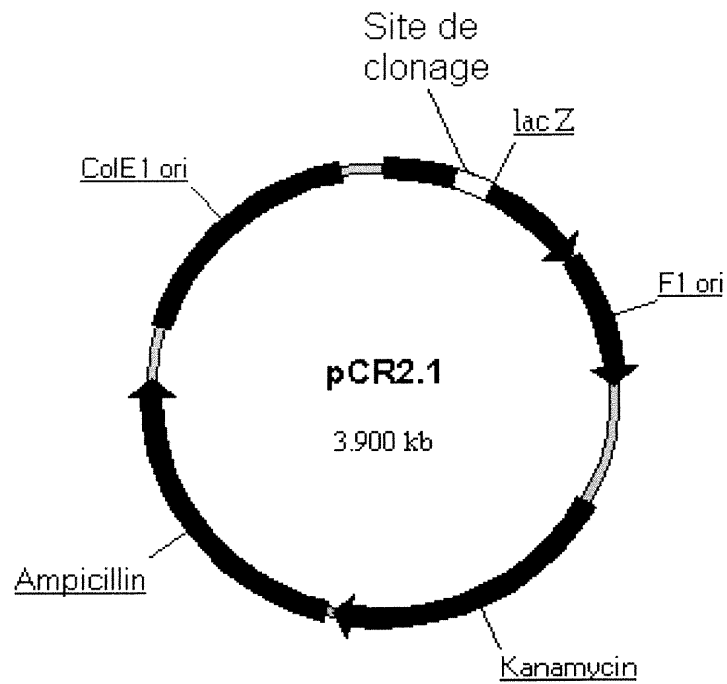


Figure 9

III. RÉSULTATS

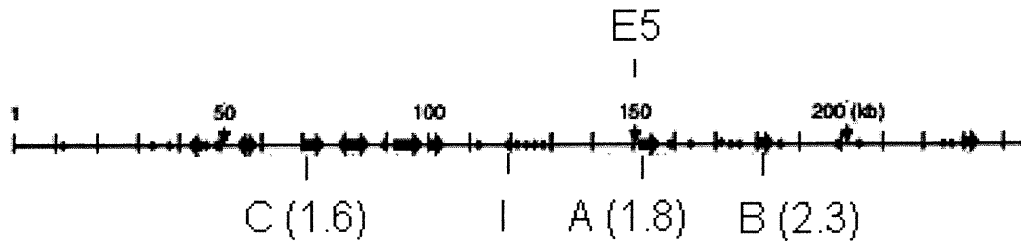
III.1 Résultats de l'approche expérimentale utilisée

Notre étude de l'évolution des éléments LINEs consistait, dans un premier temps, à faire l'analyse du comportement évolutif de ces éléments, à un endroit donné dans le génome humain au niveau du tissu sain. Par la suite, nous nous sommes intéressés au comportement de ces éléments dans le processus tumoral. Pour réaliser ce projet, nous avons d'abord sélectionné trois éléments L1 rapportés dans la littérature comme étant impliqués dans des événements de réarrangements génomiques (Inoue *et al.* 1997). Ces trois éléments, identifiés A, B et C à la figure 10A), p.35, sont situés dans la région fragile FRA3B du gène FHIT. Nous avons choisi ces éléments puisqu'ils ont été associés à des délétions qui semblaient provenir d'événements de recombinaison homologe entre le L1 C et A, et le L1 C et B. Nous voulions analyser la séquence de chacun de ces trois éléments L1 au niveau du tissu sain périphérique et du tissu tumoral, à partir de biopsies de tumeurs, chez dix individus atteints du cancer du colon. Pour faire l'analyse de chacun de ces éléments, nous avons repéré dans GenBank (voir section matériel et méthodes) la séquence totale de chacun des éléments et de leurs séquences flanquantes. L'alignement des séquences complètes des trois éléments nous a permis de mettre en évidence les différences de bases nucléiques entre les trois éléments. Nous avons choisi une région de 300 pb qui présentait le plus de différences entre les trois éléments. À la

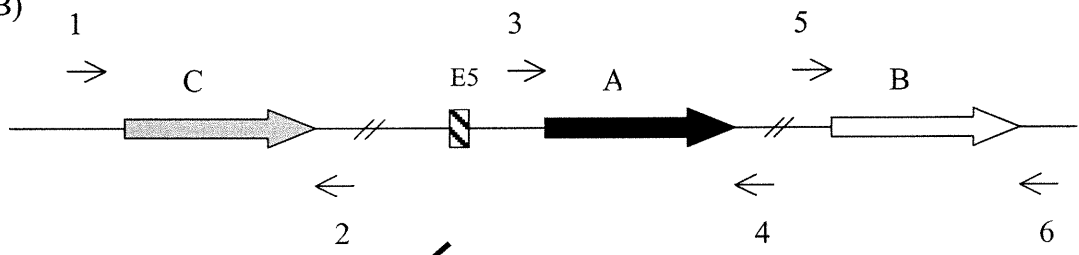
Figure 10 : Présentation schématique de l'organisation des L1 A, B et C dans la région fragile FRA3B du gène FHIT.

A) Schéma de l'organisation des éléments L1 A, B et C sélectionnés dans la région fragile FRA3B du gène FHIT située sur le chromosome 3 (3p14.2) (tiré de Inoue *et al.* 1997). Les flèches noires pleines représentent la position des 43 éléments L1 dénombrés dans cette région. Les points repère de la longueur de cette région sont identifiés en kb. La distance entre le L1 C et le L1 A est de 80 kb et entre le L1 A et le L1 B est de 30 kb, donc la distance entre le L1 C et le L1 B est de 110 kb. Le I représente la position de la séquence intronique utilisé dans notre étude par rapport aux L1 A, B et C. La longueur en kb de chacun des éléments L1 (A, B et C) est présentée entre parenthèses. E5 représente la position de l'exon 5 du gène FHIT. B) Présentation de la position des amorces (1 à 6) dans les régions uniques du génome pour l'amplification spécifique des trois éléments L1 A, B et C (flèches pleines) et présentation d'un exemple de recombinaison homologue entre les séquences L1 C et B et les produits possibles de cet échange, soit la conversion génique (échange non-réciproque de séquences d'ADN) ou le crossing-over (échange réciproque de séquences d'ADN).

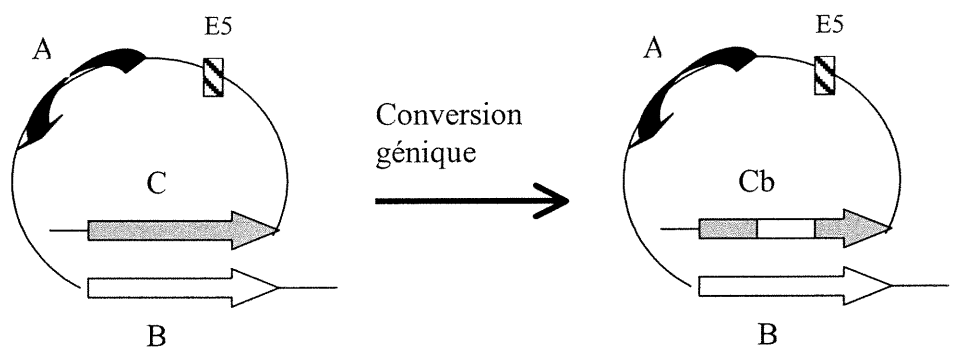
A)



B)



Alignement des séquences



Crossing-over

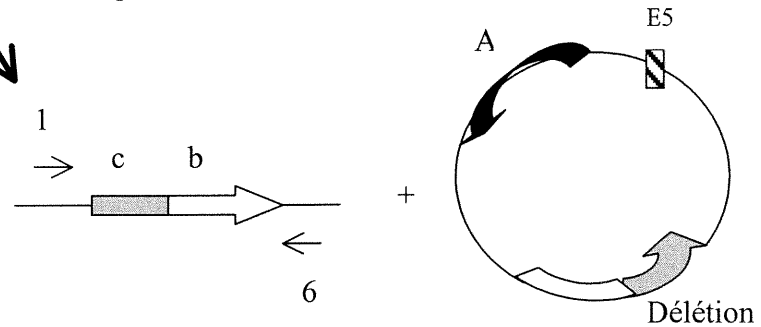


Figure 10

figure 11, p.38 nous présentons cette région de 300 pb que nous avons sélectionné. Nous avons mis en évidence (en gris) les bases distinctives de chacun des éléments L1. Cette distinction entre les trois L1 a été importante pour tenter d'identifier les partenaires de recombinaison homologue si l'événement avait lieu entre deux de ces éléments. En d'autres mots, le transfert des bases caractéristiques d'un élément L1 donné à un autre élément durant le processus de recombinaison nous permettait de repérer les éléments qui auraient pu interagir ensemble. Les séquences flanquantes, pour leur part, ont été utilisées pour choisir des amorces uniques dans le génome, afin d'amplifier de façon spécifique chacun des éléments pour en faire leur analyse.

Nous avons illustré à la figure 10B), p.35 l'organisation schématique de ces trois éléments L1 (A, B et C) dans le génome humain. Ces éléments sont situés sur le même chromosome et de part et d'autre de l'exon 5 (E5) du gène FHIT. Les interactions possibles de ces séquences homologues entre elles (A, B et C) ou avec d'autres séquences L1 dispersées dans le génome peuvent donner lieu principalement à deux types de produits de recombinaison homologue, soit la conversion génique ou le crossing-over. Nous avons schématisé un exemple d'événement de recombinaison homologue entre le L1 C et le L1 B (figure 10B), p.35). Le produit de conversion génique résulte de l'échange non réciproque de séquences d'ADN puisque le L1 donneur (L1 B) demeure inchangé. Donc le L1 receveur (L1 C) reçoit une copie identique de la séquence du L1 donneur (L1 B), c'est-à-dire qu'il porte les bases caractéristiques du donneur pour cette région de conversion. Par contre, le produit de crossing-over résulte de l'échange réciproque de séquences d'ADN puisque

Figure 11 : Alignement de la région de 300 pb sélectionnée des trois séquences L1 A B et C retrouvées dans Gene Bank.

Nous présentons les variations de bases nucléiques (en gris) caractérisant chacun des trois éléments L1. Les chiffres au dessus des séquences réfèrent aux positions nucléiques où furent identifiées les variations de bases nucléiques donnant lieu à différents types d'allèles pour chacun des L1, tel que mentionné dans les tableaux des résultats I, III et V.

		5	20	31	38	40	46	50	53
L1 A	ACAAAGAACTCAAAACAAAATTTTACAAGAAAAAAAACAAACAAACCCCATCAACAAGTGGG								
L1 B	ACAAATGAACTCAAAACAAAATTTTACAAGAAAAAAAACAAACAAACCCCATCAACAAGTGGG								
L1 C	ACAAATGAACTCAAAACAAAATTTTACAAGAAAAAAAACAAAGAAATCCCATCAACAAGTGGG								
	58	75	113						
L1 A	CGAAAAGATATGAACAGACAGTCTCAAAAAGAAAGACATTTATGGCAGCCCAACAGACACA								
L1 B	TGAAGGATATGAACAGACACACTTCTCAAAAAGAAAGACATTTATG-CAGCCAAAAGACACG								
L1 C	TGAAGGATATGAACAGACACACTTCTCAAAAAGAAAGACATTTATG-CAGCCAAAAGACACG								
	125	166							
L1 A	TGAAAAAAAT-GCTCATCATCAATGGCCATCAGAGAAAAATGCCAAATCAAAACCAACAATG								
L1 B	TGAAAAAAACGCTCATCATCACTGGCCATCAGAGAAA-TGCCAAAACAAATCCGCAATG								
L1 C	TGAAAAAAACGCTCATCATCACTGGCCATCAGAGAAA-TGCCAAATCAAAACCAACAATG								
L1 A	AGATACAAATCTCACACCAGCTAGAAATGGCGATCAGTAAAAAAGTCAGGAAAACAACAGGT								
L1 B	AGATACCATCTCTCACCCAGTTAGAATGGCAATCATTAAAAAAGTCAGGAAAACAACAGGT								
L1 C	AGATACCATCTCACACCAGTTAGAATGGTGAATCATTAAAAAAGTCAGGAAAACAACAGGT								
	239	267							
L1 A	GCTGGAGAGGATGTGGAGAAAACAGGAACAATTTTACACTGTTGCTGGGACTGTAAACT								
L1 B	GCTGGAGAGGATGTGGAGAAAATAGGAACAATTTTACACTGTTGCTGGGACTGTAAACT								
L1 C	GCTGGAGAGGATGTGGAGAAAATAGGAACAATTTTACACTGTTGCTGGGACTGTAAACT								
	294								
L1 A	AGTTCAACCA								
L1 B	AGTTCAACCA								
L1 C	ACTTCAACCA								

Figure 11

ce ne sont plus les mêmes séquences qui se retrouvent associées. Dans le cas que nous présentons à la figure 10B), p.35 la recombinaison homologue avec crossing-over entre deux séquences L1 (B et C), situées sur le même chromosome, entraîne l'association d'une partie de la séquence du L1 C et de celle du L1 B et aussi la délétion des séquences entre les deux éléments L1 incluant l'exon 5 (E5) du gène FHIT. Ce dernier produit entraîne donc un réarrangement génomique extrême pouvant mettre en jeu l'intégrité du génome.

Pour réaliser ce projet, nous devions nous assurer d'amplifier spécifiquement ces trois éléments L1, c'est pourquoi nous avons choisi des amorces situées dans les régions flanquantes uniques dans le génome (figure 10B), p.35). De plus, ces amorces ne devaient pas former de dimères entre elles afin de ne pas nuire à l'amplification. Ces amorces (décrites dans la section II.3, matériel et méthodes) ont été testées pour chacun des éléments L1 (A, B et C). La mise au point des conditions d'amplification (voir matériel et méthodes) nous a permis d'obtenir les fragments attendus (voir plus loin). Suite à l'obtention des fragments d'amplification des trois éléments L1, nous avons testé avec le L1 A, une première stratégie de clonage présentée à la figure 12, p.41 qui consistait à ajouter à l'extrémité 5' de chacune des deux amorces, un site de restriction (une amorce contenant la séquence reconnue par l'enzyme de restriction SacII et l'autre amorce contenant celle reconnue par HindIII). Ceci aurait permis de cloner le L1 complet et ses régions flanquantes uniques, lesquelles auraient pu servir pour des études subséquentes, puisque dans cette méthode les sites de clonage étaient situés aux extrémités du produit d'amplification.

Figure 12 : Illustration de la première méthode de clonage testée pour faire l'analyse des séquences L1.

L'élément L1 A (flèche pleine) est amplifié à l'aide des amorces PCR (flèches sens et anti-sens) portant chacune un site de restriction à leur extrémité (sacII et HindIII). Le produit d'amplification L1 A(S/H) est digéré par les enzymes de restriction sacII et HindIII. Ce produit de digestion est ensuite cloné dans le plasmide pBSK préalablement digéré aux sites sacII et HindIII contenus dans la région de clonage multiple (MCS). Les chiffres entre parenthèses réfèrent à la position en pb des sites de restriction dans le vecteur.

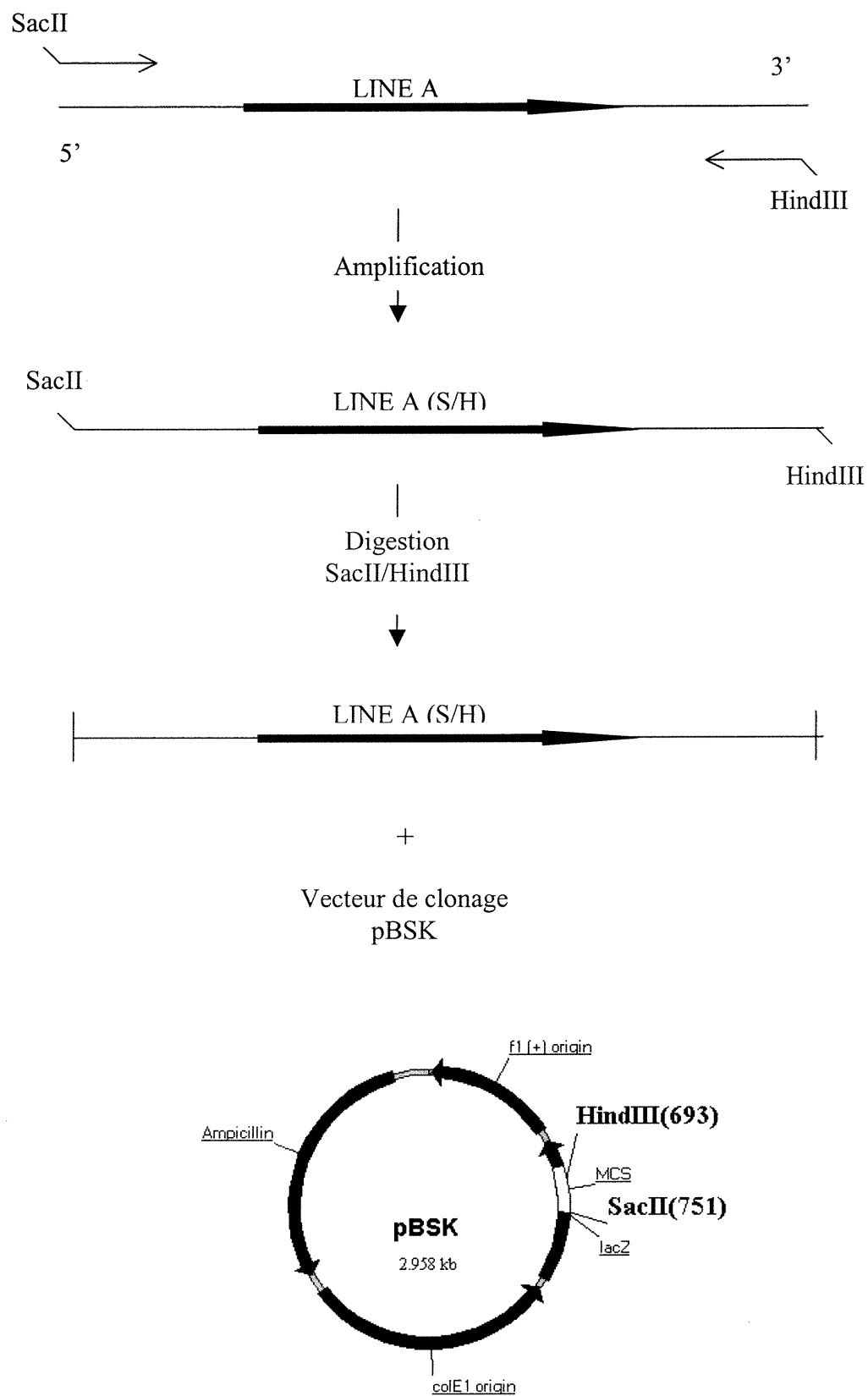


Figure 12

Cependant, tous les tentatives de ligation du fragment d'ADN amplifié dans le vecteur de clonage pBluescript (pBSK) ont échoué. Nous avons donc voulu savoir quelle étape du clonage ne fonctionnait pas. Premièrement, nous avons vérifié si c'était le vecteur qui était la cause de nos problèmes de clonage. Pour ce faire, il nous fallait un autre fragment d'ADN contenant les sites de restriction SacII/HindIII qui allait nous servir d'insert. Nous avons digéré le vecteur p3'SS par les enzymes de restriction SacII/HindIII, tel qu'illustré à la figure 13, p.44 , et le fragment libéré devenait l'insert de notre essai. La ligation de ce fragment aux sites SacII/HindIII du vecteur pBSK a très bien fonctionné. Nous avons alors émis l'hypothèse que le problème de clonage se situait au niveau de la digestion des extrémités des produits d'amplification. Nous avons donc voulu vérifier l'efficacité de digestion aux extrémités des produits PCR, tel qu'illustré à la figure 14, p.46 . Pour ce faire nous avons marqué au $\gamma^{32}\text{P}$ ATP les extrémités 5' du fragment de L1 A amplifié à l'aide des amorces contenant les sites de restriction. Cette méthode était donc basée sur le fait que la digestion des extrémités, en vue du clonage, entraînerait la perte des phosphates marqués, c'est-à-dire une baisse du signal radioactif. Le L1 A était amplifié d'une part avec deux amorces portant chacune le site de restriction SacII, menant au produit PCR L1 A(S), et d'autre part avec deux amorces dont l'une contenait le site SacII et l'autre contenait le site HindIII, menant au produit PCR L1 A(S/H). Suite au marquage des extrémités 5' des produit PCR, nous avons effectué trois digestions différentes (SacII, HindIII et SacII/HindIII) du produit PCR L1 A (S/H) et une digestion simple (SacII) du produit PCR L1 A (S). Au bas de la figure 14, p.46, nous présentons les résultats obtenus suite au marquage des extrémités du

Figure 13: Test de ligation pour vérifier l'étape de ligation de la première méthode de clonage.

Un insert provenant de la digestion SacII et HindIII du vecteur p3'SS est cloné dans le plasmide pBSK aux sites SacII et HindIII. Les chiffres entre parenthèses réfèrent à la position en pb des sites de restriction SacII et HindII dans chacun des vecteurs.

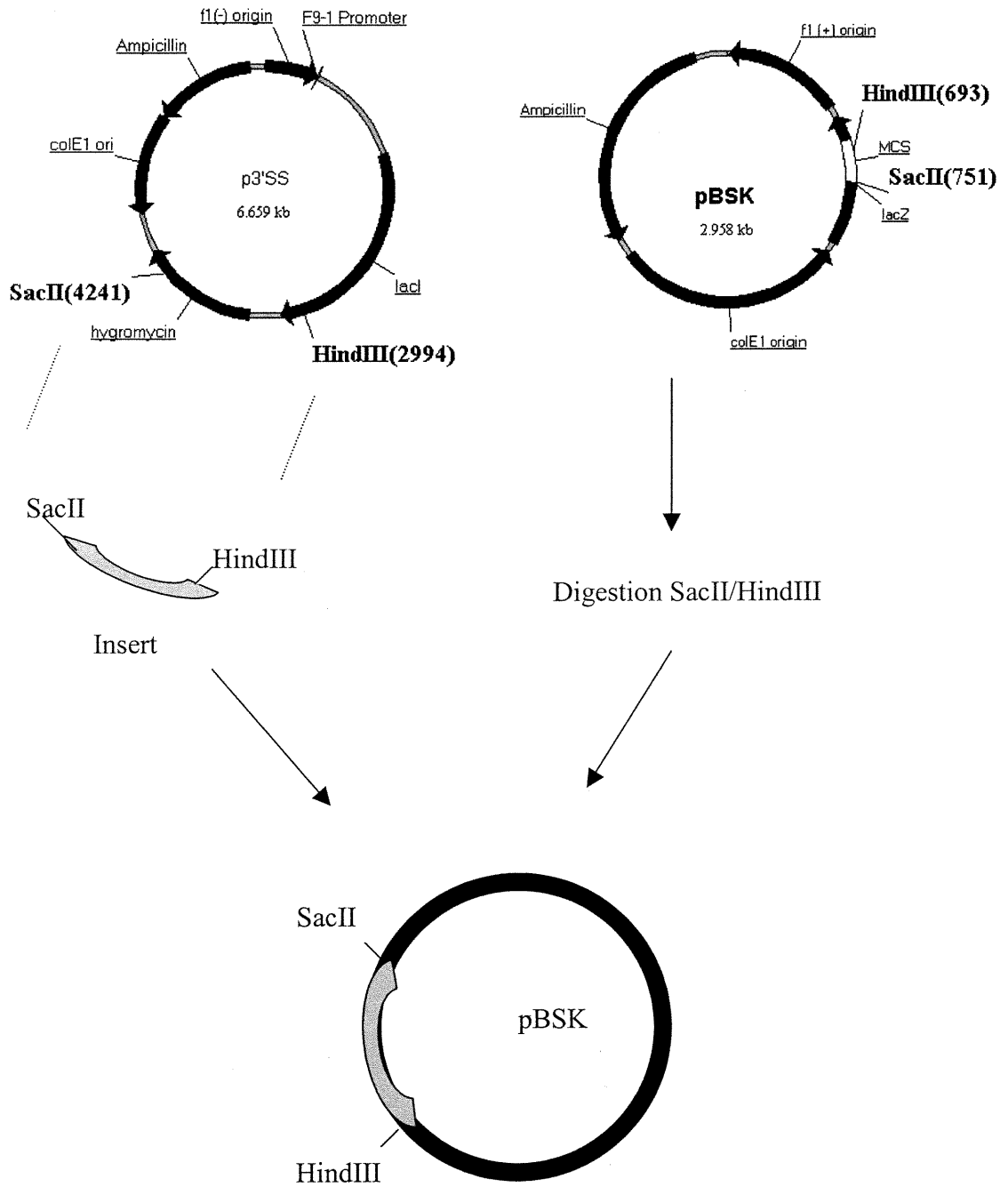


Figure 13

Figure 14: Test pour vérifier l'étape de digestion aux extrémités de la première méthode de clonage.

L'élément L1 A (flèche pleine) est amplifié d'une part à l'aide d'amorces (flèche sens et anti-sens) portant, à leur extrémité 5' les sites de restriction SacII et HindIII et donnant lieu au produit d'amplification LINE A(S/H), et d'autre part à l'aide d'amorces contenant le site de restriction SacII donnant lieu au produit LINE A(S). Chacun des produits d'amplification sont marqués au γ^{32} ATP aux extrémités 5'. Les différentes digestions (SacII, HindIII et SacII/HindIII) de ces produits sont analysées sur un synergel qui est par la suite exposé toute la nuit. Nous présentons aussi un contrôle (C) non-digéré, pour chacun des produits PCR.

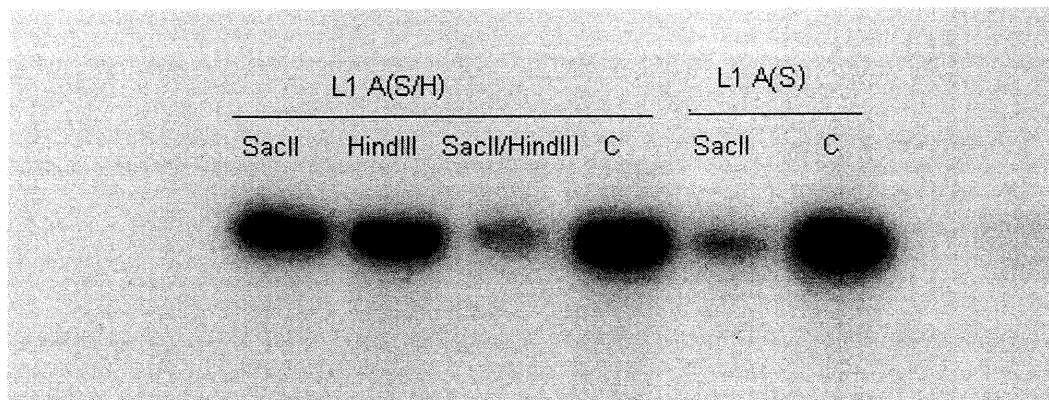
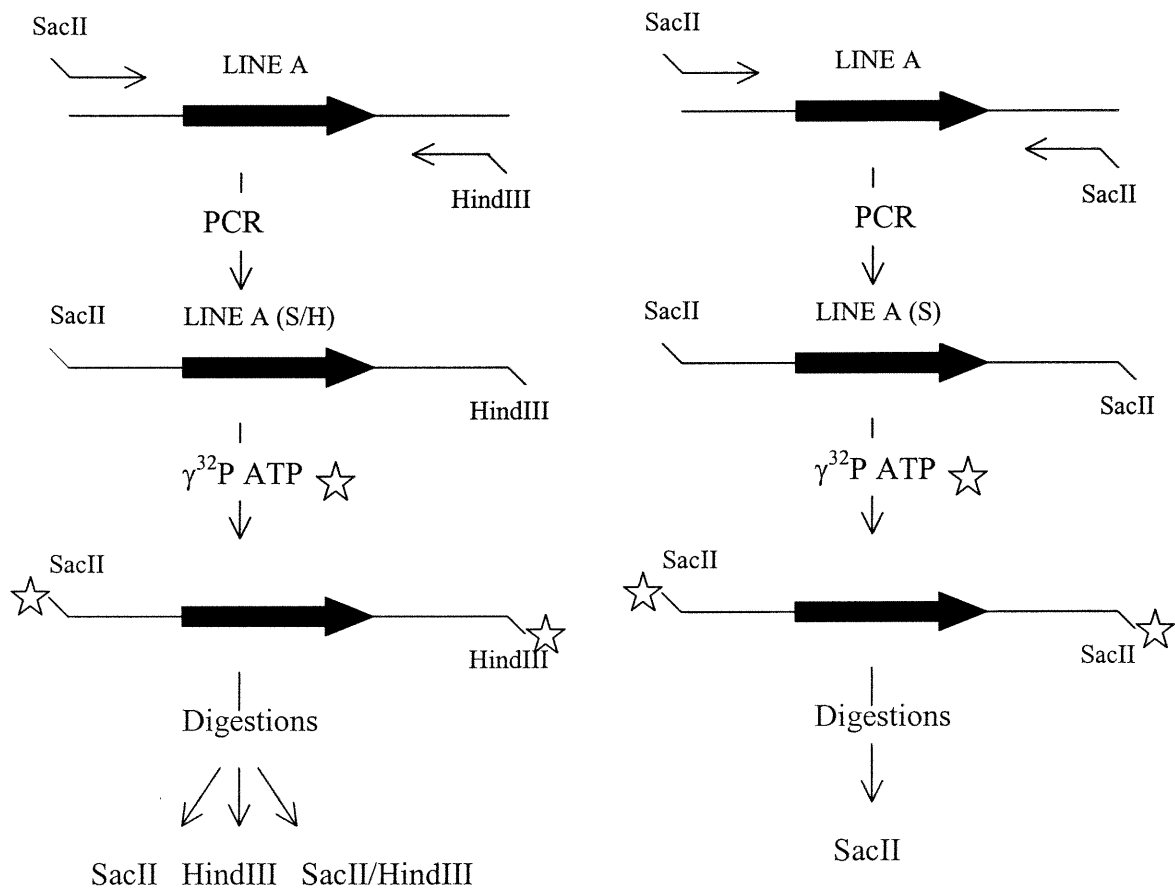


Figure 14

produit PCR L1 A, avant et après les tests de digestion. Nous pouvons observé une baisse du signal radioactif en suivant l'ordre des produits que voici : les produits PCR L1 A(S) ou L1 A(S/H) non digérés servant de contrôles, les produits PCR digérés à une seule extrémité soit L1 A(S/H) digéré par SacII ou HindIII et enfin les produits PCR digérés aux deux extrémités soit L1 A(S) digéré par SacII et L1 A(S/H) digéré par SacII/HindIII. Ces résultats montrent que la digestion aux extrémités fonctionne bien mais pas totalement. Ces deux analyses, soit celle portant sur le vecteur pBSK et celle portant sur la digestion aux extrémités des produits PCR, semblent nous indiquer que le clonage devrait être possible. Cependant, pour des raisons que nous ne pouvions nous expliquer, cette stratégie de clonage ne semblait pas appropriée pour notre étude. C'est pourquoi nous avons utilisé une autre méthode de clonage.

Cette nouvelle stratégie consistait plutôt à digérer les produits d'amplification par des enzymes dont les sites de restriction se retrouvaient à l'intérieur du produit PCR. Les sites de clonage ont été sélectionnés à partir d'une carte de restriction établie pour chacune des trois séquences L1 et leurs régions flanquantes. Nous avons conservé les mêmes séquences pour les amorces PCR en éliminant seulement les extrémités contenant le site de restriction. Afin de ne pas alourdir ce travail, nous présenterons seulement les résultats obtenus pour un seul individu, pour montrer les différentes étapes de clonage de chacun des éléments L1 A, B et C. Veuillez noter que les résultats obtenus sont les mêmes pour les 10 individus, et ce pour l'ADN normal et tumoral.

Les figures 15, 16 et 17 présentent les schémas des résultats attendus et les photos montrant un exemple de résultats obtenus aux différentes étapes de clonage pour chacun des éléments L1. Il est à noter que nous avons utilisé deux types de clonage différents, soit en utilisant le vecteur de clonage p3'SS qui nécessitait la digestion du produit PCR pour les L1 B et C, et le vecteur PCR2.1 qui permettait le clonage direct du produit PCR pour le L1 A. L'utilisation du vecteur PCR2.1 a été nécessaire pour le clonage du L1 A, puisque nous n'avons pas réussi à trouver des sites de restrictions compatibles avec les vecteurs de clonage disponibles dans notre laboratoire. Pour ce clonage, décrit à la section matériel et méthodes, le produit PCR a été mis en présence de l'enzyme polymérase Amplitaq pour ajouter des nucléotides A aux extrémités 3' du produit. Pour sa part, le vecteur topo PCR2.1 porte à ses extrémités 3' des nucléotides T. C'est l'enzyme topoisomérase liée de façon covalente aux extrémités 3' du vecteur qui permet la réaction de ligation entre le vecteur et le produit d'amplification. Le groupe hydroxyle (OH) du produit PCR attaque le lien phosphate existant entre l'enzyme et le vecteur, ce qui permet la libération de l'enzyme et la ligation du produit PCR. De plus, cette méthode est beaucoup plus rapide et très efficace.

Tout d'abord, à la figure 15, p.50 nous présentons les résultats attendus et obtenus aux différentes étapes de clonage du L1 A. En A) nous présentons le résultat de l'amplification spécifique du L1 A et de ses régions flanquantes chez l'individu 244 au poids attendu de 3342 pb. Le contrôle négatif représente une amplification sans ADN et le contrôle positif a été effectué à l'aide d'ADN génomique provenant

Figure 15: Schéma des résultats attendus et présentation des résultats obtenus aux différentes étapes de clonage du L1 A chez l'individu 244.

A) Amplification du L1 A (flèche pleine) et de ses séquences flanquantes (ligne mince). Le contrôle positif (C+) provient de l'ADN d'un individu témoin utilisé pour la mise au point des conditions d'amplification et le contrôle négatif (C-) est sans ADN. 1 Kb est l'échelle standard de poids moléculaire. B) Présentation des cinq clones (1, 2, 4, 5 et 6) choisis suite au clonage du produit d'amplification dans le vecteur PCR2.1. Circ. représente l'échelle standard de poids moléculaire d'ADN circulaire. C) Présentation des résultats des digestions caractéristiques (bandes caractéristiques pointées par les flèches noires à droite de la photo), selon l'orientation de l'insert, des cinq clones choisis. Les enzymes de restrictions utilisées sont MfeI et XhoI. Le contrôle positif (C+) représente la digestion du produit d'amplification par l'enzyme MfeI qui produit deux fragments d'ADN, un de 2600 pb et l'autre de 800 pb.

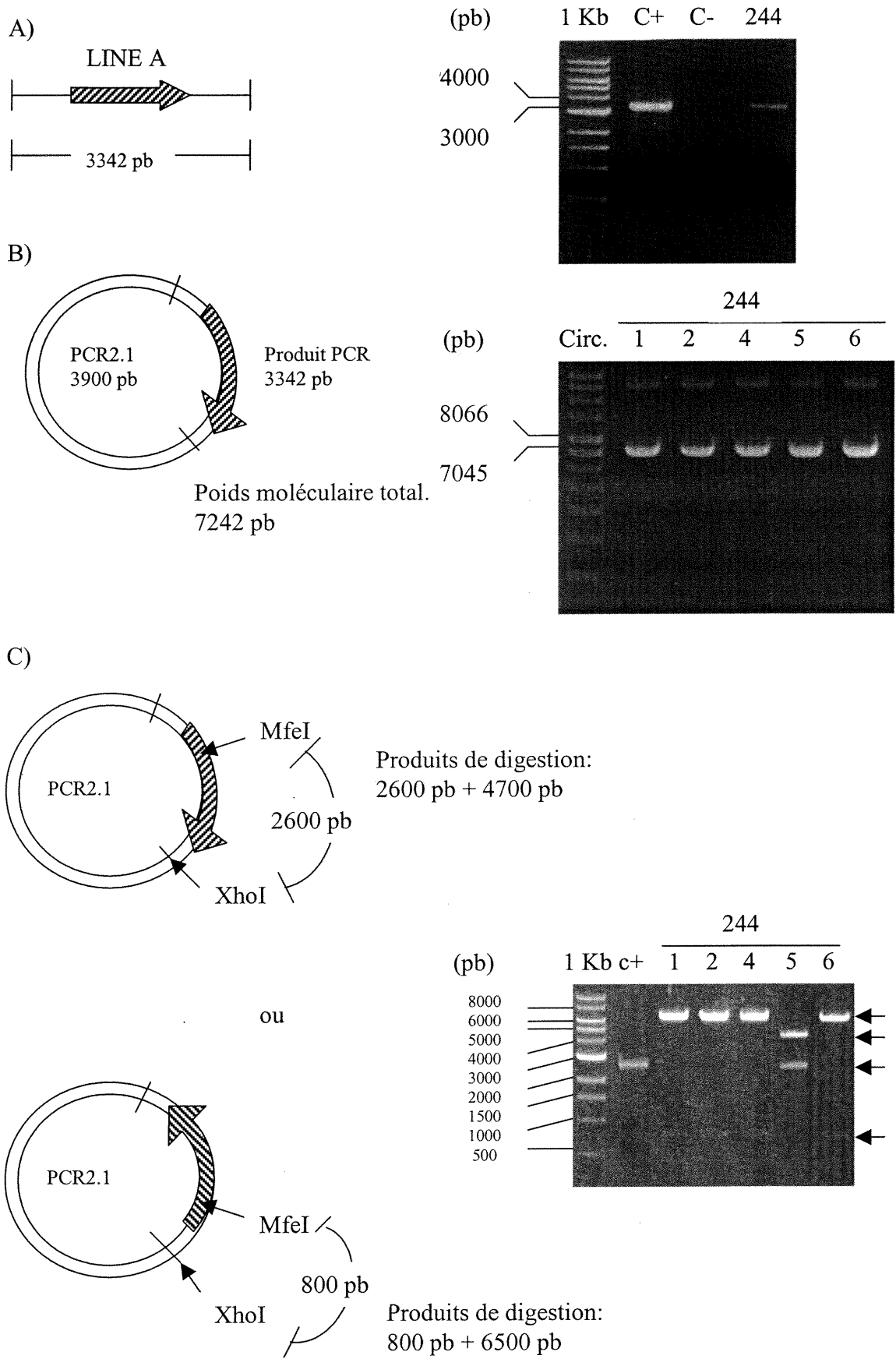


Figure 15

d'un individu témoin nous ayant servi pour mettre au point nos conditions d'amplification pour chacun des éléments L1 et pour la séquence intronique (voir plus loin). Le clonage direct du produit d'amplification du L1 A dans le vecteur PCR2.1 nous donnait un produit attendu de 7242 pb (circulaire), ce qui était le cas pour les 5 clones retenus présentés en B). Par la suite, les digestions caractéristiques de ces clones nous ont confirmé que le clonage du L1 A avait bien fonctionné. Nous présentons donc en C) les résultats des digestions des 5 clones obtenus pour l'individu 244. L'enzyme MfeI coupe une seule fois dans le fragment de L1 A et l'enzyme XhoI digère le plasmide PCR2.1 près du site de clonage de l'insert. La digestion MfeI de l'insert amplifié nous a permis de confirmer la digestion caractéristique. De plus, nous retrouvions deux types de produits de digestion, aux poids moléculaires attendus de 2600 pb + 4700 pb et 800 pb + 6500 pb, puisque l'insert pouvait se cloner dans les deux sens tel qu'illustré en C).

À la figure 16, p.53 nous présentons les résultats attendus et obtenus aux différentes étapes de clonage du L1 B. En A) nous présentons les résultats d'amplification du L1 B et de ses régions flanquantes chez l'individu 244 au poids attendu de 4173 pb. La digestion de ce produit d'amplification, par les enzymes de clonage choisies, soit KpnI et NdeI, libérait un insert de 1301 pb présenté en B). Cet insert, une fois isolé, a été cloné dans le vecteur p3'SS, préalablement digéré aux sites de clonage KpnI/NdeI, nous donnant un produit de clonage de 4673 pb dont les 5 clones choisis sont présentés en C). Les digestions caractéristiques de ces 5 clones nous ont confirmé que le clonage du fragment de L1 B avait bien fonctionné.

Figure 16 : Schéma des résultats attendus et présentation des résultats obtenus aux différentes étapes de clonage du L1 B chez l'individu 244.

A) Amplification du L1 B (flèche pleine) et de ses séquences flanquantes (ligne mince). Le contrôle positif (C+) provient de l'ADN utilisé pour la mise au point du PCR et C- est le contrôle négatif sans ADN. 1 Kb est l'échelle standard de poids moléculaire. B) Digestion du produit d'amplification par les enzymes de clonage KpnI et NdeI. Isolation de la bande d'ADN d'intérêt montrée par une flèche. C) Présentation des cinq clones (2, 3, 4, 6 et 7) choisis suite au clonage du produit de digestion dans le vecteur p3'SS préalablement linéarisé par KpnI/NdeI. Circ. représente l'échelle standard de poids moléculaire d'ADN circulaire. D) Présentation des résultats obtenus des digestions caractéristiques des cinq clones choisis (bandes caractéristiques pointées par les flèches noires à droite de la photo). Les enzymes de restrictions utilisées sont SpeI et KpnI.

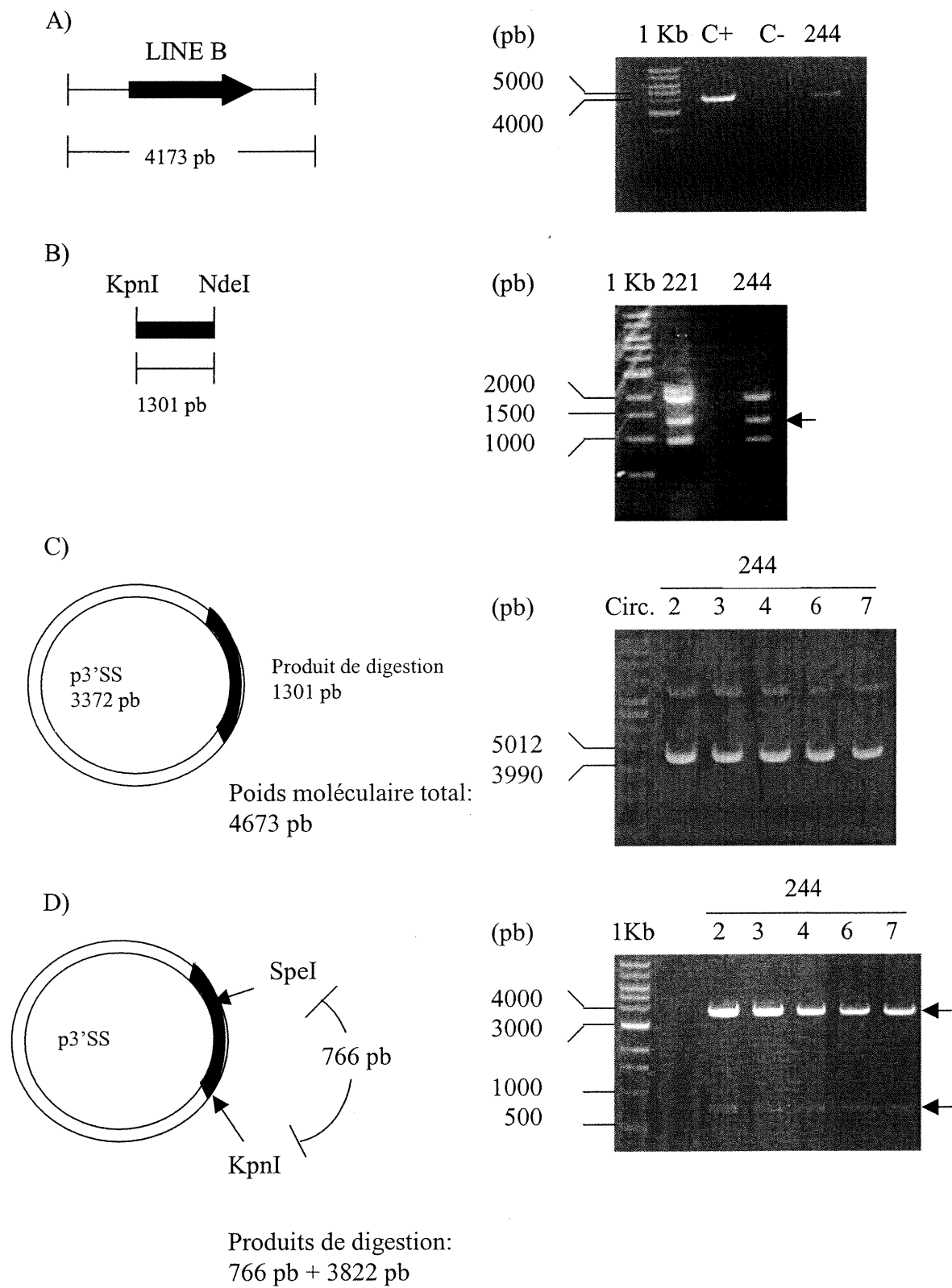


Figure 16

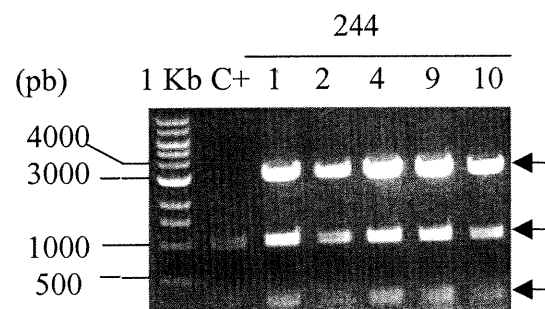
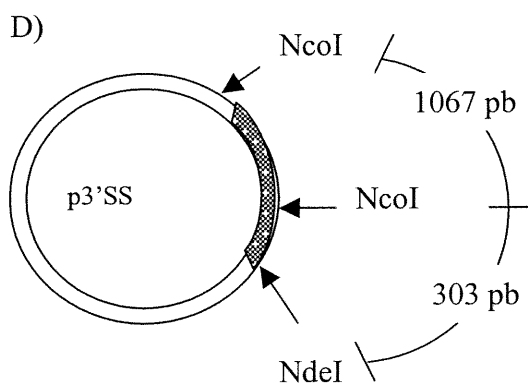
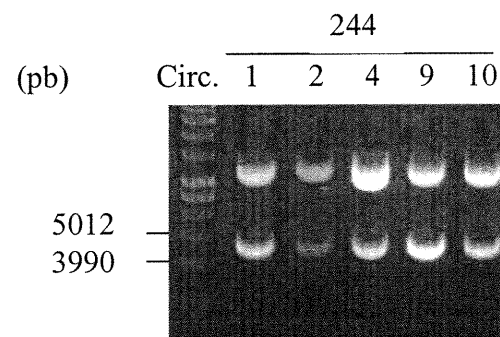
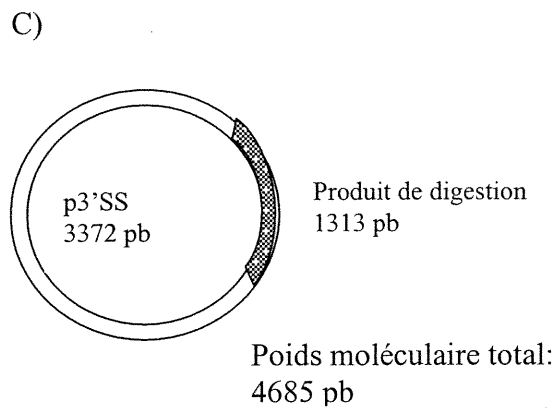
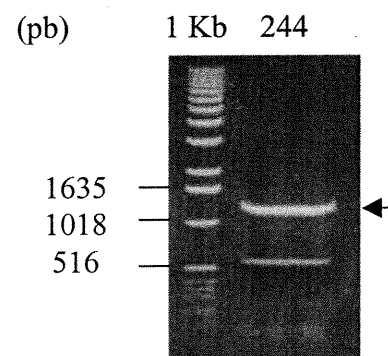
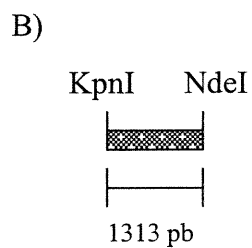
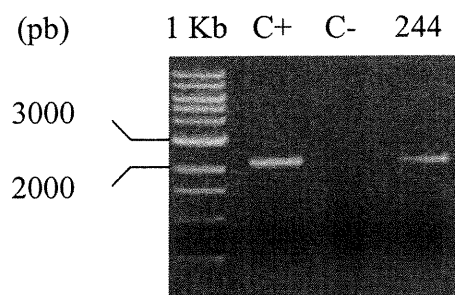
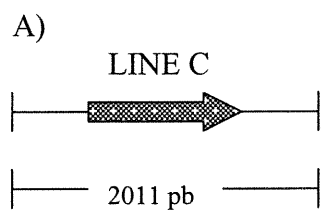
En effet, les résultats des digestions de ces clones, présentés en D), montrent bien le poids des bandes attendues, soit 766 pb + 3822 pb. L'enzyme de restriction KpnI correspond à un des sites de clonage du vecteur p3'SS et l'enzyme SpeI possède un site unique de digestion au niveau du fragment de L1 B. Il n'y avait qu'une seule orientation possible de l'insertion, puisque le clonage de l'insert était dirigé dans un sens par les deux sites de clonages différents.

Enfin, le L1 C a été cloné exactement de la même façon que le L1 B. Les résultats des différentes étapes de clonage sont présentés à la figure 17, p.56. En A) nous présentons le résultat de l'amplification du L1 C chez l'individu 244 qui correspond bien au produit attendu soit 2011 pb. La digestion du produit d'amplification par les enzymes de clonage KpnI et NdeI libère un insert de 1313 pb présenté en B). Le clonage de cet insert dans le vecteur p3'SS nous a permis d'obtenir des clones de 4685 pb parmi lesquels nous en avons choisi 5 que nous présentons en C). Ces clones ont été analysés par les enzymes de restriction NdeI, un site de clonage du vecteur p3'SS, et NcoI digérant à la fois le fragment de L1 C et le vecteur p3'SS. Ces résultats de digestions caractéristiques présentés en D), aux poids moléculaires attendus de 303 pb + 1067 pb + 3200 pb, ont confirmé le clonage du fragment L1 C.

Il est à noter que l'amplification de chacun des trois L1 (A, B et C) est spécifique, c'est-à-dire que l'on n'observe aucune bande d'ADN résultant d'amplification non-spécifique.

Figure 17. Schéma des résultats attendus et présentation des résultats obtenus aux différentes étapes de clonage du L1 C chez l'individu 244.

A) Amplification du L1 C (flèche pleine) et de ses séquences flanquantes (ligne mince). Le contrôle positif (C+) provient de l'ADN utilisé pour la mise au point du PCR et C- est le contrôle négatif sans ADN. 1 Kb est l'échelle standard de poids moléculaire. B) Digestion du produit d'amplification par les enzymes de clonage KpnI et NdeI. Isolation de la bande d'ADN d'intérêt montrée par une flèche. C) Présentation des cinq clones (1, 2, 4, 9 et 10) choisis suite au clonage du produit de digestion dans le vecteur p3'SS préalablement linéarisé par KpnI/NdeI. Circ. représente l'échelle standard de poids moléculaire d'ADN circulaire. D) Présentation des résultats obtenus des digestions caractéristiques des cinq clones choisis. Les enzymes de restrictions utilisées sont NcoI et NdeI. Le contrôle positif (C+) est représenté par la digestion de la bande d'ADN d'intérêt (1313 pb) par l'enzyme NcoI qui produit deux fragments d'ADN, un de 1010 pb et l'autre de 303 pb.



Produits de digestion:
303 pb + 1067 pb + 3200 pb

Figure 17

Par la suite, nous avons effectué le séquençage de la région de 300 pb prédéterminée, et ce pour chacun des 5 clones choisis pour chacun des dix individus (tissu sain) ainsi que pour les trois éléments L1 (A, B et C). À la figure 18, p.59 nous présentons comme exemple de lecture d'une séquence la région de 300 pb du L1 A, représentant l'haplotype 1 chez l'individu 367. Cependant, dans certains cas où la fin de la lecture de la séquence n'était pas claire, nous devions utiliser une amorce en 3' de la région de 300 pb afin de compléter la lecture.

Figure 18 : Lecture d'une séquence de la région de 300 pb d'un L1

Présentation d'un exemple de lecture d'une séquence de la région de 300 pb d'un élément L1 (L1 A allèle 1, amplifié cloné et séquencé chez l'individu 367). Les bases sont lues dans l'ordre ACGT et du côté 5' vers le côté 3' de la séquence.

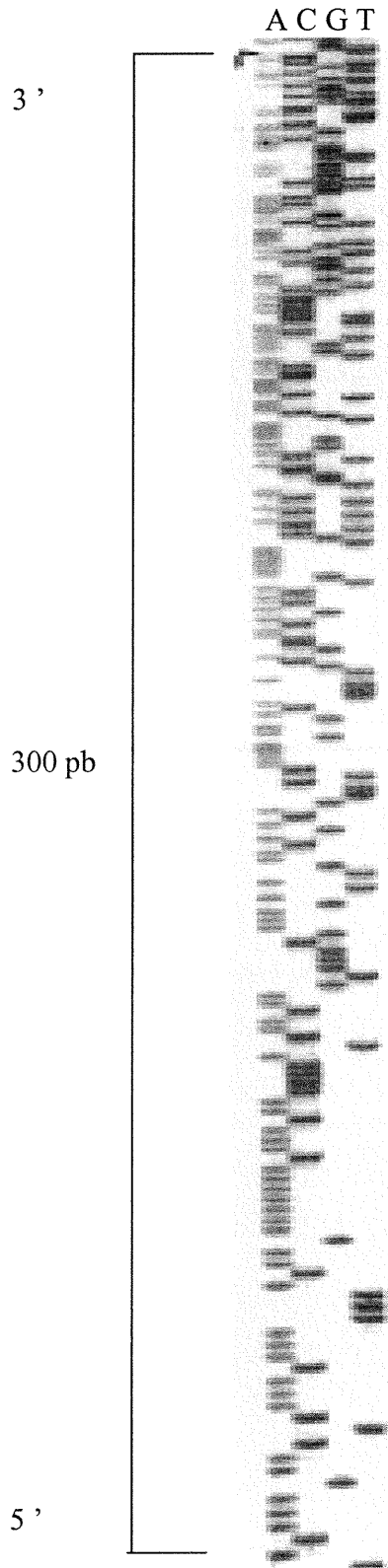


Figure 18.

III.2 Analyse de l'évolution d'un élément LINE-1 donné chez l'humain

Nous avons tout d'abord obtenu la séquence de la région de 300pb des trois éléments L1 sélectionnés, et ce pour les dix individus au niveau du tissu sain. Le séquençage de cinq clones par individu, pour chacun des L1, nous a permis de déduire les deux haplotypes (désignant les variations de bases pouvant exister entre des individus pour un locus donné) présents chez chaque individu. Par la suite, l'alignement des 20 haplotypes (2 haplotypes x 10 individus) obtenus pour chacun des L1 nous a permis de les regrouper en différents types d'haplotypes selon les différences de bases nucléiques observées (tableaux I, III, et V). Il est à noter que nous avons retrouvé les mêmes types d'haplotypes pour chaque individu au niveau de l'ADN normal et tumoral (voir section III.4), ce qui correspond à deux séries de clonages indépendants. Ceci élimine la possibilité que les variations observées soient la conséquence d'erreurs provenant de la méthode d'amplification (PCR).

Le tableau I, p.62 présente les deux types d'haplotypes observés pour le L1 A et ils diffèrent par le changement d'une seule base nucléique en position 5. On retrouve 18/20 haplotypes de type 1 et chacun des individus porte au moins un haplotype de ce type. Seulement deux individus portent un haplotype de type 2 (tableau II, p.62). En considérant que l'haplotype le plus abondant dans la population est le plus anciens, l'haplotype de type 1 pour cet élément L1 semble être le fondateur

Tableau I. Analyse comparative des différents types d'haplotypes (haplo.), basée sur les variations de bases nucléiques obtenues suite au séquençage d'une région de 300pb du LINE-1 A au niveau de l'ADN normal chez 10 individus atteints du cancer du colon, ce qui représente 20 haplotypes. La position nucléique réfère à la position de la variation observée dans la région de 300 pb séquencée. Les nombres entre parenthèses indiquent l'abondance relative de chacun des types d'haplotypes. Les variations (indiquées par la base appropriée) sont montrées par rapport à l'haplotype le plus fréquent (haplo. 1). Les séquences consensus de L1 actifs en rétrotransposition et inactifs (petits fragments de L1) furent déterminées à partir d'un alignement de 13 éléments L1 actifs et 13 éléments L1 inactifs retrouvés dans Gene Bank.

POSITION NUCLÉIQUE	5
L1 actif	T
L1 inactif	T
Haplo. 1 (18/20)	A
Haplo. 2 (2/20)	T

Tableau II. Distribution des différents types d'haplotypes du L1 A retrouvés pour chacun des 10 individus. Nous présentons entre parenthèses la provenance ethnique de chacun des individus, Canadien Français (C.F.), Canadien Italien (C.I.) et Haïtien (H). Le chiffre identifiant chacun des individus réfère au numéro attribué à l'échantillon d'ADN reçu.

INDIVIDU	TYPES D'HAPLO. ADN NORMAL	TYPES D'HAPLO. ADN TUMORAL
221 (?)	1 / 1	1 / 1
244 (C.F.)	1 / 2	1 / 2
262 (C.I.)	1 / 1	1 / 1
264 (C.F.)	1 / 1	1 / 1
266 (H)	1 / 2	1 / 2
367 (C.F.)	1 / 1	1 / 1
370 (C.F.)	1 / 1	1 / 1
374 (C.F.)	1 / 1	1 / 1
377 (C.F.)	1 / 1	1 / 1
380 (C.F.)	1 / 1	1 / 1

et l'haplotype de type 2 serait considéré comme une variation de cet élément survenue au cours de l'évolution. Cependant, il est à noter que l'haplotype de type 1 est différent du consensus de L1 actifs et inactifs qui s'apparente plutôt au type 2.

Contrairement au L1 A, le L1 B présente plusieurs variations de bases soit sept types d'haplotypes (tableau III, p.64). Chacun des types d'haplotypes porte un à trois changements de bases nucléiques et les positions nucléiques de ces changements sont dispersés à travers la région de 300pb étudiée. Ces changements semblent donc être le résultat de mutations ponctuelles s'étant accumulées au cours de l'évolution pour donner naissance à différents types d'haplotypes. On remarque aussi la fréquence élevée (13/20) de l'haplotype de type 1 et sa présence chez 9/10 individus (tableau IV, p.64), ce qui semble indiquer qu'il est l'haplotype fondateur. Les haplotypes de type 2, 3, 5, 6 et 7 représentent seulement 1/20 haplotypes, ils sont donc chacun portés par un seul individu. Pour sa part, l'haplotype 4 diffère des six autres haplotypes, mais on le retrouve en deux copies chez le même individu. Ce qui semble intéressant, c'est que nous pouvons faire des liens entre les différentes variations de bases observées, qui semblent toutes provenir de l'haplotype de type 1. Par exemple, l'haplotype de type 4 pourrait être une variation du type 1 par le changement du G en A à la position 125. Par la suite, une modification du G en A en position 239 au niveau de l'haplotype 4 aurait pu donner naissance au type 6, et ce dernier aurait pu subir le changement du G en C en position 40 pour donner naissance au type 3. On peut aussi voir le cheminement autrement en partant toujours du fait que l'haplotype 4 est une variation du type 1. Cet haplotype 4 pourrait plutôt avoir

Tableau III. Analyse comparative des différents types d'haplotypes (haplo.), basée sur les variations de bases nucléiques obtenues suite au séquençage d'une région de 300pb du LINE-1 B au niveau de l'ADN normal chez 10 individus atteints du cancer du colon, ce qui représente 20 haplotypes. Les positions nucléiques correspondent aux positions des variations nucléiques observées dans la région de 300 pb. Les nombres entre parenthèses indiquent l'abondance relative de chacun des types d'haplotypes. Les variations (indiquées par la base appropriée) sont montrées par rapport à l'haplotype le plus fréquent (haplo. 1). Le trait (-) indique qu'il n'y a pas de changement de base à cette position. Les séquences consensus de L1 actifs en rétrotransposition et inactifs (petits fragments de L1) furent déterminées à partir d'un alignement de 13 éléments L1 actifs et 13 éléments L1 inactifs retrouvés dans Gene Bank.

POSITION NUCLÉIQUE	40	125	166	239	267
L-1 actif	A	G	A	G	T
L-1 inactif	A	G	A	G	T
Haplo. 1 (13/20)	G	G	G	G	T
Haplo. 2 (1/20)	C	A	-	-	-
Haplo. 3 (1/20)	C	A	-	A	-
Haplo. 4 (2*/20)	-	A	-	-	-
Haplo. 5 (1/20)	-	A	-	-	G
Haplo. 6 (1/20)	-	A	-	A	-
Haplo. 7 (1/20)	-	-	T	-	-

*C'est le même individu qui porte ces deux allèles

Tableau IV. Distribution des différents types d'haplotypes du L1 B retrouvés pour chacun des 10 individus. Nous présentons entre parenthèses la provenance ethnique de chacun des individus, Canadien Français (C.F.), Canadien Italien (C.I.) et Haïtien (H). Le chiffre identifiant chacun des individus réfère au numéro attribué à l'échantillon d'ADN reçu.

INDIVIDU	TYPES D'HAPLO. ADN NORMAL	TYPES D'HAPLO. ADN TUMORAL
221 (?)	1 / 1	1 / 1
244 (C.F.)	4 / 4	4 / 4
262 (C.I.)	1 / 5	1 / 5
264 (C.F.)	1 / 1	1 / 1
266 (H)	1 / 3	1 / 3
367 (C.F.)	1 / 6	1 / 6
370 (C.F.)	1 / 1	1 / 1
374 (C.F.)	1 / 2	1 / 2
377 (C.F.)	1 / 1	1 / 1
380 (C.F.)	1 / 7	1 / 7

subit le changement du G en C en position 40 pour faire apparaître le type 2, qui lui-même aurait donné naissance au type 3 par le changement du G en A en position 239. Pour sa part, l'haplotype 5 pourrait être une variation du type 4 par le changement du T en G en position 267. Enfin, l'haplotype 7, suite au changement du G en T à la position 166, pourrait être une variation du type 1. Il est à noter encore que l'haplotype 1 diffère du consensus de L1 actifs et inactifs et que cette fois aucun haplotype ne correspond aux consensus.

En ce qui concerne le troisième L1, soit le L1 C, nous avons obtenu différents haplotypes, soit 5 types (Tableau V, p.66). L'haplotype de type 1 est l'haplotype dominant (10/20 haplotypes) et on le retrouve chez 8/10 individus (tableau VI, p.66), ce qui semble indiquer qu'il est le fondateur mais il diffère du consensus de L1 actifs et inactifs. Les haplotypes 2 et 3, dont la fréquence est de 1/20, portent deux variations par rapport au type 1. Il est à noter que l'haplotype 5 porte trois variations dans une région de huit bases, ce qui ne semble pas être attribuable à des variations aléatoires, et on le retrouve chez 4/10 individus (tableau VI, p.66). Nous avons analysé l'haplotype 5 en considérant que les trois changements de bases observés pourraient être le résultat d'une conversion génique puisque les trois variations sont concentrées dans une région de 8 pb. Nous avons regardé d'abord si ces trois variations existaient dans les éléments L1 A et B. Le L1 A porte un C et un G aux positions 50 et 53 respectivement (voir figure 11, p.38), tandis que le L1 B porte le G et le T aux positions 53 et 58. Nous avons ensuite soumis cette séquence de 8 pb,

Tableau V. Analyse comparative des différents types d'haplotypes (haplo.), basée sur les variations de bases nucléiques, obtenues suite au séquençage d'une région de 300pb du LINE-1 C au niveau de l'ADN normal chez 10 individus atteints du cancer du colon, ce qui représente 20 haplotypes. Les positions nucléiques correspondent aux positions des variations nucléiques observées dans la région de 300 pb. Les nombres entre parenthèses indiquent l'abondance relative de chacun des types d'haplotypes. Les variations (indiquées par la base appropriée) sont montrées par rapport à l'haplotype le plus fréquent (haplo. 1). Le trait (-) indique qu'il n'y a pas de changement de base à cette position. Le x indique l'absence de base à cette position. Les séquences consensus de L1 actifs en rétrotransposition et inactifs (petits fragments de L1) furent déterminées à partir d'un alignement de 13 éléments L1 actifs et 13 éléments L1 inactifs retrouvés dans Gene Bank.

Position nucléique	20	31	38	40	46	50	53	58	75	113	267	294
L-1 actif	T	x	C	A	T	A	G	C	C	C	T	C
L-1 inactif	T	x	C	A	T	A	G	C	C	C	T	C
Haplo. 1 (10/20)	T	A	G	A	T	A	C	C	C	C	T	C
Haplo. 2 (1/20)	-	-	-	-	-	-	-	-	T	T	-	-
Haplo. 3 (1/20)	-	-	-	-	-	-	-	-	-	-	A	T
Haplo. 4 (4/20)	C	x	C	C	G	-	G	T	-	-	-	-
Haplo. 5 (4/20)	-	-	-	-	-	C	G	T	-	-	-	-

Tableau VI. Distribution des différents types d'haplotypes du L1 C retrouvés pour chacun des 10 individus. Nous présentons entre parenthèses la provenance ethnique de chacun des individus, Canadien Français (C.F.), Canadien Italien (C.I.) et Haïtien (H). Le chiffre identifiant chacun des individus réfère au numéro attribué à l'échantillon d'ADN reçu.

INDIVIDU	TYPES D'HAPLO. ADN NORMAL	TYPES D'HAPLO. ADN TUMORAL
221 (?)	1/5	1/5
244 (C.F.)	1/1	1/1
262 (C.I.)	1/4	1/4
264 (C.F.)	1/5	1/5
266 (H)	2/5	2/5
367 (C.F.)	1/4	1/4
370 (C.F.)	3/4	3/4
374 (C.F.)	1/1	1/1
377 (C.F.)	1/5	1/5
380 (C.F.)	1/4	1/4

contenant les trois variations, à GenBank. Nous avons obtenu aucun résultat significatif puisque la zone de conversion était trop petite. Cependant, il est intéressant de noter que trois des quatre individus portant l'haplotype 5 (individus 221, 264, 377, voir tableau VI, p.66) sont aussi des porteurs homozygotes de l'haplotype 1 de B (tableau IV, p.64).

Pour sa part, l'haplotype de type 4 du L1 C, dont la fréquence est de 4/20 haplotypes, montre sept variations concentrées dans une région de 39 pb. Selon toute probabilité, ceci représente une région de conversion d'une longueur de 39 pb par rapport à l'haplotype 1. Afin d'identifier le ou les partenaires possibles de recombinaison homologue, nous avons d'abord comparé cette séquence de 39 pb aux séquences des haplotypes retrouvés pour les L1 A et B. Nous avons retrouvé deux haplotypes du L1 B (haplotype 2 et 3) portant la région de 39 pb identique à celle de l'haplotype 4 du L1 C. Nous retrouvons donc les sept bases caractéristiques de l'haplotype 4 du L1 C chez l'haplotype 2 et 3 du L1 B (voir plus loin pour la comparaison des séquences). Toutefois, nous nous sommes demandé si d'autres L1 de cette région (cette région contient 43 L1, voir figure 10 A)) ou du génome portaient ces bases caractéristiques. Nous avons alors soumis cette séquence de 39 pb à Gene Bank pour comparer cette séquence aux autres L1 du génome répertoriés dans cette banque génomique. Nous n'avons pas retrouvé d'éléments L1 possédant une région de 39 pb identiques à celles de l'haplotype 4 du L1 C. Cependant, nous avons repéré deux séquences qui présentaient 38/39 pb identiques, dont l'une correspondait à la séquence L 1 B de notre étude. En comparant cette région de 39 pb entre la

séquence L1 B et celle de l'haplotype 4 du L1 C, la base différente se trouvait en position 40. À cette position, on retrouve un C pour l'haplotype 4 du L1 C (voir tableau V, p.66), tandis que dans GenBank, la séquence L1 inconnue et la séquence L1 B portent un A à cette position. Comme nous l'avons mentionné plus haut, il est intéressant de noter que nous avons retrouvé le C en position 40 au niveau des haplotypes 2 et 3 du L1 B (voir tableau III, p.64) puisque nous avons retrouvé cette région de 39 pb identiques.

Nous avons ensuite poussé notre analyse de la séquence de l'haplotype 4 du L1 C et des haplotypes 2 et 3 du L1 B pour tenter de déterminer la longueur de la zone de conversion génique. Pour ce faire, nous avons séquencé 40 pb en 5' de cette région de 300 pb, c'est-à-dire la région située entre l'amorce et le début de la région de 300 pb, pour l'haplotype 4 du L1 C et les haplotypes 2 et 3 du L1 B. Ceci nous a permis de déterminer les bornes de cette événement de conversion qui correspondent aux différences de bases retrouvées en 5' et 3' de la région de conversion (positions –30 et 141, voir figure 19, p.69). En comparant la séquence de l'haplotype 2 ou 3 du L1 B avec la séquence de l'haplotype 4 du L1 C, on peut observer une région de conversion de 171 pb (positions –30 à 141) tel qu'illustré à la figure 19, p.69. Cette analyse nous permet donc de supposer que l'haplotype 4 du L1 C serait le résultat d'une conversion génique due à un événement de recombinaison homologue qui aurait eu lieu entre les éléments L1 B et C au cours de l'évolution. Toutefois, il faut noter qu'aucun des individus porteurs de l'haplotype 4 du L1 C n'est porteur de l'haplotype 2 ou 3 du L1 B (voir tableaux IV et VI, p.64 et 66). Ceci pourrait par

Figure19 : Présentation de la zone de conversion génique du L1 C haplotype 4.

Alignement des séquences de la région de 300 pb (comprise entre les flèches sens et anti-sens) et de la région flanquante en 5' de l'haplotype 4 du L1 C (Cb) et des haplotypes 2 et 3 du L1 B (B2 et B3) pour présenter la zone de conversion (en gris) qui s'étend de la position -30 à 141 pb où les trois séquences sont identiques. Les positions 20, 31, 38, 40, 46, 50, 53 et 58 réfèrent aux positions de bases nucléiques caractérisant l'haplotype 4 du L1 C. Les étoiles montrent les bases qui diffèrent entre le L1 B et le L1 C.

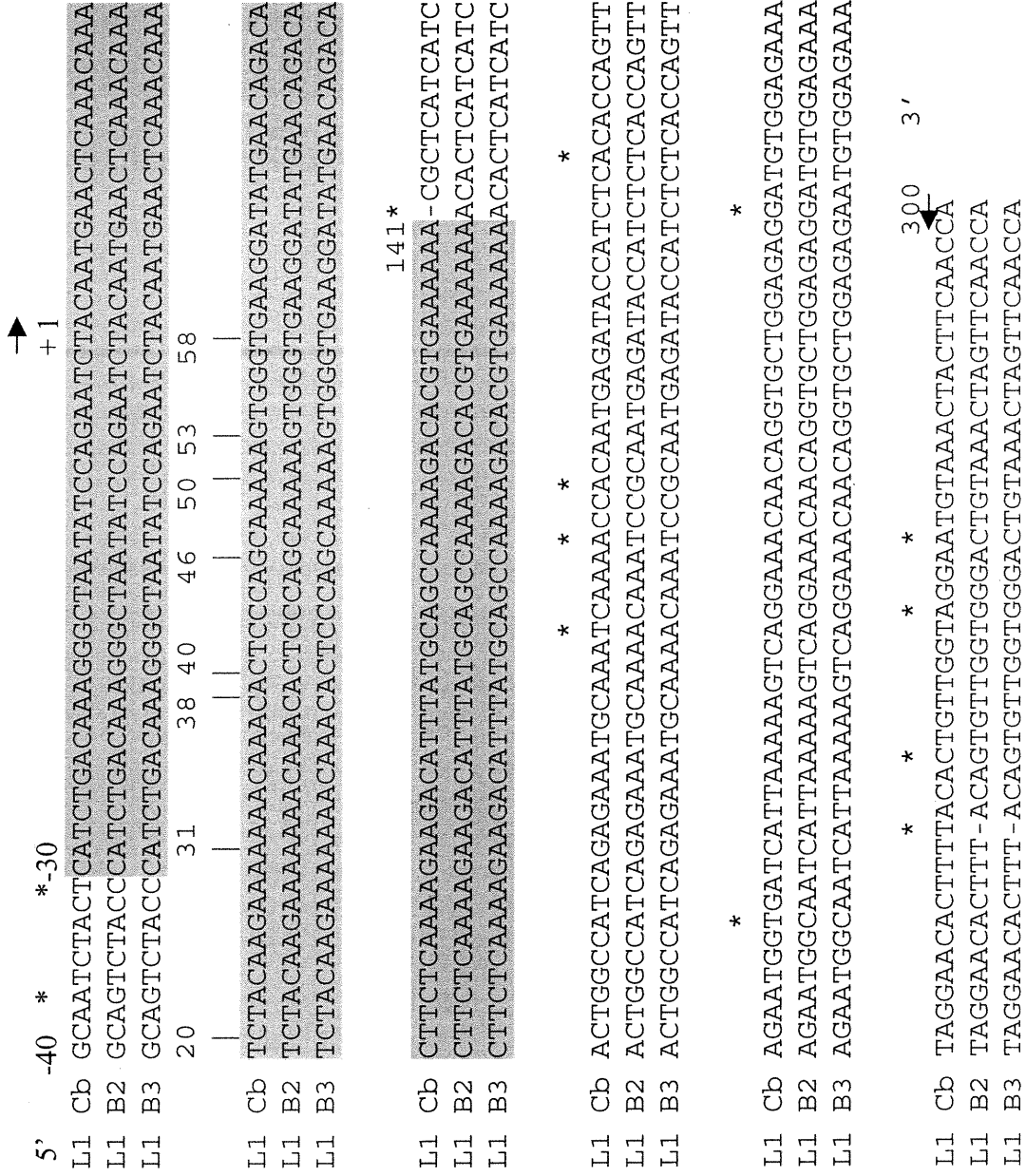


Figure 19

contre s'expliquer par le grand nombre d'haplotype combinés observés pour ces trois éléments L1 et une séquence intronique de cette région (voir discussion).

III.3 Processus d'évolution d'un élément LINE-1

Afin d'établir si les variations de bases que nous avons observées au niveau des éléments L1 étudiés proviennent de mutations ponctuelles ou d'événements de recombinaison homologue, nous avons comparé les variations de bases obtenues entre les trois éléments L1 sélectionnés et une séquence unique intronique. Cette séquence intronique (I) se situe dans la même région que les trois éléments L1 choisis et sa position génomique est illustrée à la figure 10A), p.35. Étant donné que cette séquence est unique dans le génome et qu'elle ne devrait pas subir de pression sélective, on s'attendait à ce qu'elle présente des variations de bases produites surtout par l'accumulation de mutations ponctuelles. Toutefois cette séquence peut aussi évoluer par recombinaison génétique lors de la méiose où il peut se produire des échanges de séquences d'ADN (brassage génétique) entre les chromosomes homologues. Les résultats présentés au tableau VII, p.72 suggèrent les deux types de processus d'évolution. Tout d'abord, en supposant toujours l'haplotype 1 comme haplotype fondateur présentant une fréquence de 11/20 haplotypes, nous pouvons suivre le cheminement évolutif comme suit: les changements de bases d'un A en G en position 49 et d'un T en C en position 107 de l'haplotype 1 donnent naissance au type, ensuite l'haplotype 3 subit deux changements de bases soit un C en T à la

Tableau VII. Analyse comparative des différents types d'haplotypes (haplo.), basée sur les variations de bases nucléiques obtenues suite au séquençage d'une région de 300pb d'une séquence intronique au niveau de l'ADN normal chez 10 individus atteints du cancer du colon, ce qui représente 20 haplotypes. Les positions nucléiques correspondent aux positions des variations nucléiques observées dans la région de 300 pb. Les nombres entre parenthèses indiquent l'abondance relative de chacun des types d'haplotypes. Les variations (indiquées par la base appropriée) sont montrées par rapport à l'haplotype le plus fréquent (haplo. 1). Le trait (-) indique qu'il n'y a pas de changement de base à cette position.

Position nucléique	11	49	93	107	131
Haplo. 1 (11/20)	T	A	C	T	G
Haplo. 2 (4/20)	-	G	T	C	A
Haplo. 3 (1/20)	-	G	-	C	-
Haplo. 4 (4/20)	C	G	T	C	A

Tableau VIII. Distribution des différents types d'haplotypes de la séquence intronique retrouvés pour chacun des 10 individus. Nous présentons entre parenthèses la provenance ethnique de chacun des individus, Canadien Français (C.F.), Canadien Italien (C.I.) et Haïtien (H). Le chiffre identifiant chacun des individus réfère au numéro attribué à l'échantillon d'ADN reçu.

INDIVIDU	TYPES D'HAPLO.
221 (?)	1/2
244 (C.F.)	1/1
262 (C.I.)	1/3
264 (C.F.)	1/2
266 (H)	1/4
367 (C.F.)	1/2
370 (C.F.)	1/2
374 (C.F.)	1/4
377 (C.F.)	4/4
380 (C.F.)	1/1

position 93 et un G en A à la position 131 pour donner lieu au type 2, et enfin ce dernier fait apparaître l'haplotype 4 par le changement d'un T en C en position 11. Cependant, ce modèle est difficile à expliquer puisque l'haplotype 3, qui devrait être le fondateur du type 2 et 4, est peu représenté dans la population par rapport à ces derniers. En effet, seulement 1/10 individus porte cet haplotype (tableau VIII, p.72). Cette observation peut être interprétée de deux façons, soit que l'échantillon est peu représentatif de la population ce qui implique que les fréquences observées ne sont pas les fréquences réelles, ou soit que les variations de bases sont dues à des événements de recombinaison génétique. L'accumulation de mutations ponctuelles est un processus très lent se déroulant durant l'évolution et il implique nécessairement que l'haplotype intermédiaire (ici l'haplotype 3), servant à l'apparition d'un nouveau type d'haplotype, soit bien représenté dans une population. Cet haplotype intermédiaire doit donc être suffisamment fréquent dans la population afin que celui-ci ait de bonnes chances d'être affecté par des mutations subséquentes et encore faut-il que ces dernières soient conservées dans les générations suivantes. Par contre la conversion génique ne nécessite qu'un seul événement pour donner plusieurs changements de bases dans une séquence donnée et ces variations doivent aussi être conservées au cours de l'évolution.

Ces résultats nous permettent d'établir que tout comme les séquences introniques, les séquences L1 peuvent présenter des variations de bases dues à l'accumulation de mutations ponctuelles au cours de l'évolution, tel que nous l'avons observé pour les L1 A et B. Cependant, en considérant le nombre d'interactions

possibles entre les L1 dû à leur abondance et à leur dispersion à travers le génome, nous pouvons nous attendre à observer l'apparition de variations de bases dues à des événements de recombinaison homologue. Les résultats obtenus pour le L1 C nous indiquent, selon les évidences déjà mentionnées, que les séquences L1 peuvent aussi interagir entre elles dans leur processus d'homogénéisation, et donner ainsi naissance à de nouvelles variations de bases.

III.4 Analyse de la variabilité de la région étudiée

Nous avons observé que le nombre de variations de bases des trois éléments L1 que nous avons étudiés était comparable au résultats obtenus pour la séquence unique intronique. Nous avons alors proposé que la région étudiée pourrait influencer la variabilité de la séquence des éléments L1 qui en font partie. Nous avons d'abord regardé comment les différents types d'haplotypes retrouvés pour chacun des trois L1 et pour la séquence intronique étaient combinés sur les deux chromosomes de chacun des 10 individus. Nous avons donc tenté de déterminer le nombre approximatif d'haplotypes combinés pouvant exister dans notre groupe de 10 individus. Un haplotype combiné se définit comme étant la combinaison d'haplotypes observée à des locus génétiquement liés. Pour reconstituer les haplotypes combinés, nous avons regroupé pour chacun des individus, les deux haplotypes de chacun des éléments L1 A, B et C et ceux de la séquence intronique. Par exemple, pour l'individu 221 nous avons 2 haplotypes de type 1 pour le L1 A

(voir tableau II, p.62), deux haplotypes de type 1 pour le L1 B (voir tableau IV, p.64), un haplotype de type 1 et un haplotype de type 5 pour le L1 C (voir tableau VI, p.66) et enfin un haplotype de type 1 et un haplotype de type 2 pour la séquence intronique. En considérant que l'haplotype combiné de départ était constitué des haplotypes fondateurs (type 1) pour chacun des trois éléments L1 et pour la séquence intronique, nous avons reconstitué les différents haplotype combinés existant. Par exemple, dans le cas de l'individu 221 nous avons établi qu'il y avait un chromosome qui portait les haplotypes de type 1 de chacun des éléments L1 et de la séquence intronique, et que l'autre chromosome portait les deux haplotypes de type 1 des L1 A et B, l'haplotype 5 du L1 C et l'haplotype 2 de la séquence intronique. Nous avons effectué le même exercice pour chacun des individus. Nous avons ainsi dénombré 13 haplotype combinés que nous présentons au tableau IX, p.76. Ce grand nombre d'haplotype combinés nous indique un haut degré d'instabilité de la région chromosomique que nous avons étudiée, puisque différentes combinaisons d'haplotypes existent. Nous avons aussi indiqué la fréquence de chacun des haplotype combinés dans notre groupe d'individus. Il est à noter que l'haplotype combiné #1, qui est considéré comme l'haplotype combiné préexistant, est le plus fréquent.

Par la suite, nous nous sommes intéressé au nombre de différences de bases que nous pourrions observer en comparant le génome de deux individus. Pour notre étude, nous avons séquencé une région de 300 pb pour chacun des trois éléments L1 et pour la séquence intronique. Nous avons donc séquencé un total de 1200 pb par chromosome pour chacun des individus. Nous avons ensuite tenté d'évaluer le

Tableau IX. Présentation des différents haplotypes combinés retrouvés chez les dix individus. Les haplotypes combinés correspondent aux combinaisons des différents types d'haplotypes des éléments L1 A, B, C et de la séquence intronique (I) retrouvées chez les 10 individus. Ces haplotypes combinés ont été déterminés en considérant l'haplotype combiné #1 préexistant dans l'évolution, puisqu'il est constitué de la combinaison des haplotypes fondateurs (type 1, se référer au tableaux I, III, V, VII) pour chacun des trois éléments L1 A, B et C et pour la séquence intronique. La fréquence correspond au nombre de fois que l'on retrouve cette combinaison d'haplotypes dans notre groupe de 10 individus, en considérant que nous avons reconstitué 20 combinaisons au total (2 combinaisons par individu).

Haplotypes combinés	Haplotypes L1 A	Haplotypes L1 B	Haplotypes L1 C	Haplotypes I	Fréquence
#1	1	1	1	1	6
#2	1	1	5	2	2
#3	1	5	4	3	1
#4	1	6	4	2	1
#5	1	2	1	4	1
#6	1	7	4	1	1
#7	1	1	1	4	1
#8	1	1	5	4	2
#9	1	1	4	1	1
#10	1	1	3	2	1
#11	2	3	2	1	1
#12	1	4	1	1	1
#13	2	4	1	1	1

Tableau X. Nombre de changements de bases observées en comparant la séquence totale de 1200 pb de deux individus à la fois. La séquence de 1200 pb correspond au séquençage total effectué pour chacun des individus, soit 300 pb séquencées au niveau des trois éléments L1 A, B, C et de la séquence intronique. Chacun des dix individus a été comparé aux neuf autres individus. Les résultats sont indiqués en nombres de bases.

	221	244	262	264	266	367	370	374	377
221									
244	5								
262	7	5							
264	3.5	5	7						
266	7	7	9.5	7					
367	7	7.5	7	7	9.5				
370	7	8	7.5	7	10.5	7.5			
374	5	4	7.5	5	7	7.5	8		
377	4.5	8	9	4.5	7	8	8	5	
380	6.5	5.5	6	6.5	10	7	7	7.5	9.5

nombre moyen de variations de bases sur 1200 pb séquencées pour cette région, en comparant la séquence génomique (1200 pb) de 2 individus à la fois. Nous avons comparé chacun des dix individus aux neuf autres individus. Nous présentons ces résultats au tableau X, p.77. Ainsi, nous avons observé entre 3.5 et 10.5 changements de base sur 1200 pb totales en comparant deux individus à la fois. Ensuite, nous avons calculé la moyenne de ces résultats, qui est de sept changements sur 1200 bases. Ce résultats dépassent largement le nombre moyen de variations de 1/1000 pb qui a été évalué en comparant le génome de deux individus lors des études sur les SNPs (voir discussion).

En résumé, ces analyses présentent des indications que cette région du génome (région FRA3B du gène FHIT), qui semble être plus instable ou plus variable que le reste du génome moyen, pourrait être à l'origine des variations de séquence des éléments L1 que nous avons observées.

III.5 Interactions des éléments LINEs dans un contexte d'instabilité génomique

En considérant que les éléments L1 représentent des substrats pour la recombinaison homologe, nous avons vérifié si dans un contexte d'instabilité génomique, il y avait une augmentation des interactions entre les éléments L1. Pour ce faire, nous avons comparé la séquence d'un L1 donné à un endroit donné au niveau du tissu normal et tumoral chez les dix individus. Une étude a montré que ces

trois éléments L1 (A, B et C) étaient impliqués dans des événements de recombinaison homologue menant à des délétions caractérisées dans des lignées de cellules tumorales (Inoue *et al.* 1997). Les résultats obtenus pour chacun des éléments L1 sont intégrés aux tableaux II, IV et VI, où nous présentons la distribution des différents types d'haplotypes retrouvés pour chacun des dix individus. Nous n'avons pas observé de variations de bases nucléiques au niveau de la séquence de chacun des L1 entre le tissu normal et tumoral pour chacun des individus. Donc la fréquence de conversion génique des trois éléments L1 observés, dans ce contexte d'instabilité génomique, ne semble pas suffisamment élevée ou sélective pour l'émergence d'un nouvel haplotype. Toutefois l'analyse de cinq clones pour chacun des individus pourrait ne pas être suffisante pour observer ce type d'événement, puisque l'événement de recombinaison pourrait s'être produit dans peu de cellules et ainsi être dilué dans l'échantillon d'ADN provenant du tissu tumoral. Il est à noter que l'échantillon d'ADN tumoral est composé d'au moins 70% de cellules tumorales. De plus, dans le cas d'un échantillon d'ADN normal périphérique provenant d'une biopsie de tumeur du colon, l'analyse histologique indique qu'il y a très peu ou pas de cellules tumorales (Dr. Mark Basik, communication personnelle).

De plus, nous avons testé nos échantillons d'ADN tumoral pour les événements de crossing-over entre les trois éléments L1. Pour ce faire, nous avons tenté d'amplifier ces événements avec différentes combinaisons d'amorces PCR. À la figure 10B), p.35 nous illustrons un exemple de combinaison d'amorces pour amplifier un événement de crossing-over susceptible de se produire entre les éléments

L1 B et C. La combinaison de l'amorce 1 du L1 C et l'amorce 6 du L1 B aurait permis d'amplifier cet événement de recombinaison, car sans cet événement la distance de 110 kb entre ces deux éléments est beaucoup trop grande pour permettre une amplification. Après avoir effectué une première amplification avec différentes combinaisons d'amorces (voir figure 10B), p.35) soit 1-4, 1-6 et 3-6 nous n'avons observé aucun produit d'amplification. Nous avons alors tenté une deuxième réaction d'amplification sur la première. Nous avons pu observer des produits d'amplification seulement pour la combinaison d'amorces 1-4 au niveau de l'ADN tumoral d'un seul individu qui pouvait révéler la possibilité d'événements de recombinaison homologue avec crossing-over entre le L1 C et le L1 A. Cependant, nous avons tenté de reproduire les deux réactions d'amplification sur le même ADN tumoral pour s'assurer que les produits qui nous semblaient intéressants étaient spécifiques, mais nous avons constaté que les bandes d'amplification ne se répétaient pas. De plus, différentes digestions caractéristiques furent effectuées sur les produits qui nous semblaient intéressants et aucune des digestions ne répondait au patron de digestion attendu.

IV. DISCUSSION

Nous savons maintenant que les éléments L1 occupent non seulement une grande proportion du génome humain, mais aussi qu'ils existent chez tous les mammifères. Cette constatation a poussé les chercheurs à se poser des questions concernant leur implication dans l'évolution du génome. Leur mécanisme de dispersion fut d'abord largement étudié pour constater jusqu'à maintenant que la rétrotransposition a contribué à l'augmentation de la complexité et de la longueur du génome chez les mammifères. Ces éléments constitueraient 20 % de notre génome (Int. Hum. Gen. Seq. Cons. 2001). Cependant, il s'avère que la majorité des éléments L1, que l'on retrouve dans notre génome seraient inactifs en rétrotransposition. Donc à première vue, ces éléments inactifs qui ne codent pour aucune protéine n'ont aucune fonction biologique. Toutefois, il est à noter qu'ils ont été conservés au cours de l'évolution et qu'ils conservent trois caractères importants soit la similarité de séquence, l'abondance et la dispersion. Nous avons alors émis l'hypothèse que ces éléments pourraient avoir un rôle au sein de l'organisation structurale du génome. Ceci impliquerait leur pairage homologue dans la stabilisation des domaines chromosomiques formés lors de l'organisation de la chromatine. Cette association intime entre les éléments L1 risque de former des substrats pour la recombinaison homologue. En effet, on a rapporté plusieurs évidences d'échange entre des éléments L1 soit dans des cas de maladies génétiques, des cancers ou des échanges spontanées dans des cellules en culture ou chez la souris. De plus, nous savons que les séquences

répétitives dispersées telles que les LINEs et les SINEs évoluent de façon concertée. Leur processus d'homogénéisation implique donc un mécanisme d'interaction entre les séquences d'une même famille répétitive. Le premier mécanisme est la rétrotransposition, qui permet la propagation d'un élément fondateur. Cependant le fait que la plupart des éléments soient défectifs soulève la participation possible d'un autre mécanisme d'homogénéisation. Par exemple, nous avons vu que la conversion génique serait un mécanisme secondaire dans l'évolution des éléments SINEs chez l'humain. De plus, ce mécanisme a aussi été observé par notre laboratoire chez les LINEs dans des cellules en culture et *in vivo* chez la souris.

Le but de notre projet était d'analyser le comportement évolutif d'un élément L1 donné à un endroit donné dans le génome chez l'humain. Il nous fallait d'abord sélectionner des éléments L1 dont nous connaissions la séquence de l'élément et de ses séquences flanquantes (voir section II.1 matériel et méthodes). Les régions flanquantes furent utilisées pour faire l'amplification spécifique de chacun des éléments L1 au niveau de l'ADN normal périphérique chez dix individus atteints du cancer du colon (voir plus loin). Par la suite, nous avons mis au point une stratégie de clonage pour chacun des éléments L1. Puis, le clonage et le séquençage de la région de 300 pb prédéterminée (voir section III.1 des résultats) de cinq clones par individu pour chacun des éléments L1, nous a permis de déduire les deux haplotypes de chacun des individus, pour un total de 20 haplotypes pour chacun des éléments. Enfin, l'alignement des 20 haplotypes pour chacun des L1 nous a permis de les regrouper en différents types d'haplotypes selon leur variation de bases nucléiques.

Dans un premier temps nous avons démontré qu'il existe plus d'un type d'haplotypes pour chacun des éléments L1 A, B et C (voir tableaux I, III et V, aux pages 62, 64 et 66). Plusieurs auteurs ont mentionné que les LINEs pouvaient diverger au cours de l'évolution en accumulant des mutations ponctuelles, puisqu'ils ne subissent pas la pression sélective (Voliva *et al.* 1983 ; Hardies *et al.* 1986 ; Scott *et al.* 1987 ; Pascale *et al.* 1993 ; Deberardinis et Kazazian 1998). Nous savons aussi que les mutations ponctuelles sont des changements de bases nucléiques survenant de façon aléatoire. Les résultats obtenus pour les L1 A et B présentent bien ce type de processus évolutif se traduisant par des changements de bases nucléiques aléatoires permettant l'apparition de nouveaux types d'haplotypes. Par exemple, l'apparition des différents types d'haplotypes des L1 A et B semble résulter de l'accumulation successive de mutations ponctuelles aléatoires, puisque nous pouvons faire des liens entre les différents types d'haplotypes tel que décrit dans la section des résultats (section III.2 des résultats) et ceci est comparable à la séquence intronique (voir section II.3 des résultats). Par contre, les résultats obtenus pour le L1 C nous indiquent, tel qu'attendu, que les L1 peuvent aussi évoluer par recombinaison homologue. En effet, nos résultats suggèrent un événement de conversion génique qui pourrait expliquer l'apparition de l'haplotype 4 du LINE C (tableau V, p.66). Nous nous basons sur le fait que cet haplotype présente 7 variations de bases nucléiques concentrées dans une même région de 39 pb par rapport à l'haplotype 1 (tableau V, p.66). Ceci représente donc l'équivalent d'environ un changement de base à toutes les 5 à 6 bases. Toutefois, selon certaines études, la diversité génétique

chez l'humain serait grandement attribuable aux SNPs (*Single Nucleotide Polymorphisms*) qui réfèrent aux variations de bases nucléiques simples que l'on peut observer en comparant la séquence des génomes de deux individus. Nous devrions observer, en moyenne, un changement de base à tous les 1000 bases en faisant la comparaison de deux chromosomes aléatoires (Stoneking 2001 ; Chakravarti 1999 ; Altshuler 2000 ; Int. SNP map working group 2001). L'observation de plusieurs changements de base concentrés dans une même région nous indiquerait donc que ces variations puissent provenir d'une conversion génique due à un événement de recombinaison homologue qui aurait eu lieu entre le L1 C et un autre élément L1 du génome. En effet, la recombinaison homologue est un facteur puissant d'évolution car elle produit des variantes génétiques en utilisant des blocs d'informations biologiques préexistants. De plus, une simulation informatique montre que des séquences d'ADN peuvent évoluer plus rapidement par recombinaison homologue que par mutation ponctuelle (Levinson 1994). Nous avons aussi tenter d'identifier le partenaire de recombinaison homologue du L1 C qui a permis l'apparition de l'haplotype 4. Pour ce faire, nous avons soumis la séquence de la région de 39 pb portant les variations de bases nucléiques à Gene Bank en considérant que ces variations représentent la signature d'un autre L1. Cette banque de données nous permet, entre autre, de comparer notre séquence aux autres séquences L1 du génome qui y sont répertoriées. Elle permet d'aligner notre séquence avec les autres et nous renseigne sur le niveau d'homologie entre chacune des combinaisons. Parmi les séquences les plus similaires, c'est-à-dire 38/39 bases identiques, figurait la séquence du L1 B. Curieusement, la seule base différente, soit le c en position 40 de

l'haplotype 4 du L1 C (voir tableau V, p.66), a été retrouvée chez deux types d'haplotypes du L1 B soit l'haplotype 2 et 3 (voir tableau III, p.64). Donc, en alignant la séquence des haplotypes 2 et 3 du L1 B avec la séquence de l'haplotype 4 du L1 C, tel qu'illustré à la figure 19, p.69, nous pouvons observer une région de parfaite identité de séquence de 171 pb et les bornes en 5'(position -30) et en 3' (position + 141) de cette région homologue. Nous avons schématisé cet événement de recombinaison homologue menant à la conversion génique d'un bout de séquence du L1 C par le L1 B à la figure 10B), p.35. Dans l'étude effectuée par Inoue et ses collaborateurs (1997), on a observé des interactions entre le L1 B et le L1 C dans des lignées de cellules tumorales. Il est à noter que nous présentons, dans notre modèle, la conversion du L1 C par le L1 B mais il faut considérer que cet événement aurait autant de chance de se produire dans le sens inverse, c'est-à-dire la conversion d'un bout de séquence du L1 B en C. Cependant, nous n'avons pas observé ce phénomène dans nos résultats.

Dans un deuxième temps, nous avons tenté d'établir si les variations de bases, que nous avons observées au niveau des L1 étudiés, seraient liées aux éléments L1 ou bien à la région étudiée. Nous avons donc comparé les variations de bases obtenues entre les trois éléments L1 sélectionnés et une séquence unique intronique. Si les variations que nous observions au niveau de nos séquences L1 provenaient de mutations ponctuelles, nous pouvions nous attendre à observer un taux comparable de mutations dans une séquence intronique, puisqu'elle peut accumuler des mutations sans subir de pression sélective. Nous avons donc séquencé une région de 300 pb de

cette séquence intronique, au niveau de l'ADN normal chez les dix individus, et nous avons obtenu 4 types d'haplotypes (voir tableau VII, p.72). Cependant, nous nous attendions à observer très peu de variations nucléiques en comparant la séquence des dix individus, puisque selon les études sur les SNPs, nous devrions observer en moyenne un changement sur 1000 bases en comparant le génome de deux individus. Curieusement, nous avons retrouvé 2 à 5 changements de bases nucléiques sur les 300 pb séquencées en comparant les différents types d'haplotypes (voir tableau VII, p.72). De plus, ces variations sont concentrées dans une région de 120 pb. Ces résultats suggèrent les deux types d'évolution, soit l'accumulation de mutations ponctuelles et la conversion génique. L'accumulation de mutations ponctuelles est plausible puisque nous pouvons faire des liens dans l'apparition des différents types d'haplotypes, tel que décrit dans la section III.3 des résultats. Cependant, il est difficile d'expliquer comment l'haplotype 3, peu fréquent dans la population, pourrait être à l'origine des haplotypes 2 et 4 par l'accumulation de mutations ponctuelles (voir tableau VII, p.72). Nous avons donc formulé deux explications. Soit que l'échantillon (dix individus) est peu représentatif de la population, ou bien que, les variations de bases sont dues à des événements de recombinaison génétique lors de la méiose. Cependant, ces éléments L1 qui sont en très grand nombre ont sûrement plus de chance d'interagir ensemble au cours de l'évolution que la séquence intronique qui peut seulement interagir avec son chromosome homologue. En d'autres mots, le taux de conversion génique pour un élément L1 devrait être plus élevé que celui d'une séquence unique intronique. De plus, nous ne pouvons exclure que l'échantillon de dix individus est peu représentatif en ce qui concerne la fréquence de chacun des

haplotypes dans la population. Ainsi, il est probable que les variations de bases observées pour la séquence intronique soient le résultat d'accumulation de mutations au cours de l'évolution. De plus, si nous comparons les types d'haplotypes obtenus pour le L1 C (voir tableau V, p.66) et ceux obtenus pour la séquence intronique (voir tableau VII, p.72), nous pouvons remarquer qu'il ne semble pas y avoir d'haplotype intermédiaire permettant l'apparition de l'haplotype 4 du L1 C, contrairement aux différents types d'haplotypes retrouvés pour la séquence intronique qui semblent tous reliés. Par contre, il est possible d'établir un cheminement évolutif par l'accumulation de mutations pour les L1 A et B. Les L1 A et B sont donc comparables du point de vue du cheminement évolutif, à la séquence intronique. Il est à noter que nous avons aussi tenté de trouver le partenaire de recombinaison du L1 C qui aurait pu permettre l'apparition de l'haplotype 5 (voir tableau V, p.66), en considérant que les trois changements de bases concentrés dans une région de 8 pb seraient le résultat d'une conversion génique. Nous avons retrouvé deux bases sur trois dans le L1 B et dans le L1 A (voir section II.2, p. des résultats). Toutefois la zone de conversion que l'on observe ne possède pas suffisamment de bases caractéristiques pour être soumise à Gene Bank pour l'obtention de résultats spécifiques.

Nous avons obtenu des résultats comparables, du point de vue du nombre de variations de bases, en comparant les résultats de la séquence intronique à ceux des éléments L1. Ceci nous a amené à poser l'hypothèse que cette variabilité des séquences L1 pourrait être associée à la région. Cette région FRA3B du gène FHIT

que nous avons étudiée est considérée comme une région fragile du génome (voir définition plus loin). Un site fragile présente différentes caractéristiques dont une hausse de la fréquence de recombinaison homologe induite par la présence de plusieurs points de cassure de l'ADN. Cette région serait donc plus instable que le reste du génome en moyenne. Or, Mark Stoneking (2001) et Aravinda Chakravarti (1999) mentionnent que certaines régions du génome présentent une plus grande diversité que la moyenne du génome. Par exemple, une région génomique peut montrer une plus grande diversité causée par certains aspects qui la caractérisent comme son contenu en séquences répétitives ou sa fréquence de recombinaison. Ceci se traduirait dans nos résultats par un nombre élevé de mutations de bases nucléiques et par le fait même une augmentation du nombre d'haplotypes pour les éléments L1 et la séquence intronique. En effet, nous avons observé dans nos résultats plusieurs variations de bases ainsi qu'un bon nombre de changements de bases, qui semble plus élevé que la moyenne du génome. Par ailleurs, nous avons séquencé 300 pb pour chacun des trois éléments L1 et pour la séquence intronique, ce qui fait un total de 1200 pb séquencées pour cette région génomique, pour chacun des individus. Nous avons donc tenté d'évaluer le nombre moyen de changements de bases observés sur les 1200 pb séquencées. Pour ce faire, nous avons d'abord reconstitué les liens existants entre les haplotypes des éléments L1 A, B, C et de la séquence intronique pour chacun des deux chromosomes de chaque individu. En effectuant ces différentes combinaisons, nous avons dénombré approximativement 13 haplotype combinés (voir tableau IX, p.76). Un haplotype combiné se définit comme étant la combinaison d'haplotypes observés à des locus génétiquement liés. Il existe

donc plusieurs combinaisons entre les différents types d'haplotypes des éléments L1 et de la séquence intronique étudié. Par la suite, nous avons comparé la séquence génomique (1200 pb) de deux individus à la fois pour calculer le nombre de différences de bases entre les deux individus (voir tableau X, p.77). En comparant les individus deux à la fois et en calculant la moyenne des résultats obtenus, nous avons établi qu'il y avait environ 7 changements de base sur 1200 pb entre deux individus dans cette région génomique. Ce résultat dépasse largement le nombre moyen de variations 1/1000 pb qui a été évalué en comparant le génome de deux individus lors des études portant sur les SNPs (Stoneking 2001 ; Chakravarti 1999 ; Altshuler 2000 ; Int. SNP map working group 2001). Nous observons donc plus de changements de base dans cette région que ce qui est observé en moyenne dans le génome. Selon cette observation, la fragilité de cette région semble avoir une influence sur le taux de variabilité des séquences L1 et de la séquence intronique que nous avons étudiés. D'autre part, le grand nombre d'haplotypes combinés existants pourrait expliquer le fait que nous n'avons pas retrouvé, chez le même individu, l'haplotype 4 du L1 C lié avec l'haplotype 2 ou 3 du L1 B soupçonnés d'être impliqués dans un événement de recombinaison homologue (tableaux IV et VI, p.64 et 66). On pourrait spéculer que cet événement de recombinaison pourrait s'être produit à un moment x de l'évolution et que par la suite le haut niveau de recombinaison de cette région pourrait être responsable des réarrangements de celle-ci, en ayant donné lieu à la perte du lien génétique entre les haplotypes de ces deux éléments (haplotype 4 du L1 C et haplotype 2 ou 3 du L1 B) au cours de l'évolution.

Comme nous l'avons déjà mentionné, nos résultats suggèrent qu'il y aurait eu recombinaison homologue entre deux éléments L1, menant à la conversion génique d'un élément au cours de l'évolution. Ceci implique qu'il y ait eu échange d'un bloc d'information génétique d'un élément L1 vers un autre, sans en affecter les séquences génomiques flanquantes. Cependant, ces interactions doivent être contrôlées par la cellule de façon à limiter d'abord ces échanges, et s'il y a lieu diriger la résolution de ces échanges vers la conversion génique plutôt que le crossing-over (Richardson *et al.* 1998 ; Jonhson et Jasin 2000). Mais si la cellule perd le contrôle de ces échanges pour différentes raisons, telles que l'inactivation des fonctions de réparation de l'ADN ou la perte de régulation du cycle cellulaire, est-ce que les éléments L1 ne deviendraient pas des amplificateurs du processus tumoral ? C'est pourquoi, dans un troisième temps, nous avons analysé la séquence de ces éléments L1 dans les cellules tumorales. Nous nous sommes demandé si la fréquence de recombinaison homologue entre les éléments L1 pouvait augmenter dans un contexte d'instabilité génomique, puisque selon certaines études, il y aurait une hausse de la fréquence de recombinaison homologue dans les cellules cancéreuses (Finn *et al.* 1989 ; Tlsty *et al.* 1989). Nous avons donc regardé les interactions possibles entre les éléments L1 dans les cellules tumorales où nous nous retrouvions dans un contexte d'instabilité génomique (Coleman et Tsongalis 1999), c'est-à-dire qu'il y avait dérèglement des mécanismes de suppression de la recombinaison entre les séquences homologues non alléliques. Par exemple, ceci se serait traduit dans nos résultats par l'apparition d'un nouvel haplotype au niveau de l'ADN tumoral pour un individu, en comparant la

séquence d'un élément L1 donné entre le tissu normal et tumoral. Pour ce faire, nous avons séquencé la région de 300 pb de chacun des trois L1 sélectionnés chez les mêmes 10 individus au niveau de l'ADN tumoral. Au départ, nous savions que cette région fragile FRA3B du gène FHIT était associée à plusieurs types de cancers chez l'humain, dont le cancer du colon (Ohta et al. 1996 ; Noguchi et al. 1999 ; Morikawa 2000). De plus, nous avons choisi ces éléments L1 puisqu'ils étaient impliqués dans des événements de recombinaison homologue responsables de la délétion de l'exon 5 du gène FHIT dans des lignées de cellules tumorales (Inoue *et al.* 1997 ; Mimori *et al.* 1999). Nous savions donc qu'il y avait des interactions possibles entre ces éléments L1. Cependant, nous n'avons observé aucune différence entre la séquence d'un élément L1 au niveau de l'ADN normal et tumoral, c'est-à-dire que nous avons obtenu les mêmes types d'haplotypes pour les trois L1 (voir tableaux II, IV, VI, p.62, 64 et 66). Ainsi, la fréquence de conversion génique des trois éléments L1 observée dans le tissu tumoral ne semble pas suffisamment élevée ou sélective pour l'émergence d'un nouvel haplotype. De plus, l'analyse de cinq clones pour chacun des individus pourrait ne pas être suffisante pour observer ce type d'événement, puisque l'événement de recombinaison pourrait s'être produit dans peu de cellules et ainsi être dilué dans l'échantillon d'ADN provenant du tissu tumoral.

Nous avons aussi testé nos échantillons d'ADN tumoral pour les événements de crossing-over entre les trois éléments L1. Pour ce faire, nous avons utilisé différentes combinaisons d'amorces, tel que mentionné dans la section III.4 des résultats, mais nous n'avons obtenu aucun produit de crossing-over. Il faut

mentionner que nous avons vérifié la recombinaison entre ces trois éléments L1 dans cette région fragile du gène FHIT, puisque la recombinaison entre ces séquences pourrait être un mécanisme responsable de l'instabilité de cette région dans plusieurs types de tumeurs solides chez l'humain (Inoue *et al.* 1997 ; Mimori *et al.* 1999). C'est d'abord les études cytogénétiques de certains types de cancer qui ont mené à l'observation que le bras court du chromosome 3 était souvent impliqué dans ces maladies (Naylor *et al.* 1987 ; Hibi *et al.* 1992). Ensuite, des études physiques sur les délétions du chromosome 3 ont permis d'associer l'inactivation du gène FHIT au développement de plusieurs types de cancers chez l'humain (Croce *et al.* 1999 ; Huebner *et al.* 1998). Donc, si nous avons pu montrer que la fréquence de recombinaison homologue entre les éléments L1 était augmentée dans des cellules tumorales, nous aurions pu spéculer que la recombinaison homologue entre des éléments L1 puissent être responsable, du moins en partie, du degré d'instabilité génomique caractérisant les cellules tumorales, en entraînant le changement de la relation spatio-temporelle des gènes.

D'autre part, nous avons obtenu avec nos échantillons d'ADN, la provenance ethnique de 9/10 individus tel que mentionné dans les tableaux II, IV et VI, p.62, 64 et 66. Dans une étude menée par Alec Knight (1996) et ses collaborateurs, on a comparé la séquence d'une région de 300 pb d'un élément répété de type Alu spécifique chez 60 individus de différentes populations ethniques. Ils ont montré que les variations de bases nucléiques d'une séquence répétée de type ALU, observées entre des individus de différentes populations ethniques, étaient faibles à raison d'une

différence sur un total de 300pb séquencées. De plus, ils n'ont observé que 3 types d'haplotypes pour ce locus. Cependant, dans notre étude portant sur les LINEs, 7/10 échantillons d'ADN provenaient d'une même population d'individus, soit les canadiens français, et nous avons observé dans les cas des L1 B et C, la présence d'haplotypes possédant plus de deux variations de bases nucléiques et plusieurs types d'haplotypes (tableaux III et V, p.64 et 66). Par contre, dans le cas du LINE A, nous n'avons observé que deux types d'haplotypes différant par une seule variation de base nucléique. Selon l'étude de Knight et ses collaborateurs, nous nous attendions à observer peu de variations de bases puisque la majorité des individus de notre échantillon de population avaient la même origine ethnique. Cependant, nous avons étudié une région génomique différente. Ceci suggère, une fois de plus, que la région pourrait influencer la variabilité allélique des séquences qui en font partie.

En résumé, nos résultats suggèrent d'une part que les éléments L1 puissent interagir de façon spontanée au cours de l'évolution chez l'humain, et d'autre part que la variabilité allélique de ces séquences L1 dépend aussi de la région génomique où elles se localisent. Toutefois, ces résultats soulèvent diverses questions intéressantes à explorer.

Premièrement nous pourrions nous demander si tous les éléments L1 évoluent au même taux. Par exemple, si nous comparons le nombre de variations de bases obtenu dans nos résultats pour chacun des trois L1 (voir tableaux I, III et V, p. 62, 64 et 66), on remarque que le L1 A, semble être l'élément le moins variable. Cette

séquence se comporte comme le reste du génome. Tandis que les L1 B et le L1 C sont plus variables que la moyenne du génome. Nous pourrions alors émettre l'hypothèse que la position génomique d'un élément L1 pourrait avoir une influence sur son évolution.

Ainsi nous pourrions regarder si, de façon générale, la région dans laquelle se retrouve ces éléments L1, soit la région fragile FRA3B du gène FHIT, pourrait jouer un rôle sur le comportement évolutif de ces éléments, comme nos résultats tendent à nous le suggérer. Nous savons qu'un site fragile se définit comme étant une région du génome qui présente des cassures ou des délétions chromosomiques lorsque des cellules en culture sont exposées à des agents ou des conditions inhibant la réplication ou la réparation de l'ADN (Shutherland 1991). Une région fragile présente différentes caractéristiques, dont une hausse de la fréquence de recombinaison homologue induite par la présence de plusieurs points de cassure de l'ADN. Nous pourrions alors émettre l'hypothèse que des complexes de protéiques, recrutés dans cette région pour permettre la réparation des cassures de l'ADN, pourraient augmenter la fréquence de recombinaison entre des séquences L1 appariées se trouvant à proximité du lieu de réparation. Pour vérifier si l'instabilité d'une région peut influencer la fréquence de recombinaison d'un élément L1, nous pourrions refaire la même analyse avec des éléments L1 caractérisés dans une autre région du génome tel que présenté dans l'article de Burwinkel et Kilimann (1998). Ces éléments L1 furent aussi reconnus comme participant à des événements de recombinaison homologue. En analysant la variabilité allélique de ces séquences L1

et en la comparant avec nos résultats pour les éléments L1 que nous avons étudiés, nous pourrions déterminer si la région génomique peut influencer l'évolution de ces éléments.

Une autre avenue intéressante à explorer serait d'analyser davantage le comportement des éléments L1 dans le processus tumoral. Dans notre projet, nous avons posé l'hypothèse que les éléments L1 pourraient être des amplificateurs du processus tumoral. Cependant nous n'avons pas obtenu de résultats confirmant cette hypothèse. À ce sujet, d'autres études sont présentement en cours dans notre laboratoire. Elles consistent à mesurer la contribution des interactions homologues entre des séquences répétitives à l'instabilité génomique des cellules tumorales. Dans différentes lignées cellulaires humaines normales, instables et tumorales, le système consiste à induire une cassure double brin dans un élément L1, préalablement intégré dans le génome des cellules, pour ensuite analyser la fréquence de recombinaison homologue, les mécanismes de réparation de l'ADN ainsi que le spectre des partenaires de recombinaison homologue impliqués. D'autre part, nous pourrions aussi refaire la même analyse décrite dans ce projet en utilisant d'autres types de tumeurs chez l'humain. Si ces analyses révélaient que les éléments L1 sont impliqués dans le processus tumoral, ils pourraient servir d'outils de mesure de l'instabilité génomique dans la caractérisation du stade de progression tumorale. En d'autres mots, l'analyse du niveau d'instabilité de ces éléments nous renseignerait sur l'agressivité de la tumeur.

De toute évidence, il reste plusieurs sujets à explorer concernant ces éléments L1 qui semblent maintenant prendre une importance considérable dans notre évolution. Cependant, nous voudrions approfondir nos connaissances sur le rôle de leurs interactions dans l'organisation structurale et l'évolution du génome chez l'humain.

En conclusion, nous avons observé une grande diversité allélique au niveau des éléments L1 que nous avons étudiés chez l'humain. Nous avons pu, dans certains cas, établir des liens entre les différentes variations de bases, ce qui suggère d'une part que les éléments L1 puissent évoluer par l'accumulation de mutations ponctuelles au cours de l'évolution. D'autre part, nous avons identifié une forme allélique qui présentait, selon toute probabilité, une zone de conversion génique pouvant être attribuable à un événement de recombinaison homologue ayant eu lieu entre deux éléments L1 localisés dans la même région génomique. Ceci nous indique que les séquences L1 pourraient aussi évoluer par conversion génique chez l'humain. De plus, nos résultats suggèrent que le comportement évolutif des éléments L1 pourrait être influencé par la région génomique dont il font partie. Ces résultats intéressants nous encouragent à poursuivre l'enrichissement de nos connaissances en ce qui concerne ces éléments chez l'humain.

V. BIBLIOGRAPHIE

Altshuler, D., V.J. Pollara, C.R. Cowles, W.J. Van Etten, J. Balwin, L. Linton, E.S. Lander, (2000), An SNP map of the human genome generated by reduced representation shotgun sequencing, *Nature*, 407: 513-516.

Belmaaza, A., J.C. Wallenburg, S. Brouillette, N. Gusew, P. Chartrand, (1990), Genetic exchange between endogenous and exogenous LINE-1 repetitive elements in mouse cells, *Nuc. Acids Res.*, 18: 6385-6391.

Belmaaza, A., E. Milot, J.F. Villemure, P. Chartrand, (1994), Interference of DNA sequence divergence with precise recombinational DNA repair in mammalian cells, *EMBO J.*, 13: 5355-5360.

Boeke, J.D., O.K. Pickeral, (1999), Retroshuffling the genomic deck, *Nature*, 398: 108-110.

Burwinkel, B., M.W. Kilimann, (1998), Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease, *J. Mol. Biol.*, 277: 513-517.

Chakravarti, A., (1999), Population genetics-making sense out of sequence, *Nature genetics supplement*, 21: 56-60.

Coen, E., T. Strachan, G. Dover, (1982), Dynamics of concerted evolution of ribosomal DNA and histone gene families in the melanogaster species subgroup of *Drosophila*, *J. Mol. Biol.*, 158: 17-35.

Coleman, W.B., G.J. Tsongalis, (1999), The role of genomic instability in human carcinogenesis, *Anticancer Research*, 19: 4645-4664.

Croce, C.M., G. Sozzi, K. Huebner, (1999), Role of FHIT in human cancer, *J. Clin. Onco.*, 17: 1618-1624.

DeBerardinis, R.J., H.H. Kazazian Jr, (1998), Full-length L1 elements have arisen recently in the same 1-Kb region of the gorilla and human genomes, *J. Mol. Evol.*, 47: 292-301.

Dover, G.A., S. Brown, E. Coen, J. Dallas, T. Strachan, M. Trick, (1982), The dynamics of genome evolution and species differentiation, in G.A. Dover and R.B. Flavell (eds.), *Genome evolution*, Academic Press, San Diego.

Elder Jr., J.F., B.J. Turner, (1995), Concerted evolution of repetitive DNA sequences in eucaryotes, *The quarterly review of biology*, 70: 297-320.

Fanning, T.G., (1983), Size and structure of highly repetitive BAM HI element in mice, *Nuc. Acids. Res.*, 11: 5073-5091.

Fanning, T.G., M.F. Singer, (1987), LINE-1: a mammalian transposable element, *Biochim. Biophys. Acta*, 910: 203-212.

Fin, G.K., B.W. Kurz, R.Z. Cheng, R.J. Shmookler Reis, (1989), Homologous plasmid recombination is elevated in immortally transformed cells, *Mol. Cell. Biol.*, 9: 4009-4017.

Furano, A.V., (2000), The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposon, *Prog. Nucleic Acid Res. Mol. Biol.*, 64: 255-295.

Gaillard, C., F. Strauss, (1994), Association of poly(CA)-poly(TG) DNA fragments into four-stranded complexes bound by HMG1 and 2, *Science*, 264: 433-436.

Goodier, J.L., E.M. Ostertag, H.H. Kazazian Jr., (2000), Transduction of 3'-flanking sequences is common in L1 retrotransposition, *Hum. Mol. Genet.*, 9: 653-657.

Grimaldi, G., J. Skowronski, M.F. Singer, (1984), Defining the beginning and end of KpnI family segments, *EMBO J.*, 3: 1753-1759.

Gu, Z., H. Wang, A. Nekrutenko, W.-H. Li, (2000), Densities, length proportions, and other distributional features of repetitive sequences in the human genome estimated from 430 megabases of genomic sequence, *Gene*, 259: 81-88.

Haber, J.E., (2000), Partners and pathways: repairing a double-strand break, *TIG*, 16: 259-264.

Harada, T., J. Nagayama, K. Kohno, L.A. Mickley, T. Fojo, M. Kuwano, M. Wada, (2000), Alu-associated interstitial deletions and chromosomal rearrangement in 2 human multidrug-resistant cell lines, *Int. J. Cancer*, 86: 506-511.

Hardies, S.C., S.L. Martin, C.F. Voliva, C.A. Hutchison III, M.H. Edgell, (1986), An analysis of replacement and synonymous changes in the rodent L1 repeat family, *Mol. Biol. Evol.*, 3: 109-125.

Hibi, K., T. Takahashi, K. Yamakawa, (1992), Three distinct regions involved in 3p deletions in human lung cancer, *Oncogene*, 7: 445-449.

Hickey, D.A., (1982), Selfish DNA: a sexually-transmitted nuclear parasite, *Genetics*, 101: 519-531.

Hill, A.S., N.J. Foot, T.L. Chaplin, B.D. Young, (2000), The most frequent constitutional translocation in humans, the t(11:22)(q23;q11) is due to highly specific Alu-mediated recombination, *Hum. Mol. Genet.*, 9: 1525-1532.

Huebner, K., P.N. Garrison, L.D. Barnes, C.M. Croce, (1998), The role of the FHIT/FRA3B locus in cancer, *Annu. Rev. Genet.*, 32: 7-31.

Hutchison III, C.A., S.C. Hardies, D.D. Loeb, W.R. Shehee, M.H. Edgell, (1989), LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome, pp. 593-617 in *Mobile DNA*, edited by D.E. Berg and M.M. Howe, American Society for microbiology, Washinton DC.

Inoue, H., H. Ishii, H. Alder, E. Snyder, T. Druck, K. Huebner, C.M. Croce, (1997), Sequence of the FRA3B common fragile region: implications for the mechanism of FHIT deletion, *Proc. Natl. Acad. USA*, 94: 14584-14589.

International human genome sequencing consortium, (2001), Initial sequencing and analysis of the human genome, *Nature*, 409: 860-921.

International SNP map working group, (2001), A map of human genome sequence variation containing 1,42 million single nucleotide polymorphisms, *Nature*, 409: 928-933.

Johnson, R.D., M. Jasin, (2000), Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells, *EMBO J.*, 19:3398-3407.

Jubier-Maurin, V., B.J. Dod, M. Bellis, M. Piechaczyk, G. Roizes, (1985), Comparative study of the L1 Family in the genus *Mus*: possible role of retroposition and conversion events in its concerted evolution, *J. Mol. Biol.*, 184: 547-564.

Kass, D.H., M.A. Batzer, P.L. Deininger, (1995), Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution, *Mol. Cell. Biol.*, 15: 19-25.

Kazazian Jr., H.H., (2000), L1 retrotransposons shape the mammalian genome, *Science*, 289: 1152-1153.

Khodarev, N.N., T. Bennett, N. Shearing, I. Sokolova, J. Koudelik, S. Walter, M. Villalobos, A.T.M. Vaughan, (2000), LINE L1 retrotransposable element is targeted during the initial stages of apoptotic DNA fragmentation, *J. Cell. Biochem.*, 79: 486-495.

Knight, A., M.A. Batzer, M. Stoneking, H.K. Tiwari, W.D. Scheer, R.J. Herrera, P.L. Deininger, (1996), DNA sequences of Alu elements indicate a recent replacement of the human autosomal genetic complement, *P.N.A.S. USA*, 93: 4360-4364.

- Kucherlapati, R., G.R. Smith** (Eds), (1998), Genetic recombination, *American society for microbiology*, Washington.
- Lamond, A.I., W.C. Earnshaw**, (1998), Structure and function in the nucleus, *Science*, 280: 547-553.
- Levinson, G.**, (1994), Crossovers generate random recombinants under Darwinian selection. In: Maes, R. (Ed.), *Artificial Life IV*. Mit Press, Cambridge, MA, pp. 90-101.
- Li, W.-H., Z. Gu, H. Wang, A. Nekrutenko**, (2001), Evolutionary analyses of the human genome, *Nature*, 409: 847-849.
- Liao, D.**, (1999), Concerted evolution: molecular mechanism and biological implications, *Am. J. Hum. Genet.*, 64: 24-30.
- Lilley, D.M.J.**, (1992), HMG has DNA wrapped up, *Nature*, 357: 282-283.
- Maeda, N., C.I. Wu, J. Bliska, J. Reneke**, (1988), Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences, *Mol. Biol. Evol.*, 1: 1-20.
- Makalowski, W.**, (2000), Genomic scrap yard: how genomes utilize all that junk, *Gene*, 259: 61-67.
- Mears, M.L., C.A. Hutchison III**, (2001), The evolution of modern lineages of mouse L1 elements, *J. Mol. Evol.*, 52: 51-62.
- Mimori, K., T. Druck, H. Inoue, H. Alder, L. Berk, M. Mori, K. Huebner, C.M. Croce**, (1999), Cancer-specific chromosome alterations in the constitutive fragile region FRA3B, *Proc. Natl. Acad. Sci. USA*, 96: 7456-7461.
- Moffat, A.S.**, (2000), Transposons help sculpt a dynamic genome, *Science*, 289: 1455-1457.
- Moran, J.V., R.J. DeBerardinis, H.H. Kazazian Jr.**, (1999), Exon shuffling by L1 retrotransposition, *Science* 283: 1530-1534.
- Morikawa, H., Y. Nakagawa, K. Hashimoto, M. Niki, Y. Egashira, I. Hirata, K. Katsu, Y. Akao**, (2000), Frequent altered expression of fragile histidine triad protein in human colorectal adenomas, *Biochem. Biophys. Res. Com.*, 278: 205-210.
- Naylor, S.L., B.E. Johnson, J.D. Minna**, (1987), Loss of heterozygosity of chromosome 3p markers in small cell lung cancer, *Nature*, 329: 451-454.

Noguchi, T., W. Müller, H.C. Wirtz, R. Willers, H.E. Gabbert, (1999), FHIT gene in gastric cancer: association with tumour progression and prognosis, *J. Pathol.*, 188: 378-381.

Ohta, M., H. Inoue, M.G. Cotticelli, K. Kastury, R. Baffa, J. Palazzo, Z. Siprashvili, M. Mori, P. McCue, T. Druck, C.M. Croce, K. Huebner, (1996), The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers, *Cell.*, 84:587-597.

Pascale, E., C. Liu, E. Valle, K. Usdin, A.V. Furano, (1993), The evolution of long interspersed repeated DNA (L1, LINE1) as revealed by analysis of an ancient rodent L1 DNA family, *J. Mol. Evol.*, 36: 9-20.

Pavelitz, T., D. Liao, A.M. Weiner, (1999), Concerted evolution of the tandem array encoding primate U2 snRNA (the RNU2 locus) is accompanied by dramatic remodelling of the junctions with flanking chromosomal sequences, *Embo J.*, 18: 3783-3792.

Pickeral, O.K., W. Makalowski, M.S. Boguski, J.D. Boeke, (2000), Frequent human genomic DNA transduction driven by LINE-1 retrotransposition, *Genome Res.*, 10: 411-415.

Richard, M., A. Belmaaza, N. Gusew, J.C. Wallenburg, P. Chartrand, (1994), Integration of a vector containing a repetitive LINE-1 element in the human genome, *Mol. Cell. Biol.*, 14: 6689-6695.

Richard, M., N. Gusew, A. Belmaaza, P. Chartrand, (1997), Homologous junctions formed between a vector and human genomic repetitive LINE-1 elements as a result of one-sided invasion, *Som. Cell. & Mol. Gen.*, 23: 75-81.

Richardson, C., M.E. Moynahan, M. Jasin, (1998), Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations, *Genes Dev.*, 12: 3831-3842.

Rogers, J., (1985), The origin and evolution of retroposons, *Int. Rev. Cytol.*, 93: 187-279.

Rohlf, E. M., N. Paget, M.L. Graham, B.L. Weber, J.E. Garber, C. Skrzynia, J.L. Halperin, G.M. Lenoir, L.M. Silverman, S. Mazoyer, (2000), An Alu-mediated 7.1 Kb deletion of BRCA1 exons 8 and 9 in breast and ovarian cancer families that results in alternative splicing of exon 10, *Genes, chromosomes and cancer*, 28: 300-307.

Sassaman, D.M., B.A. Dombroski, J.V. Moran, M.L. Kimberland, T.P. Naas, R.J. DeBerardinis, A. Gabriel, G.D. Swergold, H.H. Kazazian Jr., (1997), Many human L1 elements are capable of retrotransposition, *Nature genet.*, 16: 37-43.

Saxton, J.A., S.L. Martin, (1998), Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome, *J. Mol. Biol.*, 280: 611-622.

Schwartz, A., D.C. Chan, L.G. Brown, R. Alagappan, D. Pettay, C. Distèche, B. McGillivray, A. de la Chapelle, D.C. Page, (1998), Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination, *Hum. Mol. Genet.*, 7: 1-11.

Scott, A.F., B.J. Schmeckpeper, M. Abdelrazik, C. Theisen Comey, B. O'Hara, J. Pratt Rossiter, T. Cooley, P. Heath, K.D. Smith, L. Margolet, (1987), Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence, *Genomics*, 1: 113-125.

Segal, Y., B. Peissel, A. Renieri, M. de Marchi, A. Ballabio, Y. Pei, and J. Zhou, (1999), LINE-1 elements at the sites of molecular rearrangements in Alport syndrome-diffuse leiomyomatosis, *Am. J. Hum. Genet.*, 64: 62-69.

Shaffer, L.G., J.R. Lupski, (2000), Molecular mechanisms for constitutional chromosomal rearrangements in humans, *Annu. Rev. Genet.*, 34: 297-329.

Shutherland, G.R., (1991), Chromosomal fragile sites, *Genet. Anal. Tech. Appl.*, 8: 161-166.

Singer, M.F., (1982), Highly repeated sequences in mammalian genomes, *Int. Rev. Cytol.*, 76: 67-112.

Smit, A.F.A., G. Toth, A.D. Riggs, J. Jurka, (1995), Ancestral, Mammalian-wide subfamilies of LINE-1 repetitive sequences, *J. Mol. Biol.*, 246: 401-417.

Smit, A.F.A., (1996), The origin of interspersed repeats in the human genome, *Curr. Opin. Genet. Dev.*, 6: 743-748.

Smit, A.F.A., (1999), Interspersed repeats and other mementos of transposable elements in mammalian genomes, *Curr. Opin. Genet. Dev.*, 9: 657-663.

Stoneking, M., (2001), From the evolutionary past..., *Nature*, 409: 821-823.

Tlsty, T.D., B.H. Margolin, K. Lum, (1989), Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell lines as measured by Luria-Delbruck fluctuation analysis, *Proc. Natl. Acad. Sci. USA*, 86: 9441-9445.

Tremblay, A., M. Jasin, P. Chartrand, (2000), A double-strand break in a chromosomal LINE element can be repaired by gene conversion with various endogenous LINE elements in mouse cells, *Mol. Cell. Biol.*, 20: 54-60.

Tremblay, A., N. Gusew, R. Lussier, B. Houle, P. Chartrand, (2001), De novo gene conversion events between LINE elements in mouse germ line, *soumis*.

Tvrđik, T., S. Marcus, S.-M. hou, S. Fält, P. Nouri, N. Prodlutskaja, F. Hanefeld, P. Stromme, B. Lambert, (1998), Molecular characterisation of two deletion events involving Alu-sequences, one novel base substitution and two tentative hotspot mutations in the hypoxanthine phosphoribosyltransferase (HPRT) gene in five patients with Lesch-Nyhan syndrome, *Hum. Genet.*, 103: 311-318.

Voliva, C.F., C.L. Jahn, M.B. Comer, C.A. Hutchison III, M.H. Edgell, (1983), The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end, *Nuc. Acids. Res.*, 11: 8847-8859.

Wendel, J.F., S.R. Wessler, (2000), Retrotransposon-mediated genome evolution on a local ecological scale, *P. N. A. S.*, 97: 6250-6252.

Zimmer, E. A., S.L. Martin, S.M. Beverly, Y.W. Kan, A.C. Wilson, (1980), Rapid duplication and loss of genes coding for the α chains hemoglobin, *Proc. Nat. Acad. Sci. USA*, 77: 2158-2162.

Zucman-Rossi, J., M.A. Batzer, M. Stoneking, O. Delattre, G. Thomas, (1997), Interethnic polymorphism of EWS intron 6: genome plasticity mediated by Alu retroposition and recombination, *Hum. Genet.*, 99: 357-363.