

2M11-2714.5

Université de Montréal

Variabilité Génétique au Saguenay

par

Geneviève Chatel

Programme de biologie moléculaire

Faculté des études supérieures

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

Avril 1999

© Geneviève Chatel



Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :
Diversité Génétique au Saguenay

présenté par :

Geneviève Chatel

a été évalué par un jury composé des personnes suivantes :

Dr Damian Labuda

Dr Daniel G. Bichet

Dr Hélène Vézina

Mémoire accepté le :

Sommaire

La colonisation de la Nouvelle-France a commencé dans la première moitié du XVII^e siècle par une immigration venue de France. Suite à un important accroissement démographique, les régions de Charlevoix, puis celle du Saguenay-Lac-St-Jean furent colonisées. La population de la grande région de Charlevoix-Saguenay, issue d'un nombre limité de fondateurs, a longtemps été isolée et compte aujourd'hui plus de 300 000 habitants. Tout comme dans plusieurs populations ayant longtemps été isolées, un effet fondateur a été observé dans la population de Charlevoix-Saguenay, dont témoigne une augmentation importante de la fréquence des porteurs de certaines maladies héréditaires récessives. Dans cette étude, nous voulions savoir si cet effet fondateur peut être observé au niveau des sites neutres de l'ADN et si oui, dans quelle mesure cet effet fondateur a affecté la variabilité génétique dans Charlevoix-Saguenay. L'effet fondateur n'est pas le seul mécanisme pouvant augmenter la fréquence de porteurs de maladies héréditaires mais il est probablement responsable de la majorité des fréquences élevées observées dans Charlevoix-Saguenay, ce qui nous amène à postuler que cet effet fondateur serait observable au niveau des sites neutres de l'ADN nucléaire. Pour le vérifier nous avons étudié un système génétique composé de six marqueurs polymorphes situés dans des régions introniques du gène de la dystrophine (Xp21). Un haplotype a été établi à l'aide de ces marqueurs pour 106 chromosomes indépendants de la région de Charlevoix-Saguenay, ainsi que pour deux autres populations : un groupe de Canadiens-Français de la région de Montréal et un groupe de la France urbaine.

Des études sur la stabilité de ce système (mutations et recombinaisons) ont été effectuées en établissant l'haplotypage pour huit familles CEPH (Centre d'Étude du Polymorphisme Humain) et la région étudiée a montré un taux élevé de recombinaisons. La diversité allélique et la diversité haplotypique ont été comparées pour les différentes populations : certains allèles montrent une fréquence nettement plus élevée dans la population de Charlevoix-Saguenay par rapport aux autres populations, mais la diversité allélique dans son ensemble est semblable à celle des autres groupes. La population de Charlevoix-Saguenay montre un nombre élevé d'haplotypes répétés comparativement aux autres populations et sa diversité haplotypique semble légèrement diminuée. L'augmentation en fréquence de certains allèles ou haplotypes dans la population de Charlevoix-Saguenay témoigne de l'effet fondateur qui serait responsable de la diminution de la variabilité haplotypique. La diversité génétique dans son ensemble a peu été affectée par cet effet fondateur puisque la variabilité allélique n'est pas diminuée et que la diminution de la variabilité haplotypique est minime.

Table des matières

| | |
|---------------------------------------|------|
| Sommaire..... | iii |
| Liste des figures..... | vi |
| Liste des tableaux..... | viii |
| Liste des abréviations..... | ix |
| Avant propos..... | 1 |
| Chapitre 1 : Introduction..... | 2 |
| Chapitre 2 : Matériel et méthode..... | 13 |
| Chapitre 3 : Résultats..... | 25 |
| Chapitre 4 : Discussion..... | 43 |
| Conclusion..... | 51 |
| Remerciements..... | 53 |
| Bibliographie..... | 54 |
| Annexe 1 : Données brutes..... | x |

Liste des figures

| | |
|---|----|
| Figure 1 : Modèle démographique illustrant la contribution de différentes vagues de fondateurs au pool génétique de la région de Charlevoix/Saguenay-Lac-St-Jean..... | 3 |
| Figure 2 : Modèle génétique représentant la portion du gène de la dystrophine étudiée et les différents marqueurs y étant situés..... | 15 |
| Figure 3 : Fréquences alléliques du marqueur 11Pol..... | 25 |
| Figure 4 : Fréquences alléliques du marqueur DXS1238..... | 26 |
| Figure 5 : Fréquences alléliques du marqueur DXS1237..... | 27 |
| Figure 6 : Fréquences alléliques du marqueur DXS1235..... | 28 |
| Figure 7 : Fréquences alléliques du marqueur IVS44SK21..... | 29 |
| Figure 8 : Distribution de fréquence allélique du marqueur Dys44Alu..... | 29 |
| Figure 9 : Distribution de fréquence des haplotypes composés des six Marqueurs à l'étude..... | 30 |
| Figure 10 : Distribution de fréquence des haplotypes pour les populations de France, de la région de Montréal et des échantillons CEPH comparée à celle de Charlevoix-Saguenay..... | 32 |
| Figure 11 : Distribution de fréquence de l'haplotype constitué des marqueurs DXS1237 et DXS1235..... | 33 |
| Figure 12 : Moyenne des fréquences alléliques composant chaque haplotype en fonction de la fréquence absolue de chaque haplotype..... | 35 |
| Figure 13 : Représentations graphiques sous forme d'arbres des matrices de distances construites à l'aide du logiciel Phylip d'après les fréquences alléliques..... | 36 |
| Figure 14 : Diversité haplotypique calculée à l'aide du logiciel Arlequin..... | 37 |
| Figure 15 : Matrice de distances générée à l'aide du logiciel Arlequin, basée sur l'index de fixation (F_{ST})..... | 38 |
| Figure 16 : Distribution de fréquence des haplotypes dys44..... | 38 |

| | |
|--|----|
| Figure 17 : Événements de recombinaison observés lors de l'étude des huit familles CEPH..... | 40 |
| Figure 18 : Lod scores obtenus à l'aide du MLINK version 5.10, en fonction de la fraction de recombinaison entre les marqueurs distaux du système génétique..... | 42 |
| Figure 19 : Modélisation de l'introduction des différents haplotypes dans la population de Charlevoix-Saguenay à partir de la population mère.. | 48 |

Liste des tableaux

| | |
|--|-------|
| Tableau I : Taux de porteurs de diverses génopathies autosomales récessives dans la population du Saguenay..... | 5 |
| Tableau II : Conditions de PCR et amorces utilisées pour l'amplification des marqueurs étudiés..... | 15 |
| Tableau III : Séquences des sondes utilisées pour l'hybridation ASO par buvardage en taches (dot blot)..... | 19 |
| Tableau IV : Distribution de fréquences des allèles étant présents chez les quatre populations à l'étude et montrant une différence significative à $p=0.3$ chez une des populations comparativement aux trois autres..... | 43 |
| Tableau V : Fréquences alléliques des différents marqueurs étudiés..... | x |
| Tableau VI : Composition allélique des haplotypes étudiés et fréquences des haplotypes..... | xii |
| Tableau VII : Distribution de fréquence des haplotypes composés des marqueurs DXS1237 et DXS1235..... | xvi |
| Tableau VIII: Normalisation de la distribution de fréquences des haplotypes... | xvii |
| Tableau IX : Moyenne des fréquences alléliques composant chaque haplotype en fonction de la fréquence absolue de chaque haplotype..... | xviii |
| Tableau X : Lod Scores obtenus à l'aide du logiciel LINKAGE pour différentes fractions de recombinaisons..... | xix |

Avant-propos

La génétique des populations est une discipline de la biologie qui consiste en l'étude et l'observation des changements génétiques dans un contexte historique et évolutif. Cette discipline vise à étudier les variations génétiques entre et parmi les différentes populations afin d'en comprendre la signification et d'en tirer des conclusions concernant l'histoire et les mécanismes d'évolution des populations étudiées.

La génétique des populations est une discipline qui en touche plusieurs autres. En effet, la biologie moléculaire, la génétique, l'histoire, l'anthropologie et les mathématiques sont quelques exemples des nombreuses branches scientifiques qui touchent à la génétique des populations. C'est une discipline qui a également de nombreuses applications : on s'en sert en épidémiologie dans le but de développer des outils diagnostiques ou pour l'identification de gènes responsables de diverses maladies génétiques, on s'en sert également en anthropologie dans le but de reconstituer l'histoire, voire l'origine de diverses populations humaines et enfin, on s'en sert dans le but de vérifier la validité des modèles déjà existants. C'est précisément ce dernier champ d'investigation qui fait l'objet de la présente étude. La population de la grande région de Charlevoix et du Saguenay-Lac-St-Jean, dont l'histoire est connue et bien documentée depuis la colonisation de la Nouvelle-France, est étudiée et comparée à celle de la France urbaine et celle des Canadiens-Français de la région de Montréal dans le but de superposer l'histoire démographique et les observations génétiques.

Chapitre 1 : Introduction

Histoire de la région de Charlevoix-Saguenay-Lac-St-Jean

La colonisation de la Nouvelle-France a commencé dans la première moitié du XVII^e siècle par des immigrants venus de France. Les villes de Québec puis de Montréal ont été fondées par ces derniers et les berges du fleuve St-Laurent entre ces deux villes ont également été colonisées. Suite à un important accroissement de la population, la pression démographique a fait en sorte que de nouvelles aires de colonisation ont été ouvertes pour répondre à la saturation de l'espace colonisé. Ainsi, la région de Charlevoix a reçu ses premiers immigrants à la fin du XVII^e siècle (Bouchard 1991). En 1765, sa population était composée d'environ un millier d'habitants elle connut ensuite un important accroissement démographique. Cependant, la situation socio-économique de Charlevoix ne pouvait supporter qu'un nombre limité d'habitants. En 1838 cette situation fut donc à l'origine d'un mouvement migratoire de plusieurs habitants de Charlevoix vers la région du Saguenay-Lac-St-Jean (Bouchard 1991). En 1850 la population de la grande région de Charlevoix et du Saguenay s'élevait à 18 000, cet accroissement résultant presque exclusivement de l'accroissement naturel, c'est-à-dire que l'immigration y était négligeable (Bouchard 1991). Dans les premières décennies de la colonisation du Saguenay, plus de 80 % des immigrants venaient de la région de Charlevoix. Puis, l'immigration s'est ensuite diversifiée : entre 1892 et 1911 seulement 30 % des immigrants venaient de Charlevoix. Cependant, 70 % des immigrants venus de

Charlevoix s'implantaient au Saguenay contre seulement 54 % des immigrants provenant d'autres régions (Gauvreau *et al.* 1990). La population de la région de Charlevoix-Saguenay passe de 18 000 en 1850 à plus de 300 000 aujourd'hui à cause d'un taux de fécondité étant demeuré longtemps très élevé dans la région. Les régions de Charlevoix et du Saguenay sont géographiquement bien distinctes mais, à cause de leur histoire, elles ont plusieurs points communs du point de vue démographique. C'est pourquoi ces deux régions seront ici étudiées ensemble et seront désormais désignées comme la région de Charlevoix-Saguenay. La figure 1 montre la contribution de chaque vague d'immigration au pool génique de la population de Charlevoix-Saguenay, estimée à l'aide d'un modèle démographique simple (Labuda *et al.* 1996); on y remarquera que les trois premières vagues, pourtant constituées d'un petit nombre de fondateurs, ont contribué pour 75 % de la population recensée en 1849.

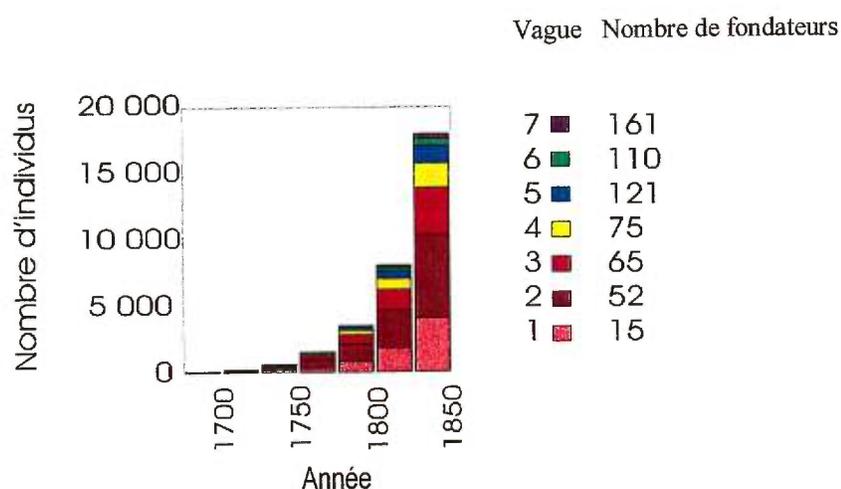


Figure 1 : Modèle démographique illustrant la contribution de différentes vagues de fondateurs au pool génétique de la région de Charlevoix/Saguenay-Lac-St-Jean (Labuda *et al.* 1996).

Effet fondateur

Ce qu'il est important de noter dans l'histoire de la population de la région de Charlevoix-Saguenay c'est qu'en fait elle représente un échantillon de la population de la Nouvelle-France qui elle-même est un échantillon de la population française. En effet, une fraction de la population de la France est venue coloniser la Nouvelle-France et a par la suite connu une expansion démographique, tout en étant isolée de la population mère. De cette nouvelle population, une poignée d'immigrants a colonisé la région de Charlevoix-Saguenay qui elle-même a par la suite connu une expansion démographique et un certain isolement du reste de la population. Il découle de ce phénomène une certaine homogénéisation de la population qui peut être démontrée entre autres par les patronymes retrouvés dans cette région. Bouchard (1991) a démontré, d'après les annuaires téléphoniques de la région, qu'environ 30 % de la population étaient porteurs des 15 patronymes les plus fréquents au Québec, contre seulement 2,2 % dans la région de Montréal et pour une moyenne de 5,6 % pour la province de Québec dans son ensemble. Cette homogénéisation peut également être responsable des taux de porteurs élevés de certaines maladies héréditaires dans la population de Charlevoix-Saguenay. Le tableau I montre quelques exemples des taux de porteurs de telles génopathies dans la région du Saguenay comparativement à d'autres populations.

On voit sur ce tableau que plusieurs maladies ont des taux de porteurs très élevés dans la population du Saguenay par rapport aux autres populations; à titre de comparaison, on observe un taux de porteurs de 1/25 pour la fibrose kystique qui est

la maladie génétique la plus fréquente dans la population caucasienne. On peut également observer le phénomène inverse avec l'ataxie de Friedreich qui est pratiquement inconnue dans la population de Charlevoix-Saguenay mais cette maladie montre un taux de porteurs élevé ailleurs au Québec et dans le monde (0,6 à 0,8 %). Le phénomène responsable de ce profil particulier de maladies génétiques est nommé effet fondateur et se définit comme étant un changement de la composition

Tableau I : Taux de porteurs de diverses génopathies autosomales récessives dans la population du Saguenay et dans d'autres populations.

| Maladie héréditaire | Fréquence des porteur(e)s | |
|------------------------------------|---------------------------|------------------------|
| | Saguenay ^a | ailleurs ^b |
| PDDR | 1/26 | très rare |
| Déficience en lipoprotéine lipase | 1/43 | très rare |
| Histidinémie | 1/32 | 1/50-1/150 |
| Polyneuropathie | 1/25 | inconnue |
| Déficience en cytochrome C oxidase | 1/28 | très rare |
| Mucopolysaccharidose II | 1/39 | rare |
| Sarcosinémie | 1/29 | 1/300 Nouv. Angleterre |
| Cystinose | 1/39 | 1/200 |
| Ataxie-spastique | 1/21 | inconnue |
| Tyrosinémie I | 1/22 | 1/165 Norvège., Suède |
| Maladie de Tay-Sachs | 0 | 1/67 ÉU, Québec |
| Déficience en pyruvate kinase | 1/64 | rare |
| <u>Ataxie de Friedreich</u> | <u>0</u> | <u>1/120-1/160</u> |

^a *De Braekeleer* (1991)

^b *Scriver et al.* (1995)

génétique d'une population par rapport à sa population mère, suite à un isolement. Mayr (1970) traite de ces changements et de leur causes et conséquences dans un ouvrage intitulé «Populations, Species and Evolution». Des motifs d'ordre religieux

ou social, ou encore la géographie de certaines régions peuvent être à l'origine d'un tel isolement. Les exemples d'effet fondateur sont nombreux parmi les populations ayant été isolées à un moment ou un autre de leur histoire : les Azoréens (Lima *et al.* 1997), les Polynésiens (Lum et Cann 1998) et les Finlandais (Kajanoja 1971) sont des populations qui ont longtemps été isolées géographiquement; les juifs ashkénaze ont été isolés pour des motifs d'ordre sociaux et religieux (Buchbinder *et al.* 1997). L'isolement de la population de Charlevoix-Saguenay est surtout un isolement géographique, quoique les traditions catholiques canadiennes-françaises aient également contribué à un certain isolement social (Bouchard 1991). Dans toutes ces populations, et dans plusieurs autres, l'effet fondateur a été observé par des augmentations de certaines génopathies par rapport aux fréquences observées dans d'autres populations. Le processus biologique de l'effet fondateur est relativement simple en soi; imaginons par exemple une population de 1000 habitants parmi laquelle une dizaine de maladies génétiques y sont observées, chacune ayant une fréquence de 1/1000. De cette population, un échantillon de 100 habitants émigre vers une région qui l'isolera de la population mère et l'échantillonnage fait que, parmi ces 100 fondateurs, un est porteur d'une des maladies génétiques. La fréquence de cette maladie est maintenant dix fois plus élevée, soit de 1/100 parmi ces fondateurs. La dérive génétique peut ensuite augmenter davantage cette fréquence, par exemple par le fait que ce fondateur peut avoir de nombreux descendants porteurs de la maladie. Chez les Canadiens-Français de Charlevoix-Saguenay, il a d'ailleurs été démontré que, même si les premières migrations furent propices à l'échantillonnage

de certaines mutations, c'est surtout l'expansion démographique suivant ces migrations qui contribua à l'établissement de l'effet fondateur (Labuda *et al.* 1996 et 1997).

Plusieurs exemples de populations ayant des fréquences élevées de porteurs de certaines maladies ont été documentés dans cette perspective. Par exemple, la phénylcétonurie chez les Juifs yéménites (Avigad *et al.* 1990) ainsi que la leukodystrophie métachromatique chez les Juifs habbanites (Zlotogora *et al.* 1980) et la maladie de Gaucher chez les habitants de la province de Norrbotten en Suède (Dahl *et al.* 1990) sont toutes des maladies héréditaires pour lesquelles les patients atteints sont homozygotes pour la même mutation. Zlotogora (1994) en conclut que ces taux élevés de porteurs sont le résultat d'un effet fondateur accompagné d'une dérive génétique. En fait, la dérive génétique contribue à l'effet fondateur ; c'est un phénomène par lequel survient un changement dans la fréquence de différents allèles dû au hasard de l'union des gamètes dans le processus de reproduction (Hartl et Clark 1989). Il a été démontré que la thalassémie chez les Juifs kurdes (Rund *et al.* 1991) est une maladie très fréquente chez cette population (20% de porteurs) mais plusieurs mutations différentes peuvent être responsables. Ceci suggère qu'il pourrait exister un avantage sélectif pour les porteurs hétérozygotes de ces mutations (Zlotogora 1994). Une étude plus poussée a clairement démontré ce phénomène d'avantage sélectif des hétérozygotes : Witchel *et al.* (1997) ont démontré que les porteurs d'une mutation dans le gène de la 21-hydroxylase (CYP21) avaient une réponse en cortisol plus élevée que les individus ne portant pas la mutation et que cette forte réponse en

cortisol pouvait mener à un retour rapide à l'homéostasie suite à un stress infectieux ou inflammatoire. L'effet fondateur n'est donc pas le seul phénomène pouvant être responsable de fréquences élevées de maladies génétiques. L'exemple classique d'avantage des hétérozygotes est celui de l'anémie falciforme en Afrique : les porteurs hétérozygotes de cette affection ne sont pas sévèrement touchés par la maladie et possèdent une certaine résistance à la malaria car le protozoaire responsable ne peut se développer dans les globules rouges mal formés des porteurs de l'anémie falciforme. Au fil des siècles une sélection positive s'est donc opérée sur le locus de l'anémie falciforme à cause de l'avantage conféré aux hétérozygotes.

Modèles génétiques et choix des marqueurs

Plusieurs modèles génétiques peuvent être utilisés pour les études de populations; le choix du modèle dépend principalement de la nature des travaux. Ainsi, certains étudieront l'ADN mitochondrial et d'autres choisiront l'ADN nucléaire. L'ADN mitochondrial est de transmission maternelle et son évolution (dans la région contrôle) est de 10 à 40 fois plus rapide que celle des parties non codantes de l'ADN nucléaire (Zischler *et al.* 1995) ; ces propriétés en font donc un outil parfait pour l'étude de l'évolution des lignées femelles et pour les populations qui sont jeunes et dont l'histoire évolutive est courte. Ceux qui choisissent de travailler avec l'ADN nucléaire savent que les systèmes génétiques y sont très diversifiés, à commencer par le choix du chromosome (autosome ou sexuel) qui est

souvent guidé par la nature des études; en effet, plusieurs d'entre eux ont des particularités les démarquant tout spécialement des autres.

Le chromosome Y par exemple est exclusivement transmis de père en fils, ce qui en fait l'outil par excellence pour les études sur l'évolution des lignées mâles. Le chromosome X étant retrouvé en une seule copie chez les mâles, ceci facilite l'haplotypage puisque tous les sujets mâles sont hémizygotés pour les allèles s'y trouvant. Il n'y a donc pas le problème de la résolution de deux haplotypes différents par individu. Hormis le choix du chromosome, la diversité des systèmes génétiques est conférée par la nature des sites de l'ADN : les régions codantes évoluent différemment des régions non codantes puisque ces dernières sont moins souvent soumises aux contraintes phénotypiques. En effet les sites neutres de l'ADN, retrouvés dans les régions non codantes, évoluent plus vite puisque les mutations y survenant n'ont pas d'effet sur le phénotype. Le choix des marqueurs est également source de diversité en termes de systèmes génétiques. Les marqueurs peuvent être de plusieurs nature : microsatellites, mutations ponctuelles, insertions/délétions font partie des différentes catégories de polymorphismes et ils évoluent tous de façons différentes, c'est-à-dire qu'ils ont des taux de mutations différents. Les microsatellites par exemple sont des répétitions en tandem de un à six nucléotides répétés jusqu'à environ 100 fois; ils sont extrêmement polymorphes (jusqu'à plusieurs dizaines d'allèles par locus, Tautz 1993) et leur taux de mutation est très élevé : de 10^{-3} à 10^{-4} /locus/génération (Zhivotovsky et Feldman 1995) comparativement à environ 2×10^{-8} /nucléotide/génération pour l'ADN nucléaire en

général. Autrement dit, pour avoir un nombre équivalent de mutations il faudrait comparer un seul microsatellite avec un locus nucléaire comportant de 5 à 50 kbs. Pour cette raison les microsatellites procurent des systèmes génétiques hautement informatifs, c'est à dire que le degré d'hétérozygotie est très élevé, et sont très utiles pour la taxonomie moléculaire, les études évolutives et la génétique des populations, surtout lorsque l'histoire de celles-ci est relativement récente.

L'étude présentée ici se base précisément sur l'allélotypage de plusieurs marqueurs microsatellites et sur l'haplotypage effectué avec ces marqueurs afin d'étudier la présence de l'effet fondateur chez les Canadiens-Français. Ces haplotypes sont exclusivement établis à partir de marqueurs génétiques situés dans des sites que l'on suppose neutres de l'ADN qui, par définition, ne peuvent conférer un avantage sélectif au porteur. La raison de l'utilisation de sites neutres réside en grande partie dans un souci d'éviter l'effet de la sélection comme étant l'explication de fréquences élevées de certains allèles.

Hypothèses

Bien que la fréquence élevée de porteurs de certaines génopathies peut quelquefois être imputable à l'avantage sélectif des hétérozygotes, il est peu probable que ce soit le cas pour toutes les génopathies qui présentent des taux de porteurs anormalement élevés dans la population de Charlevoix-Saguenay (voir tableau I). Ceci dit, l'effet fondateur est probablement responsable de la majorité de ces fréquences élevées, ce qui nous amène à postuler ce phénomène serait observable au

niveau des sites neutres de l'ADN nucléaire. Si l'on se réfère au tableau I on observe que les maladies héréditaires de fréquences élevées y figurant (qui sont autosomales récessives) ont en moyenne un taux de porteur de $1/25$ (0,04) environ, soit une fréquence de 0,02 mutation par chromosome. Notre population est constituée de 106 chromosomes X indépendants ; il y aurait donc toute les chances pour qu'au moins deux chromosomes soient porteurs d'un allèle ou d'un haplotype très rare dont l'échantillonnage serait dû à l'effet fondateur. En d'autres mots, si l'effet fondateur est le principal responsable de l'augmentation en fréquence des génopathies dans Charlevoix-Saguenay nous devrions observer, parmi les 106 chromosomes indépendants de cette région, au moins deux chromosomes portant le même haplotype qui serait très rare ailleurs. À l'inverse, les populations de la France urbaine, de la région de Montréal et des familles CEPH ne devraient pas montrer un nombre élevé d'haplotypes répétés. L'effet fondateur peut donc avoir diminué, dans une certaine mesure, la diversité génétique dans la population de Charlevoix-Saguenay.

Objectifs

La présente étude a pour premier objectif d'étudier la variabilité génétique de la population dont l'histoire démographique est connue depuis sa création. Le second vise à déterminer si l'effet fondateur dont témoigne les fréquences élevées de certaines génopathies peut être observé par l'étude de polymorphismes neutres et si oui, dans quelle mesure cet effet fondateur a entraîné une diminution de la variabilité

génétique dans cette population. Pour ce faire, un haplotype hautement informatif comprenant cinq marqueurs microsatellites et une insertion d'un élément Alu sera établi pour des échantillons d'une population de Charlevoix-Saguenay et pour des échantillons issus de la population canadienne-française de la région de Montréal, de la France et d'un groupe d'origine européenne. Des échantillons provenant des familles CEPH, décrites plus loin, seront également étudiés. Ces derniers constitueront un groupe extérieur aux autres populations de l'étude et seront comparés à ces dernières.

Chapitre 2 : Matériel et méthode

Échantillons biologiques

Trois principales populations ont été étudiées dans cette étude; la population de Charlevoix-Saguenay était représentée par 153 chromosomes X parmi lesquels 106 sont indépendants (aucun lien de parenté jusqu'à la troisième génération). Ces échantillons ont été généreusement fournis par le Dr Jack Puymirat du Centre de Recherche du CHUL. Les généalogies ascendantes de la population de Charlevoix-Saguenay sont connues, permettant éventuellement l'élaboration d'une étude généalogique en parallèle avec l'étude génétique. La deuxième population est représentée par une cinquantaine de chromosomes provenant de la France urbaine dont les échantillons nous ont gracieusement été fournis par le Dr Jean-Paul Moisan du centre de recherche de l'Université de Nantes; la troisième population comprend autant de chromosomes de Canadiens-Français de la région de Montréal (les échantillons proviennent d'individus dont le lieu de naissance est le Québec et qui sont de langue maternelle française) grâce à la collaboration des Drs Juan Carlos Feoli-Fonseca et Marie Lambert du Service de génétique du centre de recherche de l'Hôpital Ste-Justine. Des données sur plusieurs échantillons d'origine européenne provenant de la littérature et des travaux effectués au laboratoire constitueront un quatrième groupe. Il sera également question, dans l'analyse des haplotypes, d'un groupe constitué de huit familles CEPH (Centre du Polymorphisme Humain); pour plus d'informations, voir le site web <http://locus.umdj.edu/nigms>. Les individus

constituant ces familles sont d'origine européenne, six d'entre elles sont de l'Utah (États-Unis), l'une provient du Venezuela et la dernière est une famille Amish (États-Unis). Ces familles sont principalement étudiées pour les analyses de mutations et de recombinaisons au sein du système génétique établi mais puisque les haplotypes ont été constitués pour ces échantillons, ils seront également comparés aux populations de Charlevoix-Saguenay, de France et de la région de Montréal.

Modèle génétique à l'étude

Le modèle génétique (voir figure 2) est constitué des six marqueurs polymorphes suivant : quatre marqueurs microsatellites de type répétitions CA, il s'agit des marqueurs DXS1238, IVS44SK21, DXS1237 et DXS1235 ; les deux autres marqueurs sont un microsatellite de type répétition T (11POL) et une insertion d'un élément Alu (Dys44Alu). Les six marqueurs sont situés dans les régions introniques du gène de la dystrophine (Xp21) couvrant la région allant de l'intron 44 à l'intron 50. Les marqueurs DXS1238, DXS1237 et DXS1235 se sont avérés très hétérozygotes lorsqu'ils ont été décrits pour la première fois par Clemens *et al.* (1991) dans le but d'en faire un outil diagnostique pour les familles atteintes de la dystrophie musculaire de Duchenne ou de Becker. Dans le même but, Köchling *et al.* ont décrit le marqueur IVS44SK21 en 1995. Le marqueur 11POL a été décrit par Zietkiewicz *et al.* (1997). Enfin, l'insertion de l'élément Alu est un marqueur biallélique (présence ou absence de l'élément Alu au site d'insertion) et la présence

des deux allèles a été observée dans plusieurs populations (données non publiées)¹. L'ensemble de ces marqueurs forme ainsi un haplotype hautement informatif qui sera utilisé dans cette étude. Il sera également question de polymorphismes, qui avaient déjà été allélotypés par buvardage en taches pour la population de Charlevoix-Saguenay (données non publiées) et qui se situent dans une région de 8 kbs encadrant l'exon 44 du gène de la dystrophine, nommée dys44 (Zietkiewicz *et al.* 1997 et 1998). Ces polymorphismes, au nombre de 28, sont tous bialléliques et sont pour la plupart des substitutions de bases, à l'exception de deux insertions/délétions de trois nucléotides. Trois fragments amplifiés par PCR constituent l'ensemble de la région dys44 ; il s'agit des fragments 1-5, 5-11 et 12-15.

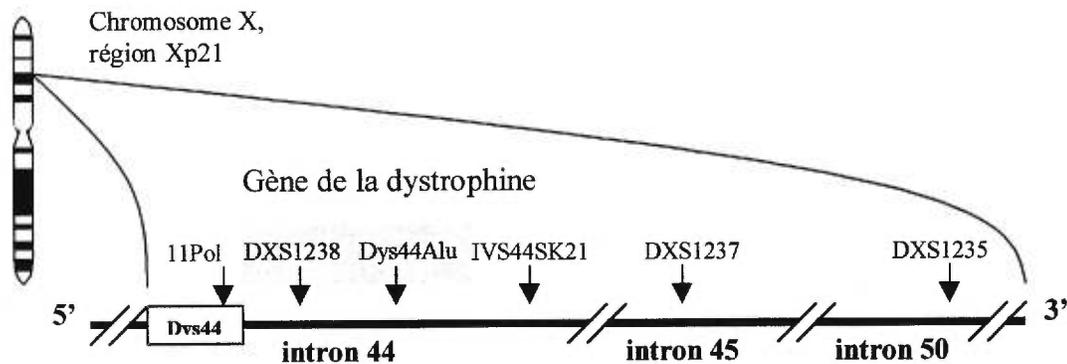


Figure 2 : Modèle génétique représentant la portion du gène de la dystrophine étudiée (chromosome X) et les différents marqueurs y étant situés; les marqueurs situés dans la zone ombragée sont ceux pour lesquels la position relative n'est pas connue avec exactitude; ainsi les marqueurs Dys44Alu et DXS1238 pourraient être dans l'ordre inversé.

¹ Les données non publiées réfèrent à des travaux effectués au laboratoire d'accueil.

Tableau II : Conditions de PCR et amorces utilisées pour l'amplification des marqueurs étudiés.

| Marqueur | Amorces (5' → 3') | Conditions de PCR |
|-------------------|---|--|
| DXS1238 | ATTCGGAGGTGATAAACATC TCAGCATTCATCACAAATAG | Dénaturation : 94°C, 3' (1 cycle) Dénaturation : 94°C, 20'' } Appariement : 48°C, 20'' } (37 cycles) Élongation : 72°C, 25'' } Élongation : 72°C, 7' (1 cycle) |
| IVS44SK21 | TAATGAAAGTAAAAAATG TTGCTCAGTGAATGAATA | Dénaturation : 94°C, 3' (1 cycle) Dénaturation : 94°C, 20'' } Appariement : 48°C, 20'' } (37 cycles) Élongation : 72°C, 25'' } Élongation : 72°C, 7' (1 cycle) |
| DXS1237 | GAGGCTATAATTCITTAACTTTG CTCTTCCCTCTTATTTCATGTTA | Dénaturation : 94°C, 3' (1 cycle) Dénaturation : 94°C, 20'' } Appariement : 48°C, 20'' } (37 cycles) Élongation : 72°C, 25'' } Élongation : 72°C, 7' (1 cycle) |
| DXS1235 | AAGGTTCCCTCCAGTAACAGATTT TATGCTACATAGTATGTCCTCAG | Dénaturation : 94°C, 3' (1 cycle) Dénaturation : 94°C, 20'' } Appariement : 53°C, 20'' } (37 cycles) Élongation : 72°C, 25'' } Élongation : 72°C, 7' (1 cycle) |
| 11Pol | GTTGAAAACCGTTGTAG GCTGAACTTAATCTCCT | Dénaturation : 94°C, 3' (1 cycle) Dénaturation : 94°C, 20'' } Appariement : 48°C, 20'' } (37 cycles) Élongation : 72°C, 25'' } Élongation : 72°C, 7' (1 cycle) |
| Dys44Alu | CATGGTAGACATGAATT TATAGTGGGTAAAAGAT | Dénaturation : 94°C, 3' (1 cycle) Dénaturation : 94°C, 30'' } Appariement : 50°C, 30'' } (37 cycles) Élongation : 72°C, 45'' } Élongation : 72°C, 7' (1 cycle) |
| Fragment : 1-5 | AGGGGGATTTGTTGAAT GCGTATATTTTTTGGTTATA | Dénaturation : 93°C, 3' (1 cycle) Dénaturation : 93°C, 30'' } Appariement : 52°C, 45'' } (39 cycles) Élongation : 72°C, 2' } |
| 5-11 | TGGTAACTTTGTTTCATATTA ATGAGGACCATTAGACATTC | |
| 12-15 | GTTATTGTTTCTACTGGCAACT | Élongation : 72°C, 7' (1 cycle) |

Marquage radioactif des amorces

La liste des amorces utilisées figure dans le tableau II. Le marquage radioactif était effectué dans un volume final de 50 μl en incubant l'amorce à 4,8 μM , l'enzyme kinase T4 à 37 °C (1 U/ μl) et le $\gamma\text{-[P}^{32}\text{]dATP}$ à une activité spécifique de 1 à 3×10^6 cpm/pmole, dans le tampon de réaction de la kinase (Imidazole 25 mM (pH 6,4) ; MgCl_2 15 mM ; 2-mercaptoéthanol 1 mM ; ADP 0,05 mM). Après une incubation d'une heure à 37 °C, 150 μl d' H_2O étaient ajoutés. Il en résultait une solution d'amorce à 1,2 μM .

Dans un volume final de 10 μl , les sondes pour le buvardage en taches d'une concentration de 8,3 μM étaient incubées avec l'enzyme kinase T4 à 37 °C (1 U/ μl) et le $\gamma\text{-[P}^{32}\text{]dATP}$ à une activité spécifique de 1 à 3×10^6 cpm/pmole, dans le tampon de réaction de la kinase (voir ci-haut). L'incubation était conduite à 37 °C pendant une heure et le volume était complété à 100 μl . La solution était ensuite passée sur colonne pour éliminer le $\gamma\text{-[P}^{32}\text{]dATP}$ non incorporé. L'activité de la sonde était vérifiée dans un compteur à scintillation Wallac 1409 et se situait entre 25 000 et 500 000 cpm / μl .

Amplification par PCR (Réaction de polymérisation en chaîne)

Les amorces utilisées, ainsi que les conditions d'amplification sont inscrites dans le tableau II. Pour les marqueurs IVS44SK21, DXS1237, DXS1236 et DXS1235, les réactions étaient conduites avec une concentration d'ADN génomique

de 0,5 ng/ μ l dans une solution contenant les deux amorces (2 μ M d'amorce froide antisens, 1,6 μ M pour d'amorce froide sens et 0,1 μ M de l'amorce sens marquée au [P^{32}]), du $MgCl_2$ 1,5 mM, de l'ADN polymérase *Taq* 0,05 U/ μ l et le tampon de PCR fournit avec l'ADN polymérase *Taq* de la compagnie GibcoBRL (Tris-HCL 20 mM pH 8,4; KCl 50 mM). Pour le marqueur 11Pol, les mêmes conditions étaient appliquées mais les concentrations d'amorces étaient deux fois moindres. Enfin, pour le marqueur Dys44Alu, les fragments 1-5, 5-11 et 12-15, les mêmes conditions étaient également appliquées mais sans utiliser d'amorces radioactives. Chaque réaction était effectuée dans un volume final de 5 μ l dans le cas des marqueurs 11Pol, IVS44SK21, DXS1237, DXS1236 et DXS1235 ; un volume final de 20 μ l dans le cas du marqueur Dys44Alu ; et un volume final de 50 μ l dans le cas des fragments 1-5, 5-11 et 12-15. Pour éviter l'évaporation des produits pendant les réactions de PCR, une goutte d'huile minérale était ajoutée dans les tubes.

Analyse des marqueurs microsatellites sur gel de polyacrylamide dénaturant

Les gels de polyacrylamide dénaturants étaient polymérisés en ajoutant 120 μ l de persulfate d'ammonium 25% et 50 μ l de TEMED (N,N,N',N'-tétraméthyléthylènediamine) à 80 ml d'une solution d'acrylamide (acrylamide :bisacrylamide 19 :1 6% ; urée 0,42 g/ml, Tris-borate 90 mM, pH 8; EDTA 2,5 mM). Les produits de PCR à analyser étaient dénaturés en mettant les échantillons à 94 °C pendant 4 minutes et en ajoutant préalablement 7,5 μ l de tampon

stop (EDTA 20 mM ; bleu de bromophénol 0,1% ; xylène cyanole 0,1% ; formamide 85%) à 5 µl de produit de PCR. Les échantillons étaient par la suite mis sur glace en attendant d'être chargés sur le gel de 35 x 50 cm. Par une prémigration à 50 milliampères, le gel de polyacrylamide était chauffé dans son tampon (Tris-borate 90 mM, pH 8 ; EDTA 2,5 mM) et lorsque la température du gel atteignait 50 °C, les produits de PCR y étaient chargés. La migration était effectuée à 50 milliampères pour une durée de 3 heures dans le cas des marqueurs IVS44SK21, DXS1237, DXS1236 et DXS1235, et pour une durée de 2 heures dans le cas du marqueur 11Pol. Les gels étaient ensuite séchés sur un papier Wattman et l'autoradiographie se faisait à la température de la pièce ou à 75 °C, de 15 à 48 heures, dépendamment de l'intensité du signal radioactif qui était préalablement estimée au compteur Geiger.

Analyse du marqueur Dys44Alu

Les produits de PCR du marqueur Dys44Alu étaient analysés sur gel d'agarose (agarose 1%, Tris 90 mM ; acide borique 89 mM ; EDTA 2,5 mM, bromure d'éthidium 0,5 µg/ml) dans un tampon TBE (Tris 90 mM ; acide borique 89 mM ; EDTA 2,5 mM). La migration se faisait à 120 volts pendant une heure. Les bandes d'ADN étaient révélées sous une lampe à ultraviolets Foto/Prep®.

Tableau III : Séquences des sondes utilisées pour l'hybridation ASO par buvardage en taches. TH : température d'hybridation; F et R : oligonucléotides complémentaires au brin - et + respectivement; les positions alléliques des sondes sont indiquées en gras; aussi souvent qu'il est possible, les sondes figurent en une seule séquence et les variants sont en minuscule et séparés par « / ».

| Fragment | Site polymorphe | Séquence de la sonde | TH (°C) |
|----------|-----------------|----------------------------------|---------|
| 1-5 | 2A-3A | F TCT GTG AAA CAG GTT | 37 |
| | 2C-3C | F TCT GTG CCA CAG GTT | |
| | 2C-3A-5 | F TCT GTG CAA CAg/a GTT | |
| 1-5 | 8 | F AGT AGC TAA a/c AG TGT | 37 |
| 1-5 | 10 | F TTC t/c GA CTC TCA ATA | 37 |
| 1-5 | 14 | F CTT c/t GG CTA CCT TCG | 38 |
| 1-5 | 20 | F ATG CTT GCT TGG a/g CA | 43 |
| 1-5 | 35 | F AGT TTC c/t TG CAT TTG | 37 |
| 5-11 | 38 | F GTT GTC atc ATT ATA | 37 |
| | | F GTT GTC --- ATT ATA TTA | |
| 5-11 | 40 | F ATA Ga/g G AAA CAG CAT | 37 |
| 5-11 | 45 | R GGC c/a TT AAA AAT TGG | 37 |
| 5-11 | 50 | F AGG TTA Ac/tc AGG GAG | 37 |
| 5-11 | 55 | F CCA TGT AGT AT a/g TAT | 37 |
| 5-11 | 58T-59 | R CCA Ac/a T TTT ACT TCT | 37 |
| | 58A-59 | R CCA Ac/a T TAT ACT TCT | |
| 5-11 | 64A-65 | F AAT ATT CTC AAc/a CCT | 37 |
| | 64C-65 | R AGG t/g CT GAG AAT ATT | |
| 11-15 | 70 | F TTT AAC Ag/aA AGC CTC | 37 |
| 11-15 | 71 | F AAC CGt/c TGT AGC ATA | 37 |
| 11-15 | 85 | F CTG GTC TAG ATC TC- --- --- -C | 38 |
| | | R G ATC TCt aga tct Cc | |
| 11-15 | 90 | F ATC CAC CAA AAc/t GTC T | 38 |
| 11-15 | 93 | F ACT GAA TTA Tc/tg TCT | 35 |

Hybridation oligonucléotide allèle-spécifique (ASO)

Cette technique avait pour but d'identifier des polymorphismes bialléliques, c'est à dire un locus sur lequel il n'y avait que deux formes alléliques possibles (Zietkiewicz *et al.* 1997). Pour ce faire, une membrane était séparée en deux parties et chaque moitié était chargée à l'aide d'un collecteur "dot blot" (buvardage en taches) avec plusieurs échantillons d'ADN (chaque échantillon en duplicata). L'ADN de la première moitié était ensuite hybridé avec une sonde d'oligonucléotide contenant la première forme allélique ; parallèlement, l'ADN de la seconde moitié de la membrane était hybridé à une seconde sonde ayant la deuxième forme allélique.

Préparation de la membrane. Une membrane de nylon Hybond™ N+ était mise sur un papier Wattman et installée dans le collecteur "dot blot". Chaque puits était rincé deux fois avec du tampon SSC 10X (NaCl 1,5 M; citrate de sodium 0,15 M ; pH ajusté à 7,0 avec NaOH). À 60 µl d'H₂O étaient ajoutés 40 µl de produit de PCR et 50 µl de cette solution étaient chargés sur chaque moitié de membrane. Les puits étaient ensuite rincés avec 100 µl de tampon SSC 10X. La membrane était ensuite submergée 10 minutes dans une solution dénaturante (NaCl 1,5 M ; TrisHCl (pH 7,2) 0,5 M ; EDTA 1 mM), et 15 minutes dans une solution neutralisante (NaCl 1,5 M ; NaOH 0,5 N). L'ADN était fixé sur la membrane en exposant celle-ci à la lumière ultraviolette (800 µjoules).

Hybridation. La membrane était coupée en ses deux parties jumelles (l'ADN ayant été chargé en duplicata) et ces deux moitiés étaient séparées et s'en suivait une

pré-hybridation à 37 °C pendant 30 minutes, dans 5 ml de solution d'hybridation : SSPE 20X (NaCl 3 M ; phosphate de sodium monohydraté 0,2 M ; EDTA 25 mM ; pH ajusté à 7,0 avec NaOH) ; NaCl 750 mM ; SDS 1% ; héparine 0,8 µg/ml ; Tris 70 µM). Une des membranes était testée pour la présence de la première forme allélique et la seconde moitié était testée pour la seconde forme allélique. Les sondes étaient ajoutées à une activité finale de 1 000 000 cpm/ml chacune et l'hybridation se faisait dans un four à hybridation pendant 30 minutes (voir tableau III pour les températures d'hybridation des différentes sondes).

Lavage. Après l'hybridation, le surplus de sondes non hybridées était éliminé par un premier lavage en submergeant la membrane quelques secondes dans la solution de lavage (SSPE 2X ; SDS 0,1%). La membrane était lavée une deuxième fois en laissant la membrane dans 10 ml de la solution de lavage pendant 10 à 15 minutes à la même température que l'hybridation.

Analyses informatiques

Les données sur les fréquences alléliques des différents marqueurs utilisés étaient compilées sur un fichier Microsoft ® Excel; les écarts types étaient calculés d'après une distribution binomiale comme suit :

$$\text{Écart type} = \sqrt{\frac{p(1-p)}{n}}$$

où p est la fréquence relative de l'allèle et n le nombre de chromosomes.

Ces données étaient ensuite analysées à l'aide de l'ensemble de programmes Phylip 3.5 (Felsenstein 1993), © 1986-1995 par Joseph Felsenstein et l'Université de Washington : les données étaient traitées à l'aide programme GENDIST (© 1986-1993 par Joseph Felsenstein et l'Université de Washington) qui établissaient les distances génétiques entre les différentes populations selon les fréquences alléliques, d'après la méthode de Cavalli-Sforza (Cavalli-Sforza et Edwards 1967) où la distance (D) était déterminée comme suit :

$$D^2 = \frac{4 \sum_m [1 - \sum_i \sqrt{p_{1mi}} \sqrt{p_{2mi}}]}{\sum_m (a_m - 1)}$$

Où p_{1mi} est la fréquence du i -ième allèle au m -ième locus dans la population 1, p_{2mi} est la fréquence du i -ième allèle au m -ième locus dans la population 2 et où a est le nombre d'allèles au m -ième locus. Les données générées par GENDIST étaient ensuite transférées sur un fichier NEIGHBOR qui élaborait les arbres de distances génétiques d'après l'algorithme Neighbor-Joining développé par Nei et Saitou (1987).

Les haplotypes constitués des variants alléliques des différents marqueurs ont été analysés à l'aide du logiciel Arlequin 1.1 (<http://acasunl.unige.ch/~laurent/arlequin/help/ARLEQUIN.html>), © 1995-1998 par Laurent Excoffier (Schneider 1997); les tests et paramètres suivants ont été étudiés : les variations inter et intra-populations étaient déterminées par le test AMOVA selon les index de fixation (F_{ST}); des matrices de distances étaient également élaborées d'après les F_{ST} des différentes population, déterminés par permutations des

haplotypes parmi les populations. La diversité génétique était définie comme étant la probabilité que deux haplotypes choisis au hasard dans une population soient différents. La diversité (D) était calculée comme suit :

$$D = (n/n-1) \left(1 - \sum_{i=1}^k p_i^2\right)$$

et la variance :

$$V(D) = (2/n(n-1)) \left\{ 2(n-2) \left[\sum_{i=1}^k p_i^3 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right] + \sum_{i=1}^k p_i^2 - \left(\sum_{i=1}^k p_i^2 \right)^2 \right\}$$

où n est le nombre de chromosomes, k est le nombre d'haplotypes et p_i est la fréquence du i-ème haplotype.

Le programme MLINK de l'ensemble LINKAGE (© 1993 par Joseph D. Terwilliger et Jurg Ott) a été utilisé pour les analyses « two-point linkage » calculant les lod scores pour un ensemble de fractions de recombinaison, ainsi que l'intervalle de support d'une unité lod défini comme étant l'intervalle des fractions de recombinaisons correspondant à plus ou moins une unité lod de la valeur maximale de lod score obtenue.

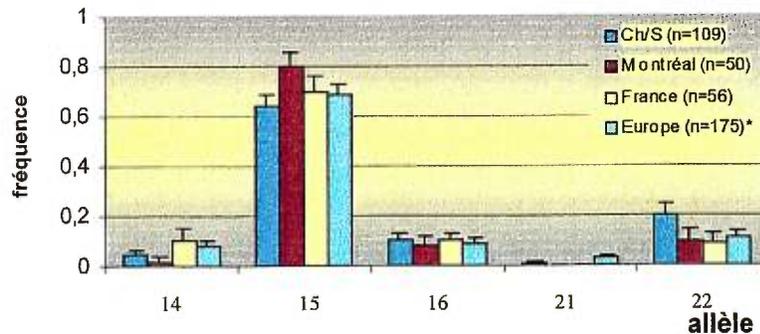
Pour corriger la différence d'échantillonnage des différentes populations (voir figure 10) nous avons utilisé un logiciel a été développé au laboratoire. Ce logiciel choisissait un nombre n d'haplotypes parmi un échantillon plus grand, au hasard et sans remise des haplotypes choisis parmi les haplotypes observés.

Chapitre 3 : Résultats

La variabilité génétique de la population de Charlevoix-Saguenay a été déterminée et comparée à celle des Français, des Canadiens-Français de la région de Montréal et des Européens d'origines variées en étudiant la variabilité allélique et la variabilité haplotypique de ces populations. L'expression de l'effet fondateur était recherchée via une ou des fréquences plus élevée(s) de certains allèles ou certains haplotypes dans le groupe de Charlevoix-Saguenay, comparativement aux autres populations.

3.1- Diversité allélique

Les fréquences alléliques des six marqueurs à l'étude ont été déterminées (voir tableau V, annexe 1) pour les échantillons de Charlevoix-Saguenay et comparées à celles des échantillons des deux autres populations : les Canadiens-Français de la région de Montréal et les Français. De plus, nous disposons de plusieurs données



*Zietkiewicz *et al.* 1997

Figure 3 : Fréquences alléliques du marqueur 11Pol observées dans la population de Charlevoix-Saguenay et trois autres populations.

trouvées dans la littérature qui ont été rapportées dans le cas des marqueurs DXS1238, DXS1237 et DXS1235. Ces données sont celles de 57 chromosomes non apparentés d'origine européenne (Clemens *et al.* 1991); Köchling *et al.* (1995) ont fait une étude semblable pour le marqueur IVS44SK21 sur 100 chromosomes indépendants ; finalement, dans les cas des marqueurs 11Pol (175 chromosomes, Zietkiewicz *et al.* 1997) et Dys44Alu (159 chromosomes, données non publiées), il s'agit également d'échantillons d'origine européenne mixte. Dans les données présentées ici, les numéros désignant les allèles des microsatellites correspondent au nombre de répétitions CA ou T, selon le marqueur, constituant le microsatellite.

Cinq allèles sont retrouvés pour le marqueur 11Pol (figure 3), l'allèle 21 n'étant présent que chez les Européens mixtes et au Charlevoix-Saguenay. Les autres allèles présentent des fréquences semblables pour tous les groupes, à l'exception de l'allèle 22 qui présente une fréquence plus élevée dans le groupe de Charlevoix-Saguenay.

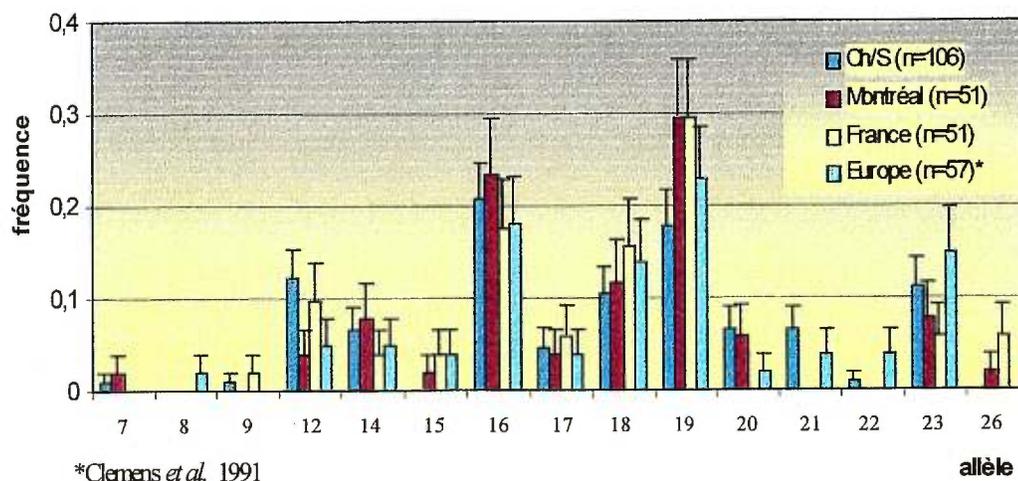


Figure 4 : Fréquences alléliques du marqueur DXS1238 observées dans la population de Charlevoix-Saguenay et trois autres populations.

Le marqueur DXS1238 (figure 4) présente une grande variabilité avec 15 allèles différents observés dans les quatre populations. Les allèles les moins fréquents sont absents dans certaines des populations : l'allèle 8 par exemple n'est présent que chez le groupe étudié par Clemens *et al.* (1991) et l'allèle 7 n'est retrouvé que chez les Canadiens-Français de Montréal et de Charlevoix-Saguenay. Les allèles les plus fréquents sont retrouvés dans toutes les populations à des fréquences relativement semblables, à l'exception de l'allèle 12 dont la fréquence est réduite dans le groupe de la région de Montréal et dans celui de l'étude de Clemens *et al.* (1991) par rapport à la population de Charlevoix-Saguenay. On remarque également l'allèle 21 qui n'est présent que chez les Européens et dans Charlevoix-Saguenay mais qui a une fréquence relativement élevée dans cette dernière population.

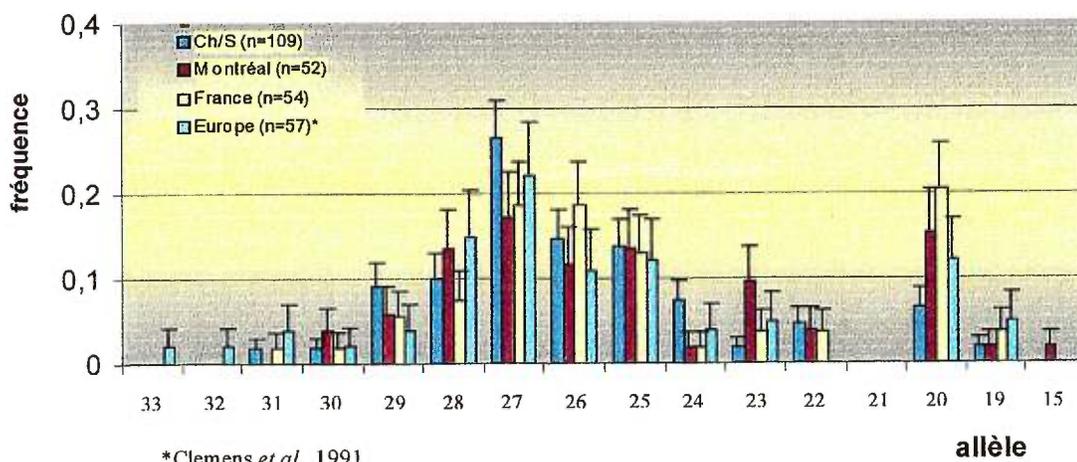


Figure 5 : Fréquences alléliques du marqueur DXS1237 observées dans la population de Charlevoix-Saguenay et trois autres populations.

La distribution des fréquences alléliques du marqueur DXS1237 (figure 5) est semblable à celle du marqueur DXS1238. Trois allèles retiennent ici l'attention : l'allèle 24 est trois fois plus fréquent au Charlevoix-Saguenay que dans la population de la région de Montréal et les Français et environ deux fois plus fréquent que chez le groupe de Clemens *et al.* (1991). L'allèle 20 par contre montre une fréquence de deux à trois fois moins élevée dans la population de Charlevoix-Saguenay que dans les trois autres populations. Enfin, l'allèle 23 présente également une diminution importante de la fréquence dans la population de Charlevoix-Saguenay : la fréquence est jusqu'à cinq fois moindre que pour la région de Montréal. La distribution de fréquences des allèles du marqueur DXS1235 (figure 6) est semblable pour toutes les populations sauf dans le cas de l'allèle 16 qui présente une fréquence diminuée de trois à cinq fois dans la population de Charlevoix-Saguenay comparativement aux trois autres populations.

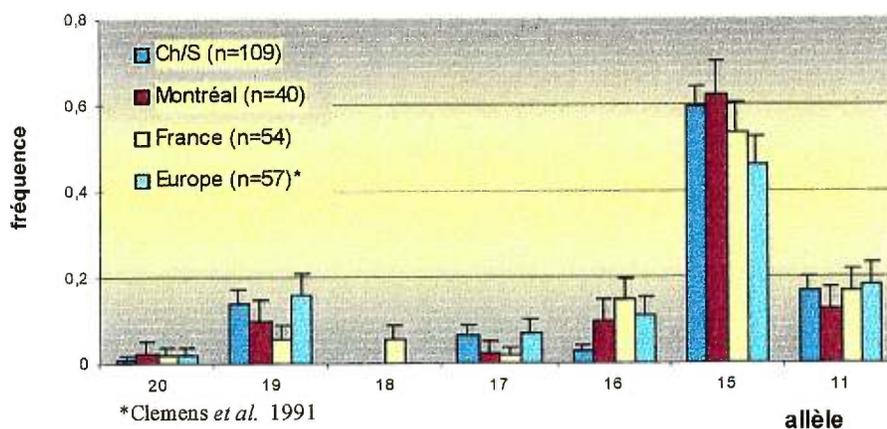
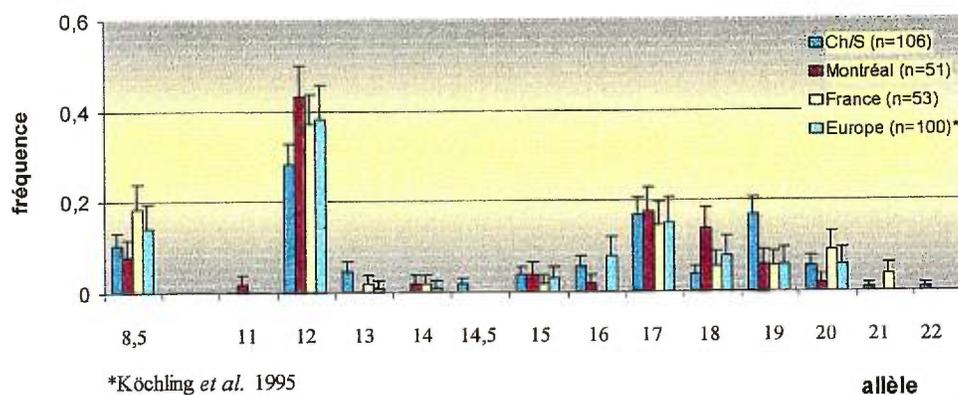


Figure 6 : Fréquences alléliques du marqueur DXS1235 observées dans la population de Charlevoix-Saguenay et trois autres populations.

La figure 7 montre la diversité allélique du marqueur IVS44SK21 pour lequel les populations à l'étude (Charlevoix-Saguenay, région de Montréal et France) sont comparées cette fois à un groupe d'origine européenne étudié par Köchling *et al.* (1995). Les allèles 8,5 et 14,5 contiennent respectivement huit et quatorze répétitions CA plus un résidu nucléotidique dont la nature demeure inconnue. L'allèle 19 de ce marqueur montre une fréquence trois fois plus élevée dans la population de Charlevoix-Saguenay que dans les trois autres populations. Tous les autres allèles, à part l'allèle 20 dont la fréquence est diminuée dans la région de Montréal par rapport



*Köchling *et al.* 1995
 Figure 7: Fréquences alléliques du marqueur IVS44SK21 observées dans la population de Charlevoix-Saguenay et trois autres populations.

aux autres populations, présentent une distribution de fréquences semblable dans les quatre populations.

Pour ce qui est du marqueur Dys44Alu (figure 8), les deux allèles se retrouvent à des fréquences semblables chez tous les groupes étudiés.

l'haplotypage pour les marqueurs de la présente étude. Cependant, nous avons effectué cet haplotypage sur huit familles CEPH afin de déterminer le taux de recombinaison au sein du système génétique à l'étude ; nous reviendrons sur ce sujet plus loin.

La figure 9 (voir tableau VI, annexe 1) illustre de façon graphique l'haplotypage effectué pour les populations étudiées. Il est important de noter que les haplotypes y sont placés arbitrairement sur l'axe des abscisses et ne sont pas nécessairement les mêmes pour les quatre populations. En fait la diversité haplotypique s'est avérée si grande que seulement quelques haplotypes étaient communs à deux populations ou plus, il n'y a donc presque pas de chevauchement et c'est pourquoi, dans un souci de clarté, nous avons regroupé les haplotypes sur l'axe des abscisses sans qu'il ne soient nécessairement identiques. L'haplotypage révèle que la population de Charlevoix-Saguenay contient de nombreux haplotypes répétés comparativement aux autres populations. Ces dernières en effet ne comprennent que des haplotypes uniques, à l'exception des familles CEPH dans lesquelles on retrouve un haplotype répété deux fois (tous les autres étant uniques). La population de Charlevoix-Saguenay par contre compte deux haplotypes répétés quatre fois, trois haplotypes répétés trois fois, huit haplotypes étant retrouvés en double et 72 haplotypes uniques. Il existe cependant une importante différence d'échantillonnage entre les populations de Charlevoix-Saguenay et les autres populations car celles-ci comptent deux à trois fois moins de chromosomes que la population de Charlevoix-Saguenay. Nous avons donc décidé de procéder à une normalisation de

l'échantillonnage dans laquelle 36 haplotypes seulement étaient choisis au hasard parmi ceux de la population de Charlevoix-Saguenay. Cette simulation avait pour but de déterminer les chances de retrouver seulement (ou presque) des haplotypes uniques comme observé dans trois autres populations. En d'autres mots, nous voulions savoir quelles étaient les probabilités de retrouver une distribution de fréquences semblable à celle observée pour la région de Montréal, par exemple, en choisissant, parmi les haplotypes de Charlevoix-Saguenay, un nombre équivalent d'haplotypes. Les résultats illustrés à la figure 10 montrent que la distribution de fréquences des 36 haplotypes choisis au hasard (voir tableau VIII, annexe 1) est

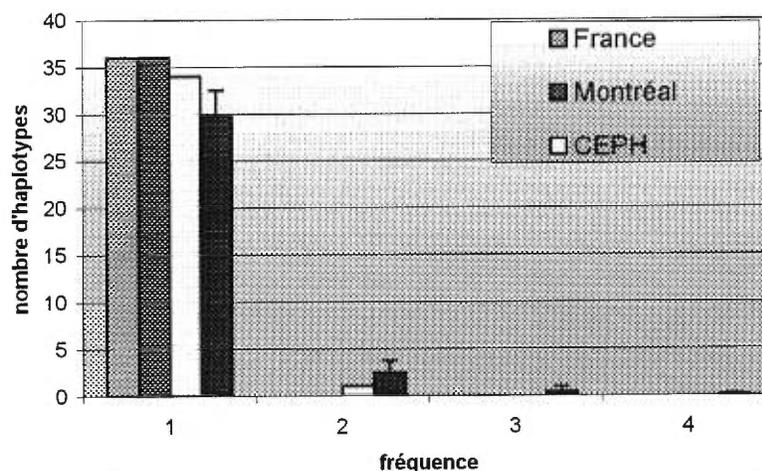


Figure 10 : Distribution de fréquences des haplotypes pour les populations de France, de la région de Montréal et des échantillons CEPH comparée à celle de Charlevoix-Saguenay. Pour avoir une comparaison équivalente, 36 haplotypes ont été choisis au hasard sans remise, et pour un total de 500 fois, parmi ceux de la population de Charlevoix-Saguenay. Les échantillons CEPH et le groupe de la région de Montréal comptaient déjà 36 chromosomes et le groupe de France a été ramené à 36 chromosomes en éliminant 16 chromosomes au hasard. L'écart type était calculé comme suit :

$$\sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

où n est le nombre d'haplotypes choisis et x le nombre de fois qu'un haplotype d'une fréquence donnée a été choisis.

nettement différente pour la population de Charlevoix-Saguenay qui a moins d'haplotypes uniques et plus d'haplotypes répétés. Toutes les populations comportaient 36 chromosomes; les populations de la région de Montréal et les échantillons CEPH comportaient déjà 36 chromosomes et celle de France a été ramenée à 36 chromosomes en excluant 16 chromosomes au hasard sans risque d'affecter la distribution de fréquences puisque tous les haplotypes y étant retrouvés sont uniques.

Étant donné la grande diversité générée par notre système génétique les données ont été analysées en réduisant les haplotypes à deux marqueurs seulement, entre lesquels aucune recombinaison n'a été observée (voir ci-bas). Ainsi plusieurs haplotypes se retrouvent communs à deux populations ou plus. À la figure 11 (voir tableau VII, annexe 1), les populations de France et de la région de Montréal ont été rassemblées en un seul groupe afin que la taille de l'échantillon soit comparable à celui de la population de Charlevoix-Saguenay. On remarque sur cette figure que la population de Charlevoix-Saguenay comporte de nombreux haplotypes dont la fréquence est plus élevée que pour les populations de France de la région de Montréal. L'haplotype 5 entre autres est répété sept fois dans la population de Charlevoix-Saguenay contre seulement deux chez le groupe de France-région de Montréal. Également l'haplotype 11 est répété quatre fois dans la population de Charlevoix-Saguenay et est absent chez le groupe de France-région de Montréal. À l'inverse, l'haplotype 18 est répété six fois chez France-région de Montréal mais est présent une seule fois dans la population de Charlevoix-Saguenay.

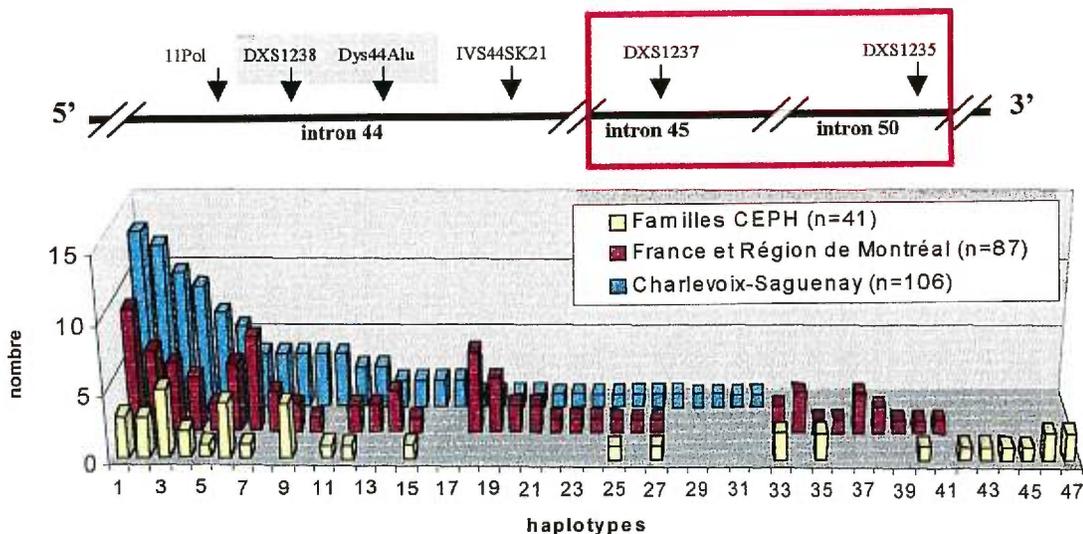


Figure 11 : Distribution de fréquences de l'haplotype constitué des marqueurs DXS1237 et DXS1235 dans la population de Charlevoix-Saguenay et trois autres populations. La portion du système génétique y étant impliquée est illustrée, encadrée de rouge, au dessus de l'histogramme.

Les différents allèles qui composent un haplotype n'ont pas tous des fréquences semblables au sein des populations étudiées (voir la partie « diversité allélique »). La fréquence de ces allèles peuvent donc influencer la fréquence des différents haplotypes ; ainsi, les haplotypes composés d'allèles fréquents sont plus susceptibles de se retrouver en plus grand nombre dans une population donnée. Dans le but de savoir si la fréquence des allèles influençait la fréquence des haplotypes, nous avons étudié les fréquences des allèles qui composaient les haplotypes fréquents et les avons comparées aux fréquences des allèles qui composent les haplotypes uniques. En d'autres mots nous voulions savoir si plusieurs haplotypes pouvaient avoir la même combinaison d'allèles simplement dû au fait que ces allèles sont très

fréquents. Si tel est le cas, les allèles composant les haplotypes fréquents devraient avoir des fréquences plus élevées que ceux composant les haplotypes uniques. Nous avons donc comparé la moyenne des fréquences des allèles composant chaque haplotype aux fréquences absolues de ces haplotypes. Les résultats (figure 12) montrent que les haplotypes répétés ne sont pas composés d'allèles plus fréquents (voir tableau IX, annexe 1).

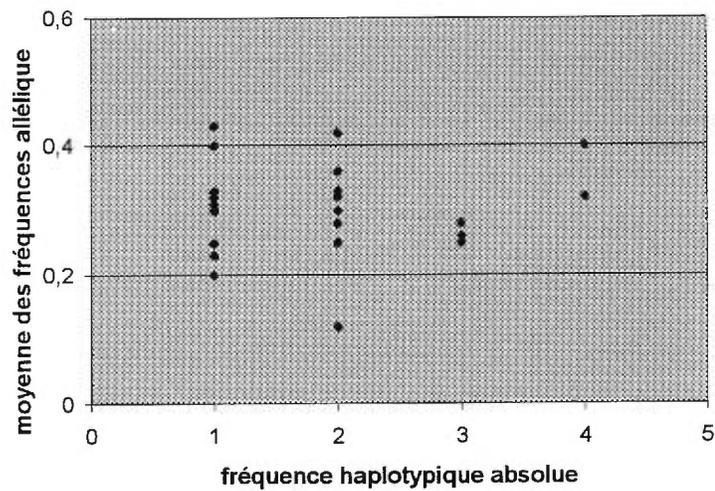


Figure 12 : Moyenne des fréquences alléliques composant chaque haplotype en fonction de la fréquence absolue de chaque haplotype.

2.3- Analyses phylogénétiques

Les distances génétiques établies par le programme GENDIST d'après les fréquences alléliques ont été converties sous forme d'arbres de distances générés par le programme Neighbor Joining. Ces arbres sont représentés à la figure 13; dans cette analyse, les groupes européens ont été exclus puisqu'ils ne sont pas toujours constitués des mêmes échantillons, selon le marqueur étudié. La population de Charlevoix-Saguenay est trouvée groupée avec le groupe de France comme illustré à la figure 13.

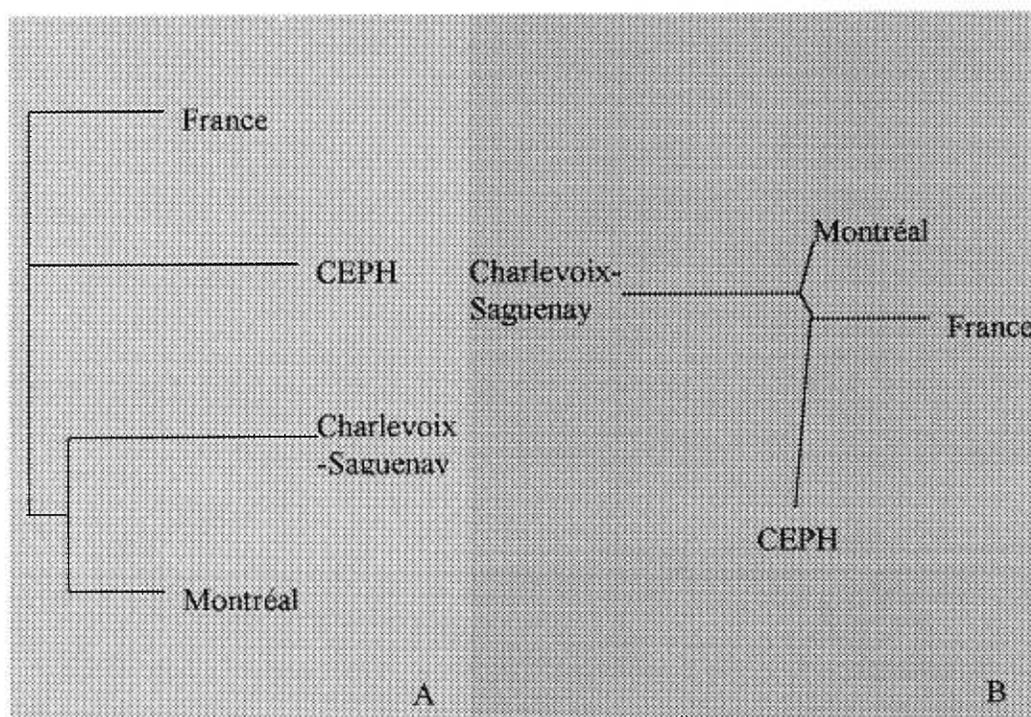


Figure 13 : Représentations graphiques sous forme d'arbres des matrices de distances construites à l'aide du logiciel Phylip selon un algorithme de « neighbour joining », d'après les fréquences alléliques des six marqueurs à l'étude pour les populations de Charlevoix-Saguenay, de France, de la région de Montréal et des échantillons CEPH. A et B sont deux formes d'arbres différentes mais issues de la même matrice de distances.

Les familles CEPH constituent le groupe le plus éloigné du groupe de Charlevoix-Saguenay et ce dernier semble plus près de la population de Montréal que celle de France. Afin de vérifier la stabilité de cette configuration nous avons exclu, tour à tour, les six marqueurs utilisés dans la construction de cette matrice. À chaque marqueur exclu, les arbres obtenus étaient toujours identiques à l'exception de légères variations dans la longueur des branches.

La diversité génétique calculée par le logiciel Arlequin et déterminée par la probabilité que deux haplotypes choisis au hasard dans une population soient différents, semble légèrement diminuée dans la population de Charlevoix-Saguenay mais cette différence n'est pas significative (figure 14). Cependant, l'importance des écarts types est due à la petite taille des échantillons de France, des familles CEPH et de la région de Montréal.

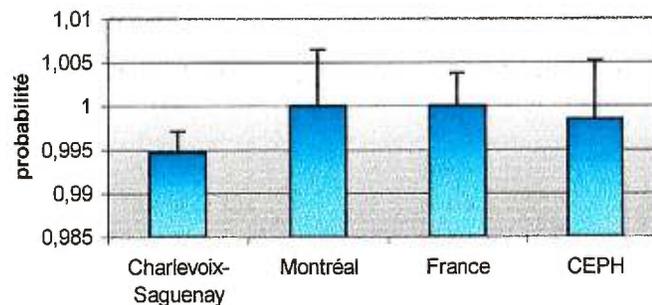


Figure 14: Diversité haplotypique calculée à l'aide du logiciel Arlequin et définie comme étant la probabilité que deux haplotypes choisis au hasard dans une population soient différents. Ces haplotypes sont ceux constitués des marqueurs microsatellites et de l'insertion Alu.

Le test AMOVA indique que 99,5% de la variabilité haplotypique est due à la variabilité intra-population, c'est-à-dire que la presque totalité des variations se retrouve au niveau individuel et non entre différentes populations.

L'index de fixation (F_{ST}), qui est la réduction d'hétérozygotie d'une sous-population due à une dérive génétique aléatoire, a été utilisé par le logiciel Arlequin afin de construire une matrice de distance entre les populations étudiées (figure 15).

| | Charlevoix-Saguenay | Montréal | France |
|----------|---------------------|----------|---------|
| Montréal | 0,00097 | | |
| France | 0,00816 | 0,00000 | |
| CEPH | 0,01303 | 0,00000 | 0,00000 |

Figure 15 : Matrice de distances générée à l'aide du logiciel Arlequin, basée sur l'index de fixation (F_{ST}) entre les différentes populations en utilisant les données sur les marqueurs microsatellites et l'insertion Alu. Les résultats significatifs sont indiqués en gras.

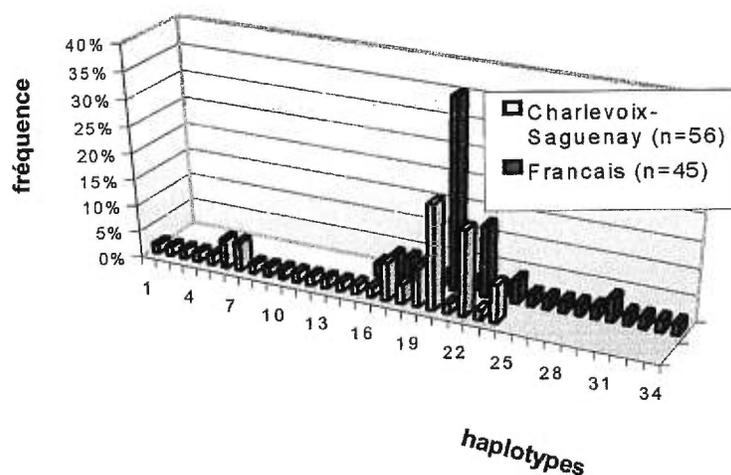


Figure 16 : Distribution de fréquences des haplotypes dys44, constitués de 28 polymorphismes bialléliques, dans la population de Charlevoix-Saguenay et en France.

Pour les haplotypes dys44 (figure 16), les deux populations possèdent chacune un bon nombre d'haplotypes endémiques mais ceux-ci sont uniques ou du moins ont des fréquences peu élevées. Les haplotypes fréquents par contre sont communs aux deux populations, et leurs fréquences sont relativement semblables sauf dans le cas de l'haplotype 20 qui est beaucoup moins fréquent dans la population de Charlevoix-Saguenay que chez les Français.

L'haplotype dys44 n'a été déterminé que pour 56 chromosomes de Charlevoix-Saguenay; nous avons relevé parmi ces 56 chromosomes ceux pour lesquels l'haplotype composé de microsatellites était répété dans la population (figure 9). À trois reprises nous avons identifié deux haplotypes identiques composé des microsatellites. Pour la première paire, les deux haplotypes dys44 sont identiques et il en va de même pour la deuxième paire à l'exception d'une position. Pour la troisième paire, les deux haplotypes dys44 sont différents.

3. Étude de recombinaisons et de mutations

Dans le but de connaître dans quelle proportion notre système génétique est stable - c'est à dire non affecté par les mutations ou recombinaisons - l'haplotypage a été effectué pour huit familles CEPH. Ces familles comprennent des généalogies à trois générations : grands-parents, parents et enfants. Chaque famille comprend en moyenne 11 enfants; en comptant les parents et les grands-parents, les huit familles comportent en tout 159 méioses (106 méioses informatives pour l'étude des recombinaisons puisque seules les méioses de la mère sont considérées étant donné qu'il s'agit du chromosome X).

Deux enfants de la famille CEPH884, les échantillons 884-5 et 884-11 (frère et sœur) présentaient l'allèle 16 qui n'était présent ni chez la mère, ni chez le père. Ces derniers étaient respectivement homozygote et hémizyote pour l'allèle 15 qui a également été observé chez les deux enfants.

Deux recombinaisons ont été observées parmi ces familles; la figure 17 illustre l'endroit, dans le système génétique, où ont été observées ces recombinaisons.

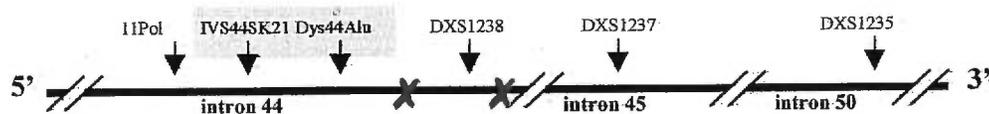


Figure 17 : Événements de recombinaison observés lors de l'étude des huit familles CEPH. La position de chacune des deux recombinaisons est indiquée par une croix rouge : la première recombinaison a été observée entre les marqueurs IVS44SK21/Dys44Alu et DXS1238 chez une des filles de la famille CEPH102. La deuxième recombinaison, entre les marqueurs DXS1238 et DXS1237, a été observée pour un des garçons de la famille CEPH1347.

Les haplotypes des huit familles ont été analysés par le logiciel LINKAGE afin d'estimer la distance génétique entre les marqueurs se trouvant aux extrémités des points de recombinaison. La figure 18 montre que les valeurs maximales de lod scores ont été obtenues entre 2,7 % et 2,9 % de recombinaisons, dans un intervalle de support d'une unité lod de 0,2 % à 5,6% (voir tableau X, annexe 1). Afin de diminuer le temps d'analyse et les complications dues à la grande variabilité des différents marqueurs, nous avons eu recours à une modification d'un procédé décrit par O'Connell et Weeks (1995) dans l'entrée des données relatives au génotypage des familles CEPH. Dans l'analyse de telles données, ces auteurs ont recours à ce qu'ils nomment «set-recoding» ou recodage des données. Une des approches consiste à réunir tous les allèles n'apparaissant pas dans une généalogie sous un seul et même allèle dont la fréquence est la somme des allèles manquants. Ainsi, les seuls allèles apparaissant dans l'entrée des données sont ceux apparaissant dans les généalogies plus un (celui regroupant ceux qui n'apparaissent pas). Lorsque nous avons entré les données, nous avons procédé de la même façon à l'exception près que nous avons considéré uniquement les allèles apparaissant dans les généalogies; l'identité des allèles n'est pas importante dans notre cas puisqu'il ne s'agit pas d'associer génotype et phénotype. Les fréquences de ces allèles n'étaient pas prises en considération dans l'entrée de données puisque le lod score est indépendant des fréquences alléliques lorsque tous les génotypes sont connus (Ott 1978), comme c'est le cas dans notre étude.

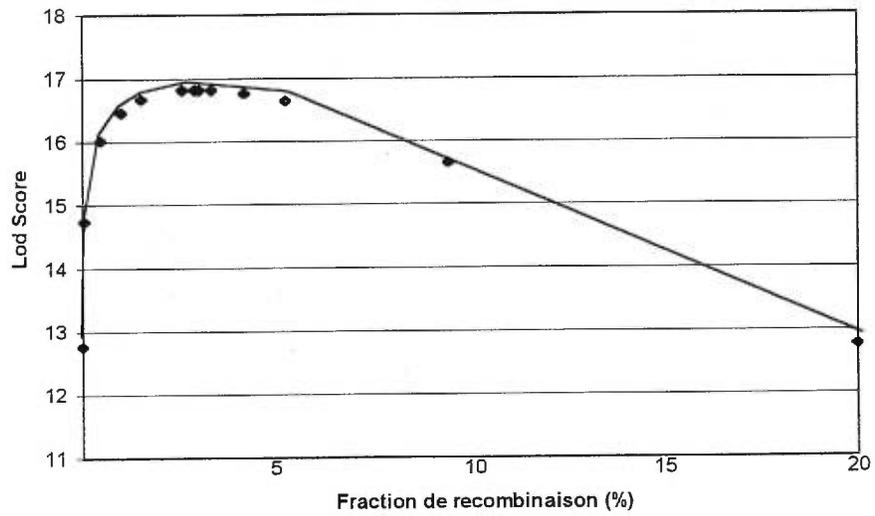


Figure 18 : Lod scores obtenus à l'aide du MLINK version 5.10, en fonction de la fraction de recombinaison entre les marqueurs distaux du système génétique, lors de l'analyse des génotypes de huit familles CEPH : CEPH102, CEPH884, CEPH1331, CEPH1332, CEPH1347, CEPH 1362, CEPH1413 et CEPH1416.

Chapitre 4 : Discussion

Diversité allélique. La diversité allélique est comparable pour toutes les populations étudiées, c'est-à-dire que le nombre d'allèles présents dans une population est sensiblement le même, autant pour celle de Charlevoix-Saguenay que pour les autres populations. Les figures 3 à 8 montrent que, pour tous les marqueurs étudiés, les allèles les plus fréquents ont en général des fréquences similaires pour toutes les populations. Les différences se font plutôt remarquer au niveau des allèles plus rares comme par exemple l'allèle 5 du marqueur DXS1235 (figure 6) qui a une fréquence beaucoup plus faible au Charlevoix-Saguenay ou l'allèle 21 du marqueur IVS44SK21 (figure 7) qui, chez cette même population, présente une fréquence beaucoup plus élevée que chez les autres groupes. Ceci s'explique par le fait que tous les groupes étudiés font partie de populations européennes ou sont issus l'un de l'autre et partagent donc de nombreux allèles dont les fréquences sont souvent très semblables. En effet, lorsqu'un groupe se détache d'une population pour en fonder une nouvelle, il est beaucoup plus probable que les changements de fréquences s'effectuent au niveau des allèles qui sont rares plutôt qu'au niveau des allèles fréquents. Autrement dit, lors de l'échantillonnage d'une population les allèles les plus fréquents seront presque sûrement retrouvés en grande quantité, tandis que les allèles rares verront souvent leur fréquence augmenter ou diminuer de façon plus significative à la suite d'un des processus stochastiques de l'échantillonnage et de la dérive génétique. Dans le tableau IV se retrouve la liste des allèles pour lesquels des différences

significatives étaient retrouvées dans une population comparativement aux trois autres. On remarque dans ce tableau que la plupart des différences sont observées dans la population de Charlevoix-Saguenay. Ce tableau n'est pas sans rappeler le tableau I (voir introduction) dans lequel les fréquences de porteurs de diverses maladies héréditaires sont listées. On peut en effet comparer la fréquence de l'allèle 9 du marqueur DXS1238 par exemple avec la fréquence de porteurs de l'ataxie

Tableau IV: Distribution de fréquences des allèles étant présents chez les quatre populations à l'étude mais montrant une différence significative à $p=0,3$ dans une des populations comparativement aux trois autres. Pour chaque allèle, la fréquence indiquée en gras est celle qui diffère des trois autres.

| Marqueur(allèle) | Fréquence | | | |
|------------------|----------------------|---------------|---------------|---------------|
| | Charlevoix | Montréal | France | Europe mixte |
| 11Pol(22) | 0.202 ± 0.040 | 0.100 ± 0.04 | 0.11 ± 0.04 | 0.110 ± 0.02 |
| DXS1238(9) | 0 | 0.021 ± 0.018 | 0.039 ± 0.027 | 0.040 ± 0.026 |
| DXS1238(14) | 0.066 ± 0.024 | 0.059 ± 0.033 | 0 | 0.020 ± 0.018 |
| DXS1237(12) | 0.046 ± 0.020 | 0.038 ± 0.026 | 0.037 ± 0.025 | 0 |
| DXS1235(5) | 0.028 ± 0.016 | 0.100 ± 0.047 | 0.148 ± 0.029 | 0.110 ± 0.041 |
| IVS44SK21(21) | 0.170 ± 0.036 | 0.059 ± 0.033 | 0.056 ± 0.032 | 0.06 ± 0.023 |

de Friedreich n'ayant pas, à ce jour, été observée dans la population de Charlevoix-Saguenay mais qui est bien représentée dans d'autres populations. Également l'allèle 21 du marqueur IVS44SK21, tout comme l'histidinémie (voir tableau I), a une fréquence jusqu'à trois fois plus élevée dans la population de Charlevoix-Saguenay que dans d'autres populations. Pour avoir une idée des chiffres, en fait de fréquences absolues ceci se traduit par 18 sur 106 pour Charlevoix-Saguenay, 6 sur 104 pour les groupes de France et de Montréal réunis et 6 sur 100 pour l'Europe.

Sur le plan de la diversité allélique, la population de Charlevoix-Saguenay est éloignée des familles CEPH et plus près du groupe de la région de Montréal que celui

de France (figures 13), ce qui est compatible avec l'histoire de ces populations: la population de Charlevoix-Saguenay est historiquement plus près des Canadiens-Français de la région de Montréal que des Français.

Diversité haplotypique. Le nombre élevé d'échantillons de la population de Charlevoix-Saguenay peut en partie être responsable du nombre élevé d'haplotypes répétés dans notre étude. Ce nombre élevé de répétitions n'est cependant pas seulement dû à la taille de l'échantillon car lorsqu'il est équivalent aux autres groupes, la population de Charlevoix-Saguenay montre toujours un nombre significativement plus élevé d'haplotypes répétés. La diversité haplotypique, telle que définie comme étant la probabilité que deux haplotypes choisis au hasard soient différents, semble légèrement diminuée dans la population de Charlevoix-Saguenay par rapport aux autres populations étudiées. Toutefois, le nombre limité de chromosomes dans les échantillons ne permet pas de conclure à une différence significative. Par contre, cette constatation ne s'applique pas nécessairement à tout le génome puisque notre système est assez instable d'un point de vue génétique à cause d'une part de la nature hautement mutationnelle des microsatellites et d'autre part du haut taux de recombinaisons dans la région étudiée (voir la partie «mutations et recombinaisons»). Lorsque les haplotypes ne comptaient que deux marqueurs (figure 11), l'augmentation en fréquence de certains haplotypes était davantage mise en évidence.

L'importance d'avoir un système génétique informatif dans une étude comme celle-ci s'avère donc cruciale pour l'interprétation des résultats. Ainsi, à la figure 11, les fréquences élevées de certains haplotypes sont en partie dues à la présence d'allèles très fréquents et n'est pas uniquement le reflet de l'effet fondateur. Les fréquences élevées de certains allèles ne peut cependant pas expliquer la présence d'haplotypes répétés lorsque cet haplotype est composé des six marqueurs car les haplotypes retrouvés les plus fréquemment ne sont pas composés d'allèles plus fréquents que les haplotypes uniques (figure 12). L'effet fondateur peut donc être observé par l'augmentation en fréquence de certains haplotypes dans une population par rapport à d'autres mais à condition que l'haplotype étudié soit informatif. Ceci dit, les haplotypes qui sont identiques, surtout lorsqu'il s'agit des haplotypes à six marqueurs, sont selon toute probabilité issus d'un seul et même haplotype fondateur. Nous savions déjà, d'après les taux élevés de porteurs de certaines maladies héréditaires, qu'il y avait une forte probabilité pour qu'au moins deux haplotypes soient identiques car issus d'un seul et même fondateur. Cette affirmation peut être vérifiée de deux façons : la première aurait été d'établir un autre haplotype informatif plus en amont ou en aval du site étudié et avec lequel notre haplotype aurait été en déséquilibre de liaison. L'haplotype dys44 qui se retrouve un peu en amont nous informe justement sur ce sujet puisque que deux haplotypes identiques composés de microsatellites se sont avérés avoir les mêmes haplotypes dys44 ; deux autres avaient également les mêmes haplotypes dys44 à l'exception d'une position. Ces deux paires d'haplotypes avaient donc probablement chacune un haplotype fondateur commun.

La deuxième façon, qui est la plus directe, aurait été de disposer des généalogies ascendantes des individus constituant notre population et de vérifier les liens possibles entre les individus ayant des haplotypes identiques. Ces généalogies existent et feront éventuellement l'objet, avec les données de cette étude, d'une étude généalogique en collaboration avec l'IREP et le Musée de l'Homme à Paris.

La matrice de distances construite par le logiciel Arlequin démontre que la population de Charlevoix-Saguenay se distingue des échantillons CEPH mais n'est pas significativement différente, d'un point de vue haplotypique, des Français et du groupe de la région de Montréal. Ceci peut en partie être dû au fait que nos échantillons sont relativement petits. Le fait que le groupe de Charlevoix-Saguenay ne se distingue pas de façon significative du groupe de la région de Montréal ou des Français n'est pas surprenant en soi puisque la population de Charlevoix-Saguenay et celle de la région de Montréal descendent toutes en majeure partie d'ancêtres venus de France au XVII^e siècle. De plus, les premiers colons français se sont établis il y a environ 17 générations, ce qui ne laisse pas beaucoup de temps pour une grande divergence.

La figure 19 illustre comment, suite au goulot d'étranglement qui a caractérisé la colonisation de la Nouvelle-France, la population de Charlevoix-Saguenay a pu voir sa diversité haplotypique diminuer sans pour autant avoir une diminution de la variabilité allélique. Chaque haplotype est une combinaison particulière de plusieurs allèles et le nombre de combinaisons possible dépasse donc largement le nombre d'allèles observés, surtout lorsque les marqueurs comportent un nombre élevé

d'allèles, comme c'est le cas dans le système étudié ici. À la figure 19 les haplotypes sont représentés par des cercles de différentes couleurs et les allèles compris dans chaque haplotype sont représentés par les symboles à l'intérieur des cercles. Cette figure illustre donc comment un nombre limité d'haplotypes est suffisant pour introduire la totalité de la variabilité allélique; la diversité haplotypique peut ultérieurement augmenter par l'avènement de mutations et de recombinaisons.

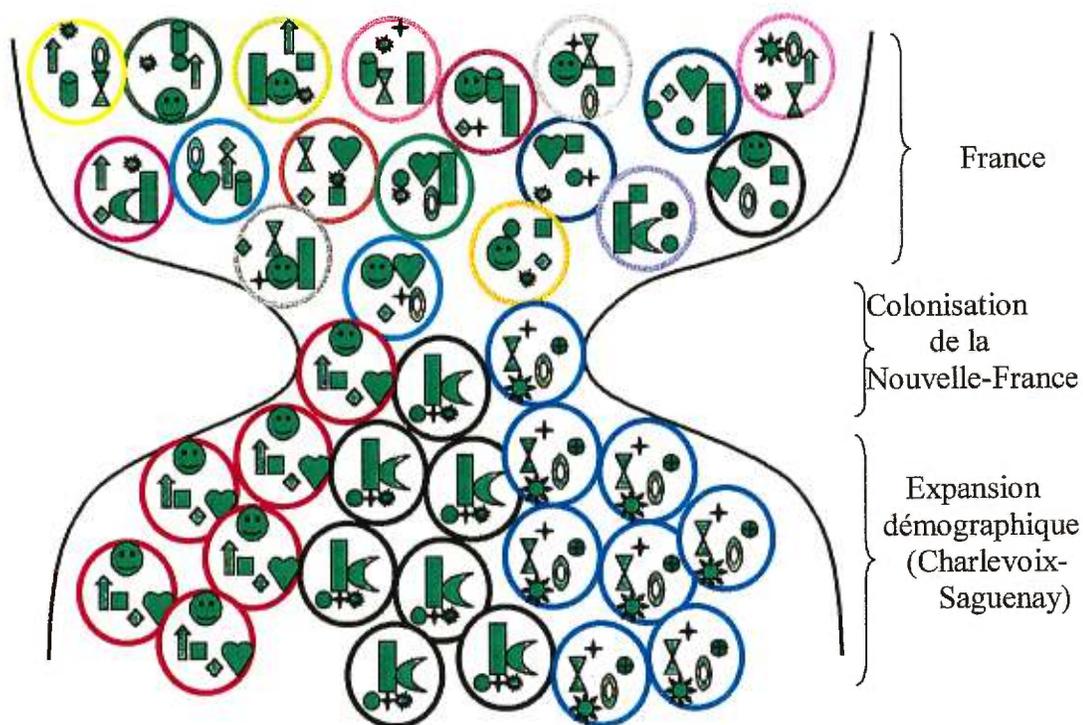


Figure 19 : Modélisation de l'introduction des différents haplotypes dans la population de Charlevoix-Saguenay à partir de la population mère. Les cercles de différentes couleurs représentent les différents haplotypes contenant divers allèles. Dans ce modèle, la France comprend un grand nombre d'haplotypes mais seuls quelques-uns parmi ces derniers sont introduits en Nouvelle-France lors du goulot d'étranglement. Ces quelques haplotypes sont cependant suffisants pour introduire la quasi-totalité des allèles retrouvés en France.

Mutations et recombinaisons. Les deux mutations observées sont exprimées chez une soeur et un frère de la famille CEPH884 mais le frère s'est révélé être hétérozygote pour la mutation et l'allèle hérité de la mère, ce qui est impossible puisqu'il s'agit du chromosome X. Un tel résultat est probablement dû à un artefact de manipulation, vraisemblablement lors de l'amplification par PCR. De plus, ces deux mutations ont été observées sur des haplotypes différents, ce qui nous amène à éliminer la mutation observée chez le garçon. Les familles CEPH étudiées ici comprennent 159 méioses en tout. Nous avons donc observé une mutation pour 159 méioses et cinq marqueurs microsatellites, ce qui donne un taux de mutation d'environ $1,2 \times 10^{-3}$ /locus/gamète/génération. Ce taux est conforme à ce qui a été observé dans la littérature pour les microsatellites, à savoir $5 \times 10^{-3} - 10^{-4}$ mutation/locus/gamète/génération (Zhivotovsky et Feldman 1995, Broman *et al.* 1998).

Le haut taux de recombinaison observé dans cette étude est en accord avec les données de plusieurs travaux concernant cette région du chromosome X. L'analyse des familles CEPH a en effet dévoilé deux événements de recombinaison entre les marqueurs IVS44SK21 et DXS1237. D'après le nombre de méioses informatives, la fraction de recombinaison calculée à l'aide du programme MLINK est d'environ 2,8 %, plus ou moins 2,6 % entre ces deux marqueurs. Or nous savons que ces marqueurs sont distants de 200 kbs (Köchling *et al.* 1995), ce qui donnent un rapport de 1 à 27 % /Mb ou 1 à 27 cM/Mb. Abbs *et al.* (1990) quant à eux ont observé un rapport de 6 cM/Mb pour le gène de la dystrophine ; Chen *et al.* (1989) ont observé un taux semblable de 5 cM/Mb. Nos observations sont donc en accord avec celles

observées dans ces travaux à savoir que la région étudiée ici comporte un taux de recombinaison nettement plus élevé que la moyenne observée pour le chromosome X dans son ensemble et qui est de 1,2 cM/Mb (Abbs *et al.* 1990). Nos observations contredisent donc les affirmations de Bertelson *et al.* (1986) qui n'ont pas observé de taux de recombinaison particulièrement élevés dans la région Xp21.

Conclusion

Lors de notre étude, l'utilisation d'un système génétique hautement informatif nous a permis d'obtenir des résultats concluants. Les résultats sur la variabilité allélique sont en accord avec les résultats obtenus pour la variabilité haplotypique. Les points majeurs, s'appliquant à ces deux types d'analyses, sont les suivants : 1) La variabilité génétique de la population de Charlevoix-Saguenay se compare à celle des autres populations étudiées ; 2) certains allèles et certains haplotypes ont une fréquence significativement augmentée dans la population de Charlevoix-Saguenay par rapport aux autres populations, témoignant ainsi de l'effet fondateur ; 3) la population de Charlevoix-Saguenay est visiblement apparentée à celles de la région de Montréal et de France mais se distingue des échantillons CEPH. La composition génétique de la population de Charlevoix-Saguenay découle de son histoire unique : une population fondée par un nombre limité de fondateurs et ayant été longtemps isolée. Ainsi partage-t-elle toujours de nombreuses similitudes avec sa population mère, tout en s'étant distinguée de ses origines de manière significative. L'effet fondateur a donc bel et bien été observé au sein de polymorphismes neutres de l'ADN mais ce phénomène n'a pas pour autant affecté la variabilité génétique de la population de Charlevoix-Saguenay.

Mis à part les conclusions concernant directement les populations étudiées, notre étude a démontré l'importance du choix du système génétique utilisé. Ainsi, le système idéal sera à la fois stable et informatif. Les populations étudiées dans les travaux ayant été présentés sont toutes proches l'une de l'autre d'un point de vue

évolutif; un système conventionnellement utilisé en génétique (polymorphismes bialléliques et taux de recombinaison «normal») n'aurait probablement pas été d'une grande utilité. Par sa nature instable, notre système s'est avéré capable d'un certain niveau de distinction entre ces populations qui sont pourtant très proches. Nous proposons donc ce type de système comme une alternative aux systèmes basés sur les polymorphismes de l'ADN mitochondrial par exemple, dans la mesure où la nature d'un système basé sur l'ADN nucléaire est plus adéquat.

Remerciements

Je tiens à remercier tout particulièrement mon directeur de recherche, Dr Damian Labuda pour m'avoir dirigée et encadrée tout au long de cette étude. Je remercie également tous les membres de son laboratoire, notamment Dr Maya Krajinowic pour l'aide apportée aux analyses faites à l'aide du logiciel LINKAGE, Dr Ewa Zietkiewicz pour son expertise dans l'analyse des données génétiques, Dr Vania Yotova qui a effectué l'haplotypage dys44, M. Alexandre Labuda qui a développé le programme utilisé pour la normalisation de l'échantillonnage, M. Jean-François Bibeau et M. Dominik Gehl pour les analyses informatiques relatives aux logiciels Arlequin et Phylip et enfin tous les étudiants du laboratoire qui m'ont apporté leur aide tout au long de mon étude : M. Nicolas Gilbert, MSc., M. Andrzej Rochowski, BSc., Mme Fanie Barnabé, ainsi que notre assistante de recherche, Mme Chantal Richer, BSc.

Bibliographie

- Abbs S., Roberts R. G., Mathew C. G., Bentley D. R., et Bobrow M. 1990. Accurate Assessment of Intragenic Recombination Frequency Within the Duchenne Muscular Dystrophy Gene. *Genomics* 7:602-606.
- Bertelson C. J., Bartley J. A., Monaco A. P., Colletti-Feener C., et Ficshbeck K., Kunkel L. M. 1986. Localisation of Xp21 Meiotic Exchange Points in Duchenne Muscular Dystrophy Families. *J. Med. Genet.* **23** :531-537.
- Bouchard G. 1991. Pour une approche historique et sociale du génome québécois *dans* Bouchard G., et de Braekeleer, M. Histoire d'un génome. Presses de l'Université du Québec. Québec (QC).
- Bouchard G., et de Braekeleer, M. 1991. Histoire d'un génome. Presses de l'Université du Québec. Québec (QC).
- Broman K. W., Murray J. C., Sheffield V. C., White R. L., et Weber J. L. 1998. Comprehensive Human Genetic Maps : Individual and Sex-Specific Variation in Recombination. *Am. J. Hum. Genet.* **63** : 861-869.
- Buchbinder J. T., Bilu Y. et Witzum E. 1997. Ethnic Background and Antecedents of Religious Conversion Among Israeli Jewish Outpatients. *Psychol. Rep.* **81** :1187-1202.

- Chen J. D., Hejtmancik J. F., Romeo G., Lindlof M., Boehm C., Caskey C. T., Kress W., Fischbeck K. H., Dreier M., Serravalle S., Grimm T., Kaariainen H., Ferrari M., Pfendner E., Meng G., De La Chapelle A., Melis M. A., Muller B., MacKinlay A. G., Muller C. R. et Denton M. J. 1989. A Genetic Linkage Map of Five Marker Loci in and around the Duchenne Muscular Dystrophy Locus. *Genomics* **4** :105-109.
- Clemens P. R., Fenwick R. G., Chamberlain, R., Gibbs, R. A., de Andrade M., Chakraborty R., et Caskey C. T. 1991. Carrier Detection and Prenatal Diagnosis in Duchenne and Becker Muscular Dystrophy Families, Using Dinucleotide Repeat Polymorphisms. *Hum. Genet.* **49** :951-960.
- Dahl N., Lagerström M., Erikson A, et Pettersson U. 1990. Gaucher Disease Type III (Norrbottnian type) is Caused by a Single Mutation in Exon 10 of the Glucocerebrosidase Gene. *Am. J. Hum. Genet* **47**:275-278.
- Drayna D., et White R. 1985. The Genetic Linkage Map of the Human X Chromosome. *Science* **230**:753-756.
- Felsenstein J. 1993. PHYLIP (PHYLogeny Inference Package), version 3.5.p, Distribué par l'auteur, Département de Génétique de l'Université Washington, Seattle.
- Gauvreau D., et Bourque, M. 1988. « Mouvements migratoires et familles : le peuplement du Saguenay avant 1911 », *Revue d'histoire de l'Amérique française*, **42**:167-192.
- Hartl D. L. et Clark A. G. 1989. *Principles of Population Genetics*, second edition. p. 61. Sinauer associates, inc. Sunderland, Massachusetts.

- Heyer E., Tremblay M. et Desjardins B. 1997. Seventeenth-Century European Origins of Hereditary Diseases in the Saguenay Population (Quebec, Canada). *Hum. Biol.* **69** :209-225.
- Kajanoja P. 1971. A study in the Morphology of the Finns and its Relation to the Settlement of Finland. *Ann. Acad. Sci. Fenn.* **146** :1-61.
- Köchling S, den Dunnen J. T., Dwornicsak B. et Horst J. 1995. Two Polymorphic Dinucleotide Repeats in intron 44 of the Dystrophin Gene. *Hum. Genet.* **95**:475-477.
- Labuda M., Labuda D., Korab-Laskowska M., Cole D. E. C., Zietkiewicz E., Weissenbach J., Popowska E. 1996. Linkage Disequilibrium Analysis in Young Populations : Pseudo-Vitamin D-Deficiency Rickets and the Founder Effect in French Canadians. *Am. J. Hum. Genet.* **59** :633-643.
- Labuda D et Zietkiewicz E. 1997. The Genetic Clock and the Age of the Founder Effect in Growing Populations : A Lesson From French Canadians and Ashkenazim. *Am. J. Hum. Genet.* **61** :768-771.
- Lum J. K et Cann R. L. 1998. mtDNA and Language Support a Common Origin of Micronesians and Polynesians in Islands South East Asia. *Am. J. Hum. Genet.* **63** :613-624.
- O'Connell J R., et Weeks D. E. 1995. The VITESSE Algorithm for Rapid Exact Multilocus Linkage Analysis via Genotype Set-Recoding and Fuzzy Inheritance. *Nature Genet.* **11** :402-408.

- Ott J. 1978. A Simple Scheme for the Analysis of HLA Linkages in Pedigrees. *Ann. Hum. Genet.* **42** :255-257.
- Rund D, Cohen T, Filon D, Dowling C E, Warren T, Barak I, Rachmilewitz E, Kazazian H H, et Oppenheim A. 1991. Evolution of a Genetic Disease in a Genetic Ethnic Isolate : β -Thalassemia in the Jews of Kurdistan. *Proc. Nat. Acad. Sci. USA.* **88** :310-314.
- Schneider S., Kueffer J.M., Roessli D et Excoffier L. 1997. Arlequin, version 1.1 : A Wofware for Population Genetics Data Analysis. Laboratoire de Génétique et de Biométrie, Université de Genève, Suisse.
- Tautz D. 1993. *Dans* Fingerprinting : State of the Science. Éds Pena S D J, Chakraborty R, Epplen J T, et Jeffreys A J. pp. 21-28.
- Witchel S. F., Lee, P. A., Suda-Hartman, M., Trucco M., et Hoffman E. P. 1997. Evidence for a Heterozygote Advantage in Congenital Adrenal Hyperplasia Due to 21-Hydroxylase Deficiency. *J. Clin. Endocrinol. Metab.* **82** :2097-2101.
- Zietkiewicz E., Yotova V., Jarnik M., Korab-Laskowska M., Kidd K. K., Mokiano D., Scozzari R., Stoneking M., Tishkoff S., Batzer M., et Labuda D. 1997. Nuclear DNA Diversity in World Wide Distributed Human Populations. *Gene* **205** :161-171.
- Zietkiewicz E., Yotova V., Jarnik M., Korab-Laskowska M., Kidd K. K., Mokiano D., Scozzari R., Stoneking M., Tishkoff S., Batzer M., et Labuda D. 1998. Genetic structure of the ancestral populations of modern humans. *J. Mol. Evol.* **47** :146-158.

- Zhivotovsky L A et Feldman M W. 1995. Microsatellite Variability and Genetic Distances. *Proc. Natl. Acad. Sci. USA.* **92** :11549-11552.
- Zischler H., Geisert H., von Haeseler A. et Pääbo S. 1995. A Nuclear « Fossil » of the Mitochondrial D-loop and the Origin of Modern Humans. *Nature* **378** :489-492.
- Zlotogora J, Bach G, Barak Y, et Elian E. 1980. Metachromatic Leukodystrophy in the Habbanite Jews : High Frequency in a Genetic Isolate and Screening for Heterozygotes. *Am. J. Hum. Genet.* **32** :663-669.
- Zlotogora J. 1994. High Frequencies of Human Genetic Diseases : Founder Effect With Genetic Drift or Selection ? *Am. J. Med. Genet.* **49** :10-13.

ANNEXE 1 : Données brutes

Tableau V: Fréquences alléliques des différents marqueurs étudiés

| Marqueur | Allèle | Fréquence | | | | | | | |
|---------------|--------|---------------------|--------|--------------------|--------|--------|--------|--------------|--------|
| | | Charlevoix-Saguenay | | Région de Montréal | | France | | Europe mixte | |
| 11Pol | 14 | 0,05 | ± 0,02 | 0,02 | ± 0,02 | 0,11 | ± 0,04 | 0,08 | ± 0,02 |
| | 15 | 0,64 | ± 0,05 | 0,8 | ± 0,06 | 0,7 | ± 0,06 | 0,69 | ± 0,03 |
| | 16 | 0,10 | ± 0,03 | 0,08 | ± 0,04 | 0,11 | ± 0,04 | 0,09 | ± 0,02 |
| | 21 | 0,01 | ± 0,01 | 0 | ± 0,00 | 0 | ± 0,00 | 0,03 | ± 0,01 |
| | 22 | 0,20 | ± 0,04 | 0,1 | ± 0,04 | 0,09 | ± 0,04 | 0,11 | ± 0,02 |
| DXS1238 | 7 | 0,01 | ± 0,01 | 0,02 | ± 0,02 | 0 | ± 0,00 | 0 | ± 0,00 |
| | 8 | 0 | ± 0,00 | 0 | ± 0,00 | 0 | ± 0,00 | 0,02 | ± 0,02 |
| | 9 | 0,01 | ± 0,01 | 0 | ± 0,00 | 0,02 | ± 0,02 | 0 | ± 0,00 |
| | 12 | 0,12 | ± 0,03 | 0,04 | ± 0,03 | 0,1 | ± 0,04 | 0,05 | ± 0,03 |
| | 14 | 0,07 | ± 0,02 | 0,08 | ± 0,04 | 0,04 | ± 0,03 | 0,05 | ± 0,03 |
| | 15 | 0 | ± 0,00 | 0,02 | ± 0,02 | 0,04 | ± 0,03 | 0,04 | ± 0,03 |
| | 16 | 0,21 | ± 0,04 | 0,24 | ± 0,06 | 0,18 | ± 0,05 | 0,18 | ± 0,05 |
| | 17 | 0,05 | ± 0,02 | 0,04 | ± 0,03 | 0,06 | ± 0,03 | 0,04 | ± 0,03 |
| | 18 | 0,1 | ± 0,03 | 0,12 | ± 0,05 | 0,16 | ± 0,05 | 0,15 | ± 0,05 |
| | 19 | 0,18 | ± 0,04 | 0,29 | ± 0,06 | 0,29 | ± 0,06 | 0,23 | ± 0,06 |
| | 20 | 0,07 | ± 0,02 | 0,06 | ± 0,03 | 0 | ± 0,00 | 0,02 | ± 0,02 |
| | 21 | 0,07 | ± 0,02 | 0 | ± 0,00 | 0 | ± 0,00 | 0,04 | ± 0,03 |
| | 22 | 0,01 | ± 0,01 | 0 | ± 0,00 | 0 | ± 0,00 | 0,04 | ± 0,03 |
| 23 | 0,11 | ± 0,03 | 0,08 | ± 0,04 | 0,06 | ± 0,03 | 0,15 | ± 0,05 | |
| 26 | 0 | ± 0,00 | 0,02 | ± 0,02 | 0,06 | ± 0,03 | 0 | ± 0,00 | |
| Dys44Alu | 1 | 0,67 | ± 0,05 | 0,78 | ± 0,07 | 0,77 | ± 0,06 | 0,74 | ± 0,03 |
| | 2 | 0,33 | ± 0,05 | 0,22 | ± 0,07 | 0,23 | ± 0,06 | 0,26 | ± 0,03 |
| IVS44SK2 1 | 8.5 | 0,1 | ± 0,03 | 0,08 | ± 0,04 | 0,19 | ± 0,05 | 0,14 | ± 0,03 |
| | 9 | 0 | ± 0,00 | 0 | ± 0,00 | 0 | ± 0,00 | 0 | ± 0,00 |
| | 11 | 0 | ± 0,00 | 0,02 | ± 0,02 | 0 | ± 0,00 | 0 | ± 0,00 |
| | 12 | 0,28 | ± 0,04 | 0,43 | ± 0,07 | 0,37 | ± 0,07 | 0,38 | ± 0,05 |
| | 13 | 0,05 | ± 0,02 | 0 | ± 0,00 | 0,02 | ± 0,02 | 0,01 | ± 0,01 |
| | 14 | 0 | ± 0,00 | 0,02 | ± 0,02 | 0,02 | ± 0,02 | 0,01 | ± 0,01 |
| | 14.5 | 0,02 | ± 0,01 | 0 | ± 0,00 | 0 | ± 0,00 | 0 | ± 0,00 |
| | 15 | 0,04 | ± 0,02 | 0,04 | ± 0,03 | 0,02 | ± 0,02 | 0,03 | ± 0,02 |
| | 16 | 0,06 | ± 0,02 | 0,02 | ± 0,02 | 0 | ± 0,00 | 0,08 | ± 0,03 |
| | 17 | 0,17 | ± 0,04 | 0,18 | ± 0,05 | 0,15 | ± 0,05 | 0,15 | ± 0,04 |
| | 18 | 0,04 | ± 0,02 | 0,14 | ± 0,05 | 0,06 | ± 0,03 | 0,08 | ± 0,03 |
| | 19 | 0,17 | ± 0,04 | 0,06 | ± 0,03 | 0,06 | ± 0,03 | 0,06 | ± 0,02 |
| | 20 | 0,06 | ± 0,02 | 0,02 | ± 0,02 | 0,09 | ± 0,04 | 0,06 | ± 0,02 |
| 21 | 0,01 | ± 0,01 | 0 | ± 0,00 | 0,04 | ± 0,03 | 0 | ± 0,00 | |
| 22 | 0,01 | ± 0,01 | 0 | ± 0,00 | 0 | ± 0,00 | 0 | ± 0,00 | |

Tableau V (suite)

| Marqueur | Allèle | Fréquence | | | |
|----------|-------------|---------------------|--------------------|-------------|--------------|
| | | Charlevoix-Saguenay | Région de Montréal | France | Europe mixte |
| DXS1237 | 33 | 0 ± 0,00 | 0 ± 0,00 | 0 ± 0,00 | 0,02 ± 0,02 |
| | 32 | 0 ± 0,00 | 0 ± 0,00 | 0 ± 0,00 | 0,02 ± 0,02 |
| | 31 | 0,02 ± 0,01 | 0 ± 0,00 | 0,02 ± 0,02 | 0,04 ± 0,03 |
| | 30 | 0,02 ± 0,01 | 0,04 ± 0,03 | 0,02 ± 0,02 | 0,02 ± 0,02 |
| | 29 | 0,09 ± 0,03 | 0,06 ± 0,03 | 0,06 ± 0,03 | 0,04 ± 0,03 |
| | 28 | 0,1 ± 0,03 | 0,13 ± 0,05 | 0,07 ± 0,04 | 0,15 ± 0,05 |
| | 27 | 0,27 ± 0,04 | 0,17 ± 0,05 | 0,19 ± 0,05 | 0,22 ± 0,05 |
| | 26 | 0,15 ± 0,03 | 0,12 ± 0,05 | 0,19 ± 0,05 | 0,11 ± 0,04 |
| | 25 | 0,14 ± 0,03 | 0,13 ± 0,05 | 0,13 ± 0,05 | 0,12 ± 0,04 |
| | 24 | 0,07 ± 0,02 | 0,02 ± 0,02 | 0,02 ± 0,02 | 0,04 ± 0,03 |
| | 23 | 0,02 ± 0,01 | 0,1 ± 0,04 | 0,04 ± 0,03 | 0,05 ± 0,03 |
| | 22 | 0,05 ± 0,02 | 0,04 ± 0,03 | 0,04 ± 0,03 | 0 ± 0,00 |
| | 21 | 0 ± 0,00 | 0 ± 0,00 | 0 ± 0,00 | 0 ± 0,00 |
| | 20 | 0,06 ± 0,02 | 0,15 ± 0,05 | 0,2 ± 0,06 | 0,12 ± 0,04 |
| 19 | 0,02 ± 0,01 | 0,02 ± 0,02 | 0,04 ± 0,03 | 0,05 ± 0,03 | |
| 15 | 0 ± 0,00 | 0,02 ± 0,02 | 0 ± 0,00 | 0 ± 0,00 | |
| DXS1235 | 20 | 0,01 ± 0,01 | 0,03 ± 0,03 | 0,02 □ 0,02 | 0,02 ± 0,02 |
| | 19 | 0,14 ± 0,03 | 0,1 ± 0,05 | 0,06 □ 0,03 | 0,16 ± 0,05 |
| | 18 | 0 ± 0,00 | 0 ± 0,00 | 0,06 □ 0,03 | 0 ± 0,00 |
| | 17 | 0,06 ± 0,02 | 0,03 ± 0,03 | 0,02 □ 0,02 | 0,07 ± 0,03 |
| | 16 | 0,03 ± 0,02 | 0,1 ± 0,05 | 0,15 □ 0,05 | 0,11 ± 0,04 |
| | 15 | 0,6 ± 0,05 | 0,63 ± 0,08 | 0,54 □ 0,07 | 0,46 ± 0,07 |
| | 11 | 0,17 ± 0,04 | 0,13 ± 0,05 | 0,17 □ 0,05 | 0,18 ± 0,05 |

Tableau VI: Composition allélique des haplotypes étudiés et fréquence des haplotypes dans la population de Charlevoix-Saguenay (C/S); dans la région de Montréal (M); dans la population de France (F) et chez les familles CEPH (C).

| ALLÉLES | | | | | | Fréquence | | | |
|-----------|---------|----------|-----------|---------|---------|-----------|---|---|---|
| Marqueurs | | | | | | C/S | M | F | C |
| 11POL | DXS1236 | Dys44Alu | IVS449K21 | DXS1237 | DXS1235 | | | | |
| 15 | 16 | 1 | 13 | 32 | 20 | | | 1 | |
| 15 | 14 | 2 | 18 | 29 | 20 | 1 | | | |
| 15 | 18 | 1 | 19 | 29 | 20 | | 1 | | |
| 16 | 22 | 2 | 15 | 32 | 19 | 2 | | | |
| 15 | 16 | 1 | 12 | 31 | 19 | | | | 1 |
| 22 | 12 | 1 | 12 | 29 | 19 | 1 | | | |
| 15 | 20 | 2 | 8,5 | 28 | 19 | 3 | | | |
| 15 | 18 | 2 | 8,5 | 28 | 19 | 1 | | | |
| 16 | 14 | 1 | 11 | 28 | 19 | | 1 | | |
| 14 | 17 | 2 | 12 | 28 | 19 | 1 | | | |
| 15 | 17 | 2 | 12 | 28 | 19 | 1 | | | |
| 15 | 17 | 2 | 18 | 28 | 19 | 1 | | | |
| 15 | 9 | 2 | 18 | 28 | 19 | | 1 | | |
| 15 | 18 | 2 | 18 | 28 | 19 | | 1 | | |
| 15 | 9 | | 18 | 28 | 19 | | 1 | | |
| 16 | 22 | 1 | 19 | 28 | 19 | 1 | | | |
| 22 | 18 | 1 | 20 | 28 | 19 | | | | 1 |
| 15 | 9 | 1 | 20 | 28 | 19 | | | 1 | |
| 15 | 15 | 1 | 21 | 28 | 19 | | | 1 | |
| 22 | 12 | 2 | 12 | 27 | 19 | 1 | | | |
| 15 | 22 | 2 | 17 | 27 | 19 | | 1 | | |
| 15 | 14 | 1 | 12 | 23 | 19 | | | 1 | |
| 15 | 17 | 1 | 20 | 23 | 19 | | | | 1 |
| 15 | 14 | 1 | 19 | 20 | 19 | 1 | | | |
| 15 | 18 | 1 | 20 | 27 | 18 | | | 1 | |
| 15 | 17 | 2 | 12 | 20 | 18 | | | 1 | |
| 15 | 15 | 1 | 17 | 20 | 18 | | | 1 | |
| 15 | 15 | 1 | 12 | 30 | 17 | 1 | | | |
| 15 | 18 | 1 | 14 | 30 | 17 | | 1 | | |
| 16 | 22 | 2 | 8,5 | 29 | 17 | | | | 1 |
| 22 | 18 | 2 | 13 | 28 | 17 | 1 | | | |
| 15 | 7 | 1 | 17 | 28 | 17 | 1 | | | |
| 22 | 18 | 2 | 17 | 28 | 17 | 1 | | | |
| 15 | 18 | 1 | 19 | 28 | 17 | | | 1 | |
| 15 | 18 | 1 | 12 | 20 | 17 | | | | 1 |
| 15 | 20 | 1 | 13 | 20 | 17 | 1 | | | |
| 15 | 15 | 1 | 12 | 19 | 17 | 1 | | | |
| 15 | 18 | 1 | 12 | 30 | 16 | | | 1 | |
| 15 | 18 | 1 | 17 | 30 | 16 | | 1 | | |
| 15 | 15 | 1 | 20 | 30 | 16 | 1 | | | |
| 15 | 15 | 1 | 12 | 29 | 16 | | | | 1 |
| 15 | 17 | 2 | 12 | 29 | 16 | | 1 | | |
| 15 | 15 | 1 | 20 | 29 | 16 | | | | 1 |
| 15 | 18 | 1 | 12 | 28 | 16 | 1 | | | |
| 14 | 17 | 1 | 12 | 28 | 16 | | | 1 | |
| 15 | 14 | 1 | 8,5 | 27 | 16 | 1 | | | |

Tableau VI (suite 1)

| 11POL | DXS1236 | Dys44Alu | IVS44Sk21 | DXS1237 | DXS1235 | C/S | M | F | C |
|-------|---------|----------|-----------|---------|---------|-----|---|---|---|
| 15 | 15 | 1 | 12 | 27 | 16 | | | 1 | |
| 15 | 18 | 1 | 18 | 27 | 16 | | | 1 | |
| 22 | | 1 | 20 | 27 | 16 | | | 1 | |
| 22 | 12 | 2 | 20 | 27 | 16 | | 1 | | |
| 16 | 22 | 1 | 12 | 26 | 16 | | | | 1 |
| 15 | 15 | 2 | 17 | 26 | 16 | | 1 | | |
| 22 | 12 | 1 | 8,5 | 20 | 16 | | | 1 | |
| 15 | 17 | 1 | 17 | 20 | 16 | | | 1 | |
| 15 | 18 | 1 | 19 | 20 | 16 | | | 1 | |
| 15 | 15 | 1 | 12 | 31 | 15 | 1 | | | |
| 15 | 15 | 2 | 17 | 31 | 15 | | 1 | | |
| 15 | 15 | 1 | 17 | 31 | 15 | | 1 | | |
| 15 | 16 | 1 | 18 | 31 | 15 | 1 | | | |
| 15 | 15 | 1 | 18 | 31 | 15 | | | | 1 |
| 16 | 17 | 1 | 12 | 30 | 15 | 1 | | | |
| 21 | 17 | 1 | 12 | 30 | 15 | 1 | | | |
| 15 | 21 | 2 | 12 | 30 | 15 | 1 | | | |
| 15 | 17 | 1 | 12 | 30 | 15 | | | | 1 |
| 16 | 15 | 1 | 12 | 30 | 15 | | | | 1 |
| 15 | 18 | 1 | 12 | 30 | 15 | | 1 | | 1 |
| 16 | 22 | 1 | 12 | 30 | 15 | | | 1 | |
| 15 | 15 | 1 | 20 | 30 | 15 | 1 | | | |
| 14 | 17 | 1 | 8,5 | 29 | 15 | | | 1 | |
| 15 | 22 | 1 | 12 | 29 | 15 | 1 | | | |
| 15 | 18 | 1 | 12 | 29 | 15 | | | | 1 |
| 15 | 12 | 1 | 12 | 29 | 15 | | | 1 | |
| 15 | 9 | 1 | 12 | 29 | 15 | | | 1 | |
| 15 | 15 | 1 | 12 | 29 | 15 | | 1 | | |
| 22 | 12 | 2 | 13 | 29 | 15 | 1 | | | |
| 15 | 18 | 1 | 17 | 29 | 15 | 1 | 1 | | |
| 15 | 17 | 1 | 17 | 29 | 15 | | | | 1 |
| 15 | 18 | 2 | 17 | 29 | 15 | | | | 1 |
| 15 | 7 | 1 | 17 | 29 | 15 | | 1 | | |
| 15 | 9 | 1 | 18 | 29 | 15 | | | | 1 |
| 22 | 12 | 2 | 19 | 29 | 15 | 3 | | | |
| 16 | 22 | 1 | 8,5 | 28 | 15 | 1 | | | |
| 15 | 15 | 1 | 8,5 | 28 | 15 | | 1 | | 1 |
| 22 | 12 | 1 | 8,5 | 28 | 15 | | | 1 | |
| 15 | 22 | 1 | 12 | 28 | 15 | 1 | | | |
| 14 | 18 | 1 | 12 | 28 | 15 | 1 | | | |
| 22 | 12 | 1 | 12 | 28 | 15 | | | | 1 |
| 15 | 20 | 1 | 17 | 28 | 15 | 1 | | | |
| 15 | 18 | 1 | 17 | 28 | 15 | 1 | | 1 | 1 |
| 15 | 16 | 2 | 17 | 28 | 15 | | | | 1 |
| 15 | 14 | 1 | 17 | 28 | 15 | | | 1 | |
| 15 | 17 | 1 | 18 | 28 | 15 | 1 | | | |
| 22 | 18 | 1 | 18 | 28 | 15 | | 1 | | |
| 15 | 18 | 1 | 20 | 28 | 15 | 1 | | | |
| 15 | 19 | 2 | 20 | 28 | 15 | 2 | | | |
| 15 | 17 | 1 | 20 | 28 | 15 | | | | 1 |

Tableau VI (suite 2)

| 11POL | DXS1238 | Dys44Alu | IVS44SK21 | DXS1237 | DXS1235 | C/S | M | F | C |
|-------|---------|----------|-----------|---------|---------|-----|---|---|---|
| 15 | 15 | 2 | 8,5 | 27 | 15 | | | | 1 |
| 15 | 18 | 2 | 8,5 | 27 | 15 | | | | 1 |
| 15 | 18 | 1 | 8,5 | 27 | 15 | | | 1 | |
| 15 | 15 | 1 | 12 | 27 | 15 | 2 | | | |
| 22 | 12 | 1 | 12 | 27 | 15 | 1 | | | |
| 22 | 12 | 2 | 12 | 27 | 15 | 1 | | | |
| 16 | 16 | 2 | 12 | 27 | 15 | | | | 1 |
| 15 | 17 | 2 | 12 | 27 | 15 | | | 1 | |
| 15 | 15 | 1 | 12 | 27 | 15 | | 1 | 1 | |
| 15 | 7 | 1 | 12 | 27 | 15 | | 1 | | |
| 15 | 19 | 2 | 13 | 27 | 15 | 1 | | | |
| 22 | 15 | 2 | 13 | 27 | 15 | 1 | | | |
| 15 | | 1 | 15 | 27 | 15 | | | 1 | |
| 22 | 15 | 2 | 17 | 27 | 15 | 3 | | | |
| 16 | 18 | 1 | 17 | 27 | 15 | 1 | | | |
| 22 | 15 | 1 | 17 | 27 | 15 | 1 | | | |
| 15 | 19 | 1 | 19 | 27 | 15 | 1 | | | |
| 15 | 16 | 1 | 8,5 | 26 | 15 | 1 | | | |
| 15 | 15 | 2 | 8,5 | 26 | 15 | | | 1 | |
| 14 | 17 | 1 | 12 | 26 | 15 | 1 | | | |
| 16 | 22 | 1 | 12 | 26 | 15 | 1 | | | |
| 15 | 18 | 1 | 12 | 26 | 15 | 1 | | | |
| 15 | 15 | 1 | 12 | 26 | 15 | | | 1 | |
| 14 | 14 | 2 | 12 | 26 | 15 | | 1 | | |
| 15 | 17 | 1 | 12 | 26 | 15 | | 1 | | |
| 15 | 18 | 1 | 16 | 26 | 15 | 1 | | | |
| 22 | 12 | 1 | 17 | 26 | 15 | 2 | | | |
| 15 | 18 | 1 | 17 | 26 | 15 | | | | 1 |
| 16 | 22 | 2 | 17 | 26 | 15 | | | 1 | |
| 15 | 17 | 1 | 17 | 26 | 15 | 1 | | | |
| 15 | 16 | 1 | 18 | 26 | 15 | | | 1 | |
| 22 | 12 | 1 | 19 | 26 | 15 | 1 | | | |
| 15 | 15 | 1 | 19 | 26 | 15 | 4 | | | 1 |
| 16 | 15 | 1 | 19 | 26 | 15 | 1 | | | |
| 15 | 18 | 1 | 19 | 26 | 15 | 1 | | | |
| 15 | 15 | 2 | 19 | 26 | 15 | 1 | | | |
| 15 | 16 | 1 | 19 | 26 | 15 | | | 1 | |
| 15 | 18 | 2 | | 26 | 15 | | | 1 | |
| 15 | 16 | 1 | 8,5 | 25 | 15 | 2 | | | |
| 15 | 16 | 2 | 8,5 | 25 | 15 | 1 | | | |
| 16 | 22 | 1 | 12 | 25 | 15 | 1 | | | |
| 15 | 26 | 1 | 12 | 25 | 15 | | 1 | | |
| 15 | 18 | 1 | 14 | 25 | 15 | | | 1 | |
| 15 | 18 | 1 | 17 | 25 | 15 | | | | 1 |
| 15 | 18 | 1 | 19 | 25 | 15 | 1 | | | |
| 15 | 22 | 1 | 19 | 25 | 15 | 1 | | | |
| 14 | 17 | 1 | 21 | 25 | 15 | 1 | | | |
| 16 | 22 | 1 | 8,5 | 24 | 15 | | | 1 | |
| 15 | 17 | 2 | 12 | 24 | 15 | | | 1 | |
| 15 | 14 | 1 | 12 | 24 | 15 | | 1 | | |
| 16 | 18 | 1 | 12 | 24 | 15 | | 1 | | |

Tableau VI (suite 3)

| II POL | DXS1236 | Dys44Alu | IVS44SK21 | DXS1237 | DXS1235 | C/S | M | F | C |
|--------|---------|----------|-----------|---------|---------|-----|---|---|---|
| 15 | 19 | 1 | 19 | 24 | 15 | 1 | | | 1 |
| 16 | 18 | 1 | 4 | 23 | 15 | | | | 1 |
| 15 | 18 | 1 | 12 | 23 | 15 | 1 | | | |
| 15 | 17 | 1 | 12 | 23 | 15 | 1 | 1 | | |
| 15 | 15 | 1 | 12 | 23 | 15 | | 1 | | |
| 14 | 17 | 1 | 22 | 23 | 15 | 1 | | | |
| 15 | 17 | 1 | 12 | 22 | 15 | | | | 1 |
| 15 | 22 | 1 | 8,5 | 20 | 15 | 1 | | | |
| 15 | 18 | 1 | 8,5 | 20 | 15 | | 1 | | |
| 22 | 22 | 1 | 8,5 | 20 | 15 | | 1 | | |
| 15 | 18 | 2 | 8,5 | 20 | 15 | | 1 | | |
| 15 | 22 | 1 | 12 | 20 | 15 | | | | 1 |
| 14 | 17 | 1 | 12 | 20 | 15 | | | 1 | |
| 15 | 17 | 1 | 12 | 20 | 15 | | | 1 | |
| 16 | 22 | 1 | 14,4 | 20 | 15 | 1 | | | |
| 22 | 18 | 1 | 16 | 20 | 15 | 1 | | | |
| 15 | 20 | 2 | 17 | 20 | 15 | 1 | | | |
| 22 | 23 | 2 | 17 | 20 | 15 | | | 1 | |
| 14 | 9 | 1 | 20 | 20 | 15 | | | 1 | |
| 22 | 15 | 1 | 12 | 19 | 15 | 1 | | | |
| 15 | 18 | 1 | 12 | 19 | 15 | | | 1 | |
| 22 | 12 | 1 | 12 | 19 | 15 | | | 1 | |
| 15 | 16 | 1 | 17 | 19 | 15 | | 1 | | |
| 15 | 14 | 1 | 12 | 30 | 11 | 4 | | | |
| 14 | 17 | 1 | 21 | 30 | 11 | | | | 1 |
| 15 | 19 | 1 | 12 | 29 | 11 | | 1 | | |
| 15 | 18 | 1 | 16 | 29 | 11 | 2 | | | |
| 15 | 18 | 1 | 17 | 29 | 11 | 1 | | | |
| 15 | 15 | 1 | 17 | 29 | 11 | | | 1 | |
| 15 | 15 | 2 | 8,5 | 28 | 11 | | | 1 | |
| 16 | 26 | 2 | 8,5 | 28 | 11 | | | 1 | |
| 15 | 17 | 1 | 12 | 28 | 11 | 1 | | | |
| 15 | 19 | 2 | 15 | 28 | 11 | 2 | | | |
| 22 | 12 | 1 | 16 | 28 | 11 | 1 | | | |
| 16 | 23 | 1 | 18 | 28 | 11 | | | 1 | |
| 14 | 18 | 1 | 8,5 | 27 | 11 | | | 1 | |
| 15 | 15 | 1 | 12 | 27 | 11 | | 1 | | |
| 15 | 15 | 2 | 17 | 27 | 11 | 2 | | | |
| 15 | 18 | 1 | 21 | 27 | 11 | | | 1 | |
| 16 | 18 | 1 | 12 | 26 | 11 | | | 1 | |
| 15 | 15 | 1 | 17 | 26 | 11 | | | | 1 |
| 15 | 15 | 2 | 17 | 25 | 11 | 1 | | | |
| 15 | 18 | 1 | 12 | 24 | 11 | | | | 1 |
| 15 | 22 | 1 | 17 | 24 | 11 | 1 | | | |
| 14 | 12 | 2 | 12 | 23 | 11 | | | 1 | |
| 22 | 12 | 1 | 14,4 | 23 | 11 | 1 | | | |
| 15 | 9 | 1 | 19 | 23 | 11 | 1 | | | |
| 15 | 18 | 1 | 12 | 22 | 11 | | | | 2 |
| 15 | 18 | 2 | 12 | 20 | 11 | | | 1 | |
| 16 | 22 | 1 | 15 | 20 | 11 | | 1 | | |
| 22 | 18 | 1 | 18 | 15 | 11 | | 1 | | |

Tableau VII: Distribution de fréquence des haplotypes composés des marqueurs DXS1237 et DXS1235 parmi les populations de Charlevoix-Saguenay, des familles CEPH, de France et de la région de Montréal.

| Haplotype | Fréquence | | |
|-----------|---------------------|---------------------------------|---------------|
| | Charlevoix-Saguenay | France et Région de Montréal | Familles CEPH |
| 1 | 13 | 9 | 3 |
| 2 | 12 | 6 | 3 |
| 3 | 10 | 5 | 5 |
| 4 | 9 | 4 | 2 |
| 5 | 7 | 2 | 1 |
| 6 | 6 | 5 | 4 |
| 7 | 4 | 7 | 1 |
| 8 | 4 | 3 | |
| 9 | 4 | 2 | 4 |
| 10 | 4 | 1 | |
| 11 | 4 | | 1 |
| 12 | 3 | 2 | 1 |
| 13 | 3 | 2 | |
| 14 | 2 | 3 | |
| 15 | 2 | 1 | 1 |
| 16 | 2 | | |
| 17 | 2 | | |
| 18 | 1 | 6 | |
| 19 | 1 | 4 | |
| 20 | 1 | 2 | |
| 21 | 1 | 2 | |
| 22 | 1 | 1 | |
| 23 | 1 | 1 | |
| 24 | 1 | 1 | |
| 25 | 1 | 1 | 1 |
| 26 | 1 | 1 | |
| 27 | 1 | 1 | 1 |
| 28 | 1 | | |
| 29 | 1 | | |
| 30 | 1 | | |
| 31 | 1 | | |
| 32 | 1 | | |
| 33 | | 2 | 2 |
| 34 | | 3 | |
| 35 | | 1 | 2 |
| 36 | | 1 | |
| 37 | | 3 | |
| 38 | | 2 | |
| 39 | | 1 | |
| 40 | | 1 | 1 |
| 41 | | 1 | |
| 42 | | | 1 |
| 43 | | | 1 |
| 44 | | | 1 |
| 45 | | | 1 |
| 46 | | | 2 |
| 47 | | | 2 |

Tableau VIII: Distribution de fréquences des haplotypes pour les populations de France, de la région de Montréal et des échantillons CEPH comparée à celle de Charlevoix-Saguenay. Pour avoir une comparaison équivalente, 36 haplotypes ont été choisis au hasard sans remplacement, et pour un total de 500 fois, parmi ceux de la population de Charlevoix-Saguenay. Les échantillons CEPH et le groupe de la région de Montréal comptaient déjà 36 chromosomes et le groupe de France a été ramené à 36 chromosome en éliminant 16 chromosomes au hasard.

L'erreur standard était calculée comme suit : $\sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$ où n est le nombre d'haplotypes choisis et x le nombre de fois qu'un haplotype d'une fréquence donnée a été choisis.

| Fréquence | Nombre d'haplotypes |
|-----------|---------------------|
| 1 | 29,7 ± 2,8 |
| 2 | 2,4 ± 1,3 |
| 3 | 0,4 ± 0,6 |
| 4 | 0,04 ± 0,2 |

Tableau IX : Moyenne des fréquences alléliques composant chaque haplotype en fonction de la fréquence absolue de chaque haplotype.

| Fréquence haplotypique | Moyenne des fréquences alléliques |
|------------------------|-----------------------------------|
| 1 | 0,2 |
| 1 | 0,32 |
| 1 | 0,3 |
| 1 | 0,23 |
| 1 | 0,43 |
| 1 | 0,4 |
| 1 | 0,4 |
| 1 | 0,25 |
| 1 | 0,31 |
| 1 | 0,33 |
| 2 | 0,12 |
| 2 | 0,33 |
| 2 | 0,42 |
| 2 | 0,32 |
| 2 | 0,36 |
| 2 | 0,3 |
| 2 | 0,25 |
| 2 | 0,28 |
| 3 | 0,26 |
| 3 | 0,25 |
| 3 | 0,28 |
| 4 | 0,4 |
| 4 | 0,32 |

Tableau X : Lod Scores obtenus à l'aide du logiciel LINKAGE pour différentes fractions de recombinaisons.

| Fraction de recombinaison (%) | Lod Score |
|-------------------------------|-----------|
| 0,05 | 12,77 |
| 0,1 | 14,74 |
| 0,2 | 15,83 |
| 0,5 | 16,02 |
| 1 | 16,47 |
| 1,5 | 16,68 |
| 2,5 | 16,82 |
| 2,8 | 16,83 |
| 2,9 | 16,83 |
| 3,2 | 16,82 |
| 4 | 16,77 |
| 5 | 16,65 |
| 10 | 15,66 |
| 20 | 12,78 |