

Université de Montréal

Régionalisation de l'effet fondateur au Québec

par

Gerbault Pascale

Département de Biologie Moléculaire

Faculté des études supérieures

Mémoire présenté à la Faculté des études supérieures

En vue d'obtention du grade de maîtrise

En biologie moléculaire

Août, 2006

©, Gerbault Pascale, 2006



QH
506
U54
2007
V.004

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :
Régionalisation de m'effet fondateur au Québec

présenté par :
Pascale Gerbault

a été évalué par un jury composé des personnes suivantes :

Dr Bernard Brais
président rapporteur

Dr Damian Labuda
Directeur de recherche

Professeure-chercheure Hélène Vézina
membre du jury



Résumé

La connaissance de la structure d'une population est primordiale pour amorcer la cartographie de gènes de maladies complexes et de maladies mendéliennes dans cette population. La population canadienne-française de la province de Québec est connue pour des fréquences plus élevées de certaines maladies génétiques. Ces six millions de Canadiens français proviennent de 10 000 fondateurs arrivés en Nouvelle-France entre 1608 et 1765. Nous avons étudié la question de la structure de la population canadienne-française du Québec à travers l'échantillonnage de trois régions : la région de Montréal, celle du Saguenay-Lac-Saint-Jean et celle de la Gaspésie, représentant des histoires de peuplement différentes au sein de la province. Afin de comparer la diversité génétique de ces régions québécoises, l'analyse de quelques régions françaises a été intégrée à cette étude. Trois systèmes de polymorphismes neutres sur Xp21 ont été utilisés pour appréhender la diversité génétique québécoise. Ces systèmes correspondent à un microsatellite composé et interrompu (DXS1238), un marqueur de polymorphismes nucléaires (*dys44*), le troisième système étant la combinaison des haplotypes de ces deux systèmes. À partir des fréquences haplotypiques de ces systèmes, différentes statistiques descriptives ont été calculées. Les taux d'hétérozygotie obtenus pour le Québec sont aussi élevés que ceux obtenus pour la population française. Les distributions des fréquences haplotypiques ainsi que les analyses de distances (ACP et F_{ST}) utilisées ont mis en évidence l'hétérogénéité du pool génique canadien-français. La Gaspésie apparaît comme un ensemble hétérogène de plusieurs populations, qui se comprend au regard de l'installation d'Européens d'origines diverses (Français, habitants des Îles britanniques, Anglo-normands) dès la fin du XVIII^e siècle. Si le Saguenay-Lac-Saint-Jean semble homogène, sa diversité génétique n'est pas inférieure aux autres populations québécoises. Enfin, la diversité génétique des Canadiens français montréalais présente elle aussi ses propres caractéristiques, liées à l'histoire du peuplement de cette région. À l'aide de ces systèmes de polymorphismes neutres, différents effets fondateurs régionaux ont été identifiés.

Mots clefs : Canadiens français, Québec, effet fondateur, chromosome X, génétique des populations

Abstract

This study of population structure is essential to the design of population base mapping of complex traits by genetic association. The French Canadian population of Quebec, currently about 6 million people, descends from about 10000 French settlers who arrived in Nouvelle-France between 1608 and 1765. To study the population structure of Quebec French Canadians, three regions were sampled: Montreal, Saguenay Lac-Saint-Jean and Gaspésie. These represent different population settling histories. The genetic diversity of these regions was compared with the genetic diversity of 5 French regions. Three polymorphic systems on Xp21 were chosen to study this genetic diversity: a microsatellite compound and interrupted (DXS1238), a marker of nuclear polymorphisms (*dys44*), and both of them were combined to form the third system (C haplotypes). Haplotype frequencies were used to calculate descriptive statistics for Quebec and French samples. Genetic diversity was similar into the populations. Distributions of haplotype frequencies and distance analysis (ACP and F_{ST}) highlighted the French Canadian gene pool heterogeneity. According to our data, Gaspé Peninsula is a heterogeneous set of different populations, genetically consistent with the installation of Europeans from different origin, known as French, Channel Islanders, and inhabitants of British Islands, starting at the end of the 18th century. Even if Saguenay-Lac-Saint-Jean is considered homogeneous, its genetic diversity is as diverse as the other Quebec populations. The Montreal French Canadian gene diversity further illustrates the demographic history of Montreal region. Using these neutral polymorphisms, different regional founder effects were identified.

Key words : French Canadian, Quebec, founder effect, X chromosome, population genetics

Table des matières

| | |
|---|-------------|
| RESUME | III |
| ABSTRACT | IV |
| LISTE DES TABLEAUX | VII |
| LISTE DES FIGURES | VIII |
| REMERCIEMENTS | XII |
| I. INTRODUCTION | 1 |
| A. Des histoires de peuplement régionaux | 3 |
| 1. Le peuplement de Montréal | 3 |
| 2. Le Saguenay-Lac-Saint-Jean | 4 |
| 3. La Gaspésie | 4 |
| B. Une étude de diversité génétique | 7 |
| C. Les microsatellites | 7 |
| D. Le chromosome X | 10 |
| II. MATERIEL ET METHODES | 11 |
| A. Les échantillons | 12 |
| 1. Échantillons québécois | 12 |
| 2. Échantillons français | 13 |
| 3. Extraction de l'ADN..... | 14 |
| B. Les marqueurs | 14 |
| 1. dys44 | 15 |
| 2. DXS1238..... | 16 |
| C. Analyses | 17 |
| 1. Les graphiques de partage d'haplotypes..... | 19 |
| 2. Les indices de diversité | 20 |
| 3. Le test de Chakraborty..... | 22 |
| 4. Les analyses de distances | 22 |
| a) Les ACPs..... | 22 |
| b) Les F_{ST} | 23 |
| III. RESULTATS | 24 |
| A. DXS1238 | 26 |

| | | |
|------------|--|------------|
| 1. | Nomenclature | 26 |
| 2. | Distribution des haplogroupes de DXS1238 | 27 |
| 3. | Les distributions des fréquences haplotypiques..... | 28 |
| 4. | Haplotypes partagés..... | 32 |
| 5. | Diversité génétique..... | 35 |
| a) | Indices de diversité | 35 |
| b) | Le test de Chakraborty..... | 37 |
| 6. | Les analyses de distance..... | 37 |
| a) | Les ACPs..... | 37 |
| b) | Les F_{ST} | 41 |
| B. | Les haplotypes de dys44..... | 43 |
| 1. | Distribution des haplotypes | 43 |
| 2. | Partage d'haplotypes | 46 |
| 3. | Diversité génétique..... | 49 |
| 4. | Les analyses de distance..... | 50 |
| C. | Les haplotypes composés (haplotypes-C) | 53 |
| 1. | Distribution des Haplotypes-C | 53 |
| a) | Vue générale..... | 54 |
| b) | Les différentes fréquences d'haplotypes-C | 54 |
| c) | Les haplotypes-C spécifiques au Québec | 59 |
| 2. | Partage d'haplotypes-C | 60 |
| 3. | Les Indices de Diversité | 62 |
| 4. | Les analyses de distance..... | 63 |
| IV. | DISCUSSION | 68 |
| A. | La question de l'échantillonnage..... | 70 |
| B. | L'effet fondateur au niveau de la province | 71 |
| C. | Diversité génétique et effet fondateur | 73 |
| D. | Une seule population canadienne-française ? | 76 |
| E. | Des effets fondateurs | 80 |
| 1. | Les Acadiens. | 80 |
| 2. | Les Anglo-normands | 84 |
| 3. | Les Canadiens français de Gaspésie..... | 85 |
| 4. | Les Loyalistes..... | 87 |
| 5. | Le Saguenay-Lac-Saint-Jean | 88 |
| 6. | Montréal | 90 |
| F. | CONCLUSION..... | 93 |
| | BIBLIOGRAPHIE..... | XCV |

Liste des tableaux

| | |
|---|----|
| <i>Tableau 1 Lieux d'origine et subdivisions géographiques des participants français.</i> | |
| | 14 |
| <i>Tableau 2 Indices de diversité du microsatellite DXS1238 dans les populations échantillonnées.</i> | |
| | 36 |
| <i>Tableau 3 Indices de diversité de dys44 dans les populations échantillonnées.</i> | 49 |
| <i>Tableau 4 Indices de diversité des haplotypes-C dans les populations échantillonnées.</i> | |
| | 63 |

Liste des figures

| | |
|---|----|
| <i>Figure 1 Chronologies des peuplements du Québec, de Charlevoix et de la Gaspésie.</i> | 6 |
| <i>Figure 2 Lieux d'origine des participants au Québec.</i> | 12 |
| <i>Figure 3 Localisation des séquences dys44 et DXS1238 sur le chromosome X (Xp21). Les positions des 35 polymorphismes et le polymorphisme de longueur (T_n) sont indiqués pour la séquence dys44, ainsi que celle du microsatellite composé et interrompu DXS1238.</i> | 15 |
| <i>Figure 4 Conditions requises pour l'amplification du microsatellite.</i> | 16 |
| <i>Figure 5 Conditions requises pour la réaction de séquence du microsatellite.</i> | 17 |
| <i>Figure 6 Gel de séquence de deux échantillons dans le sens sens (canal 800 correspondant à 795 nm).</i> | 18 |
| <i>Figure 7 Nomenclature de DXS1238.</i> | 27 |
| <i>Figure 8 Fréquences relatives des haplogroupes DXS1238 dans les populations échantillonnées.</i> | 28 |
| <i>Figure 9 Distribution des fréquences des Haplotypes de DXS1238 suivant les populations québécoises prises en compte. Les haplotypes sont classés par ordre décroissant pas rapport à la population française au complet (France). Les flèches correspondent aux haplotypes mentionnés dans le texte.</i> | 30 |
| <i>Figure 10 Distribution des fréquences des Haplotypes de DXS1238 suivant les populations françaises prises en compte. Les haplotypes sont classés par ordre décroissant pas rapport à la population française au complet (France). Les flèches correspondent aux haplotypes mentionnés dans le texte.</i> | 31 |
| <i>Figure 11 Graphiques de partages d'haplotypes DXS1238 entre différents groupes de population : (a) entre populations québécoises ; (b) entre populations québécoises</i> | |

| | |
|--|----|
| <i>et la France au complet ; (c) entre régions françaises ; (d) entre régions françaises et populations québécoises.</i> | 33 |
| <i>Figure 12 Analyse en Composantes Principales réalisée à partir des fréquences haplotypiques DXS1238 de la France et des populations québécoises.</i> | 39 |
| <i>Figure 13 Représentation MDS des F_{ST} réalisée à partir des fréquences haplotypiques DXS1238 de la France et des populations québécoises.</i> | 39 |
| <i>Figure 14 Analyse en composantes principales réalisée à partir des fréquences haplotypiques DXS1238 des régions françaises et des populations québécoises.</i> | 40 |
| <i>Figure 15 Représentation MDS des F_{ST} réalisée à partir des fréquences haplotypiques DXS1238 des régions françaises et des populations québécoises.</i> | 40 |
| <i>Figure 16 Analyse en composantes principales réalisée à partir des fréquences haplotypiques DXS1238 des régions françaises et du Québec.</i> | 41 |
| <i>Figure 17 Distribution des fréquences des Haplotypes dys44 suivant les populations considérées. Les haplotypes sont classés par ordre décroissant par rapport à la population française au complet (France). Les flèches correspondent aux haplotypes spécifiques en France (à gauche) et spécifiques au Québec (à droite).</i> | 45 |
| <i>Figure 18 Graphiques de partages d'haplotypes dys44 entre différents groupes de population : (a) entre populations québécoises ; (b) entre populations québécoises et la France au complet ; (c) entre régions françaises ; (d) entre régions françaises et populations québécoises.</i> | 48 |
| <i>Figure 19 Analyse en Composantes Principales réalisée à partir des fréquences des haplotypes dys44 des sous-populations québécoises et de la France.</i> | 52 |
| <i>Figure 20 Analyse en Composantes Principales réalisée à partir des fréquences des haplotypes dys44 du Québec et des régions françaises.</i> | 52 |
| <i>Figure 21 Distributions québécoises des fréquences des haplotypes-C communs à la France et au Québec et spécifiques en France. Les haplotypes-C sont classés par</i> | |

ordre décroissant par rapport à la France. Les flèches correspondent aux haplotypes-C mentionnés dans le texte..... 55

Figure 22 Distributions françaises des fréquences des haplotypes-C communs à la France et au Québec et spécifiques en France. Les haplotypes-C sont classés par ordre décroissant par rapport à la France. Les flèches correspondent aux haplotypes mentionnés dans le texte..... 56

Figure 23 Distributions des fréquences des haplotypes-C spécifiques au Québec. Les haplotypes-C sont classés par ordre décroissant par rapport au Québec au complet. Les flèches correspondent aux haplotypes-C mentionnés dans le texte..... 57

Figure 24 Graphiques de partages d'haplotypes-C entre différents groupes de population : a) entre populations québécoises ; (b) entre populations québécoises et la France au complet ; (c) entre régions françaises ; (d) entre régions françaises et populations québécoises..... 61

Figure 25 ACP réalisée à partir des fréquences des haplotypes-C de la France et des populations québécoises..... 64

Figure 26 ACP réalisée à partir des fréquences des haplotypes-C du Québec et des régions françaises 64

Figure 27 Représentation MDS des F_{ST} réalisée à partir des fréquences des haplotypes-C de la France et des populations québécoises..... 64

Figure 28 Représentation MDS des F_{ST} réalisée à partir des fréquences des haplotypes-C du Québec et des régions françaises..... 64

Figure 29 ACP de l'ensemble des populations québécoises et régions françaises, réalisées à partir des fréquences des haplotypes-C..... 65

Figure 30 Représentation MDS des F_{ST} de l'ensemble des populations québécoises et régions françaises, réalisées à partir des fréquences des haplotypes-C. 65

Figure 31 D'après un travail de Cathy Cauvier, apparemment intra et intergroupe dans les quatre populations gaspésiennes [1]..... 82

Liste des abréviations

- # chr** : nombre de chromosomes
haploG : nombre d'haplogroupes
haploT : nombre d'haplotypes
haploT att : nombre d'haplotypes attendu
ACP : analyse en composantes principales
ADN : acide désoxyribonucléique
ADNmt : acide désoxyribonucléique mitochondrial
AngloN : population des Gaspésiens d'origine anglo-normande
ASO : allele specific oligonucleotide
CanFr : population des Gaspésiens d'origine canadienne-française
CentreWFr : population du Centre-ouest de la France
dNTP : deoxynucléotide-5'-triphosphate
ddNTP : di-deoxynucléotide-5'-triphosphate
Fr Urb ou **Fr urbaine** : échantillon de France urbaine
hétéroZ : hétérozygotie
MDS : multidimensional scaling
N_e : effectif efficace d'une population
NordFr : population du Nord de la France
pb : paire de base
PCR : polymerase chain reaction
SudWFr : population du Sud-ouest de la France

Remerciements

Je tiens tout particulièrement à remercier mon directeur de recherche, Dr Damian Labuda, de m'avoir accueillie dans son laboratoire et permit de travailler avec lui. Je souhaite le remercier de ne jamais oublier son humanité au milieu de cet univers de chercheurs.

Je remercie grandement toute son équipe de recherche qui m'a reçue à l'hôpital Sainte-Justine, pour la précieuse aide technique et scientifique du Dr Vania Yotova et de Me Claudia Moreau, ainsi que de l'appui d'Elias Gbeha et de Stéphane Bourgeois.

Je voudrais remercier le Dr Daniel Sinnott de ses conseils, ainsi que les membres de son laboratoire pour l'excellente ambiance de travail et leur soutien.

I. INTRODUCTION

La population du Québec dépasse aujourd'hui sept millions de personnes, dont six millions descendent des premiers colons français, arrivés en Nouvelle-France entre 1608 et 1759. L'histoire du peuplement de cette province [2] est reflétée à travers un certain nombre de maladies génétiques, survenant entre autres à une fréquence accrue dans certaines populations régionales québécoises par rapport aux autres populations mondiales [3, 4]. Effet fondateur et dérive génétique sont généralement cités pour expliquer la répartition singulière de ces maladies au Québec [3-10].

L'effet fondateur définit le processus de changement des fréquences alléliques suite à la création d'une nouvelle population à partir d'un nombre restreint d'individus provenant d'une population mère. Il consiste en un échantillonnage aléatoire d'individus de la population source à la nouvelle population. Il a d'autant plus d'importance dans la nouvelle population formée que la taille de cet échantillon fondateur est réduite [11]. La dérive génétique, qui consiste en la fluctuation aléatoire des allèles d'une génération à l'autre, peut accentuer l'effet fondateur. En effet la dérive génétique a lieu dans toutes les populations, mais elle est d'autant plus drastique que l'effectif de la population est faible. Elle conduit à la perte, ou à la fixation, aléatoire de certains variants génétiques. Ainsi la nouvelle population formée sera distincte de sa population source en ce qui touche aux fréquences alléliques [12].

La croissance démographique sous le Régime français était forte, et s'est maintenue jusqu'au milieu du XIXe siècle [2]. Les territoires peuplés en bordure du Saint-Laurent furent les premiers peuplés. Saturés dès le premier quart du XIXe siècle, ce terroir canadien-français expédiera alors ses excédents démographiques vers les régions périphériques. Les migrations de la population du Québec, ainsi poussée à la colonisation de régions non peuplées, ont-elles conduit à la formation d'isolats régionaux ? Nous nous sommes particulièrement concentrés sur la région de Montréal, la Gaspésie et le Saguenay-Lac-Saint-Jean, qui illustrent des histoires de peuplement très différentes (Figure 1). L'objectif de ce projet est d'évaluer la diversité génétique dans trois régions du Québec, dont l'histoire est différente, par rapport à la diversité de la population source. La population source est représentée par l'échantillonnage de six régions françaises, dont un de « France urbaine ». Notre échantillon des régions françaises correspond à des régions de l'ouest de la France,

représentant relativement les principales régions d'origines des premiers immigrants français venus s'installer en Nouvelle-France.

La diversité génétique des Canadiens français du Québec a été appréhendée à l'aide de trois marqueurs de polymorphismes neutres situés sur le chromosome X. Le degré d'homogénéité des populations québécoises a été étudié et des liens entre ces dernières et les régions françaises échantillonnées ont pu être établis. D'après nos trois marqueurs de polymorphismes neutres, la diversité génétique au Québec n'est pas plus basse qu'en France. Les distributions des fréquences des haplotypes ont révélé l'existence d'effets fondateurs régionalement distincts, en accord avec les histoires du peuplement de ces différentes régions. Nos résultats corroborent la diversité du pool génique canadien-français québécois déjà révélée lors d'études de certaines maladies génétiques et d'études démographiques [3, 4, 13, 14].

A. Des histoires de peuplement régionaux

1. Le peuplement de Montréal

Montréal a été fondée en 1642. Dès ses débuts, la population de la ville a connu des rythmes d'accroissement divers. D'abord formée de colons français, la population montréalaise s'est développée au rythme des mouvements d'immigration des XVIIe et XVIIIe siècle, auquel s'ajoute la croissance naturelle. Au XVIIe siècle, la population urbaine peut être estimée à 1150 personnes. Entre 1650 et 1686, la population augmente rapidement. Puis la période d'incertitude et d'insécurité qui suit amène un ralentissement et la population montréalaise passe de 3000 individus en 1731 à 4000 en 1754 [15].

A partir de 1815, la Grande Bretagne permet l'émigration de ses habitants. Cette immigration massive de Britanniques d'origine entraîne une population montréalaise majoritairement anglophone dès 1832. En 1867, la ville devient de nouveau majoritairement francophone. Au milieu du XIXe siècle, Montréal reçoit la population des régions environnantes. Elle symbolise alors une aire d'attraction avec des perspectives d'emplois dans un contexte de saturation démographique des campagnes. Le recensement de 1871 dénombre 107 225 personnes à Montréal, sans les banlieues indépendantes qui l'entoure [15].

Montréal n'a cessé de connaître l'installation de communautés d'origines différentes [16]. Le recensement de 1941 compte 903 007 personnes. En 1971, la proportion relative des francophones correspond à 64% et à 11% pour les anglophones [15]. A cette période, seul un quart de la population n'est pas d'origine française ou britannique.

2. Le Saguenay-Lac-Saint-Jean

Nous connaissons le nombre élevé de maladies héréditaires au Saguenay-Lac-Saint-Jean, souvent spécifiques à cette région, et généralement attribuées à l'homogénéité de sa population fondatrice : principalement des familles originaires de Charlevoix [17, 18]. Etant donné que les fondateurs de Charlevoix mariés avant 1725 ont au moins un descendant qui sera à son tour fondateur du Saguenay, on comprend mieux la filiation étroite entre les deux populations [19, 20]. Le peuplement de Charlevoix a débuté en 1675. À la fin du XVIIe siècle on compte 200 personnes à Charlevoix, et plus de 1000 personnes en 1765, pour arriver à 8400 personnes en 1831. On considère que l'ensemble des fondateurs de Charlevoix comprend 599 personnes, soit 341 hommes et 258 femmes.

La date officielle d'installation des premiers colons de Charlevoix venus s'installer au Saguenay est 1838. Entre 1838 et 1861, 5400 personnes originaires de Charlevoix se sont établies au Saguenay ; alors qu'à la même époque Charlevoix compte 12000 habitants. L'immigration de Charlevoix vers le Saguenay est généralement de type familial, avec seulement 6% d'immigrants isolés provenant de Charlevoix contre 33% pour les autres lieux de provenance. Ainsi la similarité des patronymes les plus fréquents et des taux d'incidence et de prévalence de certaines génopathies trouvent leur origine dans ce transfert de population [19].

3. La Gaspésie

Principalement attirés par l'activité halieutique de la Gaspésie, des Européens d'origines diverses s'installent dans la péninsule gaspésienne dès la fin du XVIIe

siècle [21]. Dès 1699, l'essor de l'activité halieutique de la péninsule fût l'objet de l'installation des premiers Canadiens français. Quelques familles demeurèrent par la suite dans la région.

Puis les années qui précèdent la Conquête amènent l'installation sur les côtes de la Gaspésie des Acadiens [21]. Fuyant la Déportation de 1755, les Acadiens trouvent principalement refuge dans la Baie-des-Chaleurs, à l'embouchure de la Rivière Ristigouche. Les descendants de cette population constituent de nos jours au moins 60% de la population habitant la partie sud de la péninsule.

Les Anglo-normands arrivent en Gaspésie dès les années 1766-1770. Les premiers résidents permanents de la pointe de Forillon seraient originaires des îles anglo-normandes Jersey et Guernesey, attirés par l'activité halieutique de la péninsule. Près de la moitié des Gaspésiens compteraient des Anglo-normands parmi leurs ancêtres [21].

En 1784, fuyant les Treize Colonies, 315 Loyalistes débarquent à Paspébiac. Plusieurs des sujets britanniques qui s'installent dans cette région à ce moment sont des soldats libérés de leur engagement dans l'armée. Ils sont à l'origine de la création de paroisses comme New-Carlisle, Carleton et New-Richmond. Aujourd'hui leurs descendants constituent l'un des plus importants groupes ethniques de la Baie-des-Chaleurs [22].

Submergés numériquement par les nouveaux arrivants débarqués dans la région à la suite de la Conquête, les Canadiens français reviennent dans la péninsule par la suite. Ils en peuplent la côte nord à partir des années 1820. Une présence culturelle tranchée entre le nord et le sud de la Gaspésie contribue dès lors, et pendant longtemps, à donner une double facette à la région. Vers 1850, toutes ces communautés réunies totalisaient 20 000 habitants, et près de la moitié étaient d'expression anglaise.

Chaque communauté s'est initialement regroupée en villages; en entrant dans la Baie des Chaleurs, se succèdent en alternance le long de la route nationale : Grande-Rivière (Canadiens français), Pabos (Bretons et Normands), Newport (Irlandais), Port-Daniel (Canadiens français), Paspébiac (Basques et Jersiais), New-Carlisle (Loyalistes), Bonaventure (Acadiens), New-Richmond (Loyalistes), Maria

(Micmacs), etc [22, 23]. Ne pouvant plus compter sur les flux d'immigration de la fin du XVIIIe siècle, la proportion d'anglophones dans la péninsule subit une lente décroissance. Aujourd'hui la Gaspésie compte 80 000 habitants dont la grande majorité est d'expression française.

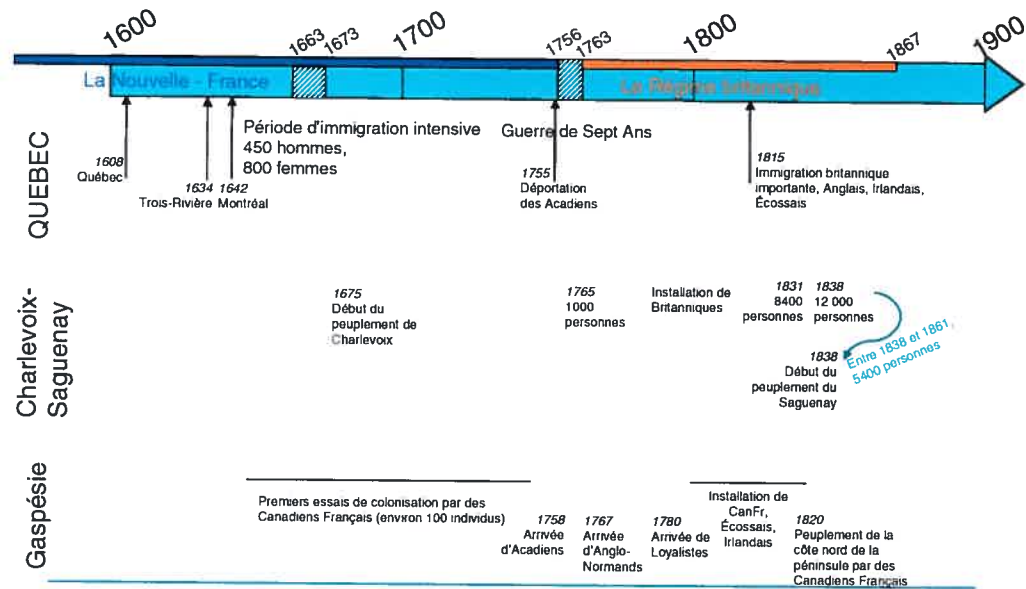


Figure 1 Chronologies des peuplements du Québec, de Charlevoix et de la Gaspésie.

La Gaspésie et le Saguenay-Lac-Saint-Jean ont connu une immigration importante à une période où les moyens de transports n'étaient pas ce qu'ils sont aujourd'hui. Ces deux régions étaient donc relativement isolées des lieux de passages et d'activités importantes que constituaient les régions de Montréal et de Québec. La région de Montréal a elle aussi représenté un lieu d'attraction au moment du développement de l'industrialisation et de l'urbanisation [16]. L'étude de la distribution régionale des patronymes ainsi que l'étude de la répartition des maladies, de même que les études généalogiques d'origine des fondateurs ont déjà illustré le paradoxe québécois, soit une certaine homogénéité régionale couplée à une hétérogénéité provinciale [3, 4, 19].

B. Une étude de diversité génétique

Les études d'association, qui tentent d'identifier des variants génétiques impliqués dans des maladies humaines communes, doivent tenir compte de la structure des populations étudiées [24, 25]. La structure d'une population détermine la transmission de sa variabilité génétique et peut affecter l'élaboration d'études d'association [26]. La structure d'une population est inférée à partir d'outils de génétique des populations, qui étudie la diversité génétique des populations.

De nombreuses populations voient leurs diversités génétiques étudiées par des marqueurs tels que l'ADN mitochondrial, retraçant les lignées maternelles, le chromosome Y pour les lignées paternelles, ou encore des systèmes autosomaux [27-29]. Ces études permettent d'établir la diversité génétique des populations humaines mondiales, de comprendre leurs liens ancestraux, et la répartition des variants alléliques de différents systèmes. De telles études ont déjà été entreprises pour différentes populations connaissant des fréquences élevées de maladies génétiques [29-32]. Notre projet a pour but d'initier les prémices d'une telle étude au Québec, en observant la diversité génétique de trois régions québécoises, soit les régions de Montréal, du Saguenay-Lac-Saint-Jean et de la Gaspésie.

C. Les microsatellites

Du fait que nous désirons appréhender la diversité de trois régions au sein d'une même province, nous ne nous attendons pas à trouver de grandes différences de fréquences au niveau de marqueurs communs. Les microsatellites sont souvent les marqueurs choisis pour ces études régionales [33-35]. Un microsatellite est une séquence d'ADN spécifique qui contient des répétitions en tandem de mono-, di-, tri-, ou tetra- nucléotides. Par exemple,

AAA AAA AAA AA sera noté $(A)_{11}$

GTGTGTGTGTGT sera noté $(GT)_6$

ACTCACTCACTCACTC sera noté $(ACTC)_4$

Les allèles d'un locus particulier peuvent différer dans le nombre de répétitions. Ils sont transmis de façon mendélienne.

Jusqu'à ce jour, les microsatellites ont été retrouvés dans tous les génomes eucaryotes. Ils sont hautement variables et peuvent être utilisés dans de nombreuses disciplines. Les microsatellites sont souvent utilisés pour l'identification légale ou les tests de parentés. Ils peuvent permettre la détection précoce de certaines maladies [36, 37], ou être utilisés dans des études de déséquilibre de liaison [38, 39]. En regardant la variation des microsatellites au niveau des populations, des inférences sur leur structure, le métissage, les migrations ou les effets de fragmentation de population peuvent être réalisées. Des événements comme la dérive génétique ou des goulets d'étranglements (*bottlenecks*), ou la date de séparation de populations [40], peuvent également être déduits à partir de ces systèmes.

Les microsatellites sont hypervariables, et leur évolution consiste en un changement de longueur au cours du temps. Dans une population, il peut exister de nombreux allèles à un même locus, les allèles se distinguant par le nombre de répétitions, soit des différences de longueurs.

Ces différents allèles sont créés par mutation. Cette caractéristique fait des microsatellites un système polymorphe à plusieurs allèles. Ils sont souvent utilisés dans des études d'évolution comprenant des intervalles de temps relativement courts (cent à mille ans), tandis que les mutations ponctuelles sont généralement plus utiles dans les études comprenant des intervalles de temps plus longs (des millions d'années) [41].

Deux hypothèses sont habituellement abordées pour expliquer le taux élevé de mutation des microsatellites [41] :

- Le glissement (*slippage*) peut apparaître lorsque le brin d'ADN natif se dissocie du brin matrice. Lorsqu'il s'agit de séquences non-répétitives, aucun problème ne se pose puisqu'il n'y a qu'une seule façon d'apparier précisément le brin natif au brin matrice. Cependant si la séquence répliquée est de nature répétitive, alors le brin natif peut s'apparier avec la répétition suivante, ou la répétition précédente, du brin matrice formant une boucle sur l'un ou l'autre brin. Quand la réplication continue

après un tel mésappariement, le brin natif résultant sera plus long ou plus court que le brin matrice, suivant le lieu de formation de la boucle.

- Quant à la recombinaison, elle peut altérer la longueur des microsatellites de deux façons : par *crossing-over* inégal ou par conversion génique. Le *crossing-over* inégal se produit entre deux chromosomes mal alignés, et survient la plupart du temps pour de longues séquences répétées en tandem, pour lesquelles la machinerie de recombinaison ne peut pas repérer son erreur facilement. Ce phénomène peut avoir lieu entre chromatides d'un même chromosome, ou entre chromosomes. La conversion génique est un transfert unidirectionnel d'information pendant la recombinaison. Probablement effective lors de dommages à l'ADN, il a été suggéré qu'elle crée de la diversité chez les minisatellites (dont les motifs sont plus longs que pour les microsatellites). Dans ce dernier cas, les loci semblent acquérir de nouvelles séquences de façon polaire.

Trois modèles de mutation sont généralement mentionnés pour tenter d'expliquer les processus mutationnels des microsatellites :

- Modèle de mutation par étape - *Stepwise Mutation Model* (SMM) – Ce modèle considère que lorsque des microsatellites mutent, le passage d'un allèle à un autre se fait par l'ajout ou la suppression d'une seule unité de répétition. D'après ce modèle, il faut tenir compte de la taille de l'allèle dans la structure de la population. La distance génétique généralement utilisée est la R_{ST} [34]. Ce modèle omet cependant de tenir compte de l'homoplasie. Lorsque deux allèles sont identiques, ils peuvent provenir du même ancêtre commun ou d'un événement de mutation, c'est à dire d'un ajout d'une unité de répétition à partir d'une longueur plus petite, ou une suppression d'une unité de répétition à partir d'un allèle plus long. Dans le premier cas les allèles sont identiques par ascendance, dans le second cas (si il y a eu mutation) ils sont seulement identiques par état. Ce phénomène est appelé homoplasie.

- Modèle des « K » allèles – *K alleles model* – ce modèle considère qu'un microsatellite peut muter de façon aléatoire en n'importe lequel des K allèles du locus.

- Modèle à nombre infini d'allèles – *Infinite Alleles Model* (IAM) – Pour ce modèle, chaque mutation peut créer aléatoirement n'importe quel allèle. Un allèle de

15 répétitions sera aussi proche qu'un allèle de dix répétitions l'est d'un allèle de 11 répétitions. D'après ce modèle, la taille importe peu dans la structure de la population, et la statistique habituellement utilisée pour étudier cette dernière est la F_{ST} .

D. Le chromosome X

Le microsatellite que nous avons choisi, DXS1238, est situé sur Xp21. Le chromosome X étant transmis par les deux parents, le biais sexuel rencontré pour l'ADN mitochondrial et le chromosome Y, respectivement transmis par la mère et par le père, disparaît. Ce mode de transmission permet aussi de limiter les effets de la dérive génétique en portant la population effective relative du chromosome X à $3/2$ par rapport aux autosomes, contre $1/2$ pour l'ADN mitochondrial et le chromosome Y. Nous avons d'abord choisi un microsatellite composé et interrompu, DXS1238, dont la nature nous permet de calibrer la diversité continentale (haplogroupes) ou régionale (haplotypes).

Compte tenu que plus on dispose de marqueurs pour étudier la diversité génétique d'une population, plus on a une idée réelle de sa structure, nous avons utilisé un autre marqueur en amont de DXS1238. Il s'agit de *dys44*, situé 8 kb en amont de DXS1238. Il a précédemment été montré que les polymorphismes de *dys44* révélaient des distributions continentales différentes [42, 43]. Etant donné le polymorphisme élevé de DXS1238, *dys44* nous a permis d'évaluer la diversité de ces populations différemment. Ces deux polymorphismes analysés ensemble comme un seul marqueur (haplotypes-C) ont apporté des informations complémentaires quant aux liens entre populations et aux particularités de ces populations.

II. MATERIEL ET METHODES

A. Les échantillons

Ce projet de maîtrise s'inscrit dans un projet plus vaste visant à étudier la diversité génétique des populations régionales du Québec. Jusqu'à présent trois régions québécoises ont été échantillonnées : Montréal, le Saguenay-Lac-Saint-Jean et la Gaspésie. Des collaborateurs français nous ont procuré des échantillons de plusieurs régions françaises, correspondant aux principaux lieux d'origine des premiers immigrants français en Nouvelle-France. Les diversités des pools géniques du Québec et de la France pourront ainsi être comparées. Les individus ont été sélectionnés sur une base non nominative et sont non apparentés (cousins germains exclus de l'échantillonnage).

1. Échantillons québécois

Les échantillons proviennent de trois régions québécoises : le Saguenay-Lac-Saint-Jean, la Gaspésie et la région de Montréal (Figure 2). L'ADN des échantillons du Saguenay-Lac-Saint-Jean (Sag, n=83) provient du laboratoire de Catherine Laprise (Université du Québec à Chicoutimi). Pour être recruté, un Saguenayen devait être Canadien français et avoir ses quatre grands-parents ancrés dans la région.

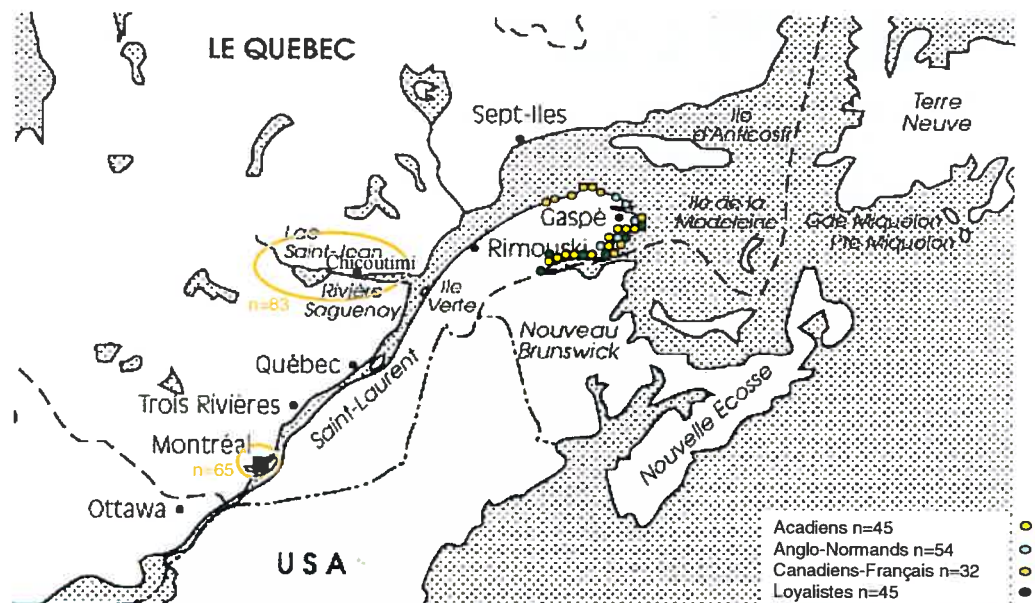


Figure 2 Lieux d'origine des participants au Québec.

En ce qui concerne la Gaspésie, les échantillons sanguins représentent quatre groupes de la péninsule (soit n le nombre de chromosomes utilisés dans cette étude) : les participants d'origine canadienne-française (CanFr, $n=33$), ceux d'origine acadienne (AC, $n=45$ chromosomes génotypés), ceux d'origine loyaliste (Loy, $n=40$ chromosomes génotypés) et ceux d'origine anglo-normande (AngloN, $n=54$). Ces quatre échantillons correspondent aux quatre ethnies principales ayant fondées la Gaspésie. Les échantillons de Gaspésie ont été sélectionnés par des recruteurs locaux pour chacune des quatre populations. Les participants gaspésiens ont eux-mêmes déclarés leur ascendance. Le recrutement a été réalisé de façon à ce que les quatre groupes présentent des effectifs comparables.

Les échantillons de Montréal proviennent du centre de prélèvement ou de la salle d'accouchement de l'hôpital Sainte-Justine à Montréal. Pour être recruté un participant montréalais devait être Canadien français installé à Montréal avec quatre grands-parents nés au Québec. Cet échantillon comprend les personnes habitants à Montréal depuis une génération comme les personnes montréalaises depuis plusieurs générations. Ainsi nous pourrions supposer que l'échantillon montréalais (Mtl, $n=65$) illustrera une diversité canadienne-française urbaine.

2. Échantillons français

Les échantillons français proviennent du laboratoire de Jean-Paul Moisan (Nantes, France), sauf les échantillons du Sud de la France qui proviennent du laboratoire d'Eric Bieth (Toulouse, France). L'échantillon de Bretagne correspond à des participants bretons dont les quatre grands-parents étaient ancrés dans la région. Nous assumons que le même mode de recrutement a été utilisé pour les autres régions. L'échantillon appelé « France urbaine » correspond à un regroupement d'échantillons provenant de plusieurs villes importantes, dont Lyon et Paris. Les grandes villes constituent une aire d'attraction pour les régions environnantes et peuvent représenter un échantillon de la diversité d'un pays.

Les régions échantillonnées, ainsi que les regroupement réalisés pour l'analyse des résultats (en gras), sont détaillés dans le tableau 1. D'un point de vue historique, les régions françaises échantillonnées correspondent aux principaux lieux d'origine

des premiers immigrants français en Nouvelle-France. Cependant au regard du tableau 1, on s'aperçoit que les tailles de ces échantillons régionaux ne sont pas équivalentes. Un regroupement des régions a alors été fait de sorte que chaque subdivision représente un compromis entre une définition historique régionale, et une taille d'effectif relativement élevé rendant nos régions comparables.

| Régions | Effectifs | Subdivisions |
|-------------------------|-----------|---------------------------------------|
| Picardie | 35 | Nord de la France (48) |
| Haute-Normandie | 13 | |
| Bretagne | 31 | Centre-Ouest de la France (35) |
| Loire-Atlantique | 66 | |
| Autres pays de la Loire | 22 | |
| Centre | 6 | |
| Poitou-Charente | 7 | |
| Limousin | 6 | Sud-Ouest de la France (33) |
| Aquitaine | 3 | |
| Auvergne | 12 | |
| Midi-Pyrénées | 12 | |
| France urbaine | 46 | |



Tableau 1 Lieux d'origine et subdivisions géographiques des participants français.

3. Extraction de l'ADN

Pour les échantillons du Saguenay et de la France, l'ADN a directement été envoyé à l'Hôpital Sainte-Justine à Montréal. Quant aux échantillons gaspésiens et montréalais, l'ADN a été extrait à partir d'échantillons de 10ml de sang, à l'aide du protocole d'extraction des kits *Genra* (PUREGENE Genra Systems).

B. Les marqueurs

Les deux systèmes de polymorphismes utilisés dans cette étude sont situés sur le chromosome X (figure 3). Seuls les individus masculins ont été génotypés, permettant une lecture directe des haplotypes.

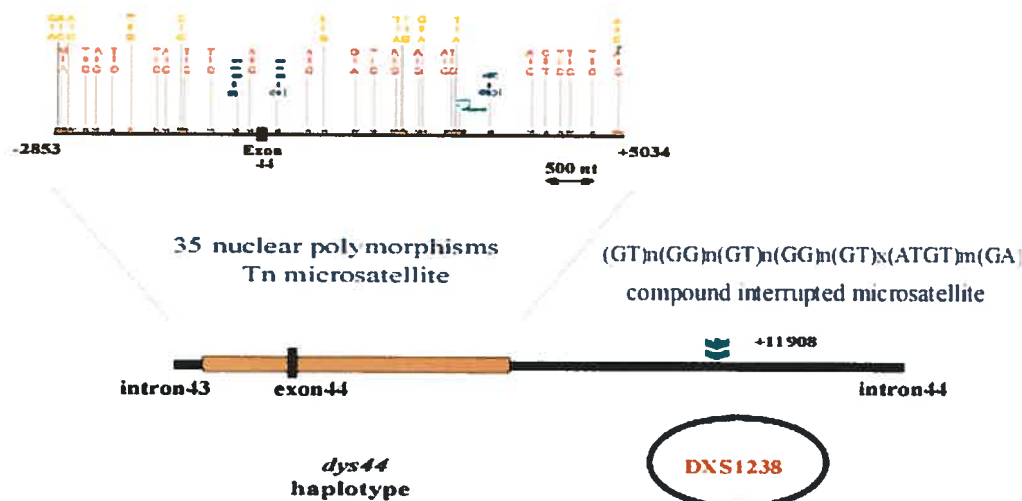


Figure 3 Localisation des séquences dys44 et DXS1238 sur le chromosome X (Xp21). Les positions des 35 polymorphismes et le polymorphisme de longueur (T_n) sont indiqués pour la séquence dys44, ainsi que celle du microsatellite composé et interrompu DXS1238.

1. *dys44*

La séquence *dys44* (U94396) s'étend sur l'exon 44 du gène de la dystrophine (ADNc positions 6499 à 6646) et les introns adjacents se trouvent entre les positions -2853 à -1 en amont et 1 à 5034 en aval. La longueur totale de cette séquence est de 8035 pb. Les polymorphismes ont été détectés par SSCP (*single strand conformational polymorphism*) et analyses d'hétéroduplexe [44-46]. Ces polymorphismes ont été caractérisés par séquençage [47]. Trente-cinq polymorphismes simples ont été identifiés, et la présence d'un microsatellite multiallélique (T)_a entre les positions 2453 et 2839 a également été observée [42, 43, 47, 48]. Les allèles ont été successivement déterminés par hybridation ASO (*allele specific oligonucleotide*) [49], à l'exception du polymorphisme de longueur, analysé par électrophorèse sur gel de polyacrilamide dénaturant [43].

Pour *dys44*, le génotypage par ASO des échantillons québécois et français a précédemment été réalisé au laboratoire par Vania Yotova, Claudia Moreau et Sylvie Bourthoumieu.

2. DXS1238

Ce fragment a été amplifié à partir de 0,5 ng d'ADN génomique. Les séquences des amorces utilisées sont $3'$ GTGAAACAAAATAGAAAATATCAGC $5'$ (amorce sens) et $3'$ AAATCACATTTCAATATAAGATTCG $5'$ (amorce anti-sens). Les PCR ont été réalisées dans un volume final de 20 μ L avec une concentration de tampon de 1X ; 1,5mM de MgCl₂ ; 0,2mM de dNTPs ; 0,25 μ M de chaque amorce et une unité de Taq platinum (*Invitrogen*). Les conditions d'amplification appliquées sont décrites dans la figure 4.

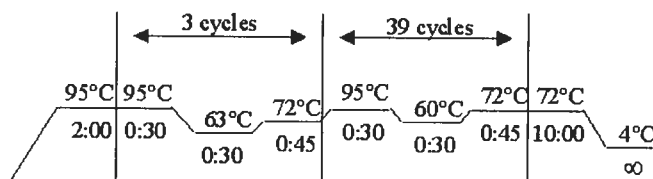


Figure 4 Conditions requises pour l'amplification du microsatellite.

A cause de sa nature complexe, le génotypage de DXS1238 a été réalisé par séquençage. L'ADN issu de cette amplification est migré sur un gel d'agarose de 1,3%. Les bandes correspondantes au fragment d'intérêt sont découpées, et l'ADN est purifié à l'aide de colonnes « *Spin - X centrifuge tube filter* » (*Corning Incorporated Costar*), avec des filtres d'acétate de cellulose de 0,22 μ m de diamètre. L'ADN amplifié est ainsi récupéré et peut alors servir pour la réaction de séquence.

Les réactions de séquence sont réalisées à l'aide des kits « *Thermo Sequenase Cycle Sequencing Kit* » (*USB corporation*). Le volume final de la réaction de séquence est de 8,75 μ L. Cette réaction est réalisée à partir de 2 μ L d'ADN issu de la première amplification. Les séquences des amorces fluorescentes utilisées sont les mêmes que pour l'amplification ; l'amorce sens marquée est utilisée à 0,10 μ M, tandis que l'amorce anti-sens est utilisée à 0,05 μ M. Quatre unités de Taq Thermosequenase sont ajoutées avec 1 μ L de *Thermosequenase reaction buffer*. Le « *Thermo Sequenase Cycle Sequencing Kit* » contient également des ddNTPs, et 0,75 μ M de chaque ddNTP est au préalable déposé dans la plaque servant à la réaction de séquence. A ces 0,75 μ M de ddNTP, 2 μ L de la réaction de séquence sont ajoutés. Pour la séquence d'un échantillon, quatre puits sont alors nécessaire, un pour

chaque nucléotide. Les conditions d'amplification de la réaction de séquence sont indiquées dans la figure 5.

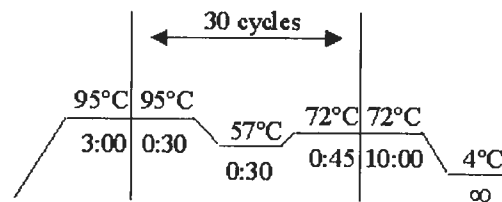


Figure 5 Conditions requises pour la réaction de séquence du microsatellite.

La réaction de séquence est achevée par 2,5 μ l de tampon d'arrêt. Avant de déposer les échantillons sur le gel, ceux-ci sont dénaturés à 95°C pendant trois minutes, et maintenus dans la glace, pour éviter que les brins complémentaires se rassemblent.

Lorsque le séquenceur a fini de détecter la séquence nous disposons de deux gels : un correspondant à notre fragment dans le sens *sens* (IRD800), l'autre correspondant à notre fragment dans le sens *anti-sens* (IRD700). La concordance de ces deux lectures permet une première validation de nos séquences. Une seconde lecture des gels par une personne supplémentaire permet de confirmer les résultats obtenus (figure 6). La nomenclature du microsatellite est décrite dans la partie des résultats.

C. Analyses

Après le génotypage des 581 chromosomes, les fréquences des haplotypes de chacun des deux marqueurs ont été calculées pour chaque population. Certains échantillons pour lesquels les résultats du génotypage n'étaient pas satisfaisants n'ont pas été inclus dans les analyses. Pour 0,5% des 581 chromosomes génotypés, les séquences obtenues n'étaient pas exploitables. Ceci explique les différences d'effectifs pour une même population entre les analyses de *dys44* et de DXS1238.

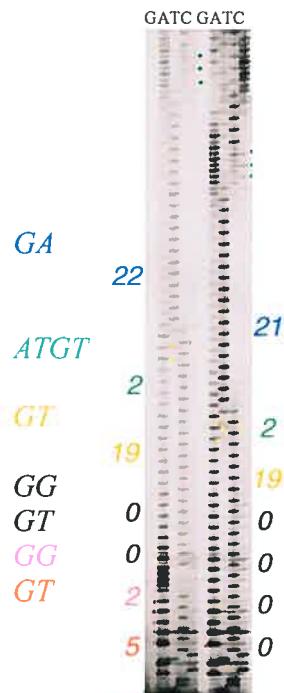


Figure 6 Gel de séquence de deux échantillons dans le sens sens (canal 800 correspondant à 795 nm).

La comparaison des populations québécoises et des régions françaises est basée sur les fréquences des haplotypes. A partir des fréquences des haplotypes des graphiques de partage d'haplotypes ont été réalisés. Ces fréquences haplotypiques nous ont également permis d'obtenir des statistiques descriptives (hétérozygotie, θ Hom, θ K) pour chaque population à l'aide du logiciel ARLEQUIN 2.0. A l'aide de ce logiciel, différents tests ont été utilisés tels que le test d'amalgamation de Chakraborty [50] et les analyses de distances F_{ST} [51] (version d'Arlequin 2.0). D'autres analyses de distances ont été réalisées à l'aide des Analyses en Composantes Principales (ACP). Les ACPs et les représentations en deux dimensions des valeurs des F_{ST} ont été obtenues à l'aide du logiciel STATISTICA 6.0.

Ces indices et ces analyses ont été réalisés pour chacun des deux marqueurs séparément, et pour les deux marqueurs ensemble. La combinaison des polymorphismes de ces deux marqueurs constitue un haplotype composé (haplotype-C). Plus de précisions concernant ces haplotypes-C seront apportées dans la partie « résultats ». Les paragraphes suivants expliquent de quelles façons les graphiques, les statistiques descriptives et les analyses de distances ont été obtenus.

1. Les graphiques de partage d'haplotypes

Les graphiques de partage d'haplotypes ont été réalisés dans le but d'observer les liens entre populations. Les graphiques de partage d'haplotypes sont obtenus à partir du nombre d'haplotypes du marqueur considéré et suivant les groupes de populations considérés. Pour chaque système, quatre graphiques ont été effectués : avec les populations québécoises seulement, avec les populations québécoises et la France au total, avec les régions françaises seulement, et avec l'ensemble des populations québécoises et des régions françaises.

Ces graphiques de partage sont constitués de trois catégories : les haplotypes uniques à une population, les haplotypes partagés avec une autre population (chaque couleur correspondant à une population), et les haplotypes partagés entre trois populations et plus, ou haplotypes communs. Le nombre d'haplotypes uniques

correspond au nombre d'haplotypes observés dans une seule population. Lorsqu'un haplotype unique est observé chez plus d'un individu dans une population, cet haplotype peut illustrer l'isolation relative de la population considérée par rapport aux autres populations comparées. Lorsqu'un haplotype est partagé par deux populations et présent chez plus d'un individu dans les deux populations, on peut penser que des migrations ont eu lieu entre ces populations. La proportion d'haplotypes communs montre la proportion du pool génique partagé entre les populations considérées.

Le nombre d'haplotypes observé dépendra de la taille de l'échantillon. Toutefois, nos échantillons québécois ont des tailles similaires, et sont donc comparables entre eux. Lorsque la France au total a été représentée dans le même graphique que les échantillons québécois, une contribution française régionale québécoise pouvait être observée, bien que la France au total ait un effectif cinq fois plus élevé que les régions québécoises échantillonnées ici.

2. Les indices de diversité

Plusieurs indices de diversité peuvent être utilisés pour comparer des populations ou des loci. La façon la plus simple est de compter le nombre d'haplotypes (K). Cependant le nombre d'haplotypes est très influencé par la taille de l'échantillon. L'hétérozygotie (HétéroZ) est une alternative au nombre d'haplotypes. L'hétérozygotie est la probabilité de prélever deux haplotypes différents dans un même échantillon. L'hétérozygotie est estimée par :

$$H = \frac{n}{n-1} \left(1 - \sum_{i=0}^k p_i^2 \right) ;$$

où n est la taille de l'échantillon, k le nombre d'haplotypes et p_i est la fréquence de l'haplotype i dans l'échantillon.

Un autre indice de diversité pour ces études de diversité génétique est le paramètre de mutation d'une population (θ). Sous un modèle d'évolution neutre, le niveau de diversité au sein d'une population peut atteindre une valeur d'équilibre au moment où l'apparition des nouveaux allèles par mutation (μ) est compensée par les allèles

perdus par dérive. Comme la dérive est inversement proportionnelle à l'effectif efficace (N_e) de la population, N_e est utilisée pour lier ces paramètres. Pour le chromosome X : $\theta = 3 \cdot N_e \cdot \mu$.

Ce paramètre peut être estimé de plusieurs façons, deux d'entre elles sont présentées ici. Le θ_k mesure la diversité basée sur le nombre d'haplotypes, tandis que le θ_{Hom} mesure la diversité à partir de l'homozygotie (soit 1-HétéroZ). En considérant le modèle des allèles infinis d'Ewens [52], le θ_{Hom} et le θ_k sont calculés comme suit :

$$E(\theta_{Hom}) = \theta \cdot \left(1 + \frac{2 \cdot (1 + \theta)}{(2 + \theta) \cdot (3 + \theta)} \right) \text{ avec l'homozygotie attendue } Hom = \frac{1}{1 + \theta}$$

D'après Chakraborty et Weiss [53] θ_{Hom} correspond à l'homozygotie attendue dans une population à l'équilibre dérive – mutation :

$$E(k) = \theta \cdot \sum_{i=0}^{n-1} \frac{1}{\theta + i} ;$$

où k est le nombre attendu d'allèles, et n la taille de l'échantillon. Les limites θ_a et θ_b de θ_k sont données (au lieu de sa variance) dans un intervalle de confiance de 95% :

$$P(\text{au moins } k \text{ haplotypes sachant que } \theta = \theta_a) = 0,025$$

$$P(\text{plus de } k \text{ haplotypes sachant que } \theta = \theta_b) = 0,025$$

Dans une population à l'équilibre, ces estimations doivent être identiques. Lorsque des différences significatives sont observées, plusieurs explications peuvent être fournies, comme des phénomènes de croissance démographique, ou de goulot d'étranglement, ou encore de dérive génétique. Il est important de prendre en compte plusieurs estimateurs de θ car deux phénomènes distincts peuvent exercer les mêmes contraintes sur un seul estimateur.

3. Le test de Chakraborty

Ce test est basé sur le modèle à nombre infini d'allèles (*infinite-allele model*) et sur la théorie d'échantillonnage des allèles neutres [52]. Le test de Chakraborty constitue un test supplémentaire pour observer la structure – ou l'homogénéité - des régions étudiées.

Le test de Chakraborty [50] se base sur l'homozygotie observée (θ_{Hom}) pour calculer un nombre d'allèles attendu. Un test de Chakraborty « positif » ou « significatif » signifie que la différence entre le nombre d'allèles observé et le nombre d'allèles attendu est significative. Ce test peut-être utilisé pour révéler un effet d'échantillonnage. Si la proportion d'allèles attendus est plus basse que la proportion d'allèles observés, alors notre échantillon analysé comme une seule population homogène, révèle en fait la présence de deux populations, au sein de cet échantillon. La mise en commun de ces deux pools géniques distincts fait alors apparaître l'échantillon comme « hétérogène », ou structuré, et le test de Chakraborty sera positif.

Plusieurs explications peuvent être fournies pour expliquer un test de Chakraborty significatif, illustrant une distribution des allèles observée différente de celle qui est attendue. Seules les probabilités significatives (S, $0,01 \leq p \text{ value} \leq 0,05$) ou très significatives (TS, $p \text{ value} > 0,01$) sont marquées dans les tableaux.

4. Les analyses de distances

Ces analyses ont été faites pour différents groupes de populations : les populations québécoises et la France au complet, les régions françaises et le Québec au complet, et enfin les populations québécoises et régions françaises ensemble. Nous utilisons ici deux méthodes : les Analyses en Composantes Principales, et les indices de différenciation F_{ST} . Ces deux méthodes seront succinctement décrites.

a) Les ACPs

Les ACPs ont été réalisées à partir des fréquences des haplotypes pour chacun des marqueurs, et en considérant différents groupes de populations. Ainsi le nombre

d'haplotypes et leur distribution dans les populations constituent les variables prises en compte par les composantes principales.

b) Les F_{ST}

Dans une population, la diversité génétique d'un locus est déterminée par l'hétérozygotie. La divergence entre populations résulte en un excès d'homozygotes et en un déficit d'hétérozygotes dans la méta-population (ou population totale) correspondante. Les F_{ST} comparent la diversité génétique moyenne trouvée dans les sous-populations (H_S) par rapport à la diversité génétique moyenne de la méta-population (H_T) :

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Les F_{ST} varient entre zéro et un. Lorsque le flux de gènes est important (beaucoup de migrations) on observe peu de différenciation entre les sous-populations et la F_{ST} tendra vers zéro. Lorsque les sous-populations sont très différenciées, la diversité génétique de la méta-population est bien plus grande que celle des sous-populations et la F_{ST} tendra vers un.

Les F_{ST} comparent les populations deux à deux, à partir des fréquences haplotypiques des marqueurs considérés. Après avoir obtenu les valeurs des F_{ST} à l'aide du logiciel ARLEQUIN 2.0, elles ont été entrées dans STATISTICA 6.0 dans le but d'obtenir une représentation en deux dimensions de ces valeurs les unes par rapport aux autres. La représentation choisie est la MDS (*multidimensional scaling*).

III. RESULTATS

La diversité régionale au Québec a été appréhendée dans trois régions : la région de Montréal, le Saguenay et la péninsule gaspésienne. Compte tenu de l'histoire de la Gaspésie, quatre sous-populations (acadienne, anglo-normande, canadienne-française, loyaliste) ont été distinguées suivant l'ascendance des participants, dans le but d'évaluer l'homogénéité de cette région. La diversité génétique du Québec a été comparée à la diversité génétique de la France. Cinq régions françaises ont été échantillonnées : le Nord de la France, la Bretagne, la Loire Atlantique, le Centre-ouest de la France, le Sud-ouest de la France, et un échantillon contrôle de la diversité française constitué par la France urbaine.

Trois systèmes sur le chromosome X nous ont permis d'étudier la diversité génétique québécoise : un microsatellite composé et interrompu, DXS1238, un marqueur de polymorphismes nucléaires, *dys44*, et la combinaison de ces deux systèmes, formant ce que nous avons appelé des haplotypes composés, ou haplotypes-C. Ces trois systèmes présentent trois niveaux de polymorphismes : *dys44* étant le système le moins polymorphe avec 25 haplotypes identifiés dans les populations de cette étude, DXS1238 procure un niveau de polymorphisme intermédiaire avec 68 haplotypes observés, et les haplotypes-C constituent le système le plus polymorphe avec 142 haplotypes observés dans les populations de cette étude.

Pour DXS1238, 319 chromosomes québécois et 259 chromosomes français ont été séquencés. Pour *dys44*, 277 chromosomes québécois et 262 chromosomes français ont été génotypés. Ces différences d'effectifs par rapport à DXS1238 s'expliquent par 42 chromosomes de moins génotypés au Québec ; tandis que du côté français trois chromosomes de plus ont été génotypés. Pour les haplotypes-C, un total de 276 chromosomes québécois et de 257 chromosomes français ont été analysés.

Pour chacun des trois systèmes, les distributions des fréquences haplotypiques par population sont présentées. A partir du nombre d'haplotypes, des graphiques de partage d'haplotypes ont été réalisés. Les fréquences haplotypiques nous ont également permis d'obtenir des indices de diversité (hétérozygotie, θ hom, θ k) pour chaque population. A partir de ces fréquences haplotypiques par population, différentes analyses ont été réalisées, dont le test d'amalgamation de Chakraborty et les analyses de distances entre populations. Ces analyses de distances sont

représentées par les ACP et les F_{ST} , nous permettant de comparer les images de ces deux analyses pour un même ensemble de populations et des marqueurs différents.

Ces indices et ces analyses ont été réalisées pour chacun des deux marqueurs séparément, et pour les deux marqueurs ensemble, la combinaison des polymorphismes de ces deux marqueurs constituant un haplotype composé (haplotype-C).

Les résultats débiteront avec une présentation du microsatellite DXS1238 et la présentation de la distribution de ses haplogroupes, puis de ses haplotypes dans les populations échantillonnées. Ensuite les résultats des partages d'haplotypes entre les populations seront exposés ; et pour finir les résultats du test de Chakraborty et des analyses de distances seront présentés. Cette organisation de la présentation des résultats sera la même pour les résultats de *dys44* et des haplotypes-C. L'ensemble de ces résultats sera discuté dans la quatrième partie du mémoire.

A. DXS1238

1. Nomenclature

La majorité des microsatellites utilisés ne comportent aucune interruption. Dans la plupart des cas, le statut interrompu du locus n'est pas connu. Les microsatellites peuvent également être composés, c'est à dire faits de répétitions en tandem de différents motifs contiguës. Notre marqueur est un microsatellite composé et interrompu situé sur le chromosome X.

Nous considérons les quatre premiers motifs comme des mutations ponctuelles stabilisant notre système (figure 7). Ces quatre premiers motifs constituent des haplogroupes, chaque haplogroupe représentant une famille de polymorphismes. Si quatre à cinq haplogroupes ont été observés jusqu'à présent en Europe, plus d'une dizaine ont été identifiés dans nos échantillons d'Afrique sub-saharienne. Ce niveau de polymorphisme nous permet de mettre en évidence des différences continentales.

Les deux microsatellites sont $(GT)_X$ et $(GA)_Y$, dont les longueurs fluctuent de part et d'autre de l'interruption ATGT. Cette interruption est notée « 0 » lorsqu'elle est

absente, et « 2 » lorsqu'elle est présente. Elle est également considérée comme une mutation ponctuelle et prise en compte au niveau des haplogroupes. Ceci dit, les seuls haplogroupes montrant l'absence de cette interruption sont spécifiquement africains.



| Haplogroupes québécois | Haplotypes – variations de | |
|------------------------|----------------------------|---------|
| | X | Y |
| a.b.c.d.X.2.Y | X | Y |
| 0.0.0.0.X.2.Y | 9 à 25 | 13 à 22 |
| 4.3.0.0.X.2.Y | 19 | 20 |
| 5.1.0.0.X.2.Y | 8 à 20 | 14 à 22 |
| 5.1.3.1.X.2.Y | 13 à 19 | 13 à 15 |
| 5.2.0.0.X.2.Y | 17 à 20 | 18 à 24 |

Figure 7 Nomenclature de DXS1238.

Au sein de chaque haplogroupe, les deux microsatellites $(GT)_X$ et $(GA)_Y$ voient leurs longueurs varier d'un individu à l'autre. A chaque individu est alors attribué un haplotype, comportant le nombre observé de chaque motif. Un même haplogroupe est ainsi constitué de plusieurs haplotypes. D'un point de vue évolutif, les haplotypes d'un même haplogroupe peuvent être considérés plus proches entre eux que deux haplotypes de différents haplogroupes. Nous comptons sur la distribution des fréquences des haplotypes pour mettre en évidence des différences au sein d'un même continent, voire au sein d'un même pays.

2. Distribution des haplogroupes de DXS1238

Après le génotypage des 578 chromosomes pour le microsatellite DXS1238, les fréquences des haplogroupes ont été calculées dans chacune des populations échantillonnées. Quatre haplogroupes ont été observés en France, et cinq haplogroupes ont été identifiés au Québec. Au niveau de la distribution des haplogroupes, que ce soit de part et d'autre de l'Atlantique, au niveau de différentes régions du Québec, ou bien au sein d'une même région au Québec, peu de

différences entre les populations sont observées. L'haplogroupe le plus fréquent est 5.1.0.0.X.2.Y, c'est aussi l'haplogroupe le plus diversifié, ensuite viennent, par ordre décroissant, les haplogroupes 0.0.0.0.X.2.Y et 5.2.0.0.X.2.Y (Figure 8).

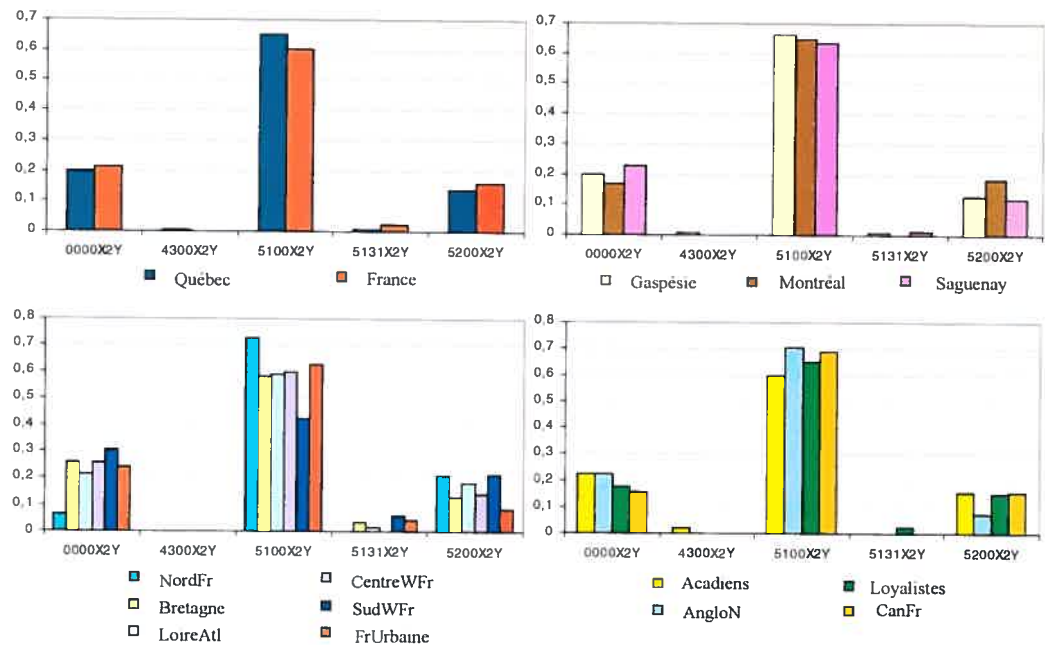


Figure 8 Fréquences relatives des haplogroupes DXS1238 dans les populations échantillonnées.

3. Les distributions des fréquences haplotypiques.

Après le génotypage des 578 chromosomes pour DXS1238, les fréquences des haplotypes pour chacune des populations échantillonnées ont été calculées. Les figures 9 et 10 représentent les fréquences haplotypiques, classées par ordre décroissant par rapport à la France au total. Ces fréquences sont relatives au nombre total de chromosomes de chaque population.

En observant cette distribution classée de façon décroissante, on réalise que les haplotypes majoritaires en France sont les mêmes qu'au Québec, à quelques différences près (Figures 9a et 10a). Quatre haplotypes ont une fréquence de 5% environ au Québec : 5.1.0.0.18.2.18 (contre une fréquence d'une fois et demie plus élevée en France), 5.1.0.0.19.2.15, 5.1.0.0.20.2.18, 0.0.0.0.19.2.22 (fréquences deux

fois plus élevées au Québec qu'en France pour ces 3 haplotypes). Notons que ces haplotypes sont relativement proches d'haplotypes très fréquents. Mais étant donné l'intervalle de temps que nous considérons l'effet des mutations peut être négligé.

Des différences plus importantes se retrouvent pour les haplotypes 0.0.0.0.19.2.20, quatre fois plus fréquent au Québec, et 5.1.0.0.19.2.20, six fois plus fréquent au Québec qu'en France. Le premier haplotype est présent à Montréal et en Gaspésie, mais plus fréquent à Montréal (Figures 9b et 10d). Cet haplotype est également présent dans trois régions françaises : la Loire Atlantique, le Sud-ouest et le Centre-ouest de la France (Figures 10d, 10e et 10f). Le second haplotype est également retrouvé en Gaspésie et à Montréal, à fréquences relativement identiques, mais n'est présent en France que dans deux des trois régions citées précédemment : le Sud-ouest et le Centre-ouest de la France (Figures 14b, 14d, 15e et 15f).

Le Québec présente 21 haplotypes spécifiques, contre 20 en France (Figures 9a et 10a). Parmi les haplotypes retrouvés au Québec uniquement, les fréquences de 5.1.0.0.19.2.16 (4%), 5.1.0.0.14.2.18 (1%) et 5.1.0.0.19.2.19 (1%) sont faibles mais pas négligeables. Étant donné que nous négligeons l'effet des mutations, ces haplotypes peuvent être perçus comme des haplotypes fondateurs. Lorsqu'on observe la distribution de ce premier haplotype, 5.1.0.0.19.2.16, au Québec, on s'aperçoit qu'il est présent en Gaspésie et au Saguenay à des fréquences comparables (Figures 9c et 9d). Cependant en Gaspésie, cet haplotype n'est observé que chez les Acadiens et les Anglo-normands (Figures 9e et 9f).

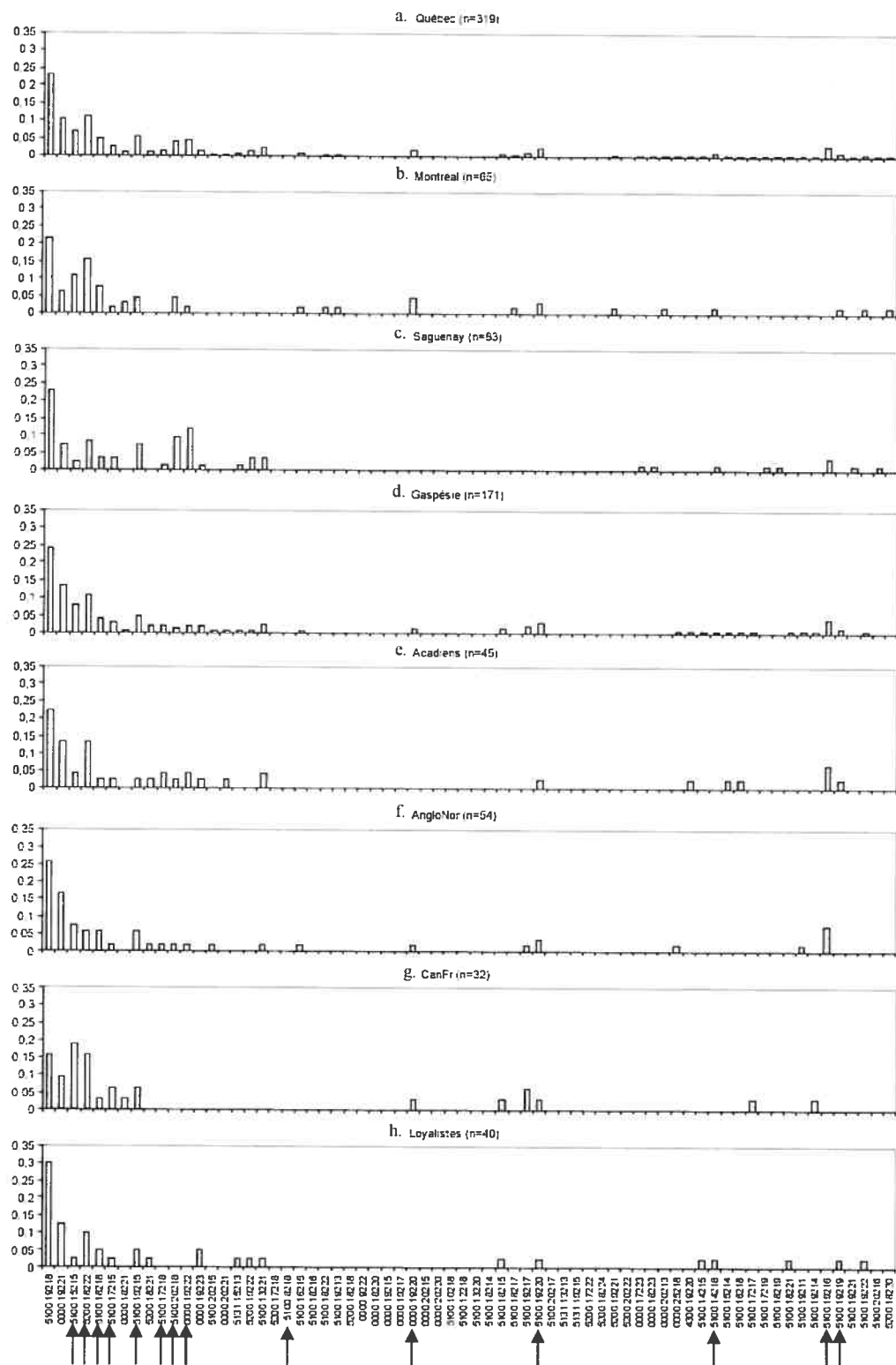


Figure 9 Distribution des fréquences des Haplotypes de DXS1238 suivant les populations québécoises prises en compte. Les haplotypes sont classés par ordre décroissant pas rapport à la population française au complet (France). Les flèches correspondent aux haplotypes mentionnés dans le texte.

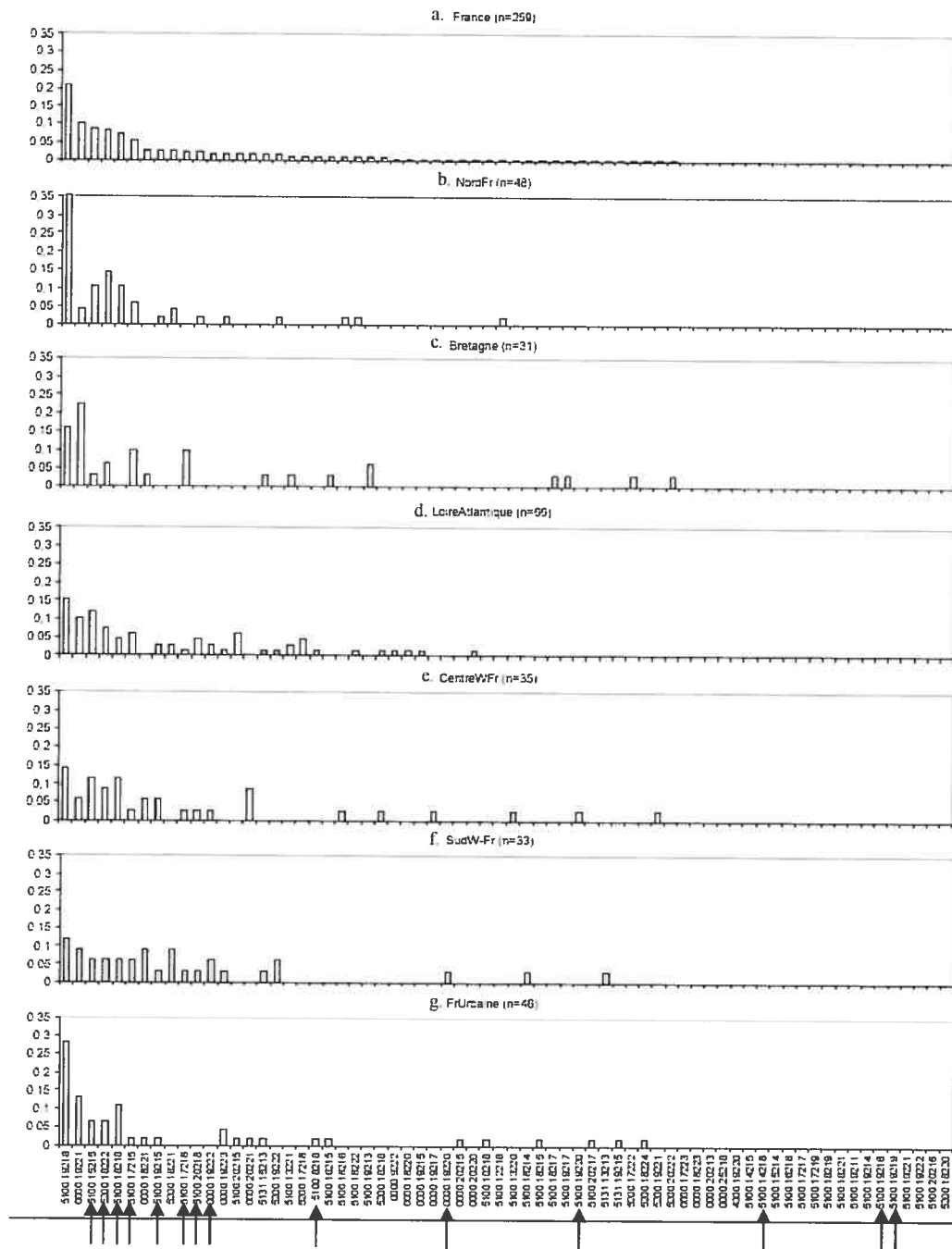


Figure 10 Distribution des fréquences des Haplotypes de DXS1238 suivant les populations françaises prises en compte. Les haplotypes sont classés par ordre décroissant pas rapport à la population française au complet (France). Les flèches correspondent aux haplotypes mentionnés dans le texte.

De même, observons la distribution des fréquences dans les sous-populations gaspésiennes (Figures 9e, 9f, 9g et 9h). D'une façon générale, l'haplotype le plus fréquent est 5.1.0.0.19.2.18 (Figures 9a et 10a). Chez les Canadiens français de Gaspésie, cet haplotype est aussi fréquent que 5.2.0.0.18.2.22, et tous deux sont relégués à la deuxième place après 5.1.0.0.15.2.15 (Figures 9g). La très haute fréquence de ce dernier haplotype dans cette sous-population gaspésienne est responsable de la similitude entre la Gaspésie et Montréal (Figures 9b et 9d). Une autre singularité des Canadiens français de Gaspésie est la fréquence de l'haplotype 5.1.0.0.17.2.15, supérieure à 5%, alors qu'elle est de 1% environ dans les autres sous-populations gaspésiennes (Figures 9e, 9f, 9g et 9h).

En ce qui concerne les subdivisions régionales de la France, la première distinction apparaît pour le Nord de la France, avec des fréquences bien plus élevées des haplotypes 5.1.0.0.19.2.18 et 5.2.0.0.18.2.22 par rapport aux autres régions (Figures 10). La Bretagne se distingue aussi par la très haute fréquence de l'haplotype 0.0.0.0.19.2.21, et de 5.1.0.0.17.2.18. Soulignons également la basse fréquence de 5.1.0.0.15.2.15 et l'absence de 5.1.0.0.18.2.18, deux haplotypes très fréquents dans les autres populations. Ajoutons également le grand nombre d'haplotypes uniques en Bretagne, dont 5.1.0.0.8.2.18 présent à hauteur de 6% et observé nulle part ailleurs.

4. Haplotypes partagés

Les graphiques de partage d'haplotypes ont été réalisés dans le but d'observer la singularité et les liens entre populations. Les graphiques de partage d'haplotypes sont effectués à partir des fréquences des haplotypes DXS1238 suivant les groupes de populations considérés (Figure 11). Pour chaque population, trois catégories d'haplotypes ont été distinguées : des haplotypes uniques, des haplotypes partagés avec une autre population et des haplotypes commun à trois populations ou plus. La proportion des trois catégories est relative au nombre total d'haplotypes de la population considérée, pour DXS1238.

Lorsque le partage d'haplotypes de DXS1238 est considéré pour les populations du Québec, le Saguenay et Montréal ressortent par leurs pourcentages d'haplotypes uniques, les plus élevés des populations québécoises (Figure 11a). En effet si ce

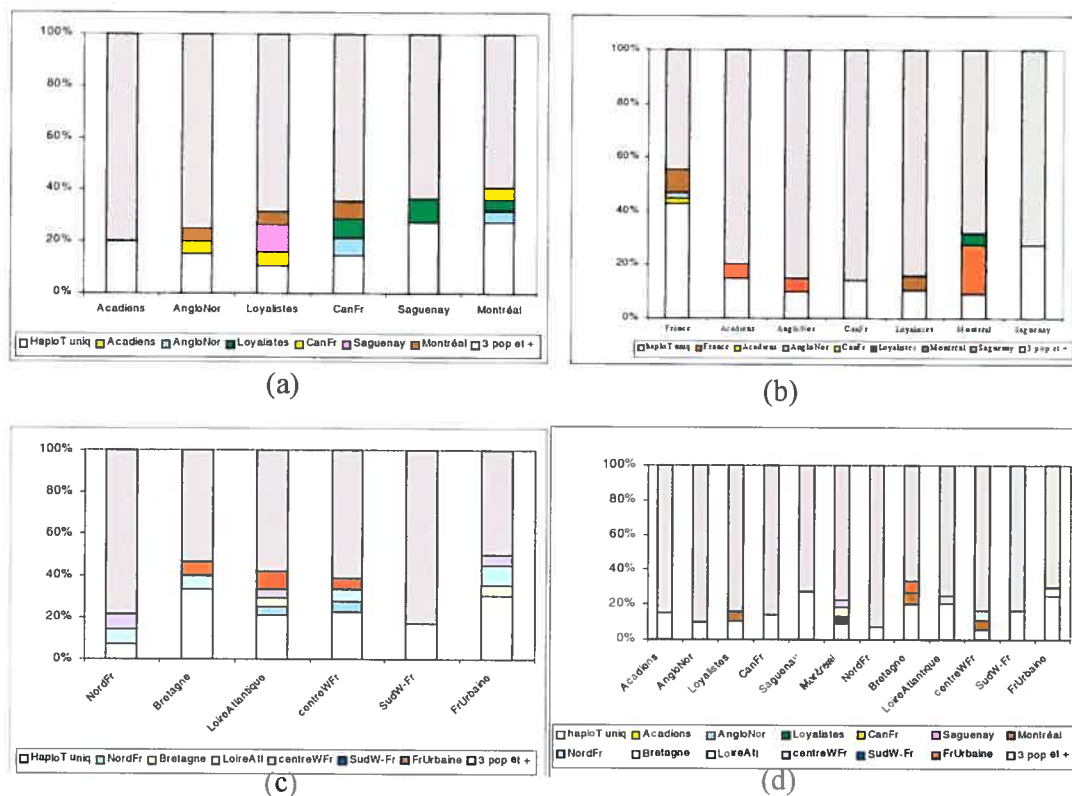


Figure 11 Graphiques de partages d'haplotypes DXS1238 entre différents groupes de population : (a) entre populations québécoises ; (b) entre populations québécoises et la France au complet ; (c) entre régions françaises ; (d) entre régions françaises et populations québécoises.

pourcentage atteint les 28% pour ces deux populations, les Acadiens viennent ensuite avec un pourcentage d'haplotypes uniques de 20% ; tandis que pour les Anglo-normands et les Canadiens français ce pourcentage est de 15% environ, et la dernière position est occupée par les Loyalistes avec seulement 10% d'haplotypes uniques.

Les Acadiens sont les seuls ayant soit des haplotypes uniques, soit des haplotypes partagés par trois populations et plus. Les Anglo-normands partagent un haplotype avec les Canadiens français et un autre haplotype avec Montréal. Un partage d'haplotype entre le Saguenay et les Loyalistes a été mis en évidence.

La figure 11c présente le partage d'haplotypes entre les régions françaises échantillonnées. Parmi nos régions françaises, la Bretagne est la plus singulière avec 32% d'haplotypes spécifiques, dont un, 5.1.0.0.8.2.18, représenté deux fois. Notre échantillon de la France urbaine a un nombre d'haplotypes spécifiques élevé avec 30%. Elle partage également 18% d'haplotypes avec une autre région française. La région montrant un taux de partage important est la Loire Atlantique avec 21% d'haplotypes partagés avec une autre région française, mettant en évidence la plaque tournante que peut constituer cette région. Le Nord de la France se remarque avec un seul haplotype spécifique, tandis que le Sud-ouest de la France ressort comme la population ayant soit des haplotypes spécifiques (18%), soit des haplotypes partagés entre trois régions et plus, dénotant un mélange de populations important.

La figure 11b représente le partage d'haplotypes DXS1238 entre les populations québécoises et la France au complet. Ce graphique montre l'importance de la contribution du génome français au Québec. En ajoutant la France au total dans le partage d'haplotypes, le nombre d'haplotypes spécifiques de Montréal diminue beaucoup : de 28% entre populations québécoises, il passe à 9%. Le Saguenay et les Canadiens-français sont les deux populations à présenter les deux seules proportions « haplotypes uniques » et « haplotypes partagés par trois populations et plus » ; le Saguenay conservant la singularité du plus grand nombre d'haplotypes spécifiques. Dans ce graphique un partage d'haplotype entre Français et Anglo-normands pourrait illustrer le lien entre population française et habitants des Iles anglo-normandes.

La figure 11d présente le partage d'haplotypes DXS1238 entre populations québécoises et régions françaises. Cette figure met en lumière les liens étroits entre

les populations québécoises et les régions françaises. Ceci est révélé d'une part par la proportion d'haplotypes partagés par deux populations qui diminue (Loyalistes, Loire Atlantique) , ou disparaît (Anglo-normands, Canadiens français, Saguenay, Nord de la France) ; et d'autre part par la diminution de la proportion des haplotypes uniques dans certaines populations (Acadiens, Montréal, Bretagne, Centre-ouest de la France, France urbaine). Ces remarques sont relatives aux figures 11a et 11c qui prennent en compte les populations québécoises seules, ou françaises seules.

5. Diversité génétique

a) Indices de diversité

A partir des fréquences des haplotypes DXS1238 de chaque population, le nombre d'haplogroupes et le nombre d'haplotypes DXS1238 pour chaque population ont été reportés et différents indices de diversité ont été calculés. L'hétérozygotie donne un indice de la diversité génétique du marqueur dans la population considérée. Le Québec (0,91) et la France (0,92) présentent des hétérozygoties similaires (*cf.* Tableau 2). Malgré la différence d'effectif entre nos échantillons de la France et du Québec, 47 haplotypes ont été observés en France contre 48 au Québec, 60 chromosomes de plus n'apportant qu'un haplotype supplémentaire au Québec.

La région française présentant la diversité génétique la plus faible est le Nord de la France (0,84) . En ce qui concerne le Québec, la diversité génétique la plus basse est retrouvée chez les Loyalistes (0,89) de Gaspésie. Ces valeurs les plus faibles ne peuvent pas être corrélées à un faible effectif puisque ces deux sous-populations ont des effectifs de 48 et 40 chromosomes respectivement, et que les populations françaises et québécoises ayant l'effectif le plus bas (31 chromosomes pour la Bretagne et 32 chromosomes pour les Canadiens français de Gaspésie) présentent des diversités génétiques plus élevées (0,92).

| populations | # chr | # haploG | # haploT | hétéroZ | θ_{Hom} | θ_k | Chakraborty | |
|-------------|-------|----------|----------|---------|----------------|------------|--------------|------|
| | | | | | | | # haploT att | Test |
| France | 259 | 4 | 47 | 0,92 | 10,04 | 16,53 | 33,52 | S |
| Bretagne | 31 | 4 | 15 | 0,92 | 9,53 | 10,8 | 14,18 | |
| CentreWFr | 35 | 3 | 18 | 0,95 | 16,59 | 14,18 | 19,16 | |
| Loire Atl | 66 | 4 | 24 | 0,94 | 14,08 | 13,12 | 24,89 | |
| NordFr | 48 | 3 | 14 | 0,84 | 4,2 | 6,28 | 11,09 | |
| SudWFr | 33 | 4 | 18 | 0,96 | 22,45 | 15,44 | 20,59 | |
| Fr urbaine | 46 | 4 | 20 | 0,89 | 7,15 | 12,91 | 14,79 | S |
| Québec | 319 | 5 | 48 | 0,91 | 8,6 | 15,45 | 31,8 | TS |
| Gaspésie | 171 | 5 | 36 | 0,90 | 7,93 | 13,63 | 25,21 | TS |
| Acadiens | 45 | 4 | 20 | 0,92 | 9,6 | 13,23 | 17,17 | |
| AngloN | 54 | 3 | 20 | 0,90 | 7,33 | 11,03 | 16,03 | |
| CanFr | 32 | 3 | 14 | 0,92 | 9,7 | 8,93 | 14,53 | |
| Loyalistes | 40 | 4 | 19 | 0,89 | 6,93 | 13,54 | 13,7 | S |
| Saguenay | 83 | 4 | 22 | 0,91 | 8,59 | 9,44 | 20,8 | |
| Montréal | 65 | 3 | 22 | 0,91 | 8,94 | 11,28 | 19,34 | |

Tableau 2 Indices de diversité du microsatellite DXS1238 dans les populations échantillonnées. Pour le test de Chakraborty, seuls les résultats significatifs (S, $0,01 \leq p \text{ value} \leq 0,05$) ou très significatifs (TS, $p \text{ value} > 0,01$) sont écrits.

Les paramètres thêta (θ) présentent également des valeurs relativement identiques entre la France et le Québec. Il est cependant intéressant de comparer les valeurs des θ_{Hom} et θ_k au sein d'une même population. Le θ_k mesure la diversité basée sur le nombre d'haplotypes, tandis que le θ_{Hom} mesure la diversité d'après la distribution des fréquences des haplotypes. Ainsi chez les Acadiens, les Anglo-normands et les Loyalistes de Gaspésie les valeurs des θ_k sont supérieures à celles des θ_{Hom} . Ceci signifie la présence de nombreux haplotypes dans ces populations, car les θ_k sont assez élevés, mais certains d'entre eux seulement sont nettement prédominants par rapport aux autres. Ceci est également constaté pour le Nord de la France et la France urbaine. La prépondérance de la fréquence sur le nombre est révélée lorsque le θ_{Hom} est supérieur au θ_k , comme c'est le cas pour le Sud-ouest de la France. Lorsque les deux θ ont des valeurs relativement proches (différence de une unité entre les deux valeurs), ceci sous-tend que les populations sont à l'équilibre. C'est le cas des Canadiens français de Gaspésie, du Saguenay et de Montréal.

b) Le test de Chakraborty

Ce test a été réalisé à partir des fréquences des haplotypes de DXS1238 pour chaque population. Pour ce marqueur, le test de Chakraborty s'est révélé significatif pour cinq populations : la France, la France urbaine, le Québec, la Gaspésie au complet et les Loyalistes. Le test de Chakraborty s'est révélé beaucoup plus significatif pour la population québécoise au complet et la Gaspésie au complet, mettant en évidence une hétérogénéité au niveau des pools géniques considérés à ces niveaux.

Le test de Chakraborty significatif pour la Gaspésie a mis en valeur le fait que les quatre groupes gaspésiens ne pouvaient être considérés comme un ensemble homogène. L'objectif n'était pas ici d'être représentatif de la Gaspésie, mais plutôt d'étudier le pool génique des trois régions échantillonnées jusqu'à maintenant. Ainsi nous nous sommes interrogés sur l'homogénéité d'une population canadienne-française idéale, construite à partir de nos échantillons de Canadiens français, soit les Montréalais, les Saguenayens, les Canadiens français de Gaspésie et les Acadiens de Gaspésie.

Afin d'observer quels pouvaient être les échantillons de Canadiens français qui participaient à l'hétérogénéité détectée dans l'échantillon du Québec au total, un autre ensemble de tests a été réalisé. Cet ensemble de tests consistait à réaliser un test de Chakraborty pour différentes combinaisons des échantillons de canadiens-français, précisés ci-dessus. Lorsque ces quatre populations ont été considérées ensemble, le test s'est révélé significatif ; et à chaque fois que l'une de ces sous-populations était retirée, le test de Chakraborty était encore significatif.

6. Les analyses de distance

a) Les ACPs

Les ACPs ont été réalisées à partir des fréquences des haplotypes DXS1238 pour différents groupes de populations. Ces analyses permettent d'observer des liens entre populations d'après la distribution des fréquences des haplotypes.

La première ACP a été réalisée en considérant la France au complet et l'ensemble des populations québécoises. La France se retrouve au centre de cette analyse, signifiant que d'après la distribution des fréquences haplotypiques elle est aussi distante des six populations québécoises (Figure 12). Montréal et les Acadiens de Gaspésie se retrouvent au centre de l'analyse également. On remarque alors une opposition au sein des populations gaspésiennes avec les Canadiens français d'une part et les Acadiens, Anglo-normands et Loyalistes d'autre part. Ces deux dernières sous-populations semblent très proches au niveau de leurs fréquences alléliques. Quant au Saguenay, c'est la population la plus éloignée des autres.

La figure 14 représente l'ACP obtenue d'après les fréquences des haplotypes DXS1238 de toutes les sous-populations québécoises et des régions françaises. Sur cette analyse, la Bretagne et le Nord de la France s'opposent, tandis que les Canadiens français sont assimilés au groupe formé par le Centre-ouest de la France, la Loire Atlantique et le Sud-ouest de la France. Les autres sous-populations gaspésiennes se trouvent au centre de l'analyse avec la France urbaine et le Saguenay. Le fait que le Saguenay soit associé à ce dernier groupe est à relativiser avec cette représentation graphique en deux dimensions d'un ensemble de variables.

Une dernière ACP a été réalisée avec les fréquences des haplotypes DXS1238 des régions françaises et du Québec au complet (figure 16). Il apparaît que la Bretagne et le Nord de la France se distinguent non seulement l'une de l'autre, mais aussi des autres régions. Il semble également que le Centre-ouest, le Sud-ouest et la Loire Atlantique forment un groupe à part.

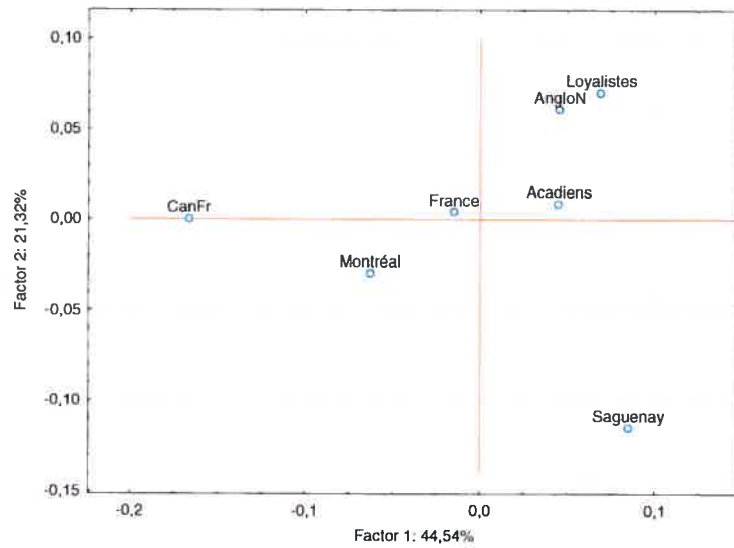


Figure 12 Analyse en composantes principales réalisée à partir des fréquences haplotypiques DXS1238 de la France et des populations québécoises.

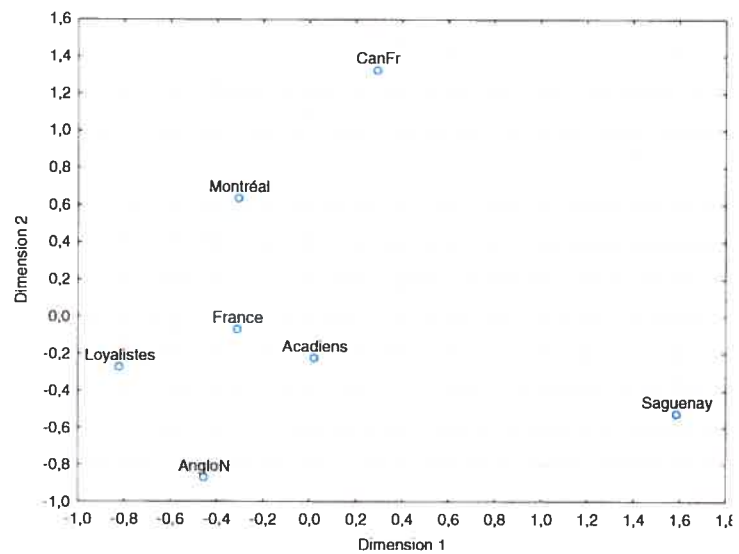


Figure 13 Représentation MDS des F_{ST} réalisée à partir des fréquences haplotypiques DXS1238 de la France et des populations québécoises.

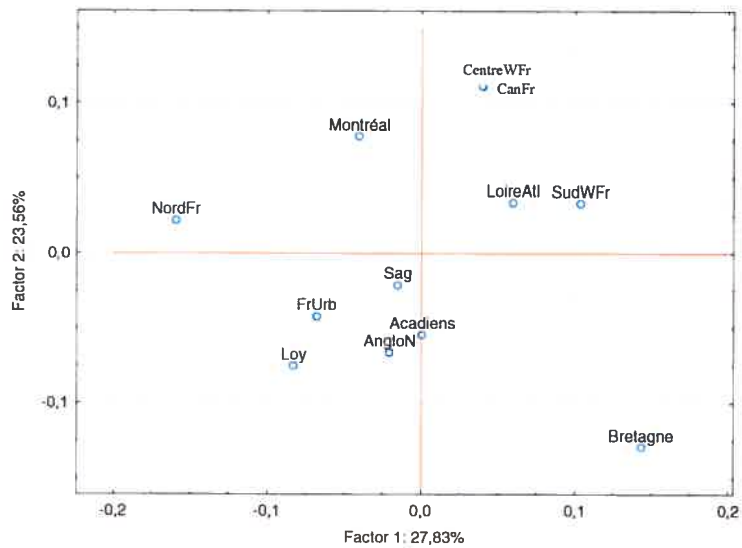


Figure 14 Analyse en composantes principales réalisée à partir des fréquences haplotypiques DXS1238 des régions françaises et des populations québécoises.

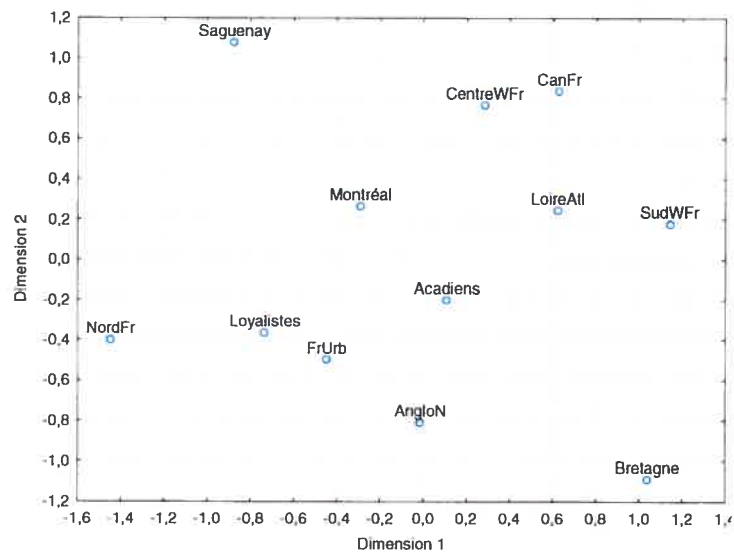


Figure 15 Représentation MDS des F_{ST} réalisée à partir des fréquences haplotypiques DXS1238 des régions françaises et des populations québécoises.

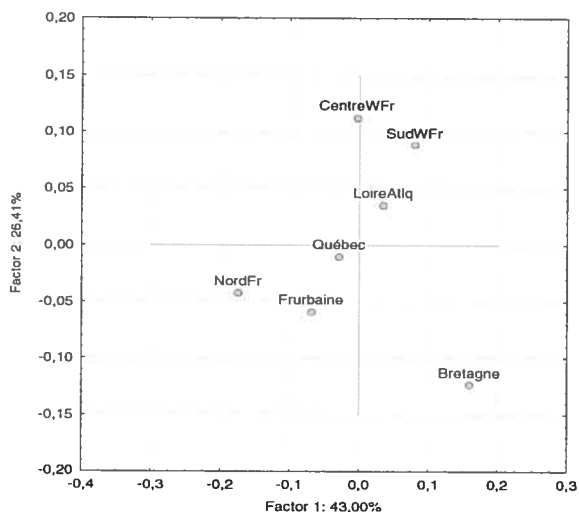


Figure 16 Analyse en composantes principales réalisée à partir des fréquences haplotypiques DXS1238 des régions françaises et du Québec.

b) Les F_{ST}

Les valeurs des F_{ST} n'ont pas été présentées. Nous avons préféré représenter ces valeurs graphiquement à l'aide de la fonction MDS. Ces graphiques représentant les F_{ST} nous permettent ainsi de comparer les distances entre population obtenues de deux façons différentes. Les F_{ST} peuvent confirmer ou infirmer l'image des liens entre populations obtenue avec les ACPs (Figures 13 et 15). Les F_{ST} procurent en plus une robustesse statistique au niveau des différences entre populations, apportant une calibration des différences visualisées en deux dimensions sur la représentation MDS ou ACP. D'une façon générale en ce qui concerne nos données, l'image obtenue entre ACPs et F_{ST} est relativement similaire. C'est pourquoi la représentation MDS des F_{ST} entre régions françaises et le Québec n'est pas présentée.

Les valeurs des F_{ST} sont généralement faibles (comprises entre 0,1% et 2,4% pour DXS1238) et peu d'entre elles sont significatives. Les différences significatives sont observées entre la France et le Saguenay, et le Saguenay et les Canadiens français de Gaspésie (Figure 13). Cette analyse, comme l'ACP, distingue les Canadiens français des autres sous-populations gaspésiennes. En effet, les F_{ST} entre sous-populations gaspésiennes sont différentes de zéro uniquement pour les comparaisons avec les Canadiens français.

La figure 15 correspond à la représentation MDS des valeurs des F_{ST} entre régions françaises et populations québécoises. D'après les F_{ST} statistiquement significatives, quatre pôles populationnels peuvent être identifiés dans cette analyse : le Saguenay, le Nord de la France, la Bretagne et le Sud-ouest de la France. Précisons toutefois que la seule valeur de F_{ST} significative entre le Sud-ouest de la France et ces trois autres sous-populations concerne la paire {Sud-ouest de la France et Nord de la France}.

Au centre de cette analyse se retrouvent les cinq dernières sous-populations. Cette position paraît normale pour la France urbaine et Montréal, puisque toutes deux sont « nos groupes témoins ». Le regroupement des trois autres populations va de paire avec les résultats de l'ACP correspondante. De plus nous savons que les Anglo-normands se sont beaucoup mélangés à d'autres sous-populations gaspésiennes. Quant à la position centrale occupée par les Loyalistes et les Acadiens, elle peut être comprise comme une absence de différence entre ces deux populations et les autres populations.

B. Les haplotypes de *dys44*

Après avoir constitué les cohortes pour DXS1238, les individus de ces mêmes cohortes ont été utilisés pour observer la distribution des polymorphismes de *dys44* dans les populations étudiées.

Pour *dys44*, l'échantillon français est constitué de 262 chromosomes, tandis que l'échantillon québécois est constitué de 277 chromosomes. En France, 18 haplotypes ont été identifiés et 21 haplotypes au Québec. La distribution des fréquences des haplotypes *dys44* a été effectuée pour chaque population. Ces premiers résultats seront présentés, puis viendront ensuite les résultats du partage d'haplotypes. Les résultats des indices de diversité et des analyses de distance seront présentés par la suite.

1. Distribution des haplotypes

Les fréquences des haplotypes sont classées par ordre décroissant par rapport à la France au complet (Figure 17). Lorsque les populations totales sont considérées, les quatre haplotypes les plus fréquents en France (B001, B003, B006 et B005) sont les mêmes qu'au Québec. Pour ce qui est des autres haplotypes *dys44*, les distributions des fréquences sont assez différentes entre la France et le Québec. Plus de singletons sont observés au Québec (sept) par rapport à la France (trois). Ces singletons correspondent à des haplotypes spécifiques. Ces haplotypes spécifiques sont B024, B030 et B042 en France ; et B009, B039, B073, B185, B186, B187 et B188 au Québec. B033 est un haplotype spécifique français, partagé par trois régions (Centre-ouest, Sud-ouest, France urbaine), et absent du Québec.

Le Saguenay est la population qui a le plus grand nombre d'haplotypes (15), mais avec seulement deux d'entre eux (B001 et B003) extrêmement fréquents (34% et 25%). Les haplotypes B186, B187 et B188 sont spécifiques au Saguenay. De même l'haplotype B067 n'est retrouvé qu'au Saguenay au Québec, mais a été également observé dans le Nord de la France. L'haplotype B013 a seulement été identifié au Saguenay et chez les Loyalistes. La fréquence de cet haplotype *dys44* est deux fois

plus élevée au Saguenay (5%) que chez les Loyalistes. Du côté français, cet haplotype est présent en Bretagne, dans le Nord de la France et en Loire Atlantique.

Chez les Loyalistes, trois haplotypes (B001, B003 et B006) sont très fréquents (32%, 20% et 18%). Pareillement aux Loyalistes, les Canadiens français présentent des fréquences des trois premiers haplotypes (B001, B003 et B006) très élevées (26%, 31% et 18%). Deux remarques peuvent être ajoutées illustrant une certaine singularité de cette population gaspésienne. Tout d'abord du fait que la fréquence de l'haplotype B005 (13%) soit très proche de celle de l'haplotype B006, pourrait nous faire considérer cet haplotype comme un haplotype prépondérant supplémentaire. Ensuite l'haplotype habituellement plus fréquent dans toutes les autres populations est B001, alors que chez les Canadiens français de Gaspésie c'est B003 qui est le plus fréquent.

On peut également ajouter que tous les haplotypes des Canadiens français de Gaspésie sont partagés avec les autres populations gaspésiennes, et même québécoises, de notre étude. Au niveau des Acadiens, un seul haplotype apparaît comme extrêmement fréquent (46%), l'haplotype B001. La distribution des haplotypes des Anglo-normands renvoie la même image que celle du Saguenay, avec deux haplotypes (B001 et B003) très fréquents (35% et 29%), par contre le nombre d'haplotypes des Anglo-normands (onze) est plus petit qu'au Saguenay.

La distribution des fréquences des haplotypes *dys44* de Montréal ressemble beaucoup à celle des Canadiens français de Gaspésie, avec trois haplotypes (B001, B003 et B006) très fréquents (22%, 22% et 25%), dont les fréquences sont relativement proches. Le fait que B006 soit l'haplotype le plus fréquent à Montréal (au lieu que ce soit B001) constitue la particularité de cette population. Le seul haplotype propre à Montréal est B009, tandis que B039 est partagé par Montréal et le Saguenay.

Si l'haplotype B008 est présent à Montréal, au Saguenay et en Gaspésie, il n'est cependant pas ubiquitaire au niveau de la Gaspésie, puisqu'il est absent chez les Acadiens. Précisons que cet haplotype est répandu dans toutes les régions françaises étudiées, avec une fréquence proche de 10%, sauf en Bretagne et en Loire Atlantique où cet haplotype est deux fois moins fréquent.

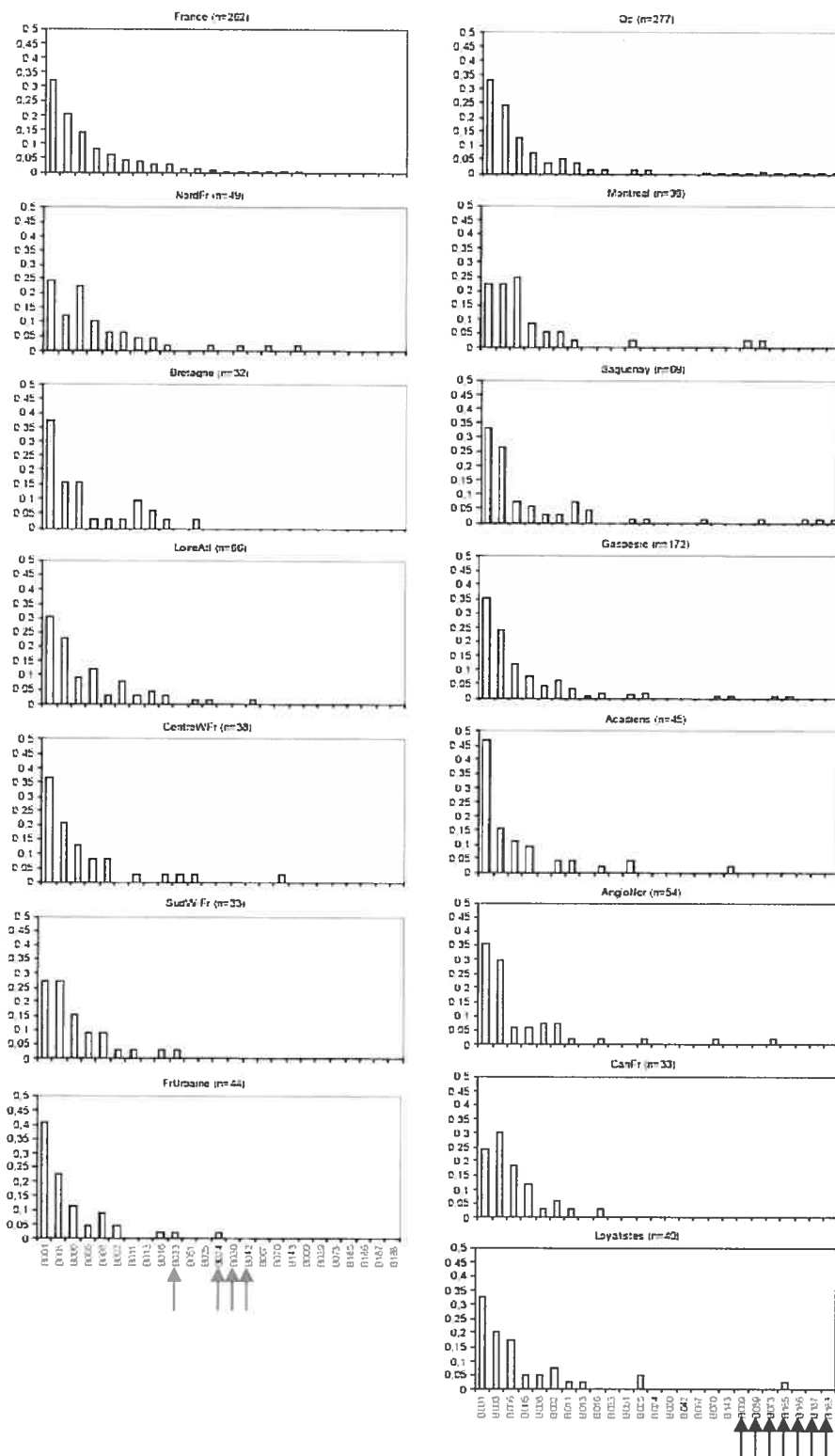


Figure 17 Distribution des fréquences des Haplotypes dys44 suivant les populations considérées. Les haplotypes sont classés par ordre décroissant par rapport à la population française au complet (France). Les flèches correspondent aux haplotypes spécifiques en France (à gauche) et spécifiques au Québec (à droite).

B016, B070, B143, B073 et B185 sont des haplotypes spécifiques à la Gaspésie au Québec. En ce qui concerne B016, il a été observé dans toutes les sous-populations gaspésiennes, sauf chez les Loyalistes. Même si sa fréquence est relativement basse, cet haplotype est ubiquitaire en France. L'haplotype B070 est uniquement observé chez les Anglo-normands et dans le Centre-ouest de la France. L'haplotype B143 est propre aux Acadiens, et a également été identifié dans le Nord de la France. L'haplotype B073 est spécifique aux Anglo-normands, tandis que l'haplotype B185 a été uniquement observé chez les Loyalistes.

Parmi les haplotypes communs à la France et au Québec, citons B051, présent uniquement en Bretagne, dans le Centre-ouest et en Loire Atlantique du côté français, tandis que cet haplotype est présent dans toutes les régions échantillonnées au Québec, soit à Montréal, au Saguenay et en Gaspésie. Cependant il a seulement été observé chez les Acadiens en Gaspésie. Citons également l'haplotype B025, retrouvé au Nord de la France et en Loire Atlantique, et identifié uniquement au Saguenay et chez les Loyalistes (où la fréquence de cet haplotype est cinq fois plus élevée) et les Anglo-normands en Gaspésie.

2. Partage d'haplotypes

Les graphiques de partage d'haplotypes ont été obtenus à partir du nombre d'haplotypes dans les populations considérées. Pour chaque population, trois catégories d'haplotypes ont été distinguées : des haplotypes uniques, des haplotypes partagés avec une autre population et des haplotypes communs à trois populations ou plus. Dans une population, la proportion de chacune de ces catégories est relative au nombre total d'haplotypes *dys44* de cette population (Figure 18).

Dès le premier aperçu de ces graphiques de partage obtenus pour les haplotypes *dys44*, on remarque que la proportion des haplotypes communs (trois populations et plus) est plus importantes que pour les haplotypes de DXS1238. Ceci est dû à la nature du marqueur *dys44* qui est moins polymorphe que DXS1238. De plus ces graphiques de partage ne prennent pas en compte la fréquence des haplotypes dans une population, mais seulement les haplotypes eux-mêmes (soit la présence ou l'absence).

La figure 18a représente le partage d'haplotypes *dys44* entre les populations québécoises. La population présentant le plus grand nombre d'haplotypes uniques est le Saguenay. Un partage d'haplotype entre Loyalistes et le Saguenay est de nouveau révélé ici. Les Acadiens et les Anglo-normands présentent seulement les catégories d'haplotypes uniques et d'haplotypes partagés par trois populations et plus, tandis que les Canadiens français de Gaspésie présentent seulement cette dernière catégorie.

Le partage d'haplotypes *dys44* entre la France au complet et les populations québécoises est illustré dans la figure 18b. La France est la population qui a le plus grand nombre d'haplotypes spécifiques (21%) et le plus grand nombre d'haplotypes partagés avec une autre population (18%). L'haplotype unique des Acadiens est devenu un haplotype partagé avec la population française (par rapport à la figure 18a). Ceci est également le cas pour un des deux haplotypes uniques aux Anglo-normands (par rapport à la figure 18a). Ce graphique montre l'importance de la contribution du génome français au pool génique québécois.

La figure 18c présente le partage d'haplotypes *dys44* entre les régions françaises. La Bretagne et le Sud-ouest de la France possèdent seulement la catégorie des haplotypes partagés par trois populations ou plus. Les deux seules populations ayant la catégorie des haplotypes partagés avec une autre population sont la Loire Atlantique et le Nord de la France. Cette dernière population a la proportion d'haplotypes uniques la plus élevée. A partir de ce graphique, nous pouvons nous rendre compte de l'homogénéité de nos régions françaises au niveau des haplotypes *dys44*.

Dans la figure 18d, le partage d'haplotypes *dys44* entre les régions françaises et les populations québécoises peut être obser. La singularité du Saguenay apparaît ici avec la proportion très élevée d'haplotypes uniques (20%) par rapport aux autres populations prises en compte dans ce graphique. En effet la population du Nord de la France voit sa proportion d'haplotypes uniques passer de 25% à 5%.

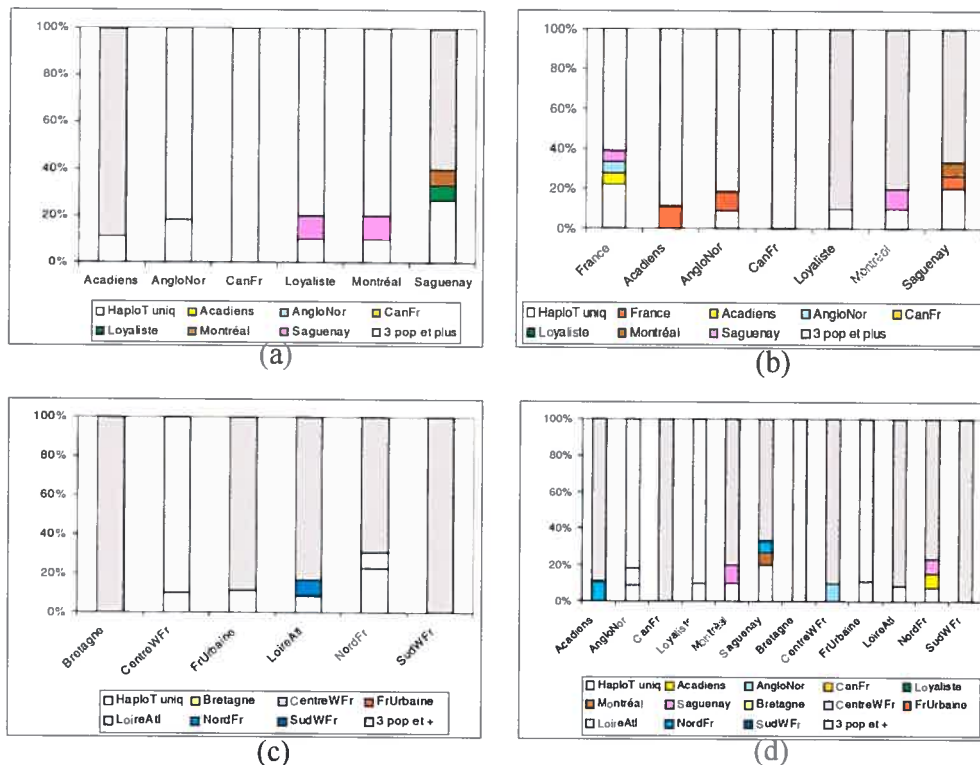


Figure 18 Graphiques de partages d'haplotypes dys44 entre différents groupes de population : (a) entre populations québécoises ; (b) entre populations québécoises et la France au complet ; (c) entre régions françaises ; (d) entre régions françaises et populations québécoises.

3. Diversité génétique

Les indices de diversité génétique ont été obtenus à partir des fréquences des haplotypes *dys44* de chaque population. Pour ce marqueur nous disposons de 262 chromosomes en France, où 18 haplotypes ont été identifiés (Tableau 3). En ce qui concerne le Québec, à partir des données de 277 chromosomes, 21 haplotypes ont été identifiés. L'hétérozygotie est équivalente en France et au Québec.

| populations | # chr | # haploT | hétéroZ | θ_{Hom} | θ_k | Chakraborty | |
|-------------|-------|----------|---------|----------------|------------|--------------|------|
| | | | | | | # haploT att | Test |
| France | 262 | 18 | 0,82 | 3,68 | 4,21 | 16,27 | |
| Bretagne | 32 | 10 | 0,81 | 3,61 | 4,6 | 8,73 | |
| CentreWFr | 38 | 10 | 0,8 | 3,34 | 4,08 | 8,9 | |
| Loire Atl | 66 | 12 | 0,83 | 4,09 | 4,03 | 12,12 | |
| NordFr | 49 | 13 | 0,86 | 5,51 | 5,44 | 13,09 | |
| SudWFr | 33 | 9 | 0,83 | 4,04 | 3,72 | 9,42 | |
| Fr urbaine | 44 | 9 | 0,77 | 2,64 | 3,14 | 8,09 | |
| Québec | 277 | 21 | 0,8 | 3,3 | 5,1 | 15,18 | S |
| Gaspésie | 172 | 15 | 0,79 | 3,04 | 3,77 | 12,85 | |
| Acadiens | 45 | 9 | 0,74 | 2,29 | 3,1 | 7,45 | |
| AngloN | 54 | 11 | 0,78 | 2,85 | 3,89 | 9,04 | |
| CanFr | 33 | 8 | 0,82 | 3,65 | 3,03 | 8,89 | |
| Loyalistes | 40 | 10 | 0,82 | 3,92 | 9,95 | 9,95 | |
| Saguenay | 69 | 15 | 0,81 | 3,48 | 5,61 | 11,08 | |
| Montréal | 36 | 10 | 0,84 | 4,48 | 4,23 | 10,32 | |

*Tableau 3 Indices de diversité de *dys44* dans les populations échantillonnées. Pour le test de Chakraborty, seuls les résultats significatifs (S, $0,01 \leq p \text{ value} \leq 0,05$) ou très significatifs (TS, $p \text{ value} > 0,01$) sont écrits.*

L'hétérozygotie la plus basse (0,77) est observée pour la France urbaine, ce qui est surprenant étant donné l'hétérogénéité supposée de cette population. Quant au Québec, cet indice a la valeur la plus basse (0,74) pour les Acadiens.

Au sein même du Québec, on peut noter une singularité observée chez les Loyalistes, et au Saguenay : le θ_{Hom} est inférieur au θ_k , donnant un indice sur le nombre d'haplotypes et leur distribution dans ces populations. En y regardant de plus près, on voit en effet que le nombre d'haplotypes observé dans ces populations est relativement important, mais seulement deux (ou trois dans le cas des Loyalistes) d'entre eux présentent des fréquences jusqu'à six fois plus élevées que les autres haplotypes, expliquant l'observation d'une hétérozygotie plus basse dans ces

populations par rapport aux autres populations. Au niveau des autres populations, ces deux indices sont très proches, dénotant des populations proches de l'équilibre concernant ce marqueur.

En observant conjointement le tableau des indices de diversité et la distribution des haplotypes, une première singularité saguenayenne saute aux yeux (Tableau 3 et Figure 17). En effet, le Saguenay est la population qui a le plus grand nombre d'haplotypes (15), mais avec seulement deux d'entre eux (B001 et B003) extrêmement fréquent, expliquant la différence entre les deux indices θ . La différence entre ces deux indices est la plus importante chez les Loyalistes. Ceci s'explique par le fait que même si trois haplotypes (B001, B003 et B006) sont très fréquents (32%, 20% et 18%), l'hétérozygotie élevée qui pourrait en résulter est contrebalancée par un nombre d'haplotypes relativement bas (dix).

Ce sont les Canadiens français de Gaspésie qui présentent le nombre d'haplotypes le plus bas (huit). La faible hétérozygotie observée est due au petit nombre d'haplotypes identifié dans cette population. Les Acadiens, avec seulement neuf haplotypes au total, présentent l'hétérozygotie la plus basse (0,74). La deuxième hétérozygotie la plus basse (0,78) apparaît pour les Anglo-normands.

4. Les analyses de distance

Les ACPs ont été réalisées à partir des fréquences des haplotypes *dys44* en considérant différents ensembles de populations, dans le but d'observer la disposition des ces populations les unes par rapport aux autres suivant la variation des groupes constitués.

L'ACP réalisée à partir des fréquences haplotypiques des populations québécoises nous permet d'identifier trois pôles populationnels qui peuvent s'expliquer par la distribution de leurs haplotypes (Figure 19). On trouve les Acadiens d'une part, Montréal d'autre part, et enfin un groupe {Saguenay, Anglo-normands}. Nous avons vu précédemment que les distributions haplotypiques du Saguenay et des Anglo-normands sont relativement identiques, avec deux haplotypes très fréquents et un

nombre d'haplotypes relativement important. Tandis que Montréal et les Acadiens présentent des distributions haplotypiques très différentes.

Les valeurs F_{ST} sont comprises entre 0,1% et 3,1% pour *dys44*. Comme pour DXS1238, peu d'entre elles sont significatives. Les valeurs F_{ST} significatives sont celles qui expriment la différence entre Acadiens et Anglo-normands de Gaspésie, et également la différence entre les Acadiens et Montréal.

Dans l'ACP réalisée à partir des fréquences des haplotypes *dys44* des régions françaises et du Québec au complet, deux populations semblent ressortir : le Nord de la France et la Bretagne (Figure 20). La singularité du Nord de la France est due au fait que le deuxième haplotype plus fréquent est B006 (23%) au lieu d'être B003 (12%), et que cet haplotype est presque aussi fréquent que B001 (25%). Tandis qu'en Bretagne l'haplotype B001 est prépondérant (36%) et que les deux haplotypes suivant (B003 et B006) ont une fréquence identique (15%), et proche d'un haplotype (B011) moins fréquent dans les autres régions françaises.

L'ACP réalisée pour l'ensemble des populations québécoises et régions françaises n'a pas semblé apporter plus d'informations, c'est pourquoi elle n'est pas représentée. Cependant les valeurs des F_{ST} présentent le Nord de la France comme significativement différent des Acadiens et des Anglo-normands.

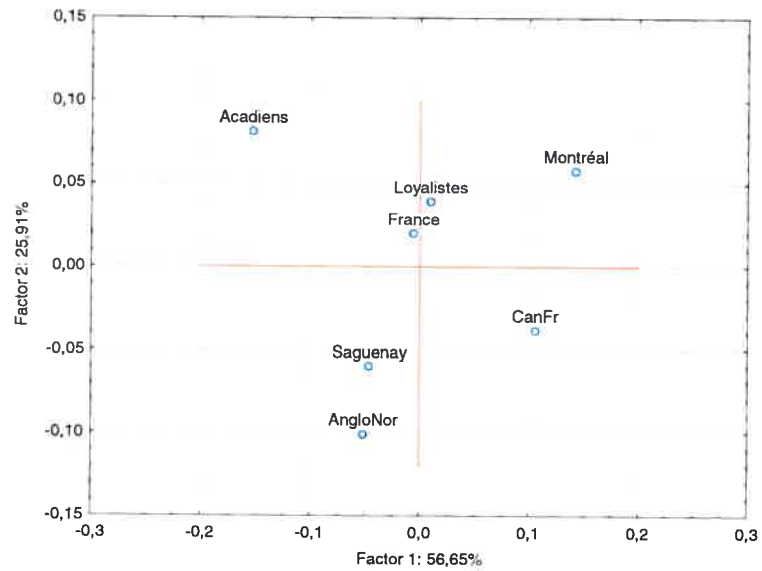


Figure 19 Analyse en Composantes Principales réalisée à partir des fréquences des haplotypes dys44 des sous-populations québécoises et de la France.

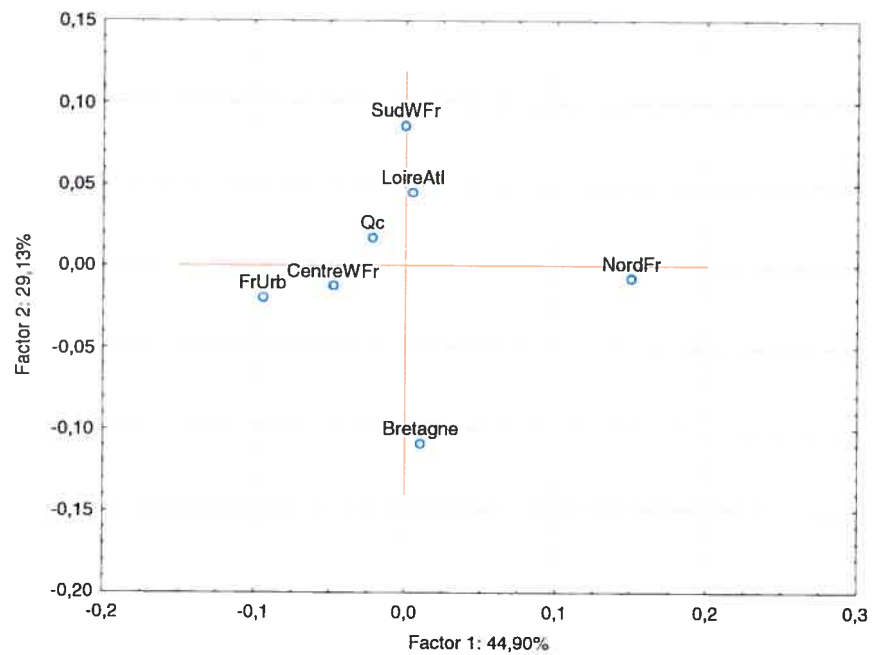


Figure 20 Analyse en composantes principales réalisée à partir des fréquences des haplotypes dys44 du Québec et des régions françaises.

C. Les haplotypes composés (haplotypes-C)

Les individus possédant un haplotype-C sont ceux qui ont été génotypés pour les deux marqueurs DXS1238 et *dys44*. Ces haplotypes-C ont été utilisés pour observer les liens entre populations. L'objectif est de comparer les combinaisons des haplotypes des deux marqueurs présentes au Québec par rapport à celles observées en France. Le degré de polymorphisme des haplotypes-C est plus important que celui des deux marqueurs séparés et pourrait mettre en évidence des relations plus subtiles entre nos populations québécoises et nos régions françaises.

La distribution des fréquences des haplotypes pour chaque population a été obtenue et est représentée par les figures 21, 22, 23. Etant donné le grand nombre d'haplotypes-C observé, la distribution des fréquences a été séparée en deux. Les figures 21 et 22 sont constituées des haplotypes observés en France et au Québec, ainsi que des haplotypes spécifiques à la France, et sont classés par ordre décroissant par rapport à cette dernière. La figure 23 est constituée des haplotypes observés uniquement au Québec, et les haplotypes sont classés par ordre décroissant par rapport au Québec au total. A partir du nombre d'haplotypes-C pour chaque population, des graphiques de partage d'haplotypes-C ont été effectués.

Le détail du nombre d'haplotypes-C observés dans chaque population est reporté dans le tableau 4 sous la mention « nombre d'haplotypes ». Ce tableau comporte également les indices de diversité pour chaque population. Comme pour les deux autres marqueurs des analyses de distances entre population ont été réalisées à l'aide d'ACP et de F_{ST} . Les résultats de la distribution des fréquences haplotypiques et des partages d'haplotypes seront d'abord présentés, puis viendront la présentation des résultats des indices de diversité et des analyses de distances.

1. Distribution des Haplotypes-C

Au total, 142 haplotypes-C ont été observés pour les 533 chromosomes concernés. 48 haplotypes-C sont spécifiques au Québec et représentent 61 chromosomes. Pour notre échantillon français, 52 haplotypes-C sont spécifiques et représentent 62 chromosomes. Les distributions des haplotypes-C présentent les populations du

Québec et de la France comme bien différente. Une vue générale de ces distributions sera présentée, puis quelques unes des combinaisons seront un peu plus détaillées. Ces combinaisons seront présentées suivant leur ordre d'apparition sur les figures.

a) Vue générale

En comparant les distributions du Québec par rapport à celles de la France on s'aperçoit de certaines différences (Figures 21 et 22). L'échantillon montréalais présente peu d'haplotypes-C, et l'haplotype-C le plus fréquent n'est pas celui rencontré habituellement. D'une façon générale, les autres populations présentent un plus grand nombre d'haplotypes-C en commun avec la France, mais des différences notables sont perçues au niveau des fréquences, dans le sens où des pointes de hautes fréquences apparaissent pour des haplotypes-C moins fréquents en France.

Chez les Canadiens français de Gaspésie, les trois premiers haplotypes-C présentent des fréquences similaires. Chez les Loyalistes la fréquence du premier haplotype-C prédomine les autres. C'est aussi le cas pour les Anglo-normands et les Acadiens, seulement la fréquence de cet haplotype-C est légèrement plus basse que chez les Loyalistes. Ces distinctions entre distributions d'haplotypes-C sont reprises au niveau des valeurs des θ (tableau 4).

Du côté français on s'aperçoit déjà que la distribution de la Bretagne n'a rien de commun avec les autres distributions. L'haplotype-C le plus fréquent n'est pas le même que dans les autres populations, et trois des haplotypes-C normalement les plus fréquents en France n'ont pas été détectés dans cette région. La distribution du Sud-ouest de la France semble très étalée, avec des fréquences similaires des haplotypes-C dans cette région, tandis que dans notre échantillon de France urbaine, la fréquence du premier haplotype-C prédomine toutes les autres.

b) Les différentes fréquences d'haplotypes-C

D'une façon générale, l'haplotype-C le plus fréquent est 5.1.0.0.19.2.18-B001. Ceci n'est pas le cas à Montréal. Soulignons que 5.1.0.0.19.2.18 est l'haplotype le plus fréquent de DXS1238 et B001 l'haplotype le plus fréquent de dys44.

L'haplotype-C 5.1.0.0.15.2.15-B005 est le deuxième haplotype en terme de fréquence en France, il a été observé dans toutes les régions, sauf en Bretagne. Du

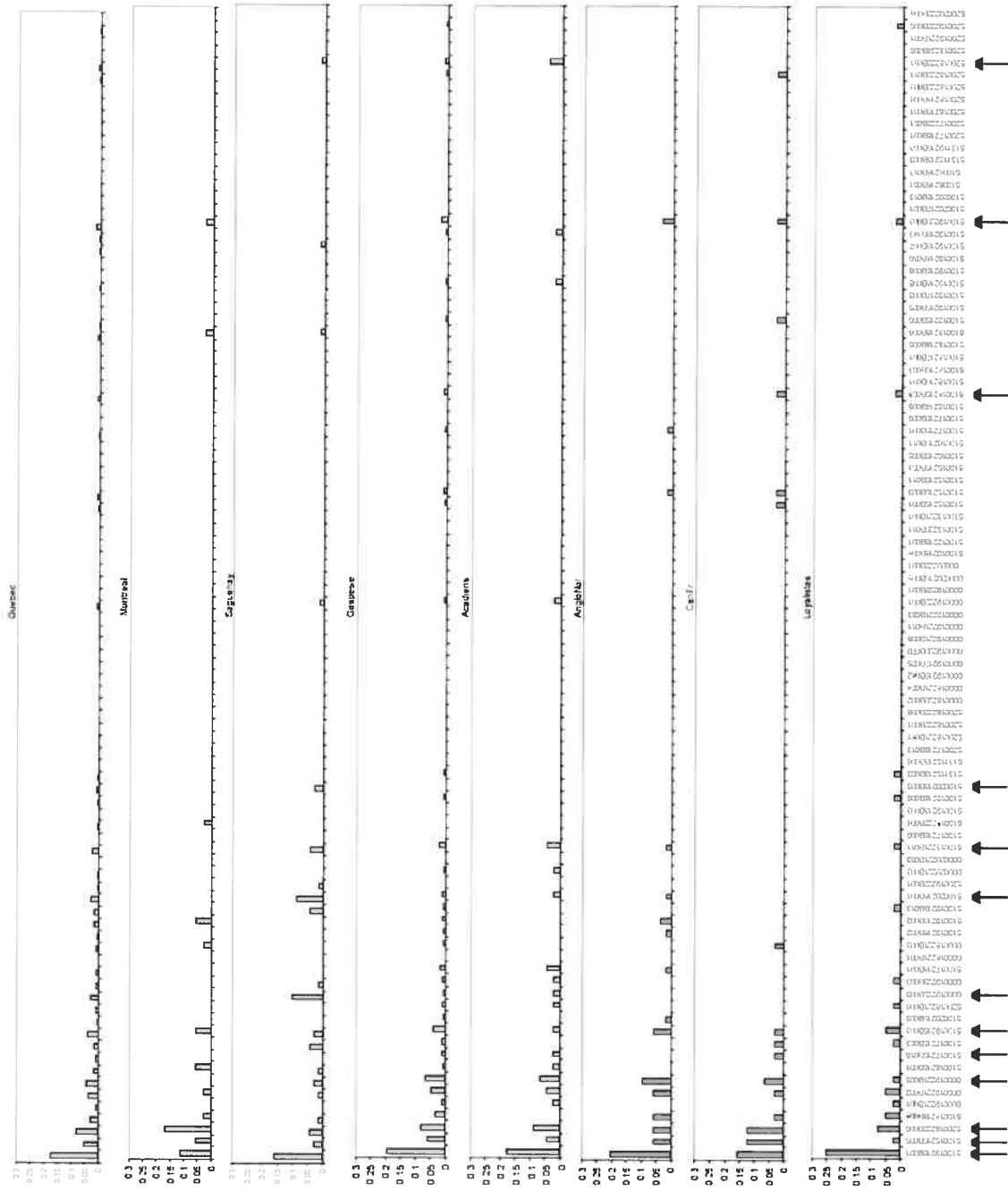


Figure 21
Distributions
québécoises des
fréquences des
haplotypes-C
communs à la
France et au Québec
et spécifiques en
France. Les
haplotypes-C sont
classés par ordre
décroissant par
rapport à la France.
Les flèches
correspondent aux
haplotypes-C
mentionnés dans le
texte.

Figure 22

Distributions

françaises des

fréquences des

haplotypes-C

communs à la

France et au

Québec et

spécifiques en

France. Les

haplotypes-C

sont classés par

ordre

décroissant par

rapport à la

France. Les

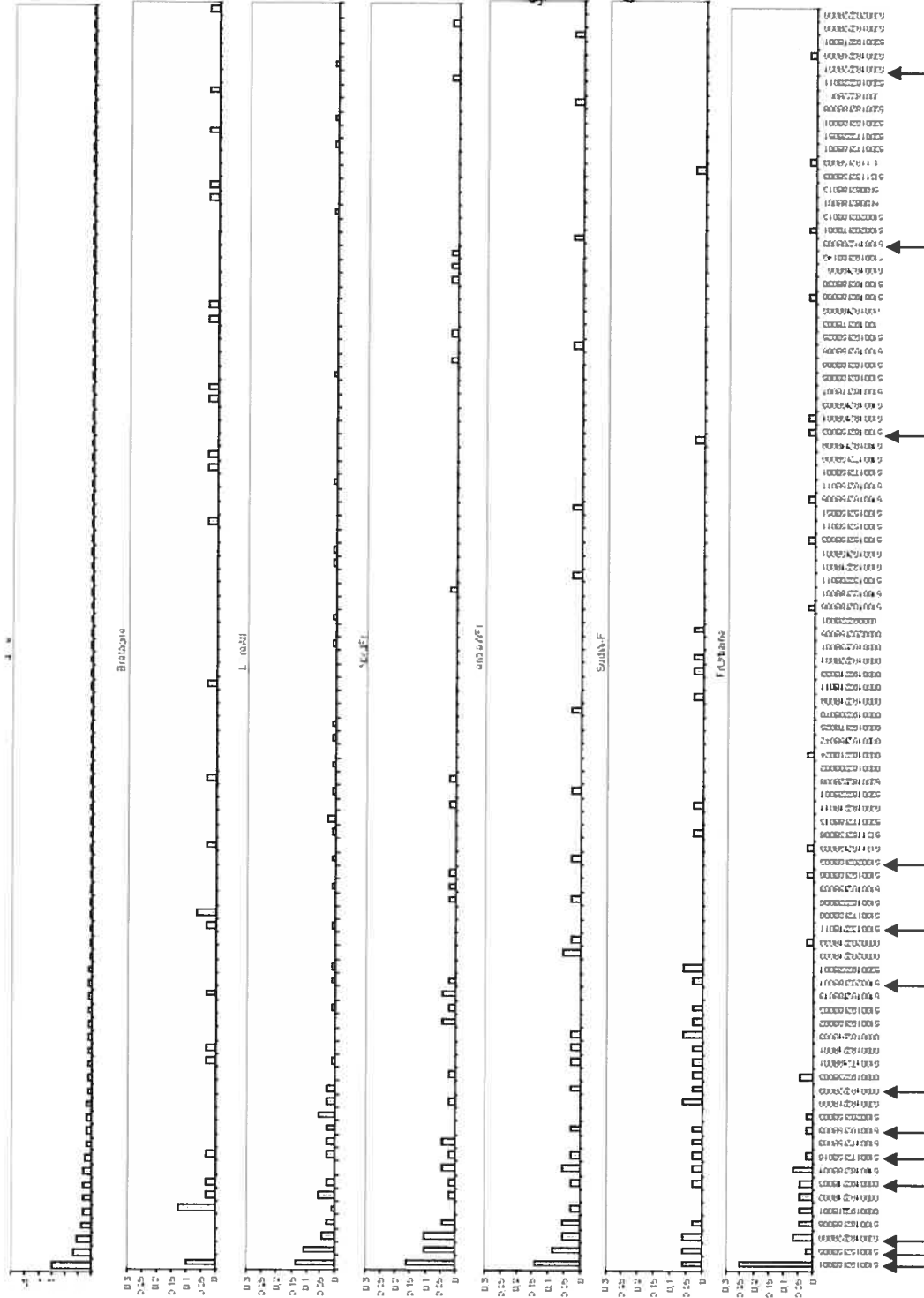
flèches

correspondent

aux haplotypes

mentionnés

dans le texte.



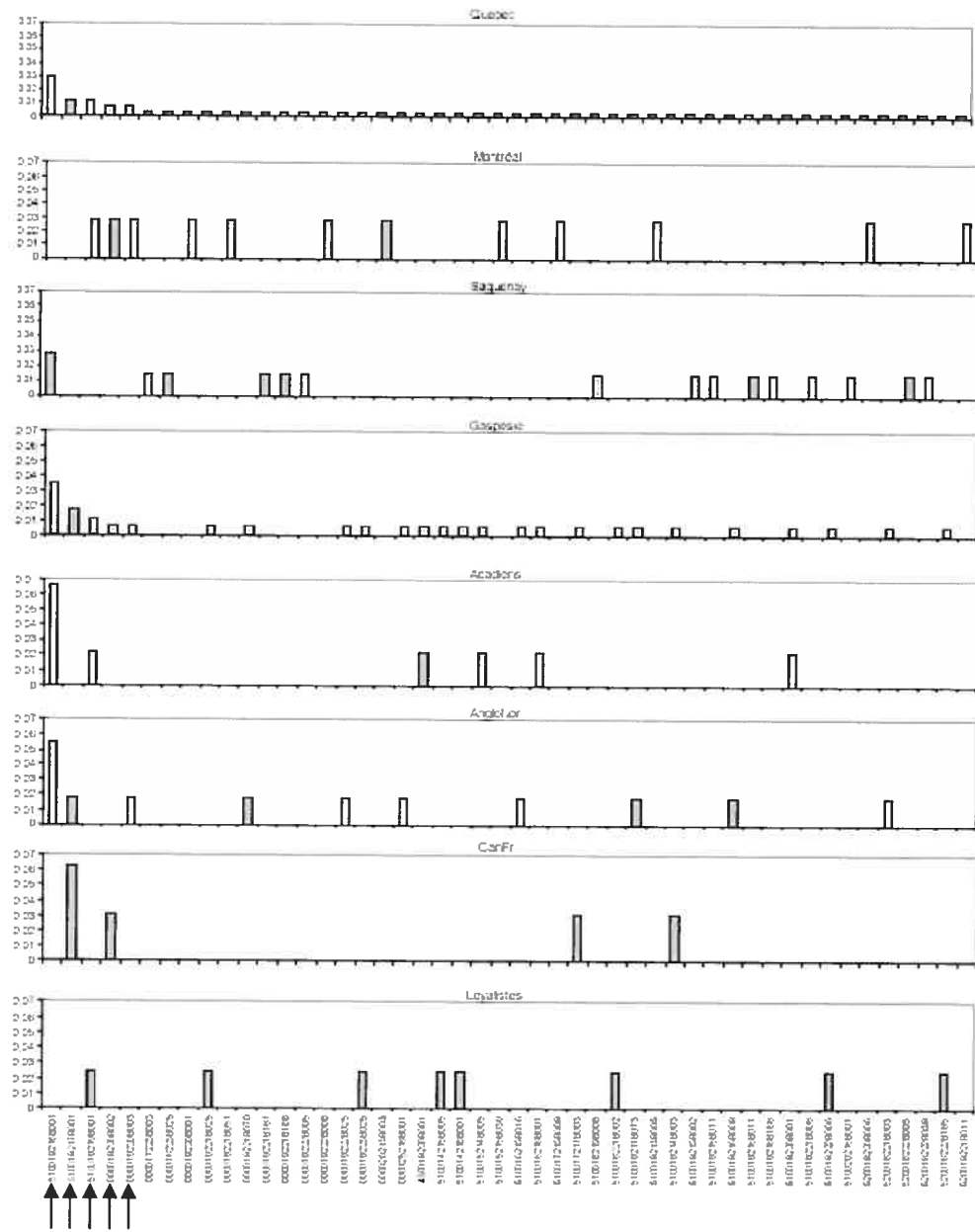


Figure 23 Distributions des fréquences des haplotypes-C spécifiques au Québec. Les haplotypes-C sont classés par ordre décroissant par rapport au Québec au complet. Les flèches correspondent aux haplotypes-C mentionnés dans le texte.

côté québécois, cet haplotype-C fait partie des plus fréquents et est retrouvé dans toutes les populations québécoises. Il fait d'ailleurs partie des trois haplotypes-C les plus fréquents des Canadiens français de Gaspésie.

L'haplotype-C 5.2.0.0.18.2.22-B006 est 1,4 fois plus fréquent au Québec qu'en France, notamment à Montréal et en Gaspésie. Des différences sont notables au niveau des sous-populations gaspésiennes, avec une fréquence de cet haplotype-C plus basse chez les Loyalistes par rapport aux trois autres sous-populations gaspésiennes, et une fréquence quatre fois supérieure à celle de la France chez les Canadiens français de Gaspésie.

L'haplotype-C 0.0.0.0.19.2.21-B003 est présent dans toutes les régions françaises et dans toutes les populations québécoises, sauf à Montréal. Cet haplotype-C est 1,5 fois plus fréquent au Québec qu'en France. Cet haplotype-C fait partie des plus fréquents en Gaspésie.

L'haplotype-C 0.0.0.0.17.2.15-B016 est également ubiquitaire en France, alors qu'au Québec il a été observé au Saguenay et en Gaspésie, et seulement chez les Canadiens français et les Acadiens de Gaspésie.

L'haplotype-C 5.1.0.0.19.2.15-B003 est observé dans quatre régions françaises sur six (Loire-Atlantique, Centre-ouest, Sud-ouest et France urbaine) alors qu'il est ubiquitaire au Québec. Cet haplotype-C est deux fois plus fréquent au Québec qu'en France.

L'haplotype-C 0.0.0.0.19.2.22-B003 a été observé dans trois des six régions françaises (Loire-Atlantique, Centre-ouest, Sud-ouest) et au Saguenay et chez les Acadiens seulement au Québec. Cet haplotype-C est deux fois plus fréquent au Québec qu'en France, et 6,5 fois plus fréquent au Saguenay qu'en France.

L'haplotype-C 5.1.0.0.20.2.21-B001 est 2,5 fois plus fréquent au Québec qu'en France. Il a été observé dans trois des six régions françaises (Loire-Atlantique, Centre-ouest, Sud-ouest) et au Saguenay et en Gaspésie seulement au Québec. En Gaspésie cet haplotype-C est partagé par les Acadiens et les Anglo-normands.

L'haplotype-C 5.1.0.0.13.2.21-B011 est 3,3 fois plus fréquent au Québec qu'en France, où il a été observé dans deux régions sur six (Bretagne et Loire Atlantique). Au Québec cet haplotype-C est présent au Saguenay et dans toutes les sous-populations gaspésiennes sauf chez les Canadiens français. Cet haplotype-C est particulier au Québec parce que tous les individus 5.1.0.0.13.2.21 sont également B011, alors qu'en France cet haplotype-C DXS1238 est également rencontré avec B001.

L'haplotype-C 5.1.0.0.20.2.18-B003 est présent dans deux régions françaises sur six (Loire Atlantique et le Centre-ouest de la France) et a été observé au Saguenay seulement au Québec.

L'haplotype-C 5.1.0.0.18.2.15-B003 a été identifié en France urbaine et en Gaspésie, où il est partagé par les Anglo-normands et les Canadiens français.

L'haplotype-C 5.1.0.0.19.2.20-B003 est 4,7 fois plus fréquent au Québec qu'en France, où il a été observé dans le Nord de la France seulement. Cet haplotype-C a été identifié à Montréal et dans toutes les populations gaspésiennes, sauf chez les Acadiens.

L'haplotype-C 5.2.0.0.18.2.22-B051 est 2,8 fois plus fréquent au Québec qu'en France, où il a été observé en Loire Atlantique seulement. Au Québec, cet haplotype-C est partagé par le Saguenay et les Acadiens.

c) Les haplotypes-C spécifiques au Québec

Parmi les 48 haplotypes spécifiques au Québec, 43 ont été observés dans une seule des populations québécoises échantillonnées (Figure 23). Ainsi cinq haplotypes-C spécifiques au Québec sont communs à deux populations au moins. Concentrons-nous sur ces haplotypes qui nous apportent des informations sur les liens entre populations québécoises.

L'haplotype-C 5.1.0.0.19.2.16-B001 a été observé au Saguenay et en Gaspésie, chez les Acadiens et les Anglo-normands.

L'haplotype-C 5.1.0.0.19.2.17-B001 a été identifié en Gaspésie seulement, chez les Anglo-normands et les Canadiens français. Cet haplotype-C est 3,4 fois plus fréquent chez les Canadiens français que chez les Anglo-normands.

L'haplotype-C 5.1.0.0.19.2.19-B001 est partagé par les Acadiens et les Loyalistes en Gaspésie, et a aussi été observé à Montréal. L'haplotype-C 0.0.0.0.19.2.22-B002 est partagé par les Canadiens français de Gaspésie et les Montréalais. L'haplotype-C 0.0.0.0.19.2.20-B003 est partagé par les Anglo-normands de Gaspésie et les Montréalais.

2. Partage d'haplotypes-C

Pour chaque population, la proportion des trois catégories du graphique de partage d'haplotypes est relative au nombre total d'haplotypes-C observé dans la population considérée.

La figure 24a représentant le partage d'haplotypes-C entre populations québécoises nous montre que les populations qui présentent les proportions d'haplotypes-C uniques les plus élevées au Québec sont le Saguenay et Montréal (42%), suivies de près par les Loyalistes et les Anglo-normands (38%). Les Canadiens français de Gaspésie présentent la proportion de partage d'haplotypes-C avec une autre population la plus élevée (30%). On s'aperçoit ici que plus on augmente la taille de l'échantillon, plus la proportion d'haplotypes uniques diminue. Ceci va de paire avec l'effort d'échantillonnage. En effet, plus on prélève d'individus appartenant à une population, et plus on a de chances de prélever les allèles les moins fréquents de la population.

Sur la figure 24b, on s'aperçoit qu'en ajoutant la France, la proportion d'haplotypes-C uniques diminue pour toutes les populations, mais surtout pour les Canadiens français de Gaspésie et les Acadiens (deux fois moins d'haplotypes-C uniques dans ces deux sous-populations gaspésiennes par rapport à la figure 24a). Ce graphique met en évidence non seulement les points communs entre les populations du Québec et la population française (à travers le nombre d'haplotypes présents dans trois populations ou plus), mais aussi la contribution génétique française spécifique à

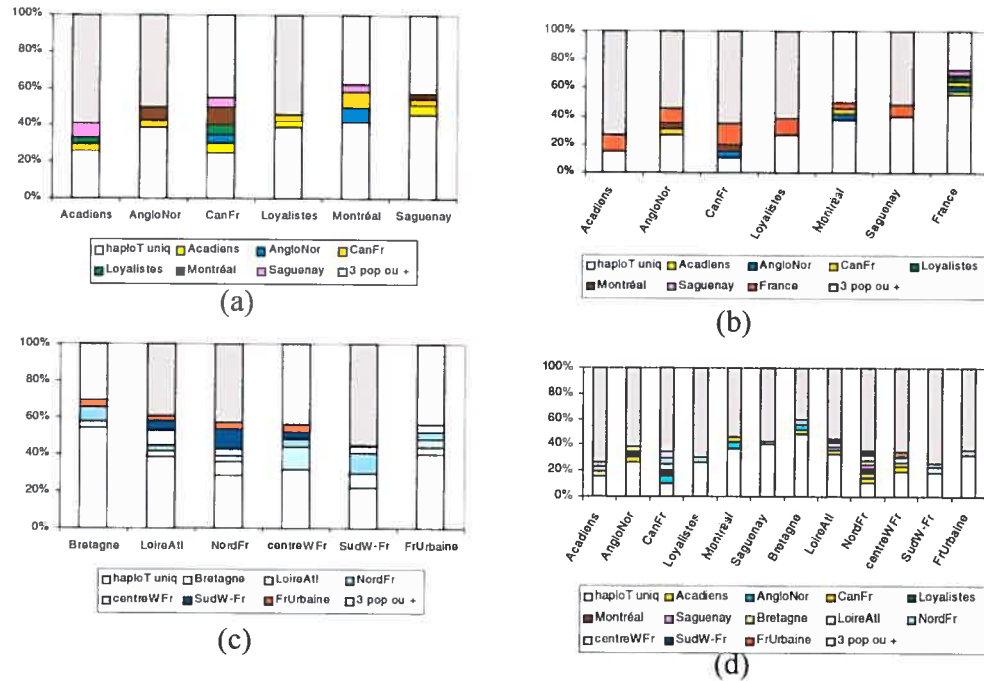


Figure 24 Graphiques de partages d'haplotypes-C entre différents groupes de population : a) entre populations québécoises ; (b) entre populations québécoises et la France au complet ; (c) entre régions françaises ; (d) entre régions françaises et populations québécoises.

chaque population québécoise (proportion d'haplotypes-C partagés entre la France et une autre population québécoise). Ce graphique nous montre aussi la singularité de la France par rapport aux populations québécoises, avec 58% d'haplotypes-C uniques.

La figure 24c présente les partages d'haplotypes-C entre les régions françaises. Ce graphique reflète les liens étroits qui existent entre les régions françaises. En effet toutes les régions françaises échantillonnées présentent une proportion d'haplotypes-C partagés avec une autre région de 20% environ. La région du Nord de la France a la proportion d'haplotypes-C partagés avec une autre population la plus élevée (25%). La Bretagne présente la proportion d'haplotypes-C uniques la plus importantes avec 56% d'haplotypes uniques. Ce graphique illustre également le lien entre les régions du Nord de la France, le Centre-ouest de la France et la Loire Atlantique.

La figure 24d présente les contributions potentielles de certaines régions françaises à certaines populations québécoises. Notamment le Nord de la France et le Centre-ouest de la France voient leur proportion d'haplotypes-C uniques diminuer de moitié (par rapport à la figure 24c). Ces deux régions françaises semblent avoir beaucoup contribué à la régionalisation génétique au Québec. Pour les populations québécoises, ce graphique montre que toutes les régions françaises n'ont pas contribué de façon identique aux pools géniques régionaux du Québec. Les proportions d'haplotypes-C uniques de Montréal et du Saguenay restent aussi élevées que dans la figure 24a, illustrant la singularité de ces populations pour les haplotypes-C.

3. Les Indices de Diversité

Pour 19 chromosomes de plus par rapport au Québec, la France présente quatre haplotypes-C supplémentaires (Tableau 4). Les hétérozygoties totales de la France (0,96) et du Québec (0,95) présentent des valeurs semblables. Bien que les Canadiens français de Gaspésie présentent le nombre de chromosomes et le nombre d'haplotypes-C les plus bas (32 chromosomes et 20 haplotypes-C), leur hétérozygotie n'est pas inférieure à celle des autres populations québécoises.

| populations | # chr | # haploT | hétéroZ | θ_{Hom} | θ_K | Chakraborty | |
|-------------|-------|----------|---------|----------------|------------|--------------|------|
| | | | | | | # haploT att | Test |
| France | 257 | 94 | 0,96 | 24,19 | 52,97 | 59,8 | TS |
| Bretagne | 31 | 25 | 0,98 | 43,67 | 57,87 | 23,6 | |
| CentreWFr | 35 | 25 | 0,97 | 32,22 | 37,7 | 23,9 | |
| Loire Atl | 66 | 36 | 0,96 | 24,1 | 31,65 | 32,1 | |
| NordFr | 48 | 28 | 0,95 | 18,64 | 27,24 | 24,1 | |
| SudWFr | 33 | 27 | 0,99 | 85,09 | 67,03 | 28 | |
| Fr urbaine | 44 | 25 | 0,93 | 11,85 | 23,19 | 18,77 | S |
| Québec | 276 | 90 | 0,95 | 17,4 | 46,05 | 49,6 | S |
| Gaspésie | 171 | 62 | 0,94 | 14,35 | 34,52 | 37,2 | TS |
| Acadiens | 45 | 26 | 0,95 | 19,35 | 24,83 | 23,6 | |
| AngloN | 54 | 26 | 0,94 | 14,28 | 19,09 | 22,7 | |
| CanFr | 32 | 20 | 0,95 | 18,03 | 21,82 | 18,7 | |
| Loyalistes | 40 | 26 | 0,93 | 12,78 | 31,1 | 18,5 | TS |
| Saguenay | 69 | 35 | 0,95 | 19,08 | 27,7 | 29,6 | |
| Montréal | 36 | 24 | 0,96 | 22,5 | 12,34 | 21,8 | |

Tableau 4 Indices de diversité des haplotypes-C dans les populations échantillonnées. Pour le test de Chakraborty, seuls les résultats significatifs (S, $0,01 \leq p \text{ value} \leq 0,05$) ou très significatifs (TS, $p \text{ value} > 0,01$) sont écrits.

D'un point de vue global, les θ_K sont supérieurs aux θ_{Hom} , sauf pour le Sud-ouest de la France et Montréal, pour qui c'est le contraire. Trois populations présentent un certain équilibre entre les valeurs de leurs θ : le Centre-ouest de la France, les Canadiens français de Gaspésie et les Anglo-normands.

Le test de Chakraborty s'est révélé positif pour la France au total et le Québec au total. Le test de Chakraborty est également significatif pour la France urbaine. Du côté québécois, les populations révélant un test de Chakraborty significatif pour les haplotypes-C sont la Gaspésie au total et les Loyalistes.

4. Les analyses de distance

Les ACPs ont été obtenues à partir des fréquences des haplotypes-C des populations considérées. Les F_{ST} ont été calculées en comparant les populations deux à deux, à partir des fréquences des haplotypes-C pour trois ensembles de populations. Les F_{ST} ont été représentées par MDS.

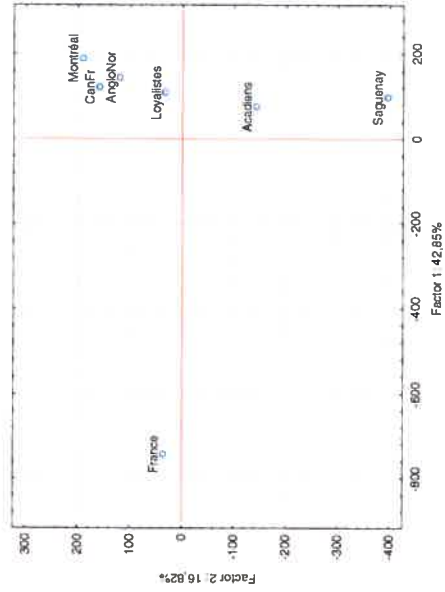


Figure 25 ACP réalisée à partir des fréquences des haplotypes-C de la France et des populations québécoises.

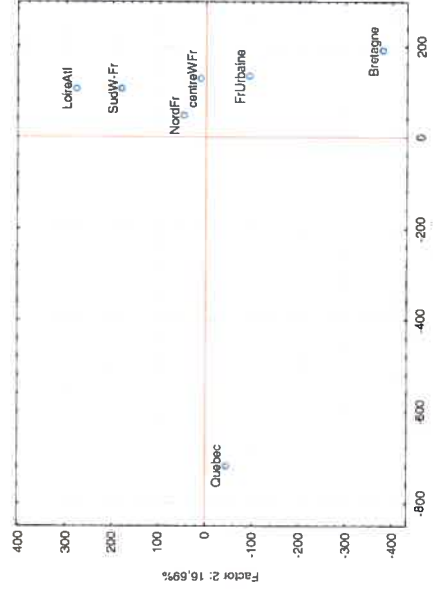


Figure 26 ACP réalisée à partir des fréquences des haplotypes-C du Québec et des régions françaises.

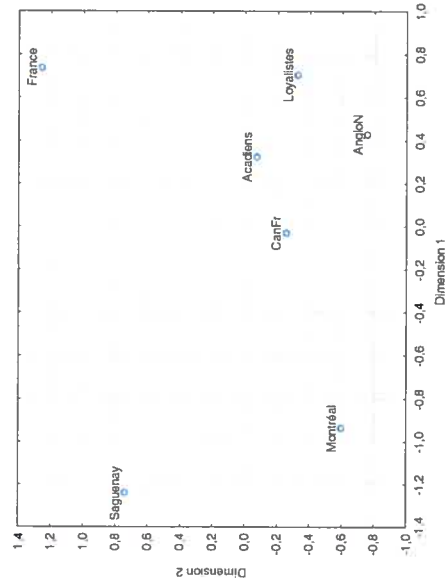


Figure 27 Représentation MDS des F_{ST} réalisée à partir des fréquences des haplotypes-C de la France et des populations québécoises.

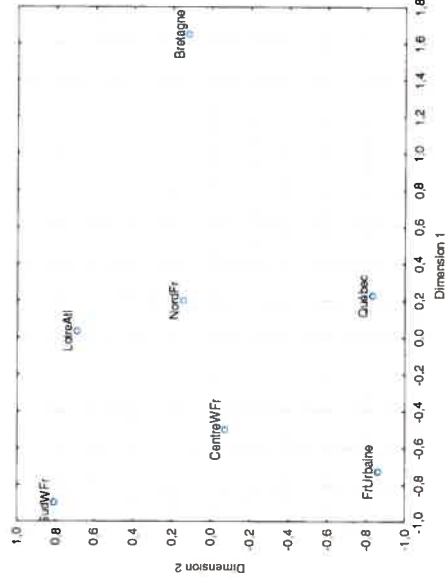


Figure 28 Représentation MDS des F_{ST} réalisée à partir des fréquences des haplotypes-C du Québec et des régions françaises.

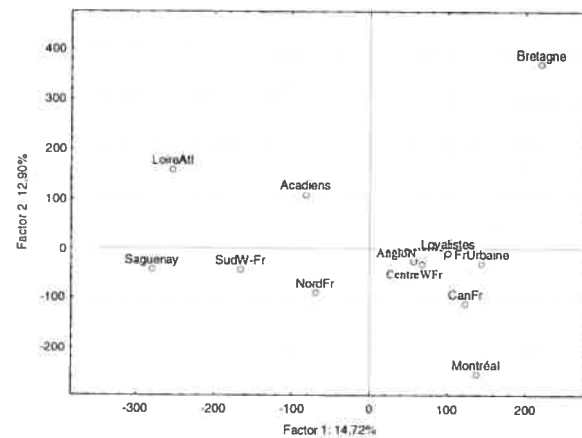


Figure 29 ACP de l'ensemble des populations québécoises et régions françaises, réalisées à partir des fréquences des haplotypes-C.

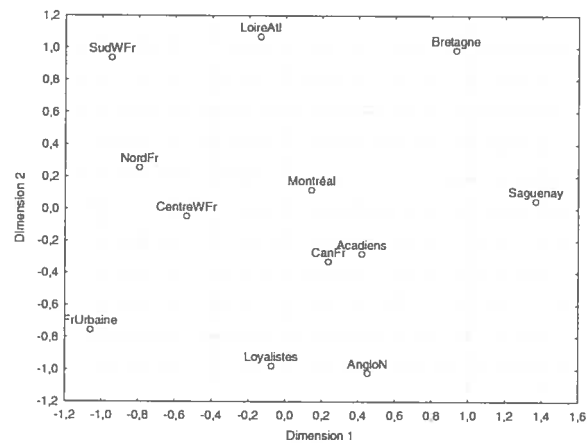


Figure 30 Représentation MDS des F_{ST} de l'ensemble des populations québécoises et régions françaises, réalisées à partir des fréquences des haplotypes-C.

La figure 25 marque une opposition entre la France au total et les populations québécoises. Cette opposition est également appuyée par la figure 26 qui révèle une opposition entre le Québec au total et les régions françaises. Au niveau des sous-populations québécoises, une opposition Saguenay/Montréal est mise en évidence, tandis que les sous-populations gaspésiennes forment un groupe central (figure 25). Au niveau des régions françaises, une opposition Loire Atlantique/Bretagne est révélée, avec la formation d'un groupe central composé du Nord de la France, du Centre-ouest et de la France urbaine (figure 26). La figure 29, représentant l'ACP pour l'ensemble des sous-populations de cette étude, met en évidence trois pôles populationnels : la Bretagne, Montréal et le Saguenay. Un groupe central composé des Loyalistes, des Anglo-normands et de la France urbaine est également formé.

Il faut toutefois relativiser les groupements inférés par cette analyse à l'aide des pourcentages des facteurs exprimés pour chacun des deux axes principaux. En effet les facteurs des axes des ACP ne sont pas très élevés (60% pour la figure 25 et 55% pour la figure 26, considérant qu'une bonne représentation des variables serait appuyée par 80%). La difficulté de séparer distinctement les populations étudiées à partir des haplotypes-C est également mise en évidence sur les représentations MDS des F_{ST} (figures 28, 29 et 30). En effet sur ces deux représentations les populations donnent une image éclatée des liens entre populations. Avec le faible pourcentage des axes principaux de l'ACP, on comprend que nos populations sont difficile à départager.

La représentation MDS des F_{ST} entre la France et les sous-populations québécoises révèle quant à elle quatre pôles populationnels : la France et le Saguenay d'une part, respectivement opposées avec l'ensemble des sous-populations gaspésiennes et Montréal (figure 27). Les valeurs des F_{ST} entre la France et les populations québécoises sont significatives entre la France et le Saguenay, et entre Montréal et le Saguenay.

Pour les haplotypes-C les valeurs des F_{ST} sont comprises entre 0,1% et 1,9% (figure 30). Ces valeurs nous indiquent que la différenciation de la Bretagne par rapport aux autres régions françaises et au Québec est significative. Ces valeurs de F_{ST} au niveau de l'ensemble des sous-populations sont significatives pour la

Bretagne et les populations québécoises, sauf dans le cas des Loyalistes et des Acadiens.

IV. DISCUSSION

La diversité génétique de trois régions du Québec a été appréhendée à l'aide de marqueurs de polymorphismes neutres situés sur le chromosome X. L'avantage d'utiliser ces polymorphismes neutres est d'éviter l'influence de la sélection naturelle sur la diversité génétique d'une population. De plus, dans une population, les autosomes sont les systèmes génétiques les moins sensibles à la dérive génétique (effectif efficace $2.N_e$), tandis que le chromosome Y et l'ADNmt sont les systèmes les plus sensibles à la dérive génétique (effectif efficace $\frac{1}{2}.N_e$). Ainsi le chromosome X (avec un effectif efficace égal à $\frac{3}{2}.N_e$) nous donne un aperçu intermédiaire de l'effet de la dérive génétique dans une population.

La dérive génétique est définie par la fluctuation aléatoire des fréquences des allèles d'une génération à l'autre. Elle a pour effet de diminuer la diversité génétique d'une population, en menant à long terme à la perte ou à la fixation d'un allèle. La dérive génétique se produit dans toutes les populations. Son effet est très marqué dans les populations isolées, qui, par définition, ne connaissent pas d'apport extérieur d'allèles. La dérive génétique agit d'autant plus drastiquement sur des populations de petite taille ayant connu un effet fondateur. Ceci s'explique par le nombre d'individus restreint à l'origine de la population, et par le fait que tous les individus d'une population n'ont pas un succès reproducteur identique.

L'effet fondateur est défini comme l'établissement d'une nouvelle population par quelques fondateurs originels, portant une petite fraction seulement de la variation génétique totale de la population mère. Il consiste en un échantillonnage aléatoire d'individus de la population source pour former la nouvelle population. Considérant un locus multiallélique, l'effet fondateur est perçu dans la nouvelle population à travers des proportions différentes des allèles de la population source. Dans le cas d'un effet fondateur, la fluctuation des fréquences alléliques se produit d'une population à l'autre.

D'une part à cause de l'échantillonnage aléatoire initial, et d'autre part à cause de la dérive génétique, population mère (France) et nouvelle population (Québec) seront différentes par leurs fréquences alléliques, et non par la nature de leurs allèles. Autrement dit un effet fondateur est illustré par des fréquences différentes d'un ensemble d'allèles dans la nouvelle population par rapport à la population source. Dans cette étude, un effet fondateur a non seulement été mis en évidence pour la

population du Québec (soit pour les trois régions étudiées), mais d'autres effets fondateurs régionalement distincts ont aussi pu être observé. Ainsi deux niveaux d'effets fondateurs ont été révélés.

A. La question de l'échantillonnage

L'objectif de ce projet est d'étudier la diversité génétique des régions du Québec. Les trois premières régions échantillonnées sont présentées ici. Le Québec est constitué de 17 régions administratives. Ces régions n'ont pas été peuplées au même moment et reflètent donc des histoires démographiques sensiblement différentes. La diversité génétique d'une population dépend de son histoire démographique. Ainsi l'étude de la diversité régionale d'une population revient à se demander quel est l'impact de l'histoire démographique de cette région sur sa diversité génétique ? D'après les histoires démographiques de ces régions, quelles seront leurs similitudes québécoises et leurs différences inter-régionales ?

Le Québec est privilégié dans ces études grâce à la disponibilité des données généalogiques. Ainsi les données généalogiques de ces trois régions ne sont pas présentées ici, mais seront accessibles ultérieurement. Les échantillons montréalais et saguenayens sont constitués de Canadiens français ancrés dans chacune de ces deux régions, et une sélection plus approfondie pourra être réalisée, après reconstruction des généalogies. D'après le recrutement fait dans ces deux régions, une certaine proportion de migrants est prise en compte, nous permettant d'inclure les migrations inter-régionales dans notre échantillonnage, processus important dans l'histoire du peuplement des régions du Québec [3, 8, 23, 54]. Ceci a d'ailleurs pu être étudié avec les graphiques de partage d'haplotypes.

Le recrutement réalisé en Gaspésie est basé sur l'histoire du peuplement de la péninsule, et répond en ce sens à notre problématique. Cependant les quatre groupes ethniques distingués ici n'ont pas eu la même contribution au pool génique gaspésien [23]. Nos quatre échantillons gaspésiens de tailles équivalentes ne peuvent donc pas être regroupés et utilisés pour étudier la population de la Gaspésie en tant qu'ensemble homogène. Ceci est d'ailleurs confirmé par les résultats du test de Chakraborty. Cependant notre objectif était dans un premier temps d'étudier et de

comparer les pools géniques des gaspésiens suivant leur ascendance, et dans une problématique de génétique des populations, des tailles d'échantillons équivalentes s'avéraient nécessaires. Ainsi des tailles d'échantillons identiques pour les quatre groupes gaspésiens nous ont permis de comparer les nombres d'haplotypes, leurs fréquences et d'observer si des structures existaient au sein de ces populations.

Dans la discussion qui va suivre les termes « France » (ou France au total) et « Québec » (ou Québec au total) seront utilisés en référence à l'ensemble des régions de l'ouest de la France et à l'ensemble des trois régions du Québec échantillonnées pour cette étude. Seule la provenance des échantillons français a été mentionnée et aucune autre information ne nous a été fournie. Il est possible de voir ici la limite de cet échantillon français. Toutefois ces échantillons ont été principalement utilisés dans l'interprétation des analyses comme une population européenne de référence plutôt qu'une analyse région par région.

Ayant conscience des limites de notre échantillonnage, les données obtenues ont été analysées et les conclusions tirées pourront être vérifiées dans la suite du projet.

B. L'effet fondateur au niveau de la province

Pour les trois systèmes, les haplotypes majoritaires en France sont les mêmes qu'au Québec. Les graphiques des haplotypes partagés ont également mis en évidence les similitudes entre la France et les populations québécoises échantillonnées (figures 11b, 18b, 24b). Le nombre d'haplotypes partagés par deux populations nous révèle des informations sur les liens entre ces populations. Le nombre d'haplotypes rencontrés dans trois populations ou plus met en évidence la proportion d'haplotypes communs aux groupes de populations considérés. A ce niveau on remarque que plus de 50% des haplotypes des trois systèmes sont partagés entre populations québécoises et populations françaises.

L'ensemble de ces graphiques révèle le lien originel entre le Québec et la France. Ce lien a été appuyé par les analyses de distances (ACPs et F_{ST}) de *DXS1238* et *dys44*, démontrant la proximité des populations totales de la France et du Québec avec, respectivement, les populations québécoises et les régions françaises

(figures 12, 13, 16, 19 et 20). En effet les différences entre les populations comparées sont subtiles et non drastiques, puisque les axes principaux des ACPs expriment entre 27% et 75% des données, et les valeurs des F_{ST} comparant les populations deux à deux sont faibles, et peu d'entre elles sont significatives. Ces analyses de distances ont été utilisées pour visualiser les relations entre les populations étudiées. L'évolution des positions des populations les unes par rapport aux autres a été observée en comparant différents groupes de populations et en réalisant deux types de représentations. Les images obtenues par ces deux méthodes étaient relativement identiques, les valeurs de F_{ST} procurant une robustesse statistique quant aux différences entre populations.

Toutefois des différences de fréquences sont observées au sein des haplotypes moins fréquents en France et plus fréquents au Québec. Parmi ces haplotypes (des trois systèmes), citons les haplotypes 0.0.0.0.19.2.23, 0.0.0.0.19.2.20, 5.1.0.0.13.2.21 et 5.1.0.0.20.2.18 pour DXS1238, ou B025 pour *dys44* ou encore, pour les haplotypes-C, 5.2.0.0.18.2.22-B006, 5.1.0.0.19.2.15-B003, 5.1.0.0.20.2.18-B001, 5.1.0.0.13.2.21-B011, 0.0.0.0.19.2.22-B003, 5.1.0.0.19.2.20-B003, 5.1.0.0.19.2.17-B001 et 5.2.0.0.18.2.22-B051 deux fois (ou plus) fréquents au Québec qu'en France (figures 9, 10, 17, 21, 22, 23). Ces haplotypes révèlent l'effet fondateur qui s'est produit au Québec.

D'autres haplotypes sont présents dans plusieurs régions françaises et faiblement représentés, voire absents, du Québec ; citons 0.0.0.0.18.2.21, 5.1.3.1.15.2.13, B033, 5.1.0.0.20.2.18-B003, 0.0.0.0.20.2.21-B033, 0.0.0.0.18.2.21-B001. Ces deux remarques ensemble illustrent l'échantillonnage aléatoire constitué par les premiers fondateurs de la Nouvelle-France à partir de la population source. On peut également ajouter que l'haplogroupe 5.1.3.1.X.2.Y présente un seul haplotype au Québec, alors que dans notre échantillon français deux haplotypes ont été identifiés ; c'est également le cas pour les haplotypes-C 5.1.0.0.13.2.21-B011 et 5.1.0.0.13.2.21-B001, tous deux observés en France, alors qu'au Québec seul l'haplotype-C 5.1.0.0.13.2.21-B011 a été identifié.

En considérant l'ensemble des fréquences des haplotypes, on s'aperçoit que les haplotypes les moins fréquents dans la population mère (0% à 15% de notre échantillon français) sont les plus affectés par l'échantillonnage aléatoire. Nos trois

systemes ont ainsi illustre l'image de l'effet fondateur quebecois, d'une part au niveau de l'echantillonnage aleatoire inherent a l'effet fondateur, et d'autre part au niveau des modifications des frequences alleliques, puisque la majorite des haplotypes sont presents de part et d'autre de l'Atlantique, mais avec des frequences differentes.

C. Diversité génétique et effet fondateur

La dérive génétique est souvent associée à l'effet fondateur. Celle-ci, en agissant plus drastiquement sur les populations de faibles effectifs, diminue la diversité génétique de ces petites populations. Elle est d'autant plus importante pour les allèles rares [12]. Les allèles rares peuvent être définis comme ceux observés chez un individu ou deux dans un échantillon [55]. Etant donné l'effet fondateur au Québec, nous pourrions nous servir de la proportion de ces allèles (ou haplotypes dans notre cas) pour évaluer la signature de la dérive. On aurait ainsi pu s'attendre à ce que la proportion d'haplotypes rares soit plus basse au Québec qu'en France à cause de la dérive. Cependant la proportion de ces haplotypes est identique en France et au Québec, pour nos trois systèmes. Au niveau de ces haplotypes rares, ceci pourrait signifier que la dérive génétique n'a pas été si importante dans la différenciation du Québec par rapport à la France.

Les Juifs et les Finnois, comme les Québécois, sont des peuples connus pour avoir un nombre élevé de maladies autosomales récessives, une mutation fondatrice majoritaire étant responsable de chacune de ces maladies [30, 32, 56, 57]. Dans ces populations, les fréquences élevées de maladies mendéliennes rares sont attribuées à un effet fondateur accentué par la dérive génétique. Les études de ces populations ont révélé une diminution de diversité génétique pour le chromosome Y, se traduisant par une diminution de l'hétérozygotie par rapport aux autres populations, mais surtout par une diminution du nombre d'allèles des marqueurs étudiés [31, 58]. Tandis que la diversité génétique de l'ADNmt était comparable à celle des populations européennes pour les Finnois.

Au Québec, les hétérozygoties obtenues à partir du chromosome Y et de l'ADNmt sont comparables aux autres populations européennes, tandis que le nombre

d'haplotypes de ces deux marqueurs est plus bas au Québec (Moreau C., *et al* en préparation). Cette diminution du nombre d'allèles (ou d'haplotypes) n'est pas observée au Québec, en ce qui concerne le chromosome X, et par rapport à la diversité observée en France (tableaux 2, 3 et 4). De plus l'hétérozygotie obtenue pour chacun des trois systèmes et le nombre d'haplotypes de ces trois systèmes sont du même ordre de grandeur au Québec et en France, de même qu'entre les régions françaises et les populations québécoises.

Du fait que la diversité génétique québécoise n'est pas plus basse qu'en France et que la dérive génétique tend à diminuer cette diversité, on peut se demander si la dérive génétique a réellement été un facteur façonnant la diversité génétique actuelle du Québec. Helgason [29] souligne que l'hétérozygotie n'est pas sensible à la dérive génétique, et que des indices de diversité prenant en compte le nombre d'haplotypes sont plus performants pour mesurer son effet dans une population. L'étude de la diversité du chromosome Y et de l'ADNmt au Québec semble présenter cette diminution des θ_k par rapport aux populations européennes (Moreau C., *et al* en préparation).

Cependant en ce qui concerne le chromosome X, le nombre d'haplotypes est équivalent pour le Québec et la France pour les trois systèmes. Des différences sont visibles au niveau des θ_k , plus faibles au Québec qu'en France pour DXS1238 et les haplotypes-C, alors que c'est le contraire pour *dys44*. Toutefois pour ces trois systèmes ces indices sont du même ordre de grandeur entre le Québec et la France. Ainsi, si une diminution de diversité génétique est observée au Québec pour les lignées paternelles et maternelles, cette diminution de diversité sous l'effet de la dérive génétique n'est pas aussi évidente au niveau des résultats du chromosome X. L'impact de la dérive génétique au Québec est ici mise en évidence au niveau de l'effectif efficace du marqueur considéré. On comprend alors que les θ_k de l'ADNmt et du chromosome Y aient pu être influencées par la dérive génétique, puisque leur effectif efficace est inférieur à celui du chromosome X.

Considérons la forte croissance de la population de la Nouvelle-France à partir d'un nombre limité d'hommes et de femmes [2]. Etant donné que tous les individus n'ont pas le même succès reproducteur, on peut penser que la dérive génétique a pu accentuer la diminution de la diversité génétique du chromosome Y et de l'ADNmt.

Ceci est effectivement observé pour ces deux marqueurs au Québec, et reflète un effet fondateur pour ces deux marqueurs. Cependant la plupart des familles canadiennes-françaises étaient constituées de huit enfants par femme en moyenne. Dans de telles familles les allèles des deux parents pouvaient être représentés dans la génération suivante, et ce sur plusieurs générations. Ainsi, malgré l'effet fondateur québécois, on ne peut pas dire que la dérive génétique ait été un facteur déterminant de la diversité génétique au Québec. Le fait que la plupart des allèles aient pu être conservés au Québec est principalement dû à certaines caractéristiques démographiques de la province.

Ainsi bien que l'effet fondateur ait un impact important sur la diversité génétique du Québec, tant au niveau des polymorphismes neutres qu'au niveau des maladies génétiques, la dérive génétique ne semble pas avoir joué un rôle important comparativement aux autres populations connues pour leur effet fondateur [29, 38, 58]. Considérons l'échantillonnage aléatoire qu'a pu constitué le départ pour la Nouvelle-France de Français du XVII^e siècle. On peut se représenter la France ancienne – de François 1^{er} à Louis XV – comme une mosaïque humaine dont les entités étaient assez nettement découpées du point de vue social, culturel, géographique et aussi génétique [54]. L'immigration incessante depuis le Moyen-Age, et la position de la ville de Paris, tant au niveau métropolitain qu'au niveau internationale, se sont ajoutés à l'hétérogénéité de cette population française. Le prélèvement migratoire qui a donné naissance à la population canadienne-française s'est fait à partir de cette mosaïque [54].

Tout d'abord, la région parisienne, l'un des deux bassins de provenance les plus importants avec la Normandie, était la plus diversifiée génétiquement parmi toutes les régions de France. Ensuite, on dénombre à 10 000 le nombre de fondateurs de la Nouvelle-France [59]. Ce nombre ne peut pas être considéré comme « restreint » comparativement aux quelques 600 fondateurs de la région de Charlevoix. De plus, parmi les 10 000 immigrants fondateurs, 350 provenaient d'autres pays européens (Suisse, Allemagne et Belgique). Et enfin, si l'immigration a d'abord été familiale (avant 1700), elle s'est révélée beaucoup plus individuelle par la suite, augmentant la probabilité d'un apport génique diversifié [2, 54]. Ces trois énoncés historiques sont

en faveur d'une diversité du pool génique des premiers Canadiens français arrivés en Nouvelle-France.

Les études des patronymes les plus fréquents et les études de certaines maladies génétiques ont souvent suggéré l'homogénéité génétique du Québec [10, 60]. Il a été montré que la consanguinité ne pouvait pas expliquer les fréquences élevées des maladies génétiques au Québec [9, 61, 62]. Effet fondateur et forte croissance démographique suffisent à expliquer les fréquences élevées de certains variants rares. En effet la contribution des premiers fondateurs au pool génique d'une région a stabilisé l'effet fondateur, permettant à l'ensemble des variants génétiques, même rares, d'être conservés dans la population [63]. Au cours du temps, cette contribution des premiers fondateurs a entraîné l'augmentation de l'apparentement au sein d'une même région, comme le Saguenay-Lac Saint-Jean, ou sous-population pour ce qui est des Acadiens en Gaspésie [1, 14]. L'augmentation de l'apparentement entre individus appartenant à une population à fort effet fondateur a pu favoriser la « rencontre » d'haplotypes récessifs responsables de maladies génétiques [9].

Les résultats obtenus avec nos marqueurs génétiques se joignent à d'autres données généalogiques et génétiques pour soutenir que la démographie et l'histoire du peuplement ont été les facteurs déterminants du maintien de la diversité génétique du Québec d'aujourd'hui [9, 54, 61, 63].

D. Une seule population canadienne-française ?

DXS1238 est un marqueur hautement polymorphe avec 68 haplotypes, tandis que *dys44* est moins polymorphe avec 25 haplotypes, le nombre d'haplotypes étant celui observé dans les populations de cette étude. Les haplotypes composés de ces deux marqueurs nous permettent d'observer la distribution de 142 haplotypes. Etant donné que nous désirons observer les variations des fréquences alléliques entre nos populations et le degré de polymorphisme de chacun de ces trois marqueurs, on comprend que chacun de ces marqueurs procurera une information complémentaire concernant nos populations.

Ainsi les valeurs des F_{ST} obtenues à partir de DXS1238 ou de *dys44* ne montraient pas de différence significative entre la population totale du Québec et les régions françaises ou entre la France au total et les populations québécoises. Cependant au niveau des valeurs des F_{ST} obtenues à partir des fréquences des haplotypes-C, les populations totales du Québec et de la France se sont significativement distinguées des régions françaises et des populations québécoises, respectivement (figures 25 et 26). Ceci peut être corrélé à la proportion des haplotypes-C spécifiques (figure 23). Sur 52 haplotypes-C spécifiques en France, neuf (17%) ne sont pas des singletons, tandis qu'au Québec, sur 48 combinaisons haplotypes-C cinq (10%) seulement ne sont pas des singletons. La proportion relativement élevée de ces haplotypes-C en France s'ajoute aux variations de fréquences haplotypiques, détaillées plus haut, différenciant les populations totales de la France et du Québec et illustrant des effets fondateurs québécois.

La distinction faite au niveau des haplotypes-C pour les populations totales du Québec et de la France atteste des différences régionales non négligeables. En effet, les haplotypes de nos trois systèmes ne sont pas distribués de la même façon dans toutes les régions québécoises étudiées.

Dans les populations juive ashkénaze et finlandaise, des effets fondateurs multiples sont mentionnés pour expliquer les fréquences régionalement différentes des mutations responsables de ces maladies, traduisant une certaine hétérogénéité génétique [57, 64]. Une distribution régionale des fréquences de certaines maladies génétiques est également observée au Québec [3, 4, 9]. Citons l'hypercholestérolémie familiale dont la proportion de porteurs est de 1/900 dans la région de Montréal, 1/167 dans la région du Bas-Saint-Laurent, et 1/122 au Saguenay Lac-Saint-Jean [3, 4, 9]. Au Québec, plusieurs mutations sont responsables de cette maladie, laissant supposer l'existence d'une diversité génétique. L'hypercholestérolémie familiale est un exemple de maladie génétique parmi beaucoup d'autres.

D'autres populations, connues pour être génétiquement distinctes et isolées, présentent une hétérogénéité génétique d'après la distribution régionale des fréquences des allèles neutres étudiés. C'est le cas des Basques, dont l'hétérogénéité spatiale est accordée à la dérive génétique, au métissage ou encore à la consanguinité

[33, 65-67]. Cette hétérogénéité est également observée en Islande [26], où la particularité des distributions des allèles est due à un excès d'haplotypes intermédiaires et un déficit d'haplotypes rares, attribuées à la dérive génétique [29]. Nous savons désormais que la consanguinité proche n'a pas eu d'influence sur la diversité génétique du Québec, et que la proportion des haplotypes rares est équivalente entre la France au total et le Québec au total.

Le Québec partage des caractéristiques démographiques communes avec la Sardaigne et la Finlande. En effet les peuplements de la Sardaigne et de la Finlande ont expérimenté des effets fondateurs et des goulots d'étranglement démographiques, entraînant aujourd'hui l'apparition de maladies génétiques rares dans les autres populations européennes. La Sardaigne, comme la Finlande, sont des isolats génétiques considérés dans leur totalité comme des populations hétérogames [68]. Ces populations présentent un nombre de sites polymorphes, des valeurs de θ et un nombre de lignées distinctes inférieures aux autres populations européennes. De plus l'étude de polymorphismes neutres a révélé l'existence de sous-populations au sein de ces isolats [57, 69, 70]. Les structures de populations observées dans ces populations sont attribuées à des effets fondateurs régionalement distincts. Ainsi les caractéristiques démographiques du peuplement du Québec se retrouvent au niveau de ces deux dernières populations, si ce n'est que la croissance de la population québécoise a empêché que la dérive génétique ait un impact important dans la différenciation génétique du Québec par rapport à d'autres populations européennes (la France dans notre cas). C'est pourquoi, en ce qui concerne notre système génétique, la diminution de la diversité génétique n'est pas observée au Québec.

La fragmentation du pool génique de la population du Québec a été démontrée en utilisant la base de données démographiques du Registre de la Population du Québec Ancien [13]. A travers l'analyse de données démographiques des XVIIe et XVIIIe siècles des regroupements de populations ont été réalisés et la diversité génétique, en terme de « contribution génétique des fondateurs par région », a été mesurée. Cette étude a révélé que l'homogénéité génétique, basée sur des données datant des XVIIe et XVIIIe siècles, suivait une stratification géographique d'est en ouest. La stratification mise en évidence pour cette période semble corrélérer avec la distribution actuelle de certaines maladies génétiques, ou de certaines mutations [13]. Ainsi des

effets fondateurs régionaux sont observés au Québec, démontrés non seulement par la régionalisation des fréquences des maladies et de leurs mutations [7-10, 12], par des données démographiques et aussi par les distributions des haplotypes des polymorphismes neutres du chromosome X.

Les distributions des fréquences des haplotypes révèlent des différences régionales. Les analyses de distances nous montrent les relations entre les populations. Le test de Chakraborty constitue un test supplémentaire pour observer la structure – ou l’homogénéité - des régions étudiées. Plusieurs explications peuvent être fournies pour expliquer un test de Chakraborty significatif. Si moins d’haplotypes ont été observés que ce qui est attendu d’après l’homozygotie, on peut penser que la population a connu un goulot d’étranglement important réduisant la diversité génétique de la population. Le cas où plus d’haplotypes ont été observés que ce qui est attendu fait généralement référence à l’effet d’échantillonnage expliqué précédemment.

Dans notre cas, les tests de Chakraborty significatifs démontraient tous que plus d’haplotypes étaient observés par rapport à ce qui était attendu. A titre de témoin considérons la population de la France au total ainsi que celle de la France urbaine. On connaît la population française comme une mosaïque culturelle et génétique. De même on sait que, considérant des individus de même « nationalité », une population urbaine est constituée d’individus provenant de régions différentes. Le fait que le test de Chakraborty ait été significatif pour la France au complet, et la France urbaine illustrerait la présence d’entités distinctes au sein de ces deux échantillons.

Est ce que la population canadienne-française du Québec peut-être considérée comme une population « homogène » ? C’est dans ce but que nous avons rassemblé les trois régions québécoises étudiées ici (Montréal, le Saguenay et la Gaspésie) pour faire un test de Chakraborty sous le nom « Québec ». Le test de Chakraborty s’est révélé significatif, et ce pour les trois systèmes.

Compte tenu que les Loyalistes et les Anglo-normands n’appartiennent pas *a priori* à la population canadienne-française, une population « canadienne-française idéale » a été formée avec les échantillons acadiens, canadiens-français de Gaspésie, saguenayens et montréalais. Cette population « canadienne-française idéale » a été

testée pour les résultats de DXS1238 seulement. Le test de Chakraborty était encore significatif. Différentes combinaisons de « population canadienne-française idéale » ont été testées à l'aide de ces quatre échantillons. Chacun de ces tests de Chakraborty s'est révélé positif. D'après les études démogénétiques et les distributions des haplotypes de nos trois systèmes, nous en avons déduit l'hétérogénéité des Canadiens français du Québec.

Ce résultat est appuyé par les analyses de distances qui présentent les Canadiens français de Gaspésie en position diamétralement opposée au Saguenayens pour DXS1238 et les haplotypes-C. Tandis que les analyses de distances réalisées à partir des fréquences de *dys44* illustre une claire opposition entre Acadiens et Saguenay. On se rend alors compte que toutes les populations gaspésiennes ne conservent pas les mêmes relations suivant le système étudié. Ceci est également vrai pour les deux autres régions québécoises étudiées.

Nous avons réalisé un test de Chakraborty pour une « population gaspésienne idéale », rassemblant sans distinction les quatre populations gaspésiennes de notre étude. Ce test s'est révélé positif. L'hétérogénéité génétique détectée par ce test est appuyée par la présence des fondateurs d'origines diverses qui se sont installés dans la péninsule au cours du XVIIIe siècle. Ces distinctions culturelles et géographiques apparaissent également au niveau génétique.

E. Des effets fondateurs

1. Les Acadiens.

Sur les analyses de distance, les Acadiens se distinguent des autres sous-populations gaspésiennes pour *dys44* uniquement (figure 19). Dans la distribution des fréquences de ce marqueur B001 apparaît largement plus fréquent chez les Acadiens par rapport à l'ensemble des distributions des autres populations étudiées (figure 17). Les haplotypes illustrant des effets fondateurs peuvent être identifiés lorsque les haplotypes concernés sont ceux qui ont des fréquences basses, voire rares, dans la population mère par rapport à la nouvelle population. Cependant B001 est l'haplotype le plus fréquent dans les populations d'une façon générale. L'hypothèse de B001 illustrant un effet fondateur dans cette population gaspésienne peut toutefois

être appuyée par trois observations. Tout d'abord le nombre de chromosomes génotypés pour cette population est assez élevé (45 chromosomes), ensuite le nombre d'haplotypes *dys44* identifiés est relativement bas (neuf haplotypes), et enfin la fréquence de B001 est nettement supérieure aux autres haplotypes.

On aurait pu penser que B001 a été introduit par un seul fondateur, dont le succès reproducteur a surpassé celui des autres fondateurs. Si tel avait été le cas, la fréquence d'un haplotype-C composé de B001 aurait été beaucoup plus élevée que celles d'autres haplotypes-C dans cette population gaspésienne. Un tel haplotype-C n'a pas été identifié. Ceci signifierait que B001 a été introduit chez les Acadiens de Gaspésie par plusieurs fondateurs.

Avant que les premiers Acadiens n'arrivent en Gaspésie, leur histoire démographique témoigne de certains événements ayant pu affecter leur diversité génétique [21]. Nous savons que les premiers Acadiens arrivés en Acadie étaient peu nombreux. Ce groupe de premiers fondateurs devait posséder, en fréquences variables suivant l'effet d'échantillonnage, au moins les cinq haplotypes les plus fréquents en France, comme c'est le cas pour l'ensemble des populations québécoises étudiées ici. Ainsi avant le Grand Dérangement, le peuple acadien a déjà pu connaître un goulot d'étranglement démographique. Nous savons qu'un goulot d'étranglement supplémentaire s'est produit pendant le Grand Dérangement. Cette succession d'événements démographiques pourrait expliquer la diminution de la diversité observée dans notre échantillon d'Acadiens de Gaspésie. De plus, l'effet fondateur des Acadiens de Gaspésie a pu être renforcé par l'immigration familiale de cette population à cette période.

Ce scénario pourrait expliquer la prépondérance de B001 par rapport aux autres haplotypes dans cette population gaspésienne. B001 étant l'haplotype le plus répandu d'une façon générale, il est difficile de mettre en évidence un haplotype-C qui rendrait compte de cet effet fondateur. Cette hypothèse est également appuyée par le fort apparemment détecté entre les Acadiens de Gaspésie eux-mêmes (figure 31).

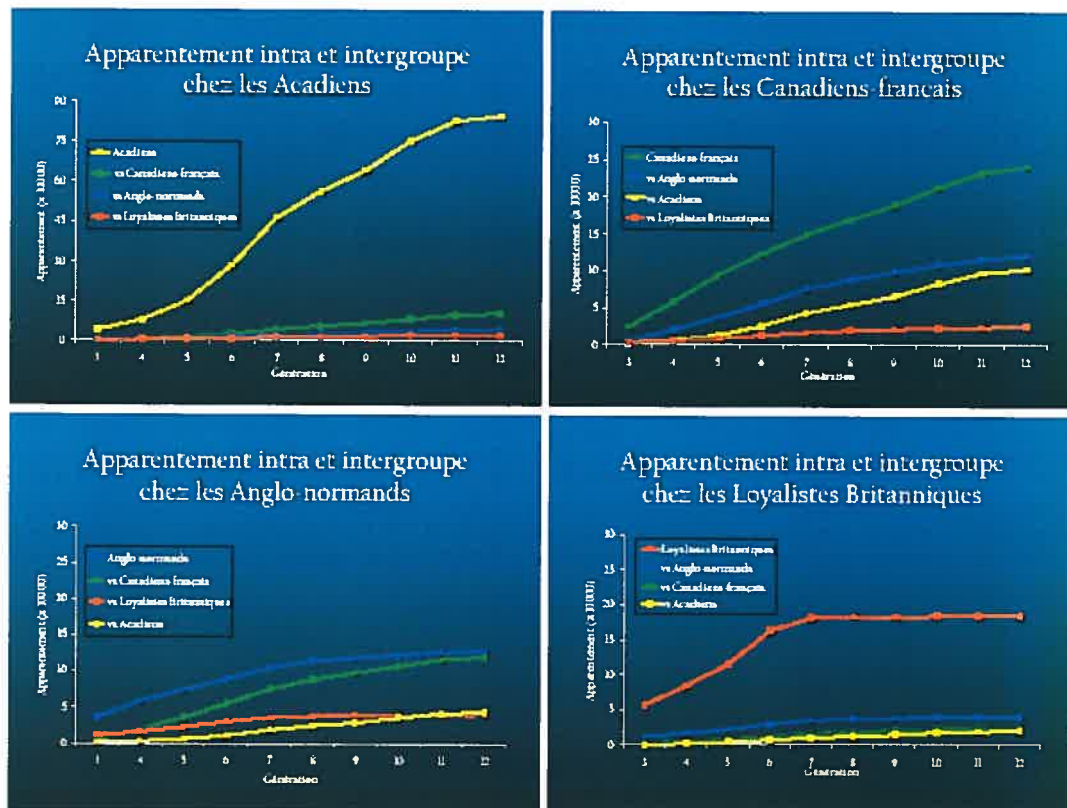


Figure 31 D'après un travail de Cathy Cauvier, appareancement intra et intergroupe dans les quatre populations gaspésiennes [1].

La diversité génétique des populations gaspésiennes a été étudiée à l'aide de données génétiques (ADNmt) et généalogiques [1]. La figure illustrant l'apparentement intra-groupe montre que ce sont les Acadiens qui présentent l'apparentement le plus élevé parmi les populations gaspésiennes. Si la démographie acadienne était similaire à celle de Charlevoix et du Saguenay-Lac-Saint-Jean [63], cet apparentement intra-acadien pourrait s'expliquer par la forte contribution génétique des premiers fondateurs au pool génique actuel.

L'hypothèse d'un goulot d'étranglement ou d'un effet fondateur extrêmement important dans la population acadienne a été révélé par l'étude de la diversité de l'ADNmt et du chromosome Y dans les populations québécoises (Moreau C., *et al* en préparation). Les populations du Québec présentent des hétérozygoties comparables aux populations européennes. Cependant les Acadiens de Gaspésie possèdent des diversités génétiques de ces deux marqueurs inférieures par rapport aux autres populations québécoises. La diversité des haplogroupes du chromosome Y, et même des haplotypes, est plus basse chez les Acadiens par rapport aux autres populations québécoises. De plus, un haplogroupe du chromosome Y est majoritaire dans cette population gaspésienne.

La population juive ashkénaze a également connu un effet fondateur important. Cependant d'après les données du chromosome Y, Behar DM *et al.* souligne que la diversité génétique de cette population diminue seulement lorsque la diversité intra-haplogroupe est prise en compte [58]. Tandis que chez les Acadiens de Gaspésie, le nombre d'haplogroupes du chromosome Y est suffisant pour percevoir la diminution de diversité génétique. Cette diminution de diversité génétique n'est observée ni pour DXS1238, ni pour *dys44*. Il est intéressant de voir ici que suivant le marqueur considéré différentes caractéristiques ressortent pour une même population.

L'ADNmt et le chromosome Y retracent les lignées maternelles et paternelles d'une population. Le chromosome X quant à lui retrace les deux lignées à la fois. Comme nous l'avons expliqué précédemment, ce n'est pas parce que la diversité d'une de ces lignées est réduite, que ceci signifie que la population est peu diversifiée. Une population est constituée d'un ensemble de gènes, chaque gène étant soumis à des contraintes propres, et dans certains cas plusieurs d'entre eux, soumis aux mêmes contraintes, racontent la même partie de l'histoire d'une population. Pour

comprendre l'intégralité de l'histoire évolutive d'une population, ses contraintes sélectives ou suivre l'histoire de ses polymorphismes neutres, plus on a de marqueurs, plus on se rapproche de l'histoire réelle de la population.

2. Les Anglo-normands

Au niveau des analyses de distance, les Anglo-normands se regroupent avec les autres populations gaspésiennes pour les trois systèmes. Les graphiques de partage d'haplotypes révèlent également un lien entre Anglo-normands et Canadiens français de Gaspésie et population de Montréal pour DXS1238 et les haplotypes-C. Au niveau des études généalogiques et démographiques, les Anglo-normands présentent un taux d'apparentement important avec les Canadiens français et les Loyalistes (figure 31).

Le moteur du peuplement de la péninsule gaspésienne était son activité halieutique. Les premiers Anglo-normands installés étaient principalement des hommes. Si les Anglo-normands sont restés dans la péninsule et ont eu des familles, la majorité d'entre eux étant des hommes, leurs femmes devaient appartenir à l'une des trois autres populations gaspésiennes. Au niveau du chromosome X, l'effectif efficace des femmes est deux fois celui des hommes. La signature du métissage est donc perçue au niveau de nos trois systèmes sur Xp21, et se comprend d'après l'histoire de la population anglo-normande en Gaspésie. Les études de l'ADNmt et du chromosome Y dans cette population ont également montré la proximité génétique entre Anglo-normands et Canadiens français de Gaspésie.

De plus, si on se replace dans le contexte de l'époque, les tensions entre francophones et anglophones, entre Français d'origine et Britanniques, étaient très vives à la Conquête. Les Iles anglo-normandes de Jersey et Guernesey sont situées sur les côtes septentrionales de la France, juste à côté de la Normandie. Elles avaient été conquises il y a fort longtemps et rattachées à la Couronne britannique. Il n'en reste pas moins que leurs habitants, à cause de leur histoire et de leur situation géographique, étaient proches de leurs voisins français. Ils connaissaient leurs coutumes et surtout leur langue, ce qui leur a permis de tisser facilement des liens

avec les descendants des Français établis le long de la Gaspésie et aux alentours du golfe Saint-Laurent [71].

L'haplotype B016 de *dys44* a été observé dans toutes les populations gaspésiennes, sauf chez les Loyalistes, et cet haplotype est ubiquitaire en France. On pourrait alors émettre l'hypothèse que cet haplotype est arrivé lors des premiers peuplements gaspésien, et qu'il est la trace de l'installation française (acadienne et canadienne-française) en Gaspésie. La présence de cet haplotype chez les Anglo-normands et son absence chez les Loyalistes pourrait illustrer la proximité des habitants des îles anglo-normandes et de la France. Un échantillon britannique confirmerait cette hypothèse, dans le cas où B016 est absent ou faiblement répandu dans cet échantillon.

La proximité entre habitants des Îles anglo-normandes et de la France pourrait être appuyée par l'haplotype B070, uniquement observé chez les Anglo-normands et dans le Centre-ouest de la France (figures 17 et 18). Il est vrai que nous ne disposons pas d'échantillons des Îles anglo-normandes, ni d'échantillons britanniques, écossais ou irlandais. Ces résultats nous permettent seulement d'émettre des hypothèses qui pourront être vérifiées ultérieurement.

Étant donné le métissage connu des Anglo-normands avec les autres sous-populations gaspésiennes il est difficile de distinguer entre haplotypes originels et haplotypes issus du métissage gaspésien. Ce métissage est mis en valeur sur les graphiques de partage d'haplotypes DXS1238 et d'haplotypes-C, où un partage d'haplotype avec Montréal et avec les Canadiens français de Gaspésie est révélé, et également au niveau de l'apparement intergroupes des Anglo-normands (figures 11, 24 et 31). Il faut cependant ajouter que sur tout les graphiques de partage d'haplotypes, les Anglo-normands conservent leur proportion d'haplotypes uniques, illustrant leur singularité au même titre que les autres populations de notre étude.

3. Les Canadiens français de Gaspésie

Bien que les Canadiens français de Gaspésie se distinguent des autres sous-populations gaspésiennes seulement pour les analyses de distances de DXS1238, les graphiques de leurs distributions haplotypiques suivent le même modèle pour les

trois systèmes (figures 9, 17, 21 et 23). C'est à dire que, pour chacun des trois systèmes, les haplotypes les plus fréquents des Canadiens français de Gaspésie ne sont pas les mêmes que dans les autres populations.

Pour DXS1238, l'haplotype le plus fréquent est 5.1.0.0.19.2.18 ; chez les Canadiens français de Gaspésie cet haplotype est aussi fréquent que 5.2.0.0.18.2.22 (normalement à la quatrième place en terme de fréquence) et la première place est occupée par 5.1.0.0.15.2.15. La fréquence de 5.1.0.0.19.2.17 est la plus élevée dans cette sous-population gaspésienne. En ce qui concerne *dys44*, l'haplotype le plus fréquent est B003 chez les Canadiens français. Pour ce marqueur cette sous-population gaspésienne présente le plus petit nombre d'haplotypes (huit), mais les quatre premiers ont des fréquences relativement élevées. L'haplotype B005 présente également une fréquence élevée dans cette population par rapport aux autres populations. B003 et B005 présenteraient des effets fondateurs propres à cette sous-population gaspésienne.

Les particularités dénotées pour ces deux marqueurs se retrouvent au niveau des haplotypes-C avec les combinaisons 5.1.0.0.15.2.15-B005 et 5.2.0.0.18.2.22-B006 aussi fréquentes que la combinaison 5.1.0.0.19.2.18-B001, habituellement prépondérante. 5.1.0.0.19.2.17-B001 est spécifique au Québec et deux fois plus fréquent chez les Canadiens français que chez les Anglo-normands. Cet haplotype-C peut illustrer un effet fondateur canadien-français gaspésien et un métissage avec les Anglo-normands. Le métissage entre Canadiens français et Anglo-normands a déjà été démontré (figure 31).

Les graphiques de partage des haplotypes-C illustrent des liens singuliers entre Canadiens français de Gaspésie et autres populations québécoises. En effet d'après la figure 24a, les Canadiens français présentent 35% d'haplotypes-C partagés avec un autre échantillon québécois ; autrement dit toutes les populations partagent un haplotype-C avec les Canadiens français de Gaspésie. De plus le passage de la figure 24a à la figure 24b montre que 17% des haplotypes-C des Canadiens français sont partagés avec la France. Cette signature française illustre encore une fois le lien étroit entre le pool génique français et le pool génique québécois francophone.

Cette sous-population gaspésienne s'est souvent trouvée proche de trois régions françaises au cours des analyses de distances ou des partages d'haplotypes : le Centre-ouest de la France, la Loire Atlantique et le Sud-ouest de la France. Ces données correspondent à la provenances d'origine de plus de 50% des premiers colons installés en Nouvelle-France [2, 54].

Les trois systèmes de Xp21 ont ainsi révélé la singularité de la population canadienne-française de Gaspésie, à travers cet effet fondateur distinct. Un haplotype seul ne justifierait pas forcément l'emploi de la notion d'effet fondateur. Cependant les distributions des fréquences des trois systèmes sur Xp21 ont montré la même signature. De plus les analyses de distances de deux marqueurs sur trois ont révélé la singularité de cette population gaspésienne. L'effet fondateur de cette population n'est pas tant illustré par la nature différente de ses haplotypes. L'ensemble des haplotypes dont les fréquences ne sont pas identiques par rapport à la population mère, démontre un effet fondateur distinct dans cette population gaspésienne.

4. Les Loyalistes.

Les Loyalistes sont la seule population gaspésienne pour laquelle le test de Chakraborty est positif pour deux marqueurs (DXS1238 et haplotypes-C) sur trois. La singularité des Loyalistes est également suggérée à travers l'absence de l'haplotype 5.1.0.0.19.2.16, dont la fréquence atteint 5% pour le Québec au total. En effet cet haplotype DXS1238 est présent dans toutes les populations québécoises, y compris chez les Anglo-normands.

Ce test positif de Chakraborty n'est pas surprenant si l'on considère que cette sous-population gaspésienne regroupe en fait l'ensemble des habitants des Iles britanniques installés dans la péninsule. Nous savons que des Irlandais, des Ecossais et des Anglais se sont installés en Gaspésie. L'hétérogénéité de l'échantillon loyaliste pourrait refléter la diversité originelle du peuplement anglophone de la Gaspésie.

Du fait de l'ascendance européenne des Loyalistes, nous ne nous attendons pas à de grandes différences d'haplotypes entre le pool génique français et le pool génique loyaliste. Cette remarque est appuyée par le fait que la majorité des haplotypes retrouvés chez les Loyalistes est partagée par d'autres populations. Il serait

intéressant d'ajouter un échantillon britannique à cette étude dans le but d'observer la contribution de cette population au pool génique des Canadiens français du Québec, de même que pour des échantillons irlandais et écossais.

Du point de vue des relations entre populations, les graphiques de partage d'haplotypes ont révélé un partage d'haplotype entre Loyalistes et le Saguenay à plusieurs reprises (figures 11a et 18a). Ceci n'est pas si surprenant lorsque les Loyalistes de Gaspésie sont considérés comme un échantillon de population britannique, et compte tenu du fait que des officiers britanniques se sont installés au Charlevoix dès la fin du XVIIIe siècle, lors du passage de la colonie sous le Régime britannique. Dès lors se produit l'installation d'officiers anglais qui obtinrent des concessions dans la région de la Malbaie. La Malbaie est la paroisse qui fournit 40% des pionniers saguenayens avant 1844 [19]. Connaissant la contribution des premiers fondateurs au pool génique saguenayen, ces partages d'haplotypes pourraient refléter la contribution britannique au pool génique saguenayen.

5. Le Saguenay-Lac-Saint-Jean

Ce lien entre Loyalistes et Saguenay est une des particularités de cette dernière région. Les autres sont :

- la proportion élevée d'haplotypes uniques dans les graphiques de partage d'haplotypes des trois systèmes ;
- le fait que cette population se distingue de façon statistiquement significative de la population française totale, alors que cette distinction est parfois observée au niveau des ACPs et des MDS pour les autres populations, mais n'est généralement pas statistiquement significative ;
- la fréquence d'haplotypes relativement rares dans d'autres populations et plus fréquents dans cette région du Québec, à l'image de certaines mutations responsables de maladies génétiques [3, 4].

Ces haplotypes ont été mis en évidence pour les trois systèmes. Les deuxièmes et troisièmes haplotypes DXS1238, en terme de fréquence, sont 0.0.0.0.19.2.22 et 5.1.0.0.20.2.18. Une singularité québécoise est observée pour DXS1238, avec la

présence de 5.1.0.0.19.2.16. Cette image québécoise se retrouve également pour *dys44*, avec B001 et B003. Au Saguenay la fréquence cumulée de ces deux haplotypes s'élève à 61%. B011 illustre la singularité de l'effet fondateur saguenayen, car cet haplotype est deux fois plus fréquent au Saguenay. Cette remarque va de paire avec la fréquence élevée de l'haplotype-C 5.1.0.0.13.2.21-B011. En ce qui concerne les haplotypes-C la distinction du Saguenay est remarquable par les fréquences élevées de 0.0.0.0.19.2.22-B003 et 5.1.0.0.20.2.18-B001.

A travers l'hétérozygotie mesurée et le nombre d'haplotypes observés, nos trois systèmes présentent le Saguenay comme une population aussi diversifiée – voire plus diversifiée – que les autres populations québécoises échantillonnées. L'étude de la diversité de l'ADNmt a également illustré cette diversité saguenayenne par rapport aux autres populations québécoises (Moreau C., *et al* en préparation). L'effet fondateur au Saguenay est perçu par deux haplogroupes de l'ADNmt dont les fréquences sont accrues au Saguenay, alors que l'étude de la diversité du chromosome Y présente seulement un haplogroupe dont la fréquence est nettement supérieure par rapport aux autres populations.

A travers l'identification d'un fondateur pour une mutation à forte prévalence au Saguenay, l'homogénéité de cette région a souvent été suggérée [72-74]. Cette première hypothèse d'homogénéité était inférée à partir d'une population partageant des phénotypes communs (les symptômes d'une maladie génétique) et donc des caractéristiques génétiques communes propres à cette maladie. Lorsque d'autres maladies génétiques sont considérées, plusieurs mutations peuvent être identifiées, représentant cette fois-ci une population hétérogène de patients [3, 4]. De même, une étude démographique faite au Saguenay-Lac-Saint-Jean [14] n'a pas identifié de structure particulière dans cette région, cette conclusion étant relative au découpage du territoire que les auteurs ont étudié. On comprend alors que lorsque la diversité d'une population est étudiée à partir d'un seul marqueur génétique, le reflet de la diversité observée sera fonction du marqueur choisit et de la population ciblée, ou de l'échantillonnage réalisé.

Ainsi à travers un marqueur de polymorphismes neutres nous avons non seulement illustré la singularité du Saguenay-Lac-Saint-Jean, comparativement aux autres

régions québécoises telle que les études des maladies génétiques la révèlent, mais également que la diversité génétique du chromosome X au Saguenay est aussi élevée que dans les deux autres régions du Québec ou dans les régions françaises étudiées. Le fait que la population saguenayenne apparaisse significativement différente de la France pour nos trois marqueurs appuie la singularité de cette population par rapport aux deux autres populations québécoises. La singularité de cette région se comprend très bien lorsque l'on considère le triple effet fondateur qui a donné lieu à cette population [54]. Nous nous apercevons ici qu'un effet fondateur n'est pas forcément synonyme d'une diminution de la diversité génétique de marqueurs de polymorphismes neutres.

Tout d'abord dès 1635 et jusqu'à la fin du régime français, l'immigration française a mené à l'établissement de 8500 fondateurs dans la vallée du Saint-Laurent. Puis dès le XVIIIe siècle, des colons prennent pied sur les deux rives du fleuve. De nombreuses migrations régionales se sont produites tout au long de l'histoire du peuplement du Québec. Les migrations contribuent à l'apport de nouveaux allèles dans une population. Ce processus a ainsi pu augmenter la proportion de variants rares de cette région, affectant la signature de l'effet fondateur.

Un de ces courants migratoires est celui qui a traversé la Côte-de-Beaupré, donnant naissance à Charlevoix. Et en troisième lieu, à partir de 1838, un troisième courant d'émigration fut à l'origine de la région du Saguenay. Ce triple effet fondateur a pu être renforcé par l'installation de ces premiers fondateurs arrivés par familles. Cependant considérant les couples reproducteurs, et étant donné le nombre d'enfants par femme des premiers Canadiens français, on peut penser que la majorité des allèles de chacun des deux parents ont été transmis, et la diversité génétique des premiers fondateurs de la région a été conservée.

6. Montréal

Compte tenu de la diversité connue des grandes villes, et à l'image de notre échantillon de la France urbaine, on aurait pu s'attendre à ce que le test de Chakraborty se révèle significatif pour l'un de nos trois systèmes pour l'échantillon montréalais. Ceci n'a pas été le cas. Cette population n'occupe pas une position très

centrale sur nos analyses de distances. Cependant pour DXS1238 et *dys44*, Montréal est toujours proche des Canadiens français de Gaspésie. A deux périodes différentes de l'histoire du Québec, la région de la Gaspésie et la région de Montréal ont attiré des Canadiens français provenant d'autres régions québécoises. Montréal, de par ses attraits socio-économiques, connaît encore aujourd'hui une immigration de Canadien français provenant d'autres régions du Québec. Tandis que la Gaspésie est actuellement caractérisée par un vieillissement de la population [75]. Malgré ces différences, une structure génétique de population commune est partagée entre les Canadiens français de Gaspésie et ceux de Montréal. Cette structure génétique commune est illustrée par les distributions des fréquences des haplotypes, et également par les analyses de distances de ces deux marqueurs. Un partage d'haplotypes avec les Canadiens français de Gaspésie vient corroborer cette remarque.

Toutefois en ce qui concerne les partages d'haplotypes de DXS1238 et des haplotypes-C, la proportion d'haplotypes uniques est toujours équivalente à celle du Saguenay (alors qu'elle est faible pour les Canadiens français de Gaspésie). La distinction de cette population est non seulement attestée par des haplotypes plus fréquents à Montréal et moins fréquents dans les autres populations québécoises ou en France, illustrant un effet fondateur propre à cette région ; mais aussi par la distribution des haplotypes-C et les analyses de distances réalisées à partir de ces haplotypes-C.

A Montréal, l'effet fondateur est illustré à travers les fréquences plus élevées des haplotypes DXS1238 5.1.0.0.18.2.18, deuxième haplotypes le plus fréquent à Montréal alors qu'il est en cinquième place en France ; et 0.0.0.0.19.2.20. Pour ce qui est de *dys44*, B006 et B003 sont aussi fréquents que B001. En ce qui concerne les haplotypes-C : 5.2.0.0.18.2.22-B006 est le plus fréquent à Montréal alors qu'il est en troisième place en France, et les haplotypes-C 5.1.0.0.18.2.18-B001, ou encore 5.1.0.0.19.2.15-B003 sont plus fréquents à Montréal qu'en France. L'haplotype-C le plus fréquent, 5.1.0.0.19.2.18-B001, est relégué à la deuxième position pour Montréal. L'haplotype-C 0.0.0.0.19.2.21-B003 est ubiquitaire au Québec, sauf à Montréal. Le nombre d'haplotypes-C identifié pour la région de Montréal est

relativement bas, et une opposition entre Montréal et le Saguenay est mise en évidence sur les analyses de distances.

D'après nos trois systèmes, la région de Montréal est généralement plus éloignée des populations acadiennes, saguenayennes et loyalistes. Etant donné que les Acadiens sont arrivés en Gaspésie par familles et qu'ils sont connus pour leur forte identité culturelle, il est compréhensible que cette sous-population ne se rapproche pas autant de Montréal que les Canadiens français de Gaspésie. Pour les Acadiens du début du XVIIIe siècle, Montréal représente plus un lieu de passage vers des zones rurales plutôt qu'un lieu d'installation [16]. Entre 1912 et 1920 la paroisse de Verdun accueille de nombreux Acadiens venus de la Gaspésie, des Iles-de-la-Madeleine et des Maritimes [16]. L'effet fondateur acadien est marqué, mais n'est pas forcément évident si les allèles fondateurs sont des allèles très communs, comme le sont B001 et 5.1.0.0.19.2.18.

L'échantillon montréalais se distingue de l'échantillon saguenayen, malgré que les graphiques de partage d'haplotypes illustrent un lien entre ces populations. Il s'agit du même pool ancestral, mais l'évolution du pool génique du Saguenay est bien distincte, et cette différence a été accentuée par le triple effet fondateur saguenayen.

En 1860, la communauté anglophone de Montréal représente 50% de la population [16]. A cette période, une immigration ininterrompue d'anglophones d'Ontario et d'Angleterre alimente la diversité du Montréal protestant. Les Irlandais catholiques représentaient alors un quart de la population de Montréal, leur intégration étant relativement aisée en raison du partage de la langue avec les protestants et de la religion avec les Canadien français [16]. D'ailleurs plusieurs d'entre eux se sont mariés dans la communauté canadienne-française.

Le bilinguisme montréalais est connu depuis le début du XIXe siècle et constitue une particularité de la région de Montréal. Cette hétérogénéité a non seulement été mise en évidence par des études démographiques [13, 76], et également au niveau des maladies génétiques. En effet peu de maladies génétiques québécoises sont associées à cette région [3]. Ceci peut s'expliquer par le fait que la plupart des habitants de la région de Montréal présentant des maladies ont des ancêtres provenant d'autres régions. En effet au XIXe siècle, Montréal représente un pôle

d'attraction pour les ruraux [15]. Il s'agit du phénomène de transfert migratoire qui s'est produit lors du retour vers l'aire du peuplement ancien et les centres urbains en développement, au moment de l'industrialisation et de l'urbanisation [16].

La diversité génétique de l'échantillon de Montréal présente une image double. Cette dualité consiste en une structure de population avec les signes d'un effet fondateur similaire à celui des Canadiens français de Gaspésie, tout en conservant une proportion d'haplotypes uniques relativement élevée par rapport aux autres populations québécoises étudiées ici. Les trois systèmes du chromosome X illustrent la singularité de la région de Montréal, au même titre que les autres populations québécoises étudiées.

F. CONCLUSION

L'étude de la diversité génétique de trois régions du Québec à l'aide d'un marqueur de polymorphismes neutres, a révélé trois résultats importants :

- Malgré l'importance de l'effet fondateur au Québec, connu par les fréquences élevées de certaines maladies génétiques rares –ou absentes- dans les autres populations, la diversité génétique du Québec, et des trois régions étudiées ici, n'est pas inférieure à la diversité génétique de la France.
- L'histoire du peuplement de chaque région a son impact sur la diversité génétique observée à partir de nos marqueurs de polymorphismes neutres. La diversité relativement élevée du Saguenay-Lac-Saint-Jean vient appuyer la contribution de fondateurs d'origines diverses au pool génique de cette région. Une structure de population a été identifiée en Gaspésie, reflétant l'importance de la prise en considération de l'ascendance des quatre principaux groupes ethniques fondateurs de la péninsule. En dépit du pôle d'attraction que Montréal a pu, et peut encore, représenter, la diversité génétique des Canadiens français montréalais reflète la singularité de cette population au même titre que les deux autres régions étudiées ici.

- Il n'existe pas une seule population canadienne-française. La distribution des maladies génétiques et des mutations responsables de ces maladies, ainsi que les études démographiques viennent appuyer nos résultats quant à la subdivision régionale des populations canadiennes-françaises. Les variations de fréquences des haplotypes de nos trois systèmes démontrent l'existence d'effets fondateurs régionaux.

Les études de déséquilibre de liaison et de la structure des populations constituent les prémices des études des maladies complexes [15, 16, 28, 30]. Dans les régions québécoises étudiées ici, différentes structures de populations ont été révélées à l'aide de nos marqueurs de polymorphismes neutres. Les études d'association menées dans les populations québécoises devront tenir compte de ces différentes structures régionales.

BIBLIOGRAPHIE

1. Cauvier, C., *Etude comparative des caractéristiques généalogiques et génétiques de quatre populations fondatrices de la Gaspésie*. 2006, Université Laval / Université du Québec à Chicoutimi: Chicoutimi. p. 174.
2. Charbonneau, H., Desjardins B., Légaré J., Denis H., *Le peuplement des basses terres*, in *Atlas historique du Québec*, S. Courville, Editor. 1996, Les presses de l'Université Laval. p. 29-60.
3. Laberge, A.M., Michaud J., Richter A., Lemyre E., Lambert M., Brais B., Mitchell G. A., *Population history and its impact on medical genetics in Quebec*. Clin Genet, 2005. **68**(4): p. 287-301.
4. Sriver, C.R., *Human genetics: lessons from Quebec populations*. Annu Rev Genomics Hum Genet, 2001. **2**: p. 69-101.
5. Roddier, K., Thomas, T., Marleau, G., Gagnon, A. M., Dicaire, M. J., St-Denis, A., Gosselin, I., Sarrazin, A. M., Larbrisseau, A., Lambert, M., Vanasse, M., Gaudet, D., Rouleau, G. A., Brais, B., *Two mutations in the HSN2 gene explain the high prevalence of HSN2 in French Canadians*. Neurology, 2005. **64**(10): p. 1762-7.
6. De Braekeleer, M., P. Hechtman, E. Andermann, F. Kaplan,, *The French Canadian Tay-Sachs disease deletion mutation: identification of probable founders*. Hum Genet, 1992. **89**(1): p. 83-7.
7. Heyer, E., *Genetic consequences of differential demographic behaviour in the Saguenay region, Quebec*. Am J Phys Anthropol, 1995. **98**(1): p. 1-11.
8. Laberge, A.M., Jomphe M., Houde L., Vezina H., Tremblay M., Desjardins B., Labuda D., St-Hilaire M., Macmillan C., Shoubridge E. A., Brais B., *A "Fille du Roy" introduced the T14484C Leber hereditary optic neuropathy mutation in French Canadians*. Am J Hum Genet, 2005. **77**(2): p. 313-7.
9. Yotova, V., Labuda, D. Zietkiewicz, E., Gehl, D., Lovell, A., Lefebvre, J. F. Bourgeois, S., Lemieux-Blanchard, E., Labuda, M. Vezina, H., Houde, L., Tremblay, M., Toupance, B., Heyer, E., Hudson, T. J., Laberge, C., *Anatomy of a founder effect: myotonic dystrophy in Northeastern Quebec*. Hum Genet, 2005. **117**(2-3): p. 177-87.
10. Heyer, E. and M. Tremblay, *Variability of the genetic contribution of Quebec population founders associated to some deleterious genes*. Am J Hum Genet, 1995. **56**(4): p. 970-8.
11. Arcos-Burgos, M. and M. Muenke, *Genetics of population isolates*. Clin Genet, 2002. **61**(4): p. 233-47.
12. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C. M., Aulchenko, Y. S., *The effect of genetic drift in a young genetically isolated population*. Ann Hum Genet, 2005. **69**(Pt 3): p. 288-95.
13. Gagnon, A. and E. Heyer, *Fragmentation of the Quebec population genetic pool (Canada): evidence from the genetic contribution of founders per region in the 17th and 18th centuries*. Am J Phys Anthropol, 2001. **114**(1): p. 30-41.
14. Lavoie, E.M., Tremblay, M., Houde, L., Vezina, H., *Demogenetic study of three populations within a region with strong founder effects*. Community Genet, 2005. **8**(3): p. 152-60.
15. Robert, J.C., *Montréal: l'histoire*, in *Montréal 1642 - 1992*, Hurturbise, Editor. 1992. p. 11 - 60.

16. Olson, S., *Le peuplement de Montréal*, in *Atlas historique du Québec*, S. Courville, Editor. 1996, Les presses de l'Université Laval. p. 81-93.
17. De Braekeleer, M., *Geographic distribution of 18 autosomal recessive disorders in the French Canadian population of Saguenay-Lac-Saint-Jean, Quebec*. *Ann Hum Biol*, 1995. **22**(2): p. 111-22.
18. De Braekeleer, M., *Hereditary disorders in Saguenay-Lac-St-Jean (Quebec, Canada)*. *Hum Hered*, 1991. **41**(3): p. 141-6.
19. Roy, R., Bouchard G., Declos M., *La première génération de Saguenayens*, in *Histoire d'un génome, Population et génétique dans l'est du Québec*, G. Bouchard, De Braekeleer M., Editor. 1991, Presses de l'Université du Québec. p. 163-186.
20. Jetté, R., Gauvreau D., Guérin M., *Aux origines d'une région: le peuplement fondateur de Charlevoix avant 1850*, in *Histoire d'un génome, Population et génétique dans l'est du Québec*, G. Bouchard, De Braekeleer M., Editor. 1991, Presses de l'Université du Québec. p. 75-106.
21. Desjardins, M., *Une société de pêcheurs (1760 - 1870) ; Le territoire et ses occupants*, in *Histoire de la Gaspésie*, I.q.d.r.s.l. culture, Editor. 1999, Les presses de l'université Laval. p. 159 - 211.
22. Desjardins, M., Frenettes Y., *L'émergence de la Gaspésie; Les Amérindiens*, in *Histoire de la Gaspésie*, I.q.d.r.s.l. culture, Editor. 1999, Les presses de l'université Laval. p. 61 - 67.
23. Desjardins, M., Frenette Y., *La Gaspésie en transition (1870 - 1965) ; Extension du peuplement et migrations*, in *Histoire de la Gaspésie*, I.q.d.r.s.l. culture, Editor. 1999, Les presses de l'université Laval. p. 351 - 382.
24. Freedman, M.L., D. Reich, K.L. Penney, G.J. McDonald, A.A. Mignault, N. Patterson, S.B. Gabriel, E.J. Topol, J.W. Smoller, C.N. Pato, M.T. Pato, T.L. Petryshen, L.N. Kolonel, E.S. Lander, P. Sklar, B. Henderson, J.N. Hirschhorn, D. Altshuler., *Assessing the impact of population stratification on genetic association studies*. *Nat Genet*, 2004. **36**(4): p. 388-93.
25. Marchini, J., Cardon, L. R., Phillips, M. S. Donnelly, P., *The effects of human population structure on large genetic association studies*. *Nat Genet*, 2004. **36**(5): p. 512-7.
26. Helgason, A., Yngvadottir B., Hrafnkelsson B., Gulcher J., Stefansson K., *An Icelandic example of the impact of population structure on association studies*. *Nat Genet*, 2005. **37**(1): p. 90-5.
27. Zerjal, T., Beckman, L., Beckman, G., Mikelsaar, A. V. Krumina, A., Kucinskias, V., Hurles, M. E., Tyler-Smith, C., *Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations*. *Mol Biol Evol*, 2001. **18**(6): p. 1077-87.
28. Jorde, L.B., Watkins W. S., Bamshad M. J., Dixon M. E., Ricker C. E., Seielstad M. T., Batzer M. A., *The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data*. *Am J Hum Genet*, 2000. **66**(3): p. 979-88.
29. Helgason, A., Nicholson G., Stefansson K., Donnelly P., *A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift*. *Ann Hum Genet*, 2003. **67**(Pt 4): p. 281-97.
30. Rund, D., Filon, D., Jackson, N., Asher, N., Oron-Karni, V., Sacha, T., Czekalska, S., Oppenheim, A., *An unexpectedly high frequency of*

- heterozygosity for alpha-thalassemia in Ashkenazi Jews*. Blood Cells Mol Dis, 2004. **33**(1): p. 1-3.
31. Sajantila, A., Salem, A. H., Savolainen, P., Bauer, K., Gierig, C., Paabo, S., *Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population*. Proc Natl Acad Sci U S A, 1996. **93**(21): p. 12035-9.
 32. de la Chapelle, A. and F.A. Wright, *Linkage disequilibrium mapping in isolated populations: the example of Finland revisited*. Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12416-23.
 33. Perez-Miranda, A.M., Alfonso-Sanchez M. A., Kalantar A., Garcia-Obregon S., de Pancorbo M. M., Pena J. A., Herrera R. J., *Microsatellite data support subpopulation structuring among Basques*. J Hum Genet, 2005. **50**(8): p. 403-14.
 34. Caglia, A., S. Tofanelli, V. Coia, I. Boschi, M. Pescarmona, G. Spedini, V. Pascali, G. Paoli, G. Destro-Bisoli, *A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F(ST) and R(ST) genetic distances*. Hum Biol, 2003. **75**(3): p. 313-30.
 35. Kashyap, V.K., Guha S., Sitalaximi T., Bindu G. H., Hasnain S. E., Trivedi R., *Genetic structure of Indian populations based on fifteen autosomal microsatellite loci*. BMC Genet, 2006. **7**: p. 28.
 36. Eichler, E.E., J.N. Macpherson, A. Murray, P.A. Jacobs, A. Chakravarti, D.L. Nelson, *Haplotype and interspersed analysis of the FMRI CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome*. Hum Mol Genet, 1996. **5**(3): p. 319-30.
 37. Liquori, C.L., Ikeda, Y., Weatherspoon, M., Ricker, K., Schoser, B. G., Dalton, J. C., Day, J. W., Ranum, L. P., *Myotonic dystrophy type 2: human founder haplotype and evolutionary conservation of the repeat tract*. Am J Hum Genet, 2003. **73**(4): p. 849-62.
 38. Varilo, T., Laan, M., Hovatta, I., Wiebe, V., Terwilliger, J. D., Peltonen, L., *Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo*. Eur J Hum Genet, 2000. **8**(8): p. 604-12.
 39. Hadj-Rabia, S., Salomon R., Pelet A., Penet C., Rotschild A., de Laet M. H., Chaouachi B., Hannachi R., Bakiri F., Brauner R., Chaussain J. L., Munnich A., Lyonnet S., *Linkage disequilibrium in inbred North African families allows fine genetic and physical mapping of triple A syndrome*. Eur J Hum Genet, 2000. **8**(8): p. 613-20.
 40. Zhivotovsky, L.A., *Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow*. Mol Biol Evol, 2001. **18**(5): p. 700-9.
 41. Hancock, J.M., *Microsatellites and other simple sequences: genomic context and mutational mechanisms*, in *Microsatellites, Evolution and Applications*, D.B. Goldstein, Schlötterer C., Editor. 1999, Oxford University Press. p. 1-9.
 42. Labuda, D., E. Zietkiewicz, and V. Yotova, *Archaic lineages in the history of modern humans*. Genetics, 2000. **156**(2): p. 799-808.
 43. Zietkiewicz, E., Yotova, V., Gehl, D., Wambach, T., Arrieta, I., Batzer, M., Cole, D. E., Hechtman, P., Kaplan, F., Modiano, D., Moisan, J. P., Michalski, R., Labuda, D., *Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity*. Am J Hum Genet, 2003. **73**(5): p. 994-1015.
 44. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T., *Detection of polymorphisms of human DNA by gel electrophoresis as single-strand*

- conformation polymorphisms*. Proc Natl Acad Sci U S A, 1989. **86**(8): p. 2766-70.
45. Zietkiewicz, E., Sinnett, D., Richer, C., Mitchell, G., Vanasse, M., Labuda, D., *Single-strand conformational polymorphisms (SSCP): detection of useful polymorphisms at the dystrophin locus*. Hum Genet, 1992. **89**(4): p. 453-6.
 46. Zietkiewicz, E., N. Akalin, and D. Labuda, *Neutral polymorphisms in the deletion-prone regions of the dystrophin gene*. Hum Hered, 1995. **45**(2): p. 80-3.
 47. Zietkiewicz, E., Yotova, V., Jarnik, M., Korab-Laskowska, M., Kidd, K. K., Modiano, D., Scozzari, R., Stoneking, M., Tishkoff, S., Batzer, M., Labuda, D., *Nuclear DNA diversity in worldwide distributed human populations*. Gene, 1997. **205**(1-2): p. 161-71.
 48. Zietkiewicz, E., Yotova, V., Jarnik, M., Korab-Laskowska, M., Kidd, K. K., Modiano, D., Scozzari, R., Stoneking, M., Tishkoff, S., Batzer, M., Labuda, D., *Genetic structure of the ancestral population of modern humans*. J Mol Evol, 1998. **47**(2): p. 146-55.
 49. Bourgeois, S., D. Labuda,, *Dynamic allele-specific oligonucleotide hybridization on solid support*. Anal Biochem, 2004. **324**(2): p. 309-11.
 50. Chakraborty, R., *Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations*. Am J Hum Genet, 1990. **47**(1): p. 87-94.
 51. Slatkin, M., *A measure of population subdivision based on microsatellite allele frequencies*. Genetics, 1995. **139**(1): p. 457-62.
 52. Ewens, W.J., *The sampling theory of selectively neutral alleles*. Theor Popul Biol, 1972. **3**(1): p. 87-112.
 53. Chakraborty, R. and K.M. Weiss, *Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium*. Am J Phys Anthropol, 1991. **86**(4): p. 497-506.
 54. Bouchard, G., De Braekeleer M., *Mouvements migratoires, effets fondateurs et homogénéisation génétique*, in *Histoire d'un génome, Population et génétique dans l'est du Québec*, G. Bouchard, De Braekeleer M., Editor. 1991, Presses de l'Université du Québec. p. 75-106.
 55. Helgason, A., Sigurardottir S., Gulcher J.R., Stefansson K., Ward R., *Sampling saturation and the European mtDNA pool: implications for detecting genetic relationships among populations*. p. 285-294.
 56. Moslehi, R., Chu, W., Karlan, B., Fishman, D., Risch, H., Fields, A., Smotkin, D., Ben-David, Y., Rosenblatt, J., Russo, D., Schwartz, P., Tung, N., Warner, E., Rosen, B., Friedman, J., Brunet, J. S., Narod, S. A., *BRCA1 and BRCA2 mutation analysis of 208 Ashkenazi Jewish women with ovarian cancer*. Am J Hum Genet, 2000. **66**(4): p. 1259-72.
 57. Varilo, T., Paunio T., Parker A., Perola M., Meyer J., Terwilliger J. D., Peltonen L., *The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories*. Hum Mol Genet, 2003. **12**(1): p. 51-9.
 58. Behar, D.M., D. Garrigan, M.E. Kaplan, Z. Mobasher, D. Rosengarten, T.M. Karafet, L. Quintana-Murci, H. Ostrer, K. Skorecki, M.F. Hammer,, *Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations*. Hum Genet, 2004. **114**(4): p. 354-65.
 59. Charbonneau, H., Desjardins B., Légaré J., Denis H., *The population of the St-Lawrence Valley, 1608-1760.*, in *A population history of North America*,

- H.R. Haines, Steckel R.H. et al., Editor. 2000, Cambridge University Press: Cambridge. p. 99-142.
60. Bouchard, G., *Pour une approche historique et sociale du génome québécois*, in *Histoire d'un génome, Population et génétique dans l'est du Québec*, G. Bouchard, De Braekeleer M., Editor. 1991, Presses de l'Université du Québec. p. 3-20.
 61. Austerlitz, F. and E. Heyer, *Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population*. Proc Natl Acad Sci U S A, 1998. **95**(25): p. 15140-4.
 62. De Braekeleer, M. and S. Gauthier, *Autosomal recessive disorders in Saguenay-Lac-Saint-Jean (Quebec, Canada): a study of inbreeding*. Ann Hum Genet, 1996. **60**(Pt 1): p. 51-6.
 63. Labuda, D., E. Zietkiewicz, and M. Labuda, *The genetic clock and the age of the founder effect in growing populations: a lesson from French Canadians and Ashkenazim*. Am J Hum Genet, 1997. **61**(3): p. 768-71.
 64. Segel, R., Silverstein, S. Lerer, I., Kahana, E. Meir, R., Sagi, M., Zilber, N., Korczyn, A. D. Shapira, Y., Argov, Z., Abeliovich, D., *Prevalence of myotonic dystrophy in Israeli Jewish communities: inter-community variation and founder premutations*. Am J Med Genet A, 2003. **119**(3): p. 273-8.
 65. Alfonso-Sanchez, M.A., Pena, J. A., Aresti U., Calderon R., *An insight into recent consanguinity within the Basque area in Spain. Effects of autochthony, industrialization and demographic changes*. Ann Hum Biol, 2001. **28**(5): p. 505-21.
 66. Alfonso-Sanchez, M.A., U. Aresti, J.A. Pena, R. Calderon, *Inbreeding levels and consanguinity structure in the Basque province of Guipuzcoa (1862-1980)*. Am J Phys Anthropol, 2005. **127**(2): p. 240-52.
 67. Iriondo, M., M.C. Barbero, and C. Manzano, *DNA polymorphisms detect ancient barriers to gene flow in Basques*. Am J Phys Anthropol, 2003. **122**(1): p. 73-84.
 68. Eaves, I.A., T.R. Merriman, R.A. Barber, S. Nutland, E. Tuomilehto-Wolf, J. Tuomilehto, F. Cucca, J.A. Todd,, *The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes*. Nat Genet, 2000. **25**(3): p. 320-3.
 69. Fraumene, C., E. Petretto, A. Angius, M. Pirastu,, *Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis*. Hum Genet, 2003. **114**(1): p. 1-10.
 70. Angius, A., D. Bebbere, E. Petretto, M. Falchi, P. Forabosco, B. Maestrale, G. Casu, I. Persico, P.M. Melis, M. Pirastu, *Not all isolates are equal: linkage disequilibrium analysis on Xq13.3 reveals different patterns in Sardinian sub-populations*. Hum Genet, 2002. **111**(1): p. 9-15.
 71. Desjardins, M., *Une société de pêcheurs (1760 - 1870); L'économie de la morue séchée*, in *Histoire de la Gaspésie*, I.q.d.r.s.l. culture, Editor. 1999, Les presses de l'université Laval. p. 219 - 295.
 72. Gagne, C., L. D. Brun, P. Julien, S. Moorjani, P. J. Lupien,, *Primary lipoprotein-lipase-activity deficiency: clinical investigation of a French Canadian population*. Cmaj, 1989. **140**(4): p. 405-11.
 73. Mathieu, J., De Braekeleer M., Prevost C., *Genealogical reconstruction of myotonic dystrophy in the Saguenay-Lac-Saint-Jean area (Quebec, Canada)*. Neurology, 1990. **40**(5): p. 839-842.

74. Pausova, Z., Jomphe M., Houde L., Vezina H., Orlov S.N., Gossard F., Gaudet D., Tremblay J., Kotchen T.A., Cowley A.W., Bouchard G., Hamet P., *A genealogical study of essential hypertension with and without obesity in French Canadians*. *Obes Res*, 2002. **10**(6): p. 463-470.
75. Mimeault, M., *Les régions du Québec - histoire en bref: La Gaspésie*, ed. I.q.d.r.s.l. culture. 2004: Les presses de l'université Laval.
76. Gagnon, A. and B. Toupance, *Testing isonymy with paternal and maternal lineages in the early Quebec population: the impact of polyphyly and demographic differentials*. *Am J Phys Anthropol*, 2002. **117**(4): p. 334-41.