

UNIVERSITÉ DU QUÉBEC À CHICOUTIMI

Mémoire présenté à
l'Université Laval
comme exigence partielle
de la maîtrise en médecine expérimentale
volet génétique offerte à
l'Université du Québec à Chicoutimi
en vertu d'un protocole d'entente avec l'Université Laval

par
JULIE ARSENAULT

**RÉGIONALISATION DU POOL GÉNIQUE QUÉBÉCOIS:
ANALYSE DU DEVENIR DES GÈNES FONDATEURS**

JUIN 1999



Mise en garde/Advice

Afin de rendre accessible au plus grand nombre le résultat des travaux de recherche menés par ses étudiants gradués et dans l'esprit des règles qui régissent le dépôt et la diffusion des mémoires et thèses produits dans cette Institution, **l'Université du Québec à Chicoutimi (UQAC)** est fière de rendre accessible une version complète et gratuite de cette œuvre.

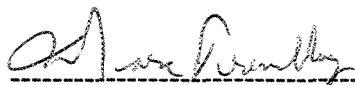
Motivated by a desire to make the results of its graduate students' research accessible to all, and in accordance with the rules governing the acceptance and diffusion of dissertations and theses in this Institution, the **Université du Québec à Chicoutimi (UQAC)** is proud to make a complete version of this work available at no cost to the reader.

L'auteur conserve néanmoins la propriété du droit d'auteur qui protège ce mémoire ou cette thèse. Ni le mémoire ou la thèse ni des extraits substantiels de ceux-ci ne peuvent être imprimés ou autrement reproduits sans son autorisation.

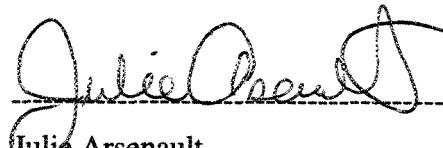
The author retains ownership of the copyright of this dissertation or thesis. Neither the dissertation or thesis, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

RÉSUMÉ

Plusieurs maladies héréditaires ont des incidences anormalement élevées dans certaines régions du Québec. L'objectif de cette étude est de vérifier la présence de différences dans l'évolution des gènes fondateurs de cinq populations du Québec: Beauce, Charlevoix, Rimouski, Saguenay et Terrebonne. À partir d'un corpus de 378 généralogies ascendantes réalisé parmi ces cinq régions, une première série de 50 000 simulations fut effectuée en utilisant la méthode du Gene-Dropping. Les résultats révèlent que les populations de Charlevoix et du Saguenay présentent les fondateurs ayant les meilleures probabilités d'atteindre différentes fréquences de porteurs parmi la population contemporaine. Les distributions de probabilités permettent de constater une évolution semblable des gènes fondateurs pour Charlevoix et le Saguenay mais distincte pour les autres populations, ceci étant attribuable à des comportements socio-reproducteurs différents.



Marc Tremblay



Julie Arsenault

AVANT-PROPOS

La réalisation de ce mémoire n'aurait pas été possible sans l'intervention de nombreuses personnes qui, de près ou de loin, par des petites ou des grandes contributions, m'ont permis d'atteindre enfin le fil d'arrivée. Je profite de cette occasion pour les remercier et leur offrir toute ma reconnaissance.

Pour leur support moral et leurs nombreux encouragements, je veux remercier ma famille et mes amis. Ils n'ont jamais cessé de croire en moi.

Pour sa patience, sa grande disponibilité et ses nombreux conseils, je tiens à remercier mon directeur Marc Tremblay. Pour toutes les connaissances et l'enthousiasme qu'elle m'a transmis ainsi que sa générosité lors de mon passage à son Laboratoire, je tiens à remercier ma co-directrice Évelyne Heyer.

Je conserve un merveilleux souvenir de toutes les personnes du projet BALSAC que j'ai eu la chance de côtoyer. Merci à Sophie Claveau, informaticienne-analyste, qui m'a accordé beaucoup de temps et de support dans de nombreuses situations. Merci également à ma professeure et conseillère, Hélène Vézina, pour les encouragements et les discussions qui m'ont aidé à voir plus clair dans mes nombreux questionnements.

Je remercie finalement l'IREP pour la bourse d'excellence qu'il m'a décernée et qui m'a beaucoup aidée, le FCAR qui m'a permis d'effectuer un stage de trois mois au Laboratoire anthropologique du Musée de L'Homme à Paris ainsi que la Fondation Asselin pour l'aide financière reçue. Merci à tous ces organismes qui m'ont appuyée et fourni le financement nécessaire afin de persévéérer.

TABLE DES MATIÈRES

| | |
|---|-----|
| RÉSUMÉ | II |
| AVANT-PROPOS | III |
| TABLE DES MATIÈRES | IV |
| LISTE DES TABLEAUX | VI |
| LISTE DES CARTES, FIGURES ET GRAPHIQUES | VII |
| INTRODUCTION | 1 |
| | |
| CHAPITRE 1: Problématique et revue de littérature | 3 |
| 1.1 Le Nord-Est québécois | 3 |
| 1.2 Choix des populations du Québec | 4 |
| 1.3 Chronologie du développement et de l'application de la méthode du Gene-Dropping | 5 |
| | |
| CHAPITRE 2: Aperçu historique des populations à l'étude | 9 |
| 2.1 Colonisation française | 9 |
| 2.2 Charlevoix | 11 |
| 2.3 Saguenay | 14 |
| 2.4 Rimouski | 15 |
| 2.5 Terrebonne | 16 |
| 2.6 Beauce | 17 |
| | |
| CHAPITRE 3: Sources des données et méthodologie | 20 |
| 3.1 Fichier-réseau BALSAC | 20 |
| 3.2 Fichier généalogique BALSAC-RETRO | 21 |
| 3.3 Obtention des données | 22 |
| 3.4 Reconstructions généalogiques | 23 |
| 3.5 Description du Gene-Dropping | 24 |
| 3.6 Gene-Dropping et Peeling | 27 |
| 3.7 Étapes de programmation | 28 |
| 3.8 Calculs des fréquences et des probabilités | 29 |

| | | |
|----------------------------|---|----|
| CHAPITRE 4: | Résultats | 30 |
| 4.1 | Complétude et profondeur des généalogies | 30 |
| 4.2 | Présentation des groupes de fondateurs de chaque population | 32 |
| 4.3 | Évaluation du nombre de fondateurs distincts communs à chaque paire de populations | 33 |
| 4.4 | Probabilités d'atteindre une fréquence de porteurs supérieure ou égale à 0,05 | 35 |
| 4.5 | Probabilités d'atteindre une fréquence de porteurs se situant entre 0,01 et 0,05 | 39 |
| 4.6 | Proportion des fondateurs de chaque population atteignant une fréquence de porteurs se situant entre 0,01 et 0,05, selon les probabilités | 44 |
| 4.7 | Probabilités d'atteindre une fréquence de porteurs se situant entre 0,035 et 0,05 | 46 |
| 4.8 | Comparaison des deux groupes d'individus pour lesquels les simulations ont été effectuées | 49 |
| 4.9 | Probabilités, pour les immigrants, d'atteindre une fréquence de porteurs située entre 0,035 et 0,05, dans chaque population | 50 |
| CONCLUSION GÉNÉRALE | | 54 |
| RÉFÉRENCES | | 56 |
| ANNEXE | | 61 |

LISTE DES TABLEAUX

Tableau

| | | |
|------|---|----|
| 3.1 | Caractéristiques des données sur les sujets par région | 22 |
| 4.1 | Indice de complétude à chaque génération pour l'ensemble des généalogies des cinq populations à l'étude et profondeur moyenne des généalogies | 31 |
| 4.2 | Nombre de fondateurs distincts obtenus dans chaque population | 33 |
| 4.3 | Proportion des fondateurs de chaque population qui sont partagés avec chacune des quatre autres populations | 34 |
| 4.4 | Les 20 probabilités les plus élevées d'atteindre une fréquence de porteurs égale ou supérieure à 0,05, dans chaque population | 36 |
| 4.5 | Caractéristiques des six fondateurs retrouvés parmi les 20 premiers de deux populations ou plus (fréquence de porteurs supérieure ou égale à 0,05) | 38 |
| 4.6 | Les 20 probabilités les plus élevées d'atteindre une fréquence de porteurs se situant entre 0,01 et 0,05, dans chaque population | 40 |
| 4.7 | Caractéristiques des huit fondateurs retrouvés parmi les 20 premiers de deux populations ou plus (fréquence de porteurs située entre 0,01 et 0,05) | 43 |
| 4.8 | Les 20 probabilités les plus élevées d'atteindre une fréquence de porteurs se situant entre 0,035 et 0,05, dans chaque population | 47 |
| 4.9 | Description des deux groupes d'individus | 49 |
| 4.10 | Comparaison des probabilités, pour les 20 plus importants fondateurs et immigrants de chaque population, d'atteindre une fréquence de porteurs se situant entre 0,035 et 0,05 | 51 |

LISTE DES CARTES, FIGURES ET GRAPHIQUES

| | | |
|------------------|--|----|
| Carte | | |
| 1 | Localisation géographique des cinq populations à l'étude | 12 |
| Figure | | |
| 3.1 | Généalogie avec deux allèles uniques assignés à chaque fondateur | 25 |
| 3.2 | Généalogie après le Gene-Dropping avec les génotypes assignés aux descendants par la ségrégation mendéienne des allèles des fondateurs | 26 |
| Graphique | | |
| 4.1 | Distribution des 20 plus importants fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs supérieure ou égale à 0,05 | 37 |
| 4.2 | Distribution des 20 plus importants fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,01 et 0,05 | 42 |
| 4.3 | Classement des fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,01 et 0,05 | 45 |
| 4.4 | Distribution des 20 plus importants fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,035 et 0,05 | 48 |
| 4.5 | Comparaison des distributions des 20 plus importants a)fondateurs et b)immigrants de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,035 et 0,05 | 53 |

INTRODUCTION

Les études touchant au domaine de la génétique des populations visent à déterminer l'ampleur de la variabilité génétique dans les populations naturelles et à en expliquer l'origine, le maintien et l'importance évolutive (Hartl, 1994). Dans cette étude sur la régionalisation du pool génique québécois, nous cherchons à mieux comprendre certains traits de l'évolution de la population québécoise.

Pour arriver à une bonne compréhension de l'évolution d'une population, on doit être conscient de la dynamique démographique qui existe au sein de cette population. On doit tenir compte de l'ensemble des comportements démographiques qui ont permis la formation et la reproduction d'une population. Ces comportements déterminent l'évolution de la population et de ses traits génétiques. Ainsi, dès qu'une mutation se produit sur un gène quelconque, entraînant un mauvais fonctionnement de ce gène et un risque de maladie, la diffusion ou la disparition de ce gène dans la population est dictée par les comportements socio-reproducteurs des personnes qui portent ces gènes mutés. Même si de nouvelles mutations sont peu fréquentes, tous les individus sont porteurs de gènes anormaux parmi tous les gènes présents dans leurs cellules (entre 4 et 5 gènes récessifs létaux) (Vogel & Motulsky, 1986).

En termes généraux, un gène est l'entité physique transmise par les parents aux enfants, lors du processus de reproduction, chaque individu étant diploïde. Pour chacun de ses gènes homologues, l'enfant reçoit une copie de son père et une copie de sa mère. Dans cette étude, nous employons la méthode Monte-Carlo (Gene-Dropping) pour simuler la transmission des gènes, en supposant que les gènes de certains fondateurs sont à l'origine de l'introduction des maladies héréditaires dans la population. Il s'agit de suivre l'évolution de la fréquence des différents allèles des fondateurs, de façon à obtenir une distribution des contributions génétiques des fondateurs à leurs descendants.

Cette méthode sera appliquée aux fondateurs du 17e siècle qui sont à l'origine de la population québécoise actuelle et plus particulièrement de celles des régions de Rimouski, Charlevoix, Terrebonne, Beauce et Saguenay. Les comparaisons régionales permettent de faire ressortir un patron d'évolution des gènes qui pourra servir, entre autres, à l'analyse des fréquences et des distributions des maladies héréditaires dans ces cinq populations du Québec. Puisque l'objectif premier de ce mémoire est de comprendre la dynamique de certaines

populations québécoises du point de vue génétique, des généalogies ascendantes sont utilisées comme matériel d'analyse.

En premier lieu, nous aborderons le problème des fréquences élevées de certaines maladies héréditaires dans l'est du Québec. Nous verrons ensuite le développement de la méthode de simulation employée et la justification du choix des cinq populations à l'étude, pour lesquelles nous présenterons un aperçu historique. Les sources de données et la méthodologie utilisées feront l'objet du 3e chapitre. Enfin, le dernier chapitre fera état des résultats touchant les simulations de transmission des gènes dans chacune des populations.

CHAPITRE 1

PROBLÉMATIQUE ET REVUE DE LITTÉRATURE

1.1 Le Nord-Est québécois

De nombreuses et diverses études ont démontré que certaines maladies héréditaires ont des incidences anormalement élevées dans certaines régions. Nous n'avons qu'à penser à la tyrosinémie et au rachitisme vitamino-dépendant au Saguenay (Bouchard et alii, 1984). Une fréquence élevée de porteurs de gènes délétères peut être expliquée par un effet fondateur (Bouchard & De Braekeleer, 1990). L'effet fondateur se produit lors de la création d'une nouvelle population à partir d'un nombre relativement restreint d'individus appartenant à une population mère (Mayr, 1995). En plus de comprendre l'échantillonnage de départ lors de la création de la population, l'effet fondateur inclut également les phénomènes stochastiques de changement de fréquences de gènes qui s'ensuivent. C'est à un effet fondateur qu'on attribue les fréquences élevées de tyrosinémie et de rachitisme vitamino-dépendant au Saguenay (De Braekeleer & Laroche, 1990; 1991). Le phénomène de dérive génétique peut aussi avoir une incidence sur les fréquences élevées de porteurs de gènes délétères (Bouchard & De Braekeleer, 1992). La dérive génétique est l'ensemble des changements aléatoires des fréquences géniques d'une génération à la suivante (Mayr, 1995).

Le Nord-Est québécois a été jusqu'à maintenant la cible de plusieurs travaux de recherche dans les domaines de la génétique, de la démographie et de l'histoire (Bouchard & De Braekeleer, 1990; Bouchard et alii, 1988; Roy et alii, 1988). La raison principale est qu'on a pu observer dans cette partie du Québec, plusieurs maladies génétiques relativement rares ou même tout à fait inconnues dans les autres populations européennes. Ceci est attribuable aux conditions dans lesquelles se sont développées ces régions mais tout d'abord à l'établissement

des premiers ancêtres porteurs de gènes délétères venus s'établir aux 17e et 18e siècles dans ces régions. Aujourd'hui, on peut supposer que les mêmes mutations sont présentes dans l'ouest du Québec suite aux mouvements migratoires.

À la lumière de ces informations, nous désirons vérifier l'existence de différences régionales en comparant l'évolution des gènes fondateurs dans cinq populations du Québec. L'hypothèse est que l'évolution des gènes fondateurs s'est effectuée très différemment d'une population à l'autre. Cependant, certaines populations peuvent présenter des parcours similaires si l'une d'elles est issue d'une autre.

1.2 Choix des populations du Québec

Il est certain qu'une représentation de toutes les régions du Québec aurait été souhaitable pour faire une analyse complète du pool génique québécois. Cependant, comme on l'a vu précédemment, cette analyse repose sur des reconstructions généalogiques et nous ne disposons pas, lorsque cette étude a été entreprise, d'un nombre suffisant de reconstructions pour chacune des régions du Québec. Nous avons donc arrêté à cinq le nombre de populations québécoises à l'étude. Ces populations ont été retenues tantôt pour des raisons de disponibilité de données, tantôt pour des raisons d'intérêt. Les populations choisies sont celles des régions du Saguenay et de Charlevoix et des comtés de Rimouski, de la Beauce et de Terrebonne.

Tout d'abord, la population du Saguenay fut retenue pour diverses raisons. Cette population a fait l'objet de plusieurs enquêtes depuis plus de 20 ans et par conséquent, une grande quantité d'informations était disponible. Par ces enquêtes, on sait maintenant que cette population présente de fortes incidences pour certaines génopathies, dont quelques-unes sont exclusives à cette région alors que d'autres, fréquentes dans les populations européennes en sont absentes (Bouchard & De Braekeleer, 1990). Récemment, la population du Saguenay a fait l'objet d'une étude visant à vérifier si les paramètres démographiques d'une population créée à partir d'un effet fondateur sont suffisants pour expliquer qu'un gène létal introduit par seulement un fondateur puisse atteindre une fréquence élevée dans la population contemporaine (Heyer, 1999). Avec tous les renseignements maintenant disponibles, la population du Saguenay est tout indiquée pour servir de point de référence.

Pour des raisons de disponibilité de données dans le fichier BALSAC-RETRO (voir section 3.2), les populations des comtés de Rimouski et de la Beauce ainsi que de la région de

Charlevoix furent choisies. Les généalogies de ces comtés et de cette région ont été utilisées dans d'autres projets de recherche s'inscrivant dans le même cadre que celui de la présente étude (Bergeron, en préparation; Gagnon, 1998). Les données sur Rimouski et la Beauce pourront donner une idée de l'évolution des gènes des fondateurs sur la rive sud du Saint-Laurent et permettront de vérifier si ces deux comtés partagent plusieurs fondateurs.

Dans le cas de Charlevoix, il y a plusieurs raisons de s'intéresser au mode de formation de cette population. Premièrement, il a été démontré que la population de cette région fut à l'origine du peuplement initial du Saguenay au milieu du 19e siècle (Pouyez et alii, 1983). De plus, il existe des maladies héréditaires communes et spécifiques aux deux régions (Bouchard & De Braekeleer, 1990). On s'attend donc à obtenir des probabilités plus élevées pour les fondateurs de Charlevoix que pour ceux du Saguenay d'atteindre certaines fréquences de porteurs puisque le Saguenay a également bénéficié d'un apport génétique extérieur à Charlevoix.

Afin d'avoir une représentation spatiale plus diversifiée et ainsi recouvrir quelque peu l'ouest du Québec, une cinquième population fut retenue, celle du comté de Terrebonne. Même si Terrebonne est la seule représentante de ce qu'on peut qualifier d'ouest québécois, nous allons tenter de comparer l'évolution des fréquences des gènes de ses fondateurs avec celle des fondateurs des régions de l'est.

Il est important de souligner que c'est le lieu de mariage ou le lieu de naissance des individus contemporains (points de départ des généalogies) qui déterminent la région d'appartenance. Dès la génération précédente, les individus apparaissant dans les généalogies peuvent provenir d'autres régions du Québec ou même de l'extérieur. Un fait persiste cependant, les quelques 5 000 Français établis au Canada avant 1700 sont à l'origine des deux tiers de la population francophone du Québec (Charbonneau et alii, 1987), population qui atteint 5 700 150 francophones en 1996 (Statistique Canada).

1.3 Chronologie du développement et de l'application de la méthode du Gene-Dropping

La méthode du Gene-Dropping est une méthode de Monte-Carlo et c'est elle qui fut retenue pour cette étude. L'application de la méthode du Gene-Dropping est le résultat de l'implication de plusieurs chercheurs dont les travaux ont eu une incidence notable. Le Gene-

Dropping s'est véritablement développé vers la fin des années soixante avec la présentation de nombreux congrès et rencontres au cours desquels des questions sur les généalogies et l'informatique étaient abordées.

De toute évidence, c'est avec le développement de l'ordinateur électronique qu'on est arrivé à faire des calculs sur des généalogies complexes au sein des populations humaines. Une des premières utilisations de l'ordinateur en génétique fut la résolution d'équations différentielles. Rapidement, des ordinateurs programmés selon les mécanismes de transmission mendélienne suivirent (Edwards, 1987).

Même si les premiers ouvrages sur les simulations informatiques en génétique n'en font pas référence, c'est à Cotterman (1940; 1941) que l'on doit le concept de l'identité par ascendance des gènes qui permet au coefficient de consanguinité d'être interprété comme une probabilité (Edwards, 1987). Ainsi, deux allèles présents à un locus sur deux chromosomes homologues peuvent être identiques par ascendance s'ils proviennent tous les deux de la réplication d'un seul allèle d'une population ancêtre (Hartl, 1994). Cotterman a également énuméré les douze états d'identité des gènes pour deux individus qui se présentent lorsqu'on considère le nombre de chemins dans lesquels les gènes maternels et paternels d'un seul individu et les gènes d'un autre individu peuvent être identiques par ascendance (Cotterman, 1940; 1941). Il a été démontré plus tard qu'une énumération distinguant aussi bien les origines parentales des gènes du second individu était plus appropriée et qu'elle conduisait à quinze états d'identité des gènes (Edwards, 1987). Il existe une puissante interaction en génétique humaine entre les simulations informatiques et l'identité des gènes par ascendance et ce, même si elles se sont développées dans des champs séparés.

C'est d'abord dans le cadre d'un atelier sur l'utilisation de l'ordinateur dans la génétique humaine tenu à l'Université du Michigan (Ann Arbor) en 1966, que Levin et MacCluer ont décrit pour la première fois de façon complète la simulation des génotypes dans des populations humaines, le Gene-Dropping (Edwards, 1987). À la suite du congrès international de Génétique Humaine tenu à Chicago et qui était le prolongement de l'atelier du Michigan, seuls Cavalli-Sforza et Zei (1967) ont présenté une étude avec simulation. Cependant, J.H. Edwards (1967) a établi une analogie entre la structure d'une population et celle d'un cristal. Il a ainsi introduit l'idée que les paramètres structuraux les plus importants dans la transmission des gènes sont les noeuds (ou plutôt les liens) reliant l'individu à ses voisins, tel un cristal.

Les généticiens mendéliens ont été habitués à faire des calculs de probabilités pour les génotypes de descendants d'individus de génotypes connus, mais un problème d'ordre plus général fut d'étendre la procédure de façon à calculer des probabilités aussi bien pour les ancêtres que les proches. Cotterman en 1940 avait déjà vu le problème mais c'est seulement avec les efforts de Hilden (1967; 1970) que ceci est devenu le défi central de Computer Genetics. Hilden présenta une définition du problème en 1967 au European Society of Human Genetics à Copenhague, sous le titre de «Formalized and Automated Pedigree Processing». Il a ainsi établi les prémisses d'une approche algébrique formelle. Thompson, en 1986, a poursuivi cette idée de l'inférence des génotypes des ancêtres sous le nom d'inférence ancestrale avec une référence spéciale à la population de Tristan da Cunha.

Au cours de l'année 1967, les idées sur la simulation et la reconstruction généalogique furent les thèmes centraux du Human Genetics Computer Project. En janvier 1967, A.W.F. Edwards présenta un aperçu des programmes informatiques AUTOKIN et SIMKIN. En résumé, AUTOKIN est conçu pour traiter un échantillon d'individus apparentés avec des génotypes connus et pour construire, uniquement à partir d'informations génétiques, une généalogie de l'échantillon. SIMKIN traite les généalogies, autant réelles que fictives, et simule la transmission de gènes à travers elles, donnant les génotypes pour les membres fondateurs (Edwards, 1987).

Avec ces deux programmes, Edwards a montré la voie à plusieurs chercheurs en leur indiquant les démarches à suivre afin d'effectuer des simulations sur des populations de structure connue.

L'utilisation du programme SIMKIN fut démontrée de façon approfondie à une rencontre de la Société Génétique (Edwards, 1968) et à une conférence à Hawaii sur les Applications de l'Ordinateur dans la Génétique (Edwards, 1969), mettant l'emphasis sur la simulation des génotypes conditionnés à une généalogie connue. Ces idées ont rapidement gagné de l'intérêt, de sorte que dans les cinq années suivantes, elles ont couramment fait partie des études en génétique humaine (Skolnick & Cannings, 1973; Thompson, 1974).

Selon A.W.F. Edwards (1987), c'est vraiment avec la puissance des ordinateurs modernes qu'on peut affronter les problèmes complexes présentés lors de l'analyse des généalogies humaines contenant une quantité substantielle de consanguinité. Cependant, toujours selon Edwards, l'investigation de la structure généalogique en suivant au hasard les chemins vers le haut et vers le bas n'est plus suffisante si l'on considère les parallèles

informatisés de la ségrégation mendélienne prenant place au moins dans toutes les directions descendantes. Déjà en 1925, Wright et McPhee suggéraient de tracer aléatoirement les chemins ancestraux dans le but d'estimer le coefficient de consanguinité lorsqu'une énumération complète était trop complexe (Edwards, 1987). Maintenant, nous avons la possibilité de simuler le processus mendélien lui-même grâce à l'informatique.

Les méthodes informatiques de simulation ont beaucoup été employées pour l'expérimentation et la prédition dans les populations d'animaux en captivité (Hausfater et alii, 1981; Dyke et alii, 1986). Les modèles de simulation ont été utilisés pour prédire la structure future des populations mais également pour clarifier les effets du passé sur la structure présente (MacCluer et alii, 1986).

C'est vers la fin des années 1980 qu'on commença à appliquer de façon régulière la simulation des génotypes basée sur des ascendances connues. Dans les 10 années précédentes, plusieurs auteurs ont plutôt simulé la structure des ascendances, c'est-à-dire, la structure des populations (Edwards, 1987).

Même si le processus de simulation des génotypes était connu, peu de chercheurs utilisaient cette méthode (Thomas, 1990). Introduite par Cannings et alii (1976; 1978), l'«exact computational technique of peeling» fut largement employée. C'est une méthode flexible pour calculer les probabilités des traits génétiques dans les généalogies. Les probabilités peuvent servir à obtenir des probabilités d'extinction des gènes. Thompson (1978; 1979) a d'ailleurs utilisé cette méthode pour analyser les fondateurs et la structure ancestrale de la célèbre population de l'île de Tristan da Cunha dans le sud de l'Atlantique, et pour vérifier l'inférence du locus du groupe sanguin MN et l'occurrence de la maladie oculaire *Retinitis pigmentosa*. Le Peeling a aussi été utilisé par Thompson en 1981 pour examiner l'instance de la maladie d'Hodgkin et l'immunodéficience généralisée dans des généalogies de Terre-Neuve (Thompson, 1981).

Aujourd'hui, les méthodes de simulations des ascendances sont fréquemment utilisées en raison des nombreux avantages qui y sont rattachés. Toutefois, cette méthode est encore considérée comme récente et il importe d'avoir une bonne connaissance de sa portée (voir section 3.5).

CHAPITRE 2

APERÇU HISTORIQUE DES POPULATIONS À L'ÉTUDE

Ce chapitre a pour but de rappeler certains éléments de l'histoire du peuplement fondateur et du mode de formation des cinq populations à l'étude. Une brève situation géographique et démographique s'impose. La qualité et la quantité d'informations disponibles sur ces populations sont cependant variables et dépendent, en grande partie, de l'intérêt que ces populations ont suscité pour les chercheurs. Mais avant d'aborder séparément chacune des populations, il serait important de retracer les grandes lignes de la colonisation française afin d'avoir une idée de l'évolution qu'a connue le Québec.

2.1 Colonisation française

À la suite de l'exploration du golfe du Saint-Laurent par Jacques Cartier (1534-1536) qui prit possession de ce territoire au nom du roi de France, le peuplement français se concentra en deux points relativement éloignés l'un de l'autre, surtout pour l'époque: sur la face sud-est du golfe, ce qui deviendra l'Acadie, et sur les rives du Saint-Laurent, ce qui deviendra le Québec. Ces deux peuplements formaient alors ce qu'on appelle la Nouvelle-France.

Ce territoire demeurera possession française pendant plus de deux siècles, jusqu'à la conquête des Anglais sur les plaines d'Abraham en 1759. Pendant toute cette période, une société française s'installe dans les deux régions. Des relations s'établissent avec les Amérindiens qui se traduiront pour les Français par une meilleure connaissance du pays, et pour les Amérindiens, en une lente assimilation. Le français du Québec garde beaucoup de traces de ces contacts dans son vocabulaire et dans la toponymie. Un exemple parmi tant

d'autres, le toponyme Natashquan évoquant la chasse à l'ours, est l'un des plus anciennement connus de la Côte-Nord. Les Français, guidés par les Amérindiens, explorent le continent, vers le sud jusqu'au golfe du Mexique et vers l'ouest, jusqu'aux Rocheuses: la prise de possession des nouveaux territoires se faisait toujours au nom du Roi de France.

La colonisation française de la province du Québec a véritablement débuté en 1608. Plus de 30 000 colons sont venus de la France (Boleda, 1984), principalement des provinces du centre et de l'ouest de la France. Cependant, on estime qu'entre 8 000 et 10 000 se sont établis de façon permanente dans la province (Charbonneau et alii, 1987).

Au début, l'établissement des colons s'est effectué le long des rives du fleuve Saint-Laurent, entre Québec et Montréal, de manière à profiter de la très grande fertilité du sol de cette vallée. Mais cette vallée passablement étroite se retrouva entièrement occupée au début du 19e siècle. Cette surpopulation stimule alors la migration vers d'autres territoires donnant naissance à de nombreuses régions (Pouyez et alii, 1983). «L'histoire de la population du Québec est à l'image de la plupart des sociétés de peuplement, c'est-à-dire, marquée par de nombreuses migrations internes permettant ainsi le peuplement des régions» (Bouchard & De Braekeleer, 1990).

Depuis trois siècles, des migrations s'effectuent, notamment, entre les régions de l'est du Québec (Gaspésie, Îles-de-la-Madeleine, Bas-Saint-Laurent, Charlevoix, Saguenay, Côte-Nord et la région de Québec). Une fois la vague d'immigration pionnière en provenance de la France passée, ces régions ont cependant reçu très peu d'apports migratoires internationaux. Les régions de l'est ont été quelque peu mises à l'écart dans les échanges et circulations qui s'effectuaient autour de la plaine de Montréal (Bouchard & De Braekeleer, 1990). L'immigration internationale (essentiellement non francophone) s'est quant à elle surtout dirigée vers l'Ouest de la province. Vers la fin du 19e siècle et le début du 20e siècle, c'est le Québec en entier qui fut touché par la forte émigration vers les États-Unis (Lavoie, 1972).

Autour de 5 000 fondateurs se sont établis avant 1700 au Québec et parmi eux, on estime que 3 500 ont des descendants jusqu'à aujourd'hui. Parmi les 8 000 à 10 000 immigrants établis pendant le Régime français, environ 5 000 d'entre eux ont des descendants jusqu'à nos jours (Charbonneau et alii, 1987). La population du Québec se situe présentement aux alentours de 7 millions d'habitants (Institut de la statistique du Québec, ISQ).

2.2 Charlevoix

La région de Charlevoix est située sur la rive nord du fleuve Saint-Laurent, à quelque 100 kilomètres au nord-est de la ville de Québec (carte 1), en plein cœur du bouclier canadien (le plus vieux sol). La superficie de la région est de 5 700 kilomètres carrés mais l'espace habité et habitable est concentré sur une frange côtière d'une centaine de kilomètres de longueur et d'une largeur variant entre 10 et 25 kilomètres (Bouchard & De Braekeleer, 1990). Charlevoix fait partie de la région administrative de Québec.

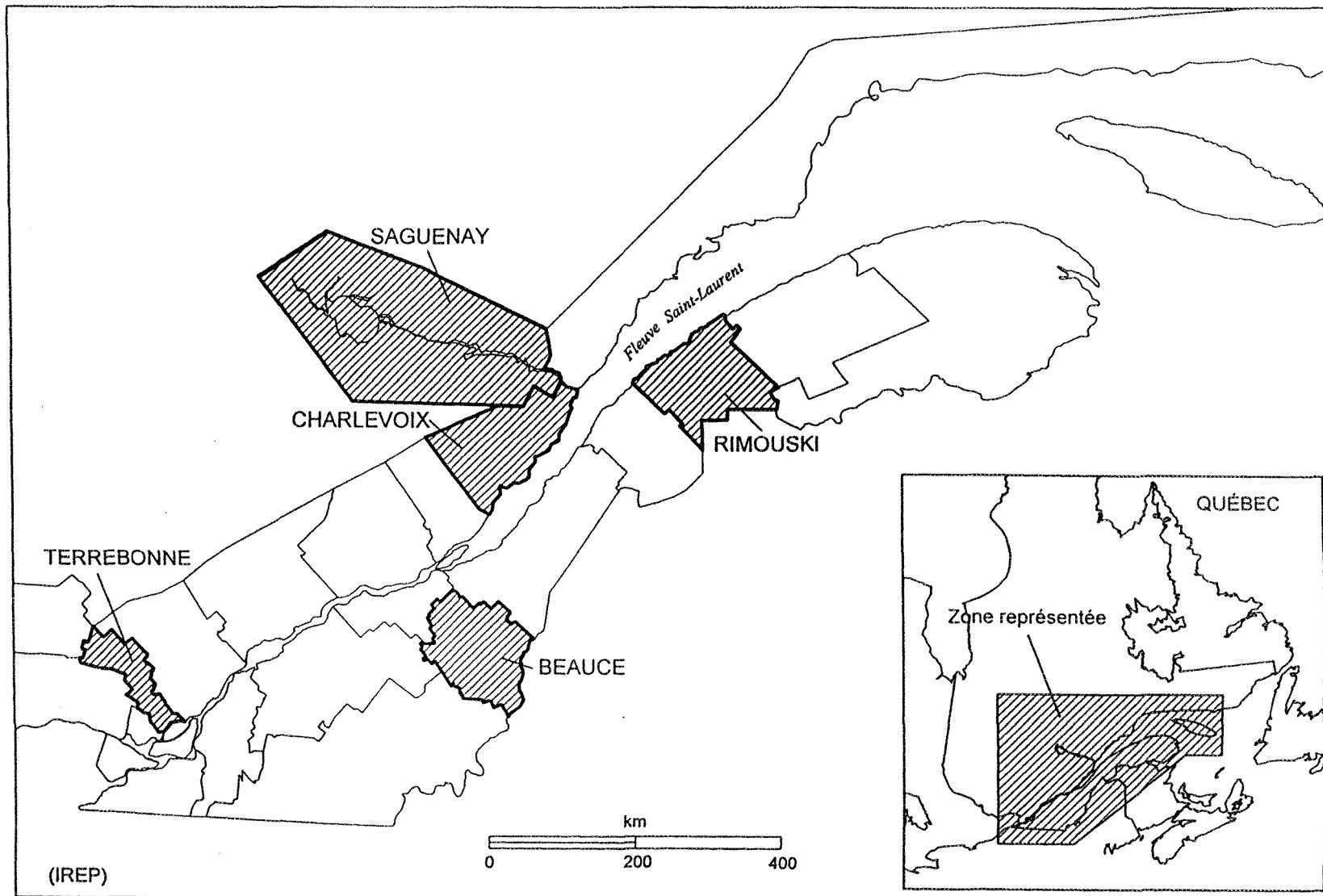
Localisée à la limite nord de la Côte-de-Beaupré, Charlevoix est séparée de cette dernière par des montagnes s'étendant sur une trentaine de kilomètres. Toute la zone habitable est pour ainsi dire encerclée. D'abord par les hautes terres et les denses forêts du plateau laurentien qui l'entourent de toutes parts, mais aussi par le fleuve Saint-Laurent au sud et la rivière Saguenay au nord (Bouchard & De Braekeleer, 1990). Charlevoix possède des paysages et des climats contrastés. Avec un climat plus clément, les basses terres seraient les plus propices à l'agriculture. Mais ces basses terres possèdent des sols de faible fertilité et des excès d'humidité à plusieurs endroits, ce qui amène à la conclusion que Charlevoix n'est pas favorisée du côté de l'agriculture. L'économie de la région repose principalement sur l'exploitation forestière et le tourisme.

Le peuplement de Charlevoix a débuté en 1675 à la Petite rivière Saint-François-Xavier. La région, de même que cette rivière, doivent leur nom au père François-Xavier de Charlevoix, jésuite et premier historien de la Nouvelle-France. Au début du peuplement, toute la population du Canada se concentrat dans la vallée du Saint-Laurent. Les immigrants venus de la ville de Québec, de l'Île-d'Orléans et surtout de la Côte-de-Beaupré auraient été parmi les premiers à fouler le sol de Charlevoix, hormis les Amérindiens (Bouchard & De Braekeleer, 1990).

Des 200 personnes que comptait Charlevoix à la fin du 17e siècle, la population dépasse les 1000 personnes en 1765 (Guérin, 1989). Suite à la conquête de 1760, on assiste au désenclavement de la vallée de la rivière Malbaie, à l'est sur le littoral. Ce phénomène est attribuable pour une bonne part à l'arrivée d'immigrants britanniques (Bluteau & Gauthier, 1984). Ainsi, vers la fin du 18e siècle, la villégiature s'amorçait au Canada, plus précisément à la Malbaie, par le biais des seigneurs écossais Malcom Fraser et John Nairne qui, dès 1760, recevaient des visiteurs dans leur manoir.

Carte 1

Localisation géographique des cinq populations à l'étude



En 1831, la population de Charlevoix compte 8400 personnes et une certaine pression démographique commence alors à s'exercer dans les territoires déjà occupés (Bouchard & De Braekeleer, 1990). De nouvelles paroisses sont érigées mais ceci n'est pas suffisant. Les familles commencent à avoir de la difficulté à établir leurs enfants sur de bonnes terres. On doit s'efforcer d'en défricher de nouvelles. C'est ainsi que s'amorce le mouvement d'émigration vers le Saguenay.

Aujourd'hui, la population de Charlevoix, en incluant celle de Charlevoix-Est, compte près de 30 000 personnes. Elle a d'abord connu une croissance importante, passant de près de 3 000 en 1790 à 15 000 en 1861. Mais avec l'émigration massive vers le Saguenay, il fallu un siècle à la population de Charlevoix pour voir doubler à nouveau ses effectifs. Depuis 1961, même si le taux de natalité chute, atteignant 8,7 pour mille en 1996 et que le solde migratoire est toujours négatif, l'effectif de la population demeure stable (ISQ).

Cependant, si de telles conditions persistaient, l'effectif ne pourrait faire autrement que de diminuer. C'est ce qui est observé dans la plupart des populations vieillissantes qui ne peuvent bénéficier d'apport nouveau.

Charlevoix tout comme le Saguenay (que nous verrons en détail dans le point suivant) présentent des maladies héréditaires qui sont caractéristiques à ces deux régions. Seules des données sur les fréquences de maladies pour le Saguenay et parfois pour les deux régions ensemble sont disponibles jusqu'à maintenant. Les maladies qui sont le plus caractéristiques sont la dystrophie myotonique (prévalence de malades=1/430), le rachitisme vitamino-dépendant (prévalence de porteurs=1/25), la tyrosinémie (prévalence de porteurs entre 1/15 et 1/20), l'acidose lactique, l'ataxie spastique (même prévalence que la tyrosinémie), la polyneuropathie sensori-motrice avec ou sans agénésie du corps calleux (prévalence légèrement inférieure à celle de la tyrosinémie et de l'ataxie spastique) et l'atrézie intestinale multiple. Les trois dernières maladies nommées, l'ataxie spastique, la polyneuropathie et l'atrézie intestinale multiple, se retrouvent exclusivement dans Charlevoix et au Saguenay (Bouchard & De Braekeleer, 1992).

2.3 Saguenay

La région du Saguenay¹ est située à 200 kilomètres au nord de la ville de Québec (carte 1). La montagne marque un répit et donne naissance à une surprenante enclave qui est devenue au fil du temps, un lieu convoité par plusieurs personnes. Cette entrave est l'héritage du passage des derniers glaciers et de leur retrait vers le golfe Saint-Laurent et l'Atlantique au cours du Quaternaire (Giroux & Joyal, 1984).

Avant que sa colonisation débute, la région du Saguenay faisait partie d'un immense territoire de chasse parcouru par les Amérindiens et contrôlé par la Compagnie de la Baie d'Hudson. On y dénombrait peu de résidants permanents, à l'exception de quelques employés des deux postes de traite de fourrures, l'un situé à Chicoutimi, l'autre en bordure sud-est du lac Saint-Jean. On y recensait également environ 200 Montagnais (Pouyez et alii, 1983).

Le Saguenay, territoire voisin de Charlevoix, apparaît propice à l'agriculture et à l'exploitation forestière. Cependant, il faut attendre la fin du bail que la Compagnie de la Baie d'Hudson détient jusqu'en 1842 pour que le Saguenay s'ouvre officiellement à la colonisation. Mais en réalité, l'année 1838 correspond à celle de l'ouverture de la région au peuplement blanc et à la colonisation. Précédemment, en 1837 à la Malbaie, se forme la Société des Vingt et Un. Cette société comprenait en plus des 21 sociétaires, 15 associés qui avaient été mandatés pour partir défricher le territoire du Saguenay. Le contingent a d'abord débarqué à l'Anse-Saint-Jean et non à La Baie comme nous l'avons longtemps pensé. Il demeure encore des imprécisions sur le nombre réel des membres de la Société qui se sont établis au Saguenay mais un fait persiste, ils sont les vrais pionniers du Saguenay, ceux qui ont commencé les défrichements et assuré les débuts du peuplement en fondant familles et foyers (Bouchard & De Braekeleer, 1992).

C'est donc à partir de la localité de la Malbaie que les premiers colons et leurs familles émigrent vers le Saguenay mais rapidement, le mouvement d'émigration s'étend à toutes les paroisses de Charlevoix. Les émigrants se déplaçant en famille sont motivés par le besoin de s'établir sur des terres de meilleure qualité mais surtout plus vastes. Les habitants de Charlevoix voyaient le Saguenay comme une extension de leur territoire (Bouchard & De Braekeleer, 1990).

1. Le nom Saguenay réfère ici à l'ensemble du territoire régional (Saguenay et Lac-Saint-Jean).

À ses débuts, le Saguenay s'est donc constitué à même un prélèvement migratoire de Charlevoix. D'ailleurs, de 1838 à 1871, Charlevoix a fourni 75% des immigrants du Saguenay et de 1872 à 1891, 57% (Gauvreau & Bourque, 1988). Par conséquent, il est établi que la population du Saguenay est issue en bonne partie de celle de Charlevoix.

L'économie inhérente à la colonisation a fait place à de nombreuses activités parallèles comme par exemple la coupe forestière, l'industrie domestique et la pêche mais l'agriculture est demeurée l'axe intégrateur. Aujourd'hui, le Saguenay compte plusieurs industries dans les domaines des pâtes et papiers, de l'aluminium et de l'hydroélectricité.

Selon les estimations de l'Institut de la statistique du Québec, la population du Saguenay-Lac-Saint-Jean est de 289 600 habitants en 1998 et son taux de natalité pour l'année 1996 se situe à 10,3 pour mille.

2.4 Rimouski

Le comté de Rimouski se situe à mi-chemin entre Québec et Gaspé (carte 1). Il est considéré comme étant la capitale de la région administrative du Bas-Saint-Laurent et la capitale océanographique du Québec. Situé au sud du golfe Saint-Laurent, Rimouski prend appui sur la chaîne des Appalaches, le long des basses terres de l'estuaire du golfe. Au sud, Rimouski est borné par la frontière de la province du Nouveau-Brunswick.

À ses débuts, l'économie de Rimouski reposait presque exclusivement sur l'agriculture et l'exploitation forestière et ce, en dépit du fait que les Appalaches soient constituées d'un sol plutôt acide accompagné d'un taux d'humidité défavorable à une bonne fertilité des terres (Giroux & Joyal, 1984). Ceci ne semble pas déranger la forêt mixte composée de conifères, de bouleaux blancs et d'éables qui s'accommode parfaitement de ce sol.

Le toponyme Rimouski remonte selon Monseigneur Laflèche à un mot micmac «animouski» signifiant demeure et chien dans la langue des Saulteux. On explique la référence au chien tantôt par la présence ancienne des chiens errants, tantôt par le mécontentement des Amérindiens qui devaient attendre la marée montante pour atteindre la Côte-du-Sud et l'embouchure de la rivière où ils habitaient l'été, ce qui les rendaient agressifs comme des chiens (?) (Commission de toponymie du Québec, 1996).

La première seigneurie de Rimouski fut concédée originellement à un dénommé Augustin Rouer de Villeray et de la Cardonnière, en 1688. Ce n'est qu'en 1696 que René Lepage en prend possession. Cependant, les premiers colons tardèrent à s'installer dans ce comté en réponse à la publicité favorisant les comtés limitrophes de Matapedia et de Témiscouata (Blanchard, 1935). La première paroisse à être fondée dans le comté de Rimouski fut celle de Rimouski en 1701. La majorité des immigrants provenaient des anciennes seigneuries de la région de Québec et près de la moitié des immigrants venus sur la Côte-du-Sud durant la période de peuplement colonisateur étaient nés en France. Le comté de Rimouski tout comme Charlevoix, bénéficia de la venue de nombreux réfugiés acadiens après leur déportation en 1755 (Bouchard & De Braekeleer, 1990).

La municipalité régionale de comté (MRC) Rimouski-Neigette comptait en 1997 près de 54 000 habitants et le taux de natalité était de 9,9 pour mille en 1996 (ISQ). Deux maladies autosomales récessives semblent relativement fréquentes dans le comté de Rimouski, il s'agit de l'ataxie de Friedreich et la maladie de Tay-Sachs (Bouchard et alii, 1979; Andermann et alii, 1977). L'ataxie de Friedreich est une maladie dégénérative spinocérébelleuse progressive tandis que la maladie de Tay-Sachs est causée par une accumulation de graisses dans le cerveau et entraîne une rapide dégradation physique et mentale (cécité, convulsions, perte des contrôles moteurs...) (Bouchard & De Braekeleer, 1992).

2.5 Terrebonne

Le comté de Terrebonne fait partie de la MRC Les Moulins de la région administrative Lanaudière. Situé au nord de Laval et donnant sur la rivière des Mille-Îles, Terrebonne est distant d'à peine douze kilomètres de Montréal (carte 1). Cette proximité a d'ailleurs eu beaucoup d'influence sur l'expansion de la région toute entière. La population de Terrebonne comptait un peu plus de 42 000 habitants en 1997 (ISQ). La MRC, dont fait partie Terrebonne, bénéficie d'une forte croissance de sa population. Si on considère l'ensemble de la population de Lanaudière, c'est dans cette MRC que l'augmentation de la population fut la plus élevée (53%) pour la période 1981-1991. À titre comparatif, pour la même période, le taux de croissance de la population du Québec n'était que de 7,8% (Guillemette, 1997).

En 1996, le taux de natalité pour l'ensemble de la MRC était de 13,3 pour mille (ISQ). La fécondité lanaudoise est cependant différente de celle du Québec. L'observation des taux de fécondité de cette région pour une période de quinze ans (1979-1993) indique que cette région

se distingue du Québec par des taux plus élevés parmi les femmes les plus fécondes (20-29 ans) et par des taux plus faibles chez les moins fécondes (moins de 20 ans et 35 ans et plus) (Guillemette, 1996). Ces données sur la fécondité et la natalité confirment bien la croissance observée dans cette population.

L'endroit tient son nom de la seigneurie de Terrebonne concédée sur les bords de la rivière des Mille-Îles en 1673 à un résidant de Paris, André Daulier-Deslandes, par la Compagnie des Indes Occidentales, fondée en 1664 par le roi Louis XIV. Rien ne prouve cependant que le sieur Daulier-Deslandes soit effectivement venu au pays (Gareau, 1927).

À cette époque, tout le territoire était encore couvert d'épaisses forêts. L'endroit était idéal pour l'établissement d'une colonie, étant donné la richesse du sol qui longeait la rivière. La dénomination Terrebonne fut donc attribuée en raison de la fertilité de ses terres. L'établissement des premiers colons se fit entre 1695 et 1702. La seigneurie de Terrebonne changea plusieurs fois de propriétaires avant d'être acquise par le sieur Louis Lepage de Sainte-Clair en 1720. Louis Lepage était le fils du premier seigneur de Rimouski. C'est en 1722 que les limites de la première paroisse furent fixées, celle de Saint-Louis-de-Terrebonne. En 1860, on comptait déjà 2 000 habitants (Gareau, 1927).

Lieu florissant à la fin du 19e siècle, Terrebonne a connu un grand incendie en 1922 qui rasa une partie de la localité. Aujourd'hui, l'économie industrielle repose sur une centaine d'entreprises manufacturières dont les plus importantes sont celles du bois de construction, l'alimentation et les produits métalliques. Les grands espaces faiblement urbanisés offrent également de nombreuses possibilités pour le développement récréo-touristique (Commission de toponymie du Québec, 1996; Guillemette & St-Cyr, 1995).

En raison du peu d'études encore réalisées sur la population de Terrebonne, il est difficile d'identifier des maladies qui pourraient lui être caractéristiques.

2.6 Beauce

La Beauce est située à une trentaine de kilomètres au sud de la ville de Québec (carte 1). S'étendant vers le sud-est jusqu'à la frontière américaine, elle est bornée au nord-est par la vallée de la rivière Etchemin et au sud-ouest par les régions de Lotbinière, de l'Amiante et de l'Estrie. La superficie de la Beauce est de 4 753 kilomètres carrés. Sa partie nord s'inscrit

dans les Basses-Terres du Saint-Laurent tandis que la majeure partie de son territoire se rattache à la région administrative Chaudière-Appalaches. La Beauce est drainée en grande partie par la rivière Chaudière qui prend source dans le lac Mégantic, coule vers le nord, traverse la Beauce et se jette dans le fleuve Saint-Laurent après un parcours de 200 kilomètres. La rivière Chaudière est connue, depuis le 18e siècle, pour les inondations qu'elle provoque (Bélanger et alii, 1990).

Lors de l'ouverture de la Beauce, à partir de 1737, les premiers colons s'installent sur un territoire pratiquement, voire entièrement inhabité. Les Abénaquis originaires de la Nouvelle-Angleterre séjournent seulement à l'occasion en Beauce à l'exception de quelques rares individus qui s'y établissent à la fin du 18e siècle. Avant 1736, les Français se sont attardés à mieux connaître la région de la Beauce. Le premier Européen à traverser la Beauce serait un homme envoyé par Champlain, en 1629, au pays des Abénaquis. Diverses concessions de seigneuries ont été effectuées mais aucune de ces propriétés situées sur le territoire beauceron ne semble avoir été exploitée (Bélanger et alii, 1990).

L'ouverture de la Beauce se fait dans le but d'établir une zone tampon pour servir de rempart face à l'expansion anglaise. Ainsi, selon le système seigneurial d'attribution des terres, le marchand de Québec Joseph Fleury de la Gorgendiére et ses gendres François-Pierre Rigaud de Vaudreuil et Thomas-Jacques Taschereau obtiennent, le 23 septembre 1736, trois seigneuries contiguës situées au cœur de la Beauce. Suivra l'attribution de trois autres seigneuries. Le succès des débuts de la colonisation de la Beauce sont imputables à Joseph Fleury de la Gorgendiére qui a su mettre en place une infrastructure de services attrayante. Le nom même de Nouvelle-Beauce, que l'on retrouve dès 1739 dans les documents, fait allusion à la Beauce française, une plaine grande productrice de blé renommée pour sa fertilité. Ce nom a sans doute été donné pour inciter les pionniers à venir s'y établir (Bélanger et alii, 1990).

Provenant principalement de la Côte de Beaupré, de l'Île d'Orléans ou de la seigneurie de Lauzon, les premiers habitants de la Beauce ont souvent des liens de parenté. Des noyaux et des réseaux se créent de façon à faciliter l'intégration et l'enracinement des pionniers. Les individus isolés ont peu de chance de réussir et la plupart de ceux-ci quittent la Beauce peu de temps après leur arrivée. Les premiers pionniers beaucerons sont surtout des Canadiens de naissance. Toutefois, phénomène rare pour le 18e siècle, un contingent important de Français se retrouve parmi eux. Certains d'entre eux s'installent mais la plupart sont enclins à quitter la région à plus ou moins brève échéance (Bélanger et alii, 1990).

En 1739, la région comptait déjà 262 âmes. Entre 1762 et 1831, la population de la Beauce est multipliée par 15, passant de 730 à 10 816 individus. Le succès de la colonisation revient au fait que la Beauce joue le rôle d'une terre d'accueil pour les réfugiés et les migrants venus de l'Acadie ou de la vallée du Saint-Laurent pendant la Guerre de Sept ans (1756-1763) comme durant la décennie suivante. La Beauce est devenue une région importante du Québec avec une population de près de 156 000 habitants en 1997 et un taux de natalité de 10,4 pour mille en 1996 (ISQ). Cette population est considérée comme homogène, d'origine française et généralement enracinée sur son territoire depuis plusieurs générations (Bélanger et alii, 1990).

Aucune maladie n'est réellement rattachée à la population beauceronne. Une étude suggère cependant que la Beauce pourrait être le centre géographique d'une mutation du gène de la lipoprotéine lipase associé à l'hyperchoylo-micronémie (De Braekeleer et alii, 1991).

CHAPITRE 3

SOURCES DES DONNÉES ET MÉTHODOLOGIE

Les données sur les cinq populations étudiées proviennent de sources différentes et ont été obtenues de façon différente. Il est donc bien important de distinguer ces sources et les procédures conduisant à leur obtention afin de permettre une analyse adéquate des résultats issus de ces données. Pour les régions du Saguenay, de Terrebonne, de Charlevoix et de Rimouski, les individus de départ ont été choisis au hasard dans le fichier-réseau BALSAC. Une description de ce fichier s'impose maintenant.

3.1 Fichier-réseau BALSAC

Le fichier-réseau BALSAC est une banque de données informatisées qui contient des informations sur les individus de diverses populations du Québec. Les données nominatives de ce fichier proviennent de sources diverses comme les registres paroissiaux (actes de naissance, mariage, sépulture), les recensements nominatifs ou d'autres listes ou répertoires. Le fichier est assorti de logiciels de jumelage automatique à l'aide desquels les informations relatives à un même individu sont regroupées dans un même enregistrement. Il est également possible de créer les liens généalogiques entre les individus.

Les débuts de ce fichier remontent en 1972 et les travaux ont d'abord porté sur la région du Saguenay. La construction de la banque pour cette région a été réalisée à partir de quelque 700 000 actes de baptême, mariage et sépulture. En 1985, on procéda à l'extension du fichier en y ajoutant des données sur les régions du nord-est québécois. Ainsi, après le Saguenay, les régions de Charlevoix, la Côte-Nord, le Bas-Saint-Laurent, la Côte-de-Beaupré et la Côte-du-

Sud furent incorporées d'où le sigle BALSAC formé des lettres initiales des régions et sous-régions initialement choisies (Bouchard, 1987; 1990).

Désormais, la saisie des données de BALSAC vise toute la population du Québec, depuis le 19e siècle jusqu'à aujourd'hui (actes de mariages seulement). Avec de telles informations, il sera possible de consulter et d'analyser par ordinateur les généalogies de tous les Québécois de souche française et d'une partie des Québécois non francophones. Évidemment, ce fichier a un caractère confidentiel et c'est pour cette raison qu'un protocole d'exploitation a été mis sur pied en 1976. En date du printemps 1998, 1 937 818 actes étaient compris dans le fichier BALSAC (Bouchard, 1998). Il y a encore des régions qui ne sont toujours pas incluses dans le fichier mais elles sont en voie de l'être.

3.2 Fichier généalogique BALSAC- RETRO

Ce fichier fut créé dans le but de permettre l'accumulation des généalogies remontant au-delà de 1840 ou reliées à des proposants non-saguenayens et reconstruites dans le cadre de divers projets du volet génétique.

BALSAC-RETRO est essentiellement une base de données généalogiques. Les informations qui y sont contenues se rapportent à des individus et à leur(s) acte(s) de mariage uniquement. Il ne contient cependant aucune donnée à caractère médical ou génétique (Jomphe & Casgrain, 1997). Le fichier BALSAC-RETRO peut servir à la fois à l'entrée des données généalogiques et aussi comme outil de recherches généalogiques (Jomphe & Casgrain, 1997). Cette base de données ne contient pas la totalité de la population québécoise. Elle est plutôt constituée au fur et à mesure selon les besoins des chercheurs. À ce jour, 222 000 individus et 115 000 couples sont enregistrés dans le fichier BALSAC-RETRO (Bouchard, 1998).

Sous forme de requêtes, il est possible d'interroger la base de données BALSAC-RETRO et d'effectuer, par exemple, des recherches d'ancêtres communs ou des calculs de coefficients de consanguinité et d'apparentement, de profondeur généalogique, de complétude, et de contribution génétique (Jomphe & Vigneault, 1992).

3.3 Obtention des données

Au total, 378 généalogies ascendantes ont été utilisées dans le cadre de ce projet, réparties comme suit: Saguenay (100), Charlevoix (80), Rimouski (80), Terrebonne (75) et Beauce (43) (tableau 3.1). Les individus de départ des généalogies du Saguenay, de Charlevoix, de Rimouski et de Terrebonne proviennent du fichier BALSAC. Pour les populations du Saguenay et de Charlevoix, les actes sont complètement dépouillés et jumelés jusqu'en 1971. Pour Rimouski et Terrebonne, le dépouillement n'est pas finalisé. Les individus de départ des généalogies ont donc été choisis au hasard parmi les naissances ou mariages disponibles dans chacune des régions.

Puisque certaines généalogies utilisées ont été réalisées pour divers projets de recherche, les années de naissance ou de mariage des individus de départ sont variables. Ainsi pour les régions du Saguenay et de Charlevoix, les individus de départ sont nés entre 1950 et 1971. Pour le comté de Rimouski, les couples de départ se sont mariés entre 1900 et 1948 tandis que pour Terrebonne, les dates de mariages se situent entre 1945 et 1965.

Pour les généalogies de la région de la Beauce, les individus de départ sont les couples-témoins des parents de patients atteints de psychose maniaco-dépressive ou de schizophrénie (Bergeron, 1999). Ces couples-témoins ont été choisis dans le répertoire des mariages de la Beauce (Société de généalogie de Québec, 1995). Ils ont été choisis selon le lieu et l'année de leur mariage qui devaient correspondre à ceux des parents des patients atteints de PMD ou de schizophrénie. Aucune caractéristique génétique ou médicale n'était associée à ces individus mariés entre 1905 et 1971.

Tableau 3.1
Caractéristiques des données sur les sujets par région

| Région | Type d'actes | Nb de sujets | Étendue des dates d'événement | Année moyenne |
|------------|--------------------|--------------|-------------------------------|---------------|
| Rimouski | mariage | 80 | 1900-1948 | 1925 |
| Charlevoix | baptême | 80 | 1950-1971 | 1960 |
| Saguenay | baptême | 100 | 1950-1971 | 1959 |
| Beauce | mariage | 43 | 1905-1971 | 1939 |
| Terrebonne | mariage | 75 | 1945-1965 | 1954 |
| Ensemble | mariage et baptême | 378 | 1900-1971 | 1947 |

Source : fichier BALSAC-RÉTRO

La reconstruction généalogique s'effectue à partir des mariages. Ainsi, pour les régions de Charlevoix et du Saguenay (naissances entre 1950-1971), les mariages des parents des sujets furent utilisés et ils se situent entre 1930 et 1970.

3.4 Reconstructions généalogiques

Les sources utilisées pour la reconstruction généalogique sont nombreuses, de fiabilité variable et non exemptes d'erreurs. Évidemment, la source la plus fiable en généalogie est l'acte de mariage lui-même. Cependant, comme il n'est pas possible de consulter tous les actes de mariage nécessaires pour chaque généalogie, il faut utiliser d'autres sources plus accessibles. À titre d'exemple, les répertoires de mariages, les fichiers informatisés, les dictionnaires généalogiques ou encore les recensements permettent de retracer les parents d'un individu. Entre le mariage de deux personnes et la reconstruction généalogique, il y a de nombreuses étapes qui impliquent toutes des sources possibles d'erreurs. Tout d'abord, le prêtre doit transcrire l'information orale dans l'acte officiel de mariage et peut introduire des erreurs suite à une mauvaise compréhension ou interprétation. Le dépouilleur de cet acte peut éprouver des difficultés à déchiffrer l'écriture ou encore l'état de conservation du registre peut rendre la lecture plus difficile. Enfin, des erreurs peuvent encore se produire à la transcription des registres, à la préparation du manuscrit et à la dactylographie, dans le cas d'un répertoire publié.

En plus des répertoires de mariage par paroisse, diocèse, ville ou comté, plusieurs autres sources ont été consultées lors de la reconstruction. En voici une brève description:

- 1) Fichier-réseau informatisé BALSAC: permet la consultation de tous les actes relevés. Très fiable.
- 2) Dictionnaire généalogique des familles du Québec (Jetté, 1983): couvre tout le Québec des origines à 1730. Très fiable, il rassemble au nom de la personne les faits de son histoire généalogique.
- 3) Fichier des mariages de René Jetté: contient tous les mariages du Québec pour la période de 1730 à 1825. Très fiable.

4) Répertoire des mariages des Canadiens Français 1760-1935 (Institut généalogique Drouin): très fiable, il couvre la plupart des paroisses du Québec.

5) Registre des mariages des paroisses ou du greffe de l'état civil: présente les actes originaux de baptême, mariage et sépulture. Très fiable; les actes peuvent être consultés aux Archives Nationales.

6) Dictionnaire généalogique des familles canadiennes (Tanguay, 1975): c'est un répertoire d'actes de baptême, de mariage et de sépulture classés en famille. Il couvre tout le Québec des débuts de la colonie aux débuts de 1800 mais contient de nombreuses erreurs.

7) Fichier Antonin Loiselle: sous forme de micro-fiches et disponible aux Archives Nationales, ce fichier contient 15 diocèses catholiques du Québec, du Madawaska et de l'est ontarien, depuis la fondation de ces paroisses jusqu'au début 1900. Fiabilité moyenne.

Le nombre de générations dépend de plusieurs facteurs. Il arrive parfois que des branches d'une généalogie s'interrompent assez tôt, car un ancêtre (ou un couple-ancêtre) n'a pu être retracé parce qu'il s'est marié à l'extérieur du Québec ou simplement que l'acte de mariage est manquant. Il fut possible de justifier un certain nombre d'interruptions de généalogies en ajoutant la variable statut qui donne de l'information supplémentaire sur un individu (voir annexe). Enfin, certaines erreurs de reconstruction ne peuvent être complètement évitées, par exemple dans des cas d'adoption, de fausse paternité ou de mauvaise attribution d'un lien généalogique lors de la présence d'homonymie. Toutefois, les généalogies font l'objet d'une série de vérifications et de validations qui aident à réduire les risques d'erreurs.

Puisque dans cette étude nous désirons évaluer l'évolution des gènes des fondateurs, toutes les généalogies ont été reconstruites jusqu'à l'arrivée des premiers ancêtres européens au 17e siècle.

3.5 Description du Gene-Dropping

La méthode du Gene-Dropping est appliquée dans diverses perspectives comme, par exemple, pour estimer les coefficients de consanguinité, pour évaluer l'étendue de la variabilité génétique dans les colonies d'animaux en captivité, et pour prédire le risque de la perte des

gènes issus des fondateurs (MacCluer et alii, 1986). Il est très important de comprendre que ce n'est pas la population qui est simulée mais bien la transmission des gènes autosomaux d'une population dont l'ascendance est connue.

La procédure du Gene-Dropping débute par l'attribution de deux allèles hypothétiques uniques à chaque fondateur de la population. Il est utile de rappeler que la notion de fondateur réfère dans ce cas-ci à un individu apparaissant dans une généalogie alors que ses parents n'y apparaissent pas et qui possède un ou plusieurs descendants dans la généalogie. La figure 3.1 illustre le processus d'attribution au sein d'une ascendance toute simple d'un seul individu et de ses ancêtres.

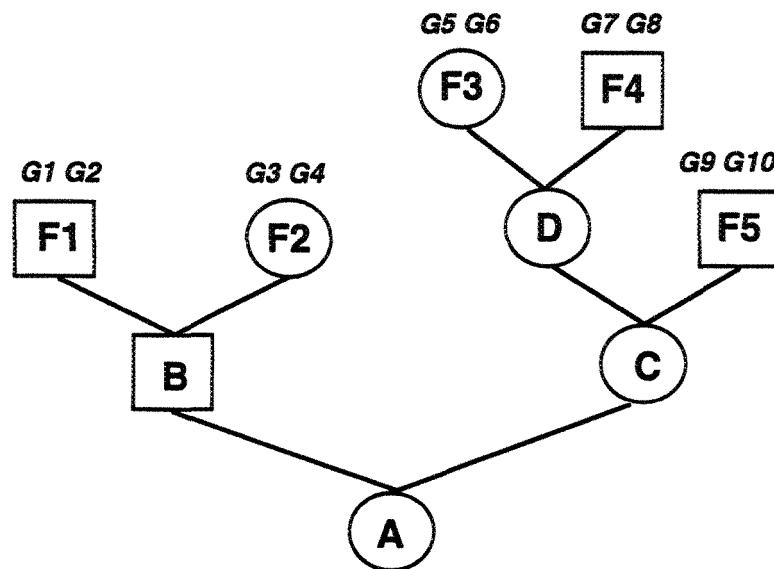


Figure 3.1 Généalogie avec deux allèles uniques assignés à chaque fondateur.

Un générateur de nombres aléatoires est utilisé pour déterminer quel gène de chaque fondateur est transmis à chaque descendant. À la fin d'une première simulation, tous les individus de la généalogie ont un génotype (fig. 3.2).

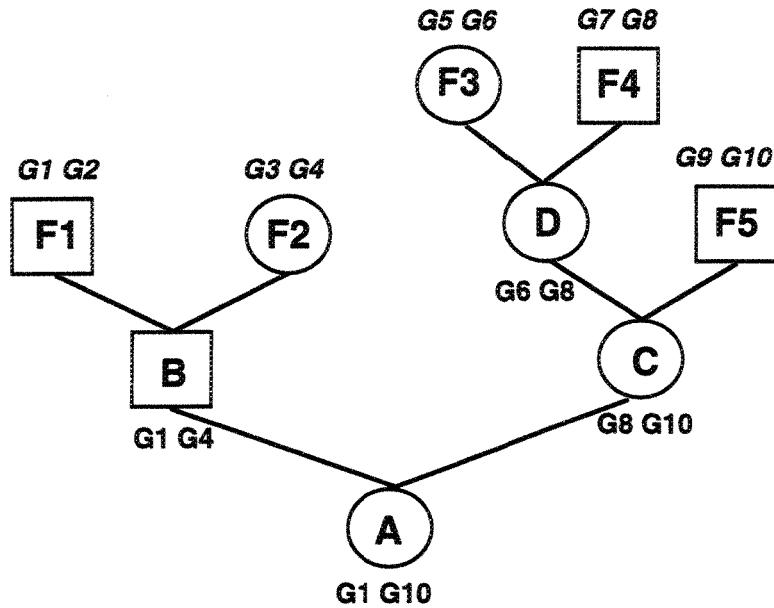


Figure 3.2 Généalogie après le Gene-Dropping avec les génotypes assignés aux descendants par la ségrégation mendéienne des allèles des fondateurs.

La procédure toute entière est répétée plusieurs fois et le nombre de gènes (ou d'allèles) distincts parmi la population contemporaine est calculé afin de générer des distributions de probabilités d'occurrence de gènes (ou d'allèles) des fondateurs.

Ainsi, la méthode et les principes fondamentaux du Gene-Dropping sont simples. Sachant que chaque individu possède deux gènes à un locus et qu'un seul de ces deux gènes sera transmis, chaque gène a donc une chance sur deux d'être transmis aux descendants et ce, indépendamment de l'héritage de l'autre parent et des autres descendants. C'est simplement la simulation de la transmission mendéienne (Thomas, 1990).

Avec l'application de la méthode du Gene-Dropping, certaines hypothèses simplificatrices peuvent être émises afin de simuler une situation précise. Dans le présent cas, toutes les transmissions de gènes sont indépendantes les unes par rapport aux autres, ainsi le déséquilibre de liaison (linkage) a dû être ignoré. De plus, dans la procédure réalisée, la sélection à un locus simulé n'a pas été incluse puisque nous désirions vérifier uniquement l'effet du hasard dû à la transmission mendéienne des gènes. Il importe que ces simplifications soient prises en considération afin d'obtenir une analyse valable et concluante des résultats des simulations.

Toutefois, les populations réelles, représentées comme des systèmes, sont d'une toute autre complexité. Les populations n'ont pas une taille infinie et leurs effectifs sont rarement constants. En plus de pouvoir fluctuer au hasard, les fréquences alléliques sont soumises à des forces évolutives telles les migrations, les mutations et la sélection naturelle, entraînant au sein des populations, une évolution directrice et non aléatoire des fréquences alléliques (Hartl, 1994). Avec la méthode employée, la migration est incluse lors de la reconstruction généalogique. Les mutations quant à elles ne sont pas abordées mais si elles sont prises à l'échelle du temps du Québec, elles sont négligeables. Pour la force de la sélection, Heyer (1999) a démontré en comparant un gène récessif neutre et un létal que dans le cas de maladies récessives, les résultats changeaient peu.

3.6 Gene-Dropping et Peeling

D'implantation et de compréhension faciles, la méthode du Gene-Dropping est de plus en plus utilisée. En dépit de sa grande efficacité, le Peeling demeure lourd informatiquement, en plus d'être difficile à comprendre et à planter. La plus grande restriction du Peeling concerne le stockage des résultats intermédiaires, ceux-ci augmentant exponentiellement avec la complexité de l'ascendance. Avec le Gene-Dropping, le stockage des résultats ne pose habituellement aucun problème (augmentation linéaire). La méthode du Gene-Dropping peut donc être utilisée pour des ascendances plus grandes et plus complexes (Thomas, 1990).

Lorsqu'on compare les résultats obtenus à partir des deux méthodes, la qualité des approximations du Gene-Dropping ressort davantage. Cependant, si on considère l'efficacité relative de ces deux méthodes en termes de temps informatique, le temps requis pour le Peeling est négligeable, comparé à celui du Gene-Dropping, dans des cas où les ascendances ne sont pas trop complexes. Le plus gros avantage du Gene-Dropping est qu'il peut toujours être exécuté, quelle que soit la complexité de l'ascendance, ce qui n'est pas le cas avec le Peeling (Thomas, 1990).

Le Peeling et le Gene-Dropping sont deux outils valables pour l'analyse d'ascendances, le premier pour des réponses exactes et rapides sur des ascendances simples, le second pour de bonnes approximations sur des ascendances complexes dans un temps acceptable. Parfois, quantifier la complexité d'une ascendance n'est pas chose facile. La solution la plus pragmatique serait d'essayer le Peeling et ensuite le Gene-Dropping si on constate que l'exécution est impossible.

Dans le cadre de cette étude, seule la méthode du Gene-Dropping fut utilisée. Cette méthode a permis de connaître les probabilités pour un gène fondateur d'atteindre une fréquence donnée dans chaque population.

3.7 Étapes de programmation

Comme matériel de base, nous avions un premier fichier contenant la liste de tous les numéros d'individus de départ ainsi qu'un autre fichier contenant les informations sur les 378 généalogies (#individu, #père, #mère, sexe, statut, région, date de mariage). La toute première étape était d'ordonner tous ces individus de façon à ce que chaque individu soit précédé de ses parents dans la nouvelle liste.

Par la suite, on a assigné à tous les fondateurs, c'est-à-dire les individus dont les deux parents sont absents dans la généalogie, une paire de gènes unique. Les simulations de la transmission des gènes furent ensuite effectuées en utilisant un générateur de nombres aléatoires afin de décider quel gène du père et quel gène de la mère étaient transmis à l'individu. Au total, 50 000 simulations furent réalisées sur les cinq populations avec les fondateurs (si on veut, les bouts de généalogies) et une autre série de 50 000 simulations a été effectuée avec les fondateurs démographiques. Un fondateur démographique est un ancêtre qui a introduit un gène dans une population (Bouchard & De Braekeleer, 1990). Les immigrants sont, par le fait même, considérés comme fondateurs démographiques à condition cependant que leurs parents ne soient jamais venus. Les simulations produites à partir de cette définition ne sont utilisées qu'à titre comparatif puisque cette définition s'appuie sur l'attribution de statuts (annexe) encore sujets de contradiction ou de perception (voir section 3.4).

Une fois les simulations produites, la fréquence à laquelle les gènes d'un fondateur reviennent dans les simulations fut calculée et ce, pour les fondateurs de chaque population. Les résultats présentés dans la section suivante ont tous été obtenus de cette façon. Les simulations de même que leur analyse furent programmées en langage C++ en utilisant les logiciels Borland C++ et Visual Studio.

3.8 Calculs des fréquences et des probabilités

La fréquence de porteurs est déterminée par le nombre d'individus de départ dans chacune des populations pour lesquels nous avons simulé la transmission de deux de leurs gènes.

fréquence de porteurs = nb d'individus porteurs de gènes ancêtres / nb d'individus de départ

Pour calculer des fréquences géniques, nous divisons le nombre de fois que le gène est trouvé dans notre échantillon par deux fois la taille l'échantillon. Ce calcul est fait pour chaque simulation. Le nombre de simulations où le gène d'un fondateur atteint une fréquence donnée nous permet de connaître la probabilité pour ce fondateur qu'un de ses gènes atteigne cette fréquence. Pour chaque fondateur, nous avons fait la moyenne pour ses deux gènes homologues. Lorsque le nombre de simulations est élevé, la probabilité pour un gène du fondateur est quasiment identique à celle obtenue pour son gène homologue. Si $P(Y)$ est la probabilité pour un fondateur donné d'atteindre une fréquence donnée dans la population et $P(X)$ est la probabilité pour ce fondateur d'atteindre une fréquence cible dans l'échantillon, la formule suivante nous permet de calculer pour chaque fondateur la probabilité qu'un de ses deux gènes atteigne une fréquence cible dans la population:

$$P(Y) = P(Y|X) \cdot P(X)$$

où $P(Y|X)$ est obtenue de la distribution normale de moyenne X et d'écart-type $X \cdot (1-X)/n$, n étant la taille de l'échantillon. Afin de connaître $P(Y)$, nous sommes les $P(Y|X) \cdot P(X)$ pour toutes les $P(X)$ obtenues à partir des simulations et qui se situent dans le même intervalle visé par $P(Y)$. Prenons par exemple le fondateur #118 de Rimouski et calculons sa probabilité d'atteindre une fréquence supérieure ou égale à 0,05. Directement des simulations, nous prenons toutes les $P(X)$ non nulles pour les fréquences supérieures ou égales à 0,05. Pour ce fondateur, aucune fréquence supérieure à 0,05 n'a été obtenue. Ainsi:

$$\begin{aligned} P(Y \geq 0,05) &= P(0,05 | X) \cdot P(0,05) \\ &= 1 \cdot 0,0004 \\ &= 0,0004 \end{aligned}$$

Nous retrouvons cette probabilité de 0,0004 pour le fondateur #118 au tableau 4.4. Cette méthode a déjà été employée dans une autre étude sur le Saguenay (Heyer, 1999).

CHAPITRE 4

RÉSULTATS

4.1 Complétude et profondeur des généalogies

L'information fournie par une généalogie ascendante se mesure par un indice de complétude (ou d'exhaustivité) qui rapporte le nombre d'ascendants connus au nombre d'ascendants attendus, à chaque génération x. La formule est la suivante (Cazes & Cazes, 1996):

$$C_x = \frac{\text{nb ascendants connus}}{\text{nb ascendants attendus}}$$

où le nombre d'ascendants attendus à la génération x est donné par la formule 2^x , la génération des parents étant la première. Le tableau 4.1 présente les indices de complétude des généalogies utilisées. Ces indices ont été calculés pour l'ensemble des généalogies d'un échantillon; ce sont donc des indices de complétude moyens.

Les généalogies sont complètes jusqu'à la deuxième génération pour Charlevoix, Rimouski et Saguenay, jusqu'à la troisième génération pour la Beauce et jusqu'à la quatrième génération pour Terrebonne. Aux générations suivantes, l'indice de complétude diminue lentement d'abord puis rapidement à partir de la 9e génération (surtout pour Rimouski). La complétude devient pratiquement nulle après 15 générations, ce qui correspond au début du peuplement du Québec. Pour les premières générations, les mariages sont pour la plupart retrouvés dans la même région que celle du mariage du proposant. Par la suite, l'aire du lieu de mariage des ancêtres s'agrandit et parfois, il nous est impossible de retracer un mariage, ce qui entraîne la fin d'une branche de la généalogie et une complétude qui se voit diminuer. Bien entendu, la qualité des sources disponibles a un rôle majeur à jouer dans cette partie de la recherche.

Tableau 4.1

Indice de complétude (en %) à chaque génération pour l'ensemble des généalogies des cinq populations à l'étude et profondeur moyenne des généalogies

| Génération | Rimouski | Charlevoix | Saguenay | Beauce | Terrebonne |
|---------------------------------------|----------|------------|----------|--------|------------|
| 0 | 100,00 | 100,00 | 100,00 | 100,00 | 100,00 |
| 1 | 100,00 | 100,00 | 100,00 | 100,00 | 100,00 |
| 2 | 100,00 | 100,00 | 100,00 | 100,00 | 100,00 |
| 3 | 99,38 | 99,38 | 99,50 | 100,00 | 100,00 |
| 4 | 97,66 | 98,75 | 97,38 | 99,71 | 100,00 |
| 5 | 96,64 | 97,89 | 96,19 | 98,98 | 98,86 |
| 6 | 94,45 | 96,95 | 95,13 | 98,62 | 97,44 |
| 7 | 91,82 | 95,86 | 94,09 | 98,15 | 96,02 |
| 8 | 87,20 | 94,24 | 92,11 | 96,68 | 93,54 |
| 9 | 79,83 | 91,87 | 89,25 | 92,04 | 89,81 |
| 10 | 61,45 | 87,47 | 84,48 | 80,64 | 79,53 |
| 11 | 31,49 | 72,11 | 68,05 | 53,81 | 48,29 |
| 12 | 9,14 | 35,69 | 33,21 | 19,46 | 16,49 |
| 13 | 1,53 | 9,06 | 8,30 | 3,82 | 2,96 |
| 14 | 0,12 | 1,24 | 1,16 | 0,47 | 0,30 |
| 15 | 0,00 | 0,10 | 0,10 | 0,03 | 0,01 |
| 16 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| Profondeur généalogique moyenne | 9,51 | 10,81 | 10,59 | 10,42 | 10,50 |
| Écart-type | 1,98 | 1,84 | 2,10 | 1,47 | 1,72 |

Source: fichier BALSAC-RETRO

La profondeur généalogique moyenne représente quant à elle, un coefficient moyen de complétude et se veut une synthèse de l'analyse générationnelle dans une généalogie ascendante (Létourneau & Mayer, 1988). La formule pour calculer cette profondeur moyenne est la suivante:

$$P = \sum_i \frac{F_i}{T_i}$$

où i est le niveau de génération, F_i le nombre de fondateurs à la génération i et T_i le nombre d'ascendants attendus à la même génération i .

Lorsque les profondeurs généalogiques moyennes sont comparées pour les cinq populations, on remarque que la population de Rimouski présente une valeur légèrement inférieure à celles des autres populations. Ceci est attribuable au fait que les points de départ pour Rimouski (mariages qui ont eu lieu entre 1900 et 1948) sont plus anciens, en moyenne,

que ceux de Terrebonne (mariages entre 1945 et 1965), du Saguenay et de Charlevoix (naissances entre 1950 et 1971, donc mariages des parents entre 1930 et 1970) et de la Beauce (mariages entre 1905 et 1971) (tableau 3.1). En moyenne, ces quatre populations ont sensiblement la même profondeur généalogique moyenne, celles du Saguenay et de Charlevoix étant les plus élevées. En fait, c'est surtout à partir de la dixième génération que les complétudes des généalogies du Saguenay et de Charlevoix se démarquent de celles des autres régions.

4.2 Présentation des groupes de fondateurs de chaque population

Les 50 000 simulations de la transmission des gènes furent réalisées sur l'ensemble des 60 789 individus distincts apparaissant dans les 378 généalogies. Les résultats nous ont permis d'obtenir une paire d'allèles pour les 378 individus de départ. Un décompte des gènes des fondateurs distinctement étiquetés fut ensuite effectué de façon à déterminer la fréquence à laquelle les gènes de chaque fondateur se retrouvaient parmi la population contemporaine.

Le terme fondateur auquel nous faisons référence pour ces calculs concerne les individus au-delà desquels on ne possède plus d'information lors de la reconstruction généalogique. Dans ce groupe de fondateurs, il est possible de retrouver des individus qui ne sont jamais venus au Québec et d'autres qui y sont venus (les immigrants). L'information disponible sur les origines d'un immigrant est toutefois variable d'un cas à un autre. Par exemple, parfois nous possédons de l'information sur la date de mariage des parents des immigrants alors que d'autres fois, aucune information n'est disponible. Cependant, lorsque nous avons de l'information sur les origines des parents et que nous prenons en considération cette information, nous évitons d'attribuer deux paires d'allèles distincts à des frères ou soeurs qui, contrairement à leurs parents, sont venus au Québec et peuvent transmettre les mêmes allèles à leurs descendants.

De façon à bien cerner les différences entre les échantillons de chaque population, nous présentons au tableau 4.2 le nombre total de fondateurs distincts et le rapport de masculinité de même que le nombre moyen de fondateurs distincts retrouvé par ascendance dans chaque population.

Tableau 4.2
Nombre de fondateurs distincts obtenus dans chaque population

| Population | Nombre d'ascendances | Nombre de fondateurs distincts | | Nb de fondateurs distincts/ Nb d'ascendances |
|-------------------|---------------------------------|---|-----------|---|
| | | Total | Rm | |
| Rimouski | 80 | 3286 | 1,03 | 41,08 |
| Charlevoix | 80 | 3053 | 1,02 | 38,16 |
| Saguenay | 100 | 5498 | 1,02 | 54,98 |
| Beauce | 43 | 2624 | 1,02 | 61,02 |
| Terrebonne | 75 | 6468 | 1,04 | 86,24 |

Source : fichier BALSAC-RÉTRO

Rm: rapport de masculinité

À partir de ce tableau, on remarque que c'est dans la population de Terrebonne que le plus grand nombre de fondateurs distincts par ascendance est retrouvé, soit 86,24; à la suite on retrouve la Beauce avec 61,02, le Saguenay avec 54,98, Rimouski avec 41,08 et finalement Charlevoix avec 38,16. Pour la population de Terrebonne, ce résultat laisse entrevoir une plus grande variabilité génique, plus du double de celle présentée dans la population de Charlevoix. Les Rm de chaque population sont tous légèrement supérieurs à 1, ce qui indique qu'il y a davantage de fondateurs que de fondatrices dans les cinq populations. Généralement, pour chaque individu venu au Québec, le couple de parents était enregistré et le terme fondateur au sens généalogique était assigné aux deux parents. Toutefois, plusieurs situations peuvent expliquer le fait que les Rm ne soient pas égaux à 1. Par exemple, une fondatrice peut être la mère de deux enfants issus de deux fondateurs différents ou encore, un fondateur peut avoir marié une femme canadienne ou immigrante donc les parents de celle-ci sont les fondateurs. Ces situations font nécessairement augmenter les Rm.

4.3 Évaluation du nombre de fondateurs distincts communs à chaque paire de populations

Afin d'évaluer la similitude des divers corpus de fondateurs, nous avons calculé les proportions de fondateurs communs entre chaque paire de populations (tableau 4.3).

Tableau 4.3

Proportion des fondateurs de chaque population qui sont partagés avec chacune des quatre autres populations

| Populations où on retrouve les fondateurs | Populations avec lesquelles les fondateurs sont partagés | | | | |
|---|--|------------|----------|--------|------------|
| | Rimouski | Charlevoix | Saguenay | Beauce | Terrebonne |
| Rimouski | X | 0,65 | 0,87 | 0,60 | 0,84 |
| Charlevoix | 0,70 | X | 0,91 | 0,62 | 0,87 |
| Saguenay | 0,52 | 0,50 | X | 0,43 | 0,81 |
| Beauce | 0,75 | 0,73 | 0,91 | X | 0,87 |
| Terrebonne | 0,43 | 0,41 | 0,69 | 0,35 | X |

Source : fichier BALSAC-RÉTRO

Comme il est possible de le constater au tableau 4.3, c'est avec la population du Saguenay que les autres populations possèdent le plus de fondateurs en commun. Pour la Beauce et Charlevoix, 91% des fondateurs sont retrouvés au Saguenay, pour Rimouski, 87% et pour Terrebonne, 69%. 81% des fondateurs du Saguenay sont également retrouvés à Terrebonne.

Ces résultats ont été calculés à partir du nombre absolu de fondateurs retrouvés dans les généalogies de chaque population. Le nombre de généalogies reconstituées pour chaque population n'est donc pas pris en considération. Toutefois, sans tenir compte réellement des proportions de fondateurs communs obtenues, il est quand même possible de voir que plusieurs fondateurs du Saguenay sont présents dans les quatre populations tandis que plusieurs fondateurs de Terrebonne sont également retrouvés dans les généalogies du Saguenay. Les résultats suivants nous permettront également de faire des rapprochements entre les différentes populations.

4.4 Probabilités d'atteindre une fréquence de porteurs supérieure ou égale à 0,05

Le tableau 4.4 présente, pour chaque population, les 20 fondateurs ayant les probabilités les plus élevées d'atteindre une fréquence de porteurs supérieure ou égale à 0,05, soit au moins un individu sur 20 qui serait porteur d'un des deux gènes d'un fondateur. Cet intervalle de fréquence permet de faire ressortir les fondateurs importants dans chacunes des populations.

C'est dans Charlevoix que l'on retrouve les probabilités les plus élevées pour un des deux gènes d'un fondateur d'atteindre une fréquence de porteurs de 0,05. En effet, Charlevoix présente quatre fondateurs (#38, #239, #316 et #323) avec des probabilités d'environ 22% d'être présents dans la population contemporaine avec une fréquence de porteurs supérieure ou égale à 0,05. Dans la population du Saguenay, les quatre mêmes fondateurs de Charlevoix sont présents avec cette fois-ci, des probabilités se situant autour de 10%. Ainsi, Charlevoix et le Saguenay sont les deux populations présentant les meilleures chances pour leurs fondateurs d'atteindre une fréquence de porteurs supérieure ou égale à 0,05.

Fait encore plus intéressant, parmi les 20 fondateurs les plus importants de Charlevoix, 19 sont les mêmes que ceux du Saguenay. De tels résultats démontrent bien le transfert de gènes qui s'est produit entre ces deux populations au début de la colonisation du Saguenay et qui a débuté par l'arrivée en sol saguenayen de représentants de Charlevoix. La présence de maladies spécifiques, telles l'ataxie spastique ou la polyneuropathie sensori-motrice avec ou sans agénésie du corps calleux dans ces deux populations, est une conséquence de fondateurs communs.

Les probabilités calculées pour tous les fondateurs de Charlevoix sont cependant toujours supérieures à celles calculées pour les mêmes fondateurs du Saguenay. Ceci constitue une autre conséquence du fait qu'une grande partie de la population saguenayenne est issue d'ancêtres provenant de la région de Charlevoix (voir section 2.3). On estime qu'environ 60% du pool génique du Saguenay proviendrait de Charlevoix (communication personnelle d'E. Heyer et M. Tremblay, 1999).

Tableau 4.4

Les 20 probabilités les plus élevées d'atteindre une fréquence de porteurs égale ou supérieure à 0,05, dans chaque population

| | Rimouski | | Charlevoix | | Saguenay | | Beauce | | Terrebonne | |
|----|----------|-----------------------|------------|--------|----------|-----------------------|--------|-----------------------|------------|------------------------|
| | fond | prob | fond | prob | fond | prob | fond | prob | fond | prob |
| 1 | 118 | 0,0004 | 38 | 0,2162 | 239 | 0,0978 | 2301 | 0,0043 | 1161 | 1,9 x10 ⁻⁵ |
| 2 | 70 | 0,0003 | 239 | 0,2155 | 38 | 0,0972 | 326 | 0,0034 | 1155 | 1,8 x10 ⁻⁵ |
| 3 | 110 | 2,0 x10 ⁻⁶ | 316 | 0,2150 | 316 | 0,0957 | 295 | 0,0032 | 1406 | 1,9 x10 ⁻⁸ |
| 4 | 54 | 1,3 x10 ⁻⁶ | 323 | 0,2144 | 323 | 0,0940 | 70 | 0,0021 | 823 | 1,3 x10 ⁻⁸ |
| 5 | 13 | 1,1 x10 ⁻⁶ | 217 | 0,0553 | 205 | 0,0083 | 189 | 0,0014 | 1394 | 1,1 x10 ⁻⁸ |
| 6 | 211 | 1,1 x10 ⁻⁶ | 7 | 0,0531 | 180 | 0,0079 | 198 | 0,0012 | 791 | 9,2 x10 ⁻¹⁵ |
| 7 | 148 | 1,1 x10 ⁻⁶ | 197 | 0,0473 | 197 | 0,0076 | 571 | 0,0009 | 827 | 5,4 x10 ⁻¹⁵ |
| 8 | 71 | 7,9 x10 ⁻⁷ | 180 | 0,0471 | 206 | 0,0075 | 118 | 0,0007 | * | |
| 9 | 39 | 7,1 x10 ⁻⁷ | 205 | 0,0454 | 217 | 0,0054 | 372 | 0,0007 | | |
| 10 | 301 | 3,9 x10 ⁻⁷ | 206 | 0,0442 | 7 | 0,0050 | 297 | 0,0002 | | |
| 11 | 522 | 2,0 x10 ⁻⁷ | 295 | 0,0223 | 295 | 0,0036 | 5167 | 0,0002 | | |
| 12 | 460 | 9,6 x10 ⁻⁸ | 326 | 0,0218 | 326 | 0,0032 | 251 | 0,0002 | | |
| 13 | 521 | 8,6 x10 ⁻⁸ | 297 | 0,0186 | 297 | 0,0021 | 196 | 0,0002 | | |
| 14 | 295 | 7,9 x10 ⁻⁸ | 204 | 0,0163 | 196 | 0,0020 | 3574 | 0,0002 | | |
| 15 | 487 | 7,6 x10 ⁻⁸ | 196 | 0,0158 | 73 | 0,0012 | 406 | 0,0001 | | |
| 16 | 326 | 6,2 x10 ⁻⁸ | 8 | 0,0153 | 204 | 0,0012 | 219 | 0,0001 | | |
| 17 | 350 | 4,6 x10 ⁻⁸ | 74 | 0,0150 | 74 | 0,0011 | 2591 | 0,0001 | | |
| 18 | 338 | 3,6 x10 ⁻⁸ | 73 | 0,0148 | 8 | 0,0009 | 482 | 9,9 x10 ⁻⁵ | | |
| 19 | 227 | 9,5 x10 ⁻⁹ | 33 | 0,0015 | 184 | 6,7 x10 ⁻⁵ | 2593 | 9,8 x10 ⁻⁵ | | |
| 20 | 451 | 8,9 x10 ⁻⁹ | 261 | 0,0013 | 261 | 4,2 x10 ⁻⁵ | 5618 | 9,3 x10 ⁻⁵ | | |

Source : fichier BALSAC-RÉTRO

* Seulement sept fondateurs ont une probabilité non nulle d'atteindre une fréquence de porteurs égale ou supérieure à 0,05 dans Terrebonne

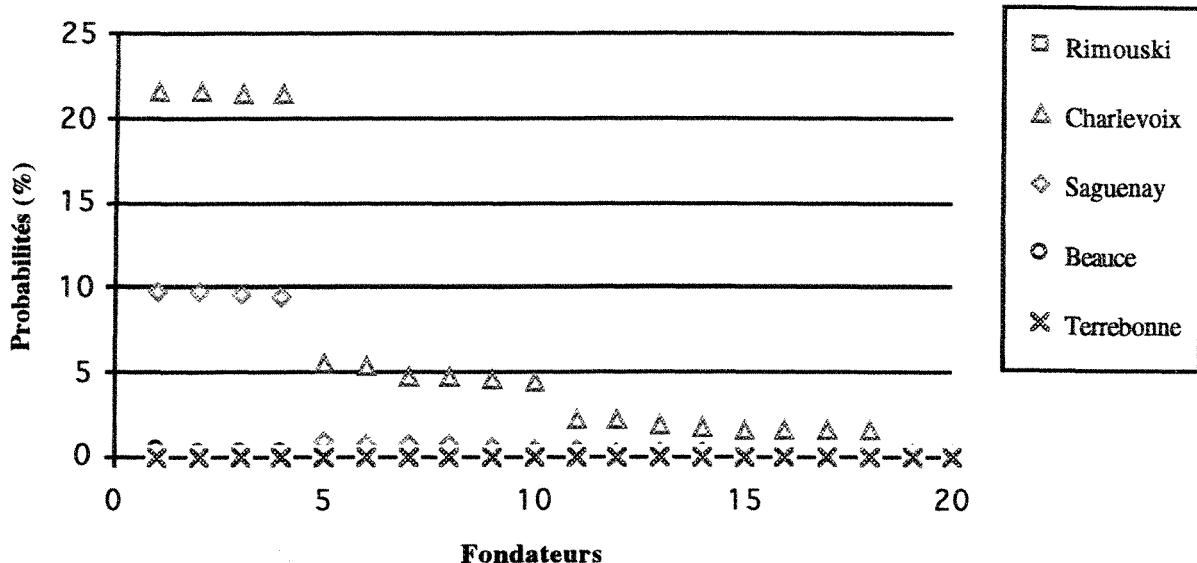
En ombragé, fondateurs retrouvés dans plus d'une population, sont exclus les fondateurs communs à Charlevoix et au Saguenay.

Dans Rimouski, les probabilités de deux fondateurs (#118 et #70) se démarquent des probabilités des autres fondateurs. Les probabilités que les gènes de ces deux fondateurs atteignent une fréquence de porteurs de 0,05 ou plus ne sont cependant que de 0,03-0,04%. Pour la Beauce, la probabilité la plus élevée pour un fondateur d'atteindre cette fréquence est de 0,43%. Les probabilités pour l'ensemble des fondateurs de cette population diminuent

lentement et aucun fondateur ne semble réellement se démarquer des autres. C'est dans la population de Terrebonne que les probabilités les plus faibles d'atteindre une fréquence de porteurs de 0,05 sont observées, soit 0,0019% pour le fondateur le plus important.

Le graphique 4.1 permet maintenant de visualiser la distribution des 20 fondateurs les plus importants de chaque population selon leur probabilité d'atteindre une fréquence de porteurs de 0,05 et plus.

Graphique 4.1. Distribution des 20 plus importants fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs supérieure ou égale à 0,05



On remarque une distribution presque identique pour Charlevoix et le Saguenay: quatre fondateurs se démarquent et ensuite les probabilités des autres fondateurs diminuent graduellement. Charlevoix présente six fondateurs dont les probabilités d'atteindre la fréquence cible sont supérieures à 5% dont quatre fondateurs avec des probabilités supérieures à 20%. Le Saguenay n'en présente que quatre avec des probabilités légèrement inférieures à 10%. Aucune des trois autres populations n'a obtenu de probabilités supérieures à 1%. Ces résultats nous permettent de constater qu'il y a très peu de chances que des gènes de fondateurs puissent être présents dans les cinq populations avec une fréquence de porteurs supérieure ou égale à 0,05 si bien sûr, on considère le hasard comme seul responsable.

Nous avons pointé les fondateurs (cellules ombragées du tableau 4.4) qui se retrouvent parmi les 20 premiers fondateurs de plus d'une population (à l'exception de ceux qui sont communs à Charlevoix et au Saguenay seulement) afin de comparer leur probabilités dans chacune des populations. À noter que Terrebonne ne présente aucun fondateur dans ses 20 premiers qui sont également retrouvés dans une autre population. Les fondateurs #295 et #326 apparaissent dans toutes les populations (sauf Terrebonne), les fondateurs #70 et #118 sont communs pour la Beauce et Rimouski et les fondateurs #196 et #297 pour Charlevoix, le Saguenay et la Beauce. Le tableau 4.5 présente les caractéristiques de ces six fondateurs.

Tableau 4.5

Caractéristiques des six fondateurs retrouvés parmi les 20 premiers de deux populations ou plus (fréquence de porteurs supérieure ou égale à 0,05)

| Fondateur | | | Probabilité d'atteindre une fréquence de porteurs $\geq 0,05$ | | | |
|-----------|------|-----------------|---|-----------------------|-----------------------|--------|
| # | sexe | date de mariage | Rimouski | Charlevoix | Saguenay | Beauce |
| 70 | F | 1616 | 0,0003 | $2,8 \times 10^{-11}$ | $2,5 \times 10^{-10}$ | 0,0021 |
| 118 | F | 1634 | 0,0004 | 0,0000* | 0,0000* | 0,0007 |
| 196 | H | 1618 | 0,0000* | 0,0158 | 0,0020 | 0,0002 |
| 297 | F | 1618 | 0,0000* | 0,0186 | 0,0021 | 0,0002 |
| 295 | F | 1620 | $7,9 \times 10^{-8}$ | 0,0223 | 0,0036 | 0,0032 |
| 326 | H | 1620 | $6,2 \times 10^{-8}$ | 0,0218 | 0,0032 | 0,0034 |

Source : fichier BALSAC-RÉTRO

* indique que la probabilité de ce fondateur ne s'est pas classée parmi les 20 premières dans cette population

Les fondatrices #70 et #118 sont retrouvées dans Rimouski avec une probabilité de 0,03-0,04% de transmettre un de leurs deux gènes jusqu'à aujourd'hui avec une fréquence de 0,05 et plus et dans la Beauce, leurs probabilités sont de 0,21% et 0,07% respectivement. Même si dans Rimouski les probabilités sont plus faibles, il faut se rapporter au tableau 4.4 pour voir que ces deux fondatrices sont néanmoins les plus importantes pour cette population. Dans la Beauce, ces fondatrices se retrouvent respectivement au 4e et au 8e rang. La présence de ces deux fondatrices dans ces deux régions voisines nous renseignent sur une migration qui s'est effectuée sur le versant sud du fleuve Saint-Laurent. On remonte également à ces fondatrices dans les autres populations mais les probabilités qu'un de leurs gènes atteigne une fréquence de 0,05 et plus sont négligeables.

Les fondateurs #196 et #297 forment un couple marié en 1618. Les gènes de ce couple obtiennent les meilleures probabilités dans les populations de Charlevoix, du Saguenay et de la Beauce mais des probabilités plutôt faibles dans celle de Rimouski . C'est dans Charlevoix que leurs probabilités sont les plus fortes, soit 1,58% et 1,86% des chances d'être retrouvés avec une fréquence de porteurs supérieure ou égale à 0,05. Les probabilités obtenues au Saguenay sont de 0,2% et dix fois moindres pour la Beauce.

Dans le cas des fondateurs #295 et #326, il s'agit à nouveau d'un couple marié en 1620 et qui a certainement eu une descendance nombreuse puisque nous pouvons retrouver les gènes de ces fondateurs dans quatre populations avec des probabilités classées parmi les 20 premières. C'est encore une fois dans Charlevoix que leurs probabilités sont les plus élevées (2,2%); suivent ensuite des probabilités presque identiques pour le Saguenay et la Beauce (0,3%) et pratiquement nulles dans Rimouski (moins de 0,00001%).

Le fait que les résultats obtenus pour les deux individus d'un même couple (#196-297 et #295-326) soient presque identiques nous confirme que 10 000 simulations étaient suffisantes. Si au contraire, de trop grandes variations avaient été observées pour un même couple de fondateurs, cela aurait indiqué que le nombre de simulations effectuées n'était pas assez élevé. Si on retrouve des probabilités plus élevées pour l'un des deux conjoints, on peut anticiper la présence de remariages et c'est d'ailleurs ce qui a été constaté dans les cas des fondatrices #70 et #118.

4.5 Probabilités d'atteindre une fréquence de porteurs se situant entre 0,01 et 0,05

Dans les résultats suivants, la fréquence cible de porteurs a été modifiée. On vise maintenant une fréquence de porteurs se situant entre 0,01 et 0,05, soit entre un individu sur 100 et un individu sur 20 qui est porteur d'un des deux gènes d'un fondateur particulier. Cet intervalle de fréquences a été retenu puisque plusieurs fréquences de porteurs de maladies génétiques se situent à l'intérieur de cet intervalle et dans le but de comparer les probabilités atteintes avec celles issues d'un intervalle plus restreint de porteurs (section 4.9).

Le tableau 4.6 présente les 20 fondateurs ayant obtenu les probabilités les plus élevées d'atteindre une fréquence de porteurs se situant entre 0,01 et 0,05. Ces fondateurs ne sont pas nécessairement les mêmes que ceux présents au tableau 4.4.

Tableau 4.6

Les 20 probabilités les plus élevées d'atteindre une fréquence de porteurs se situant entre 0,01 et 0,05, dans chaque population

| | Rimouski | | Charlevoix | | Saguenay | | Beauce | | Terrebonne | |
|----|----------|--------|------------|--------|----------|--------|--------|--------|------------|--------|
| | fond. | prob. | fond. | prob. | fond. | prob. | fond. | prob. | fond. | prob. |
| 1 | 118 | 0,4519 | 217 | 0,5060 | 323 | 0,5014 | 2301 | 0,4432 | 1161 | 0,3633 |
| 2 | 70 | 0,4303 | 7 | 0,5055 | 316 | 0,5001 | 295 | 0,4237 | 1155 | 0,3618 |
| 3 | 110 | 0,3165 | 197 | 0,4823 | 38 | 0,4996 | 326 | 0,4222 | 823 | 0,2410 |
| 4 | 7611 | 0,3145 | 206 | 0,4816 | 239 | 0,4984 | 70 | 0,4179 | 1406 | 0,2253 |
| 5 | 7815 | 0,3099 | 205 | 0,4796 | 217 | 0,4778 | 571 | 0,3836 | 1394 | 0,2184 |
| 6 | 71 | 0,3030 | 180 | 0,4795 | 7 | 0,4772 | 118 | 0,3781 | 791 | 0,1038 |
| 7 | 54 | 0,2764 | 295 | 0,4654 | 197 | 0,4701 | 189 | 0,3760 | 827 | 0,1019 |
| 8 | 148 | 0,2732 | 326 | 0,4649 | 180 | 0,4689 | 198 | 0,3721 | 110 | 0,0988 |
| 9 | 39 | 0,2728 | 297 | 0,4509 | 205 | 0,4668 | 372 | 0,3672 | 1162 | 0,0908 |
| 10 | 13 | 0,2682 | 196 | 0,4477 | 206 | 0,4650 | 5167 | 0,3493 | 295 | 0,0873 |
| 11 | 211 | 0,2679 | 204 | 0,4457 | 295 | 0,4471 | 3574 | 0,3464 | 326 | 0,0843 |
| 12 | 301 | 0,2657 | 74 | 0,4429 | 326 | 0,4447 | 406 | 0,3324 | 1157 | 0,0830 |
| 13 | 522 | 0,2619 | 73 | 0,4426 | 297 | 0,4127 | 251 | 0,3311 | 1154 | 0,0817 |
| 14 | 295 | 0,2539 | 8 | 0,4423 | 196 | 0,4105 | 433 | 0,3271 | 1156 | 0,0796 |
| 15 | 326 | 0,2505 | 323 | 0,4196 | 73 | 0,3907 | 196 | 0,3262 | 70 | 0,0725 |
| 16 | 460 | 0,2486 | 316 | 0,4194 | 204 | 0,3902 | 297 | 0,3261 | 1163 | 0,0700 |
| 17 | 521 | 0,2484 | 239 | 0,4191 | 74 | 0,3884 | 219 | 0,3214 | 2912 | 0,0689 |
| 18 | 487 | 0,2469 | 38 | 0,4181 | 8 | 0,3849 | 482 | 0,3208 | 4905 | 0,0670 |
| 19 | 4092 | 0,2375 | 71 | 0,4124 | 71 | 0,3615 | 2591 | 0,3186 | 1138 | 0,0670 |
| 20 | 451 | 0,2371 | 236 | 0,3898 | 184 | 0,3442 | 348 | 0,3162 | 2953 | 0,0668 |

Source : fichier BALSAC-RÉTRO

En ombragé, fondateurs retrouvés dans plus d'une population, sont exclus les fondateurs communs à Charlevoix et au Saguenay.

À l'exception des populations de Charlevoix et du Saguenay, les fondateurs qui avaient obtenu les meilleures probabilités pour une fréquence de porteurs supérieure ou égale à 0,05 sont les mêmes qui obtiennent les probabilités les plus élevées avec ce nouvel intervalle de fréquences. Dans l'ensemble, à cette fréquence de porteurs, les fondateurs les plus importants de chaque population atteignent des probabilités relativement élevées. Charlevoix obtient à nouveau les probabilités les plus élevées pour un fondateur d'atteindre une fréquence de

porteurs située entre 0,01 et 0,05, donc entre un individu sur 100 et un individu sur 20 qui devrait être porteur d'un des deux gènes d'un fondateur. Les probabilités atteignent 51% pour les deux plus importants fondateurs de Charlevoix (#217 et #7). À la suite de celles obtenues pour Charlevoix, on retrouve celles du Saguenay (50%), celles de Rimouski (45%), celles de la Beauce (44%) et finalement celles de Terrebonne (36%).

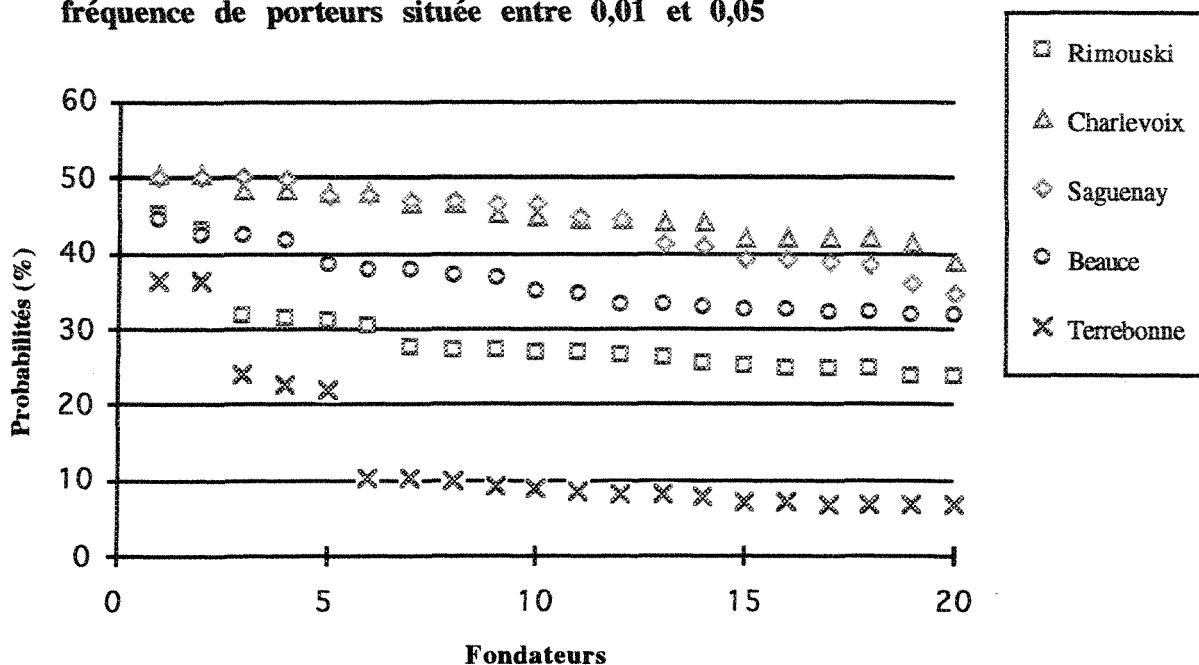
Ainsi, si un gène délétère était associé à l'un ou l'autre des principaux fondateurs, les chances que ce gène soit porté jusqu'à aujourd'hui dans les cinq populations choisies pour cette étude avec une fréquence de porteurs située entre 0,01 et 0,05 varieraient de 36 à 51%.

Les deux principaux fondateurs du Saguenay ne sont pas les mêmes que ceux de Charlevoix. D'ailleurs à cet intervalle de fréquences de porteurs, certains fondateurs du Saguenay obtiennent des probabilités supérieures à celles obtenues dans la population de Charlevoix, c'est le cas pour quatre fondateurs incluant les deux principaux (#323, 316, 38 et 239). À cet intervalle, on retrouve encore 19 fondateurs sur 20 parmi les plus importants qui sont communs au Saguenay et à Charlevoix.

Le fondateur le plus important de la Beauce (#2301) présente une probabilité de 44% d'atteindre une fréquence de porteurs se situant entre 0,01 et 0,05. Il s'agit d'une fondatrice s'étant mariée en 1654. C'est dans la population de Terrebonne que les probabilités les plus faibles associées à cet intervalle de fréquences ont été calculées. Dans la population de Terrebonne, les gènes de deux fondateurs (#1161 et #1155), un couple marié en 1656, ont 36% de chances d'être portés avec une fréquence se situant entre 0,01 et 0,05. Cette population présente trois autres fondateurs importants (#823: une femme mariée en 1629, #1406 et #1394: un couple marié en 1651) pour lesquels les probabilités se démarquent de celles des autres fondateurs de Terrebonne avec les deux fondateurs précédents.

Le graphique 4.2 montre la distribution des 20 plus importants fondateurs de chaque population selon leurs probabilités d'atteindre une fréquence de porteurs située entre 0,01 et 0,05.

Graphique 4.2. Distribution des 20 plus importants fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,01 et 0,05



Une première constatation est que les probabilités d'atteindre d'une fréquence de porteurs se situant entre 0,01 et 0,05 sont nettement supérieures à celles associées à une fréquence supérieure ou égale à 0,05 (graphique 4.1). Les populations de Rimouski, de Charlevoix, du Saguenay et de la Beauce présentent toutes des probabilités supérieures à 20% pour leurs 20 premiers fondateurs. De toute évidence, Terrebonne présente les probabilités les plus faibles d'atteindre cet intervalle de fréquences.

Toujours pour la population de Terrebonne, on remarque que les probabilités des cinq premiers fondateurs se démarquent. Les probabilités des deux principaux fondateurs de Rimouski se démarquent également mais l'écart est moins grand que celui observé pour la population de Terrebonne. Une distribution semblable est observée pour ces deux populations mais les probabilités atteintes par les fondateurs de Rimouski sont plus élevées. Dans les trois autres populations, les probabilités des fondateurs diminuent plutôt de façon progressive. La distribution des plus importants fondateurs de Terrebonne est celle qui se démarque le plus des quatre autres distributions.

La grande majorité des probabilités est concentrée dans l'intervalle de probabilités de 20 à 50%. Seules des probabilités des fondateurs de Terrebonne se situent au-dessous de cette

limite inférieure de 20% de chances d'atteindre une fréquence de porteurs située entre 0,01 et 0,05.

Afin de vérifier à nouveau le comportement de fondateurs communs à plus d'une population, nous avons reporté au tableau 4.7 les caractéristiques de huit fondateurs, dont six se retrouvaient au tableau 4.5 (ici encore, les fondateurs communs à Charlevoix et au Saguenay seulement ont été exclus).

Tableau 4.7

Caractéristiques des huit fondateurs retrouvés parmi les 20 premiers de deux populations ou plus (fréquence de porteurs située entre 0,01 et 0,05)

| Fondateur | | | Probabilité d'atteindre une fréquence de porteurs située entre 0,01 et 0,05 | | | | |
|-----------|------|-----------------|---|------------|----------|---------|-------------------------|
| # | sexe | date de mariage | Rimouski | Charlevoix | Saguenay | Beauce | Terrebonne |
| 70 | F | 1616 | 0,4303 | 0,1575* | 0,2093* | 0,4179 | 0,0725 |
| 71 | H | 1637 | 0,3030 | 0,4124 | 0,3615 | 0,1481* | 0,0007* |
| 110 | H | 1611 | 0,3165 | 0,1242* | 0,1878* | 0,2490* | 0,0988 |
| 118 | F | 1634 | 0,4519 | 0,0553* | 0,0804* | 0,3781 | 0,0167* |
| 196 | H | 1618 | 0,0053* | 0,4477 | 0,4105 | 0,3262 | 4,0 x10 ⁻⁵ * |
| 297 | F | 1618 | 0,0054* | 0,4509 | 0,4127 | 0,3261 | 3,1 x10 ⁻⁵ * |
| 295 | F | 1620 | 0,2539 | 0,4654 | 0,4471 | 0,4237 | 0,0873 |
| 326 | H | 1620 | 0,2505 | 0,4649 | 0,4447 | 0,4222 | 0,0843 |

Source : fichier BALSAC-RÉTRO

* indique que la probabilité de ce fondateur ne s'est pas classée parmi les 20 premières dans cette population

Les fondateurs #70, 110 et 118 ont des probabilités plus fortes dans Rimouski d'atteindre une fréquence de porteurs située entre 0,01 et 0,05, par la suite, leurs probabilités diminuent lentement dans la Beauce, Charlevoix et Terrebonne. Le fondateur #71 atteint la meilleure probabilité dans Charlevoix (41%) et au Saguenay (36%). Les couples #295-326 et #196-297 présentent toujours les probabilités les plus élevées dans Charlevoix pour ensuite diminuer dans les populations du Saguenay, de la Beauce, de Rimouski et de Terrebonne.

Avec une telle représentation des probabilités obtenues par différents fondateurs dans les cinq populations, on peut s'intéresser aux variations en fonction du découpage géographique. Dans le cas du fondateur #71, on observe un gradient de l'est vers le centre

puis vers l'ouest. Ainsi, des probabilités très élevées pour ce fondateur ont été obtenues dans les populations plus à l'est telles Charlevoix, Saguenay et Rimouski (entre 41 et 30%), diminuant graduellement dans la Beauce (15%) et encore davantage dans Terrebonne (0,07%). Les fondateurs # 70, 110 et 118 sont davantage présents sur la rive sud du Saint-Laurent. Pour les deux couples #196-297 et #295-326, les probabilités qu'ils atteignent dans Charlevoix, le Saguenay et la Beauce sont beaucoup plus élevées que celles qu'ils obtiennent dans Rimouski et dans Terrebonne. Dans tous les cas, les probabilités calculées pour la population de Terrebonne sont inférieures.

Ainsi, on peut voir clairement que des fondateurs sont plus associés à certaines populations. Les liens unissant les populations qui présentent des probabilités similaires pour un fondateur résultent d'un échange migratoire certain entre ces populations.

Jusqu'à maintenant, les distributions de probabilités visaient uniquement les 20 plus importants fondateurs de chaque population, et ce pour deux intervalles de fréquences de porteurs. Mais est-ce que la répartition de l'ensemble des probabilités des fondateurs est propre à chaque population ?

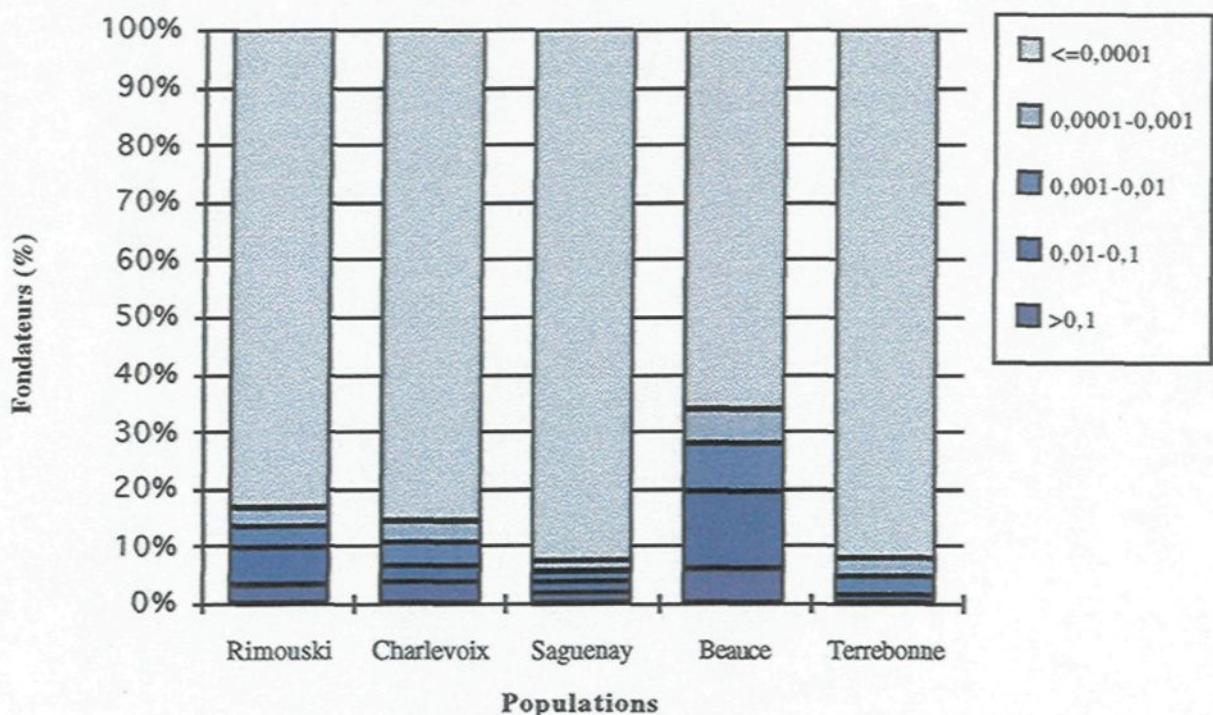
4.6 Proportion des fondateurs de chaque population atteignant une fréquence de porteurs se situant entre 0,01 et 0,05, selon les probabilités

Le graphique 4.3 montre les fondateurs de chaque population (effectifs du tableau 4.2) classés selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,01 et 0,05. Ce graphique nous permet de vérifier si les proportions de fondateurs dans chaque classe de probabilités sont les mêmes pour les cinq populations. Si tel est le cas, on pourrait s'attendre à ce que l'évolution des gènes dans ces populations soit semblable.

Il est possible de constater que la très grande majorité des probabilités atteintes par les fondateurs de chaque population sont inférieures ou égales à 0,0001. Ces fondateurs ont donc moins d'une chance sur 10 000 d'atteindre une fréquence de porteurs située entre 0,01 et 0,05. La Beauce, Charlevoix et Rimouski présentent, dans cet ordre, les proportions les plus élevées de fondateurs pour lesquels un de leurs deux gènes pourrait atteindre cette fréquence cible dans les meilleures probabilités (plus élevées que 0,1). Si on considère uniquement les deux plus importantes classes de probabilités, soit celle plus élevée que 0,1 et celle incluant les probabilités situées entre 0,01 et 0,1, la Beauce est la population dans laquelle on retrouve les

plus grandes proportions de fondateurs. Près de 20% de tous les fondateurs de la Beauce ont des probabilités plus élevées que 0,01 d'atteindre une fréquence de porteurs située entre 0,01 et 0,05. Dans les autres populations et pour les deux mêmes classes de probabilités, des proportions plus faibles sont calculées ; Rimouski (10%), Charlevoix (7%), Saguenay (3%) et Terrebonne (1%).

Graphique 4.3. Classement des fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,01 et 0,05



Source : fichier BALSAC-RÉTRO

Si on tente de faire des comparaisons, on remarque que Rimouski et Charlevoix sont les deux populations dont les profils de distributions de probabilités se ressemblent le plus. À quelques écarts près, les proportions dans chaque classe sont les mêmes pour ces deux populations. Un plus grand écart pour la classe de probabilités 0,01-0,1 est toutefois observable.

Pour les autres populations, il est plutôt difficile de faire correspondre les proportions de gènes fondateurs dans chaque classe de probabilités. Une transmission différente pour les gènes de ces populations pourrait expliquer ce résultat.

4.7 Probabilités d'atteindre une fréquence de porteurs se situant entre 0,035 et 0,05

Les probabilités d'atteindre un troisième intervalle de fréquences de porteurs furent calculées et sont présentées au tableau 4.8. À nouveau, uniquement les 20 fondateurs ayant obtenu les probabilités les plus élevées sont présentés. La fréquence de porteurs se situe maintenant entre 0,035 et 0,05, soit entre un individu sur 28 et un individu sur 20. Le choix de cet intervalle est intéressant puisque les fréquences de porteurs des maladies autosomales récessives les plus communes à la population du Saguenay se situent à l'intérieur de cet intervalle, si on exclut la fréquence de porteurs d'un individu sur 15 dans le cas de la fibrose kystique au Saguenay (Heyer, 1999).

Les populations de Charlevoix et du Saguenay sont celles qui obtiennent les meilleures probabilités d'atteindre une fréquence de porteurs située entre 0,035 et 0,05 pour leurs fondateurs. Comme dans le cas de la fréquence de porteurs supérieure ou égale à 0,05 (tableau 4.4), quatre fondateurs communs à ces deux populations (#38, 239, 316 et 323) obtiennent des probabilités nettement supérieures à toutes celles des autres fondateurs. Les probabilités atteintes sont presque les mêmes.

Les probabilités calculées pour la Beauce arrivent au troisième rang, suivent ensuite celles de Rimouski et de Terrebonne. Dans ces deux dernières populations, on remarque la présence de deux fondateurs (#118 et 70 pour Rimouski ; #1161 et 1155 pour Terrebonne) obtenant des probabilités beaucoup plus élevées comparées à celles des autres fondateurs de ces populations.

Tableau 4.8

Les 20 probabilités les plus élevées d'atteindre une fréquence de porteurs se situant entre 0,035 et 0,05, dans chaque population

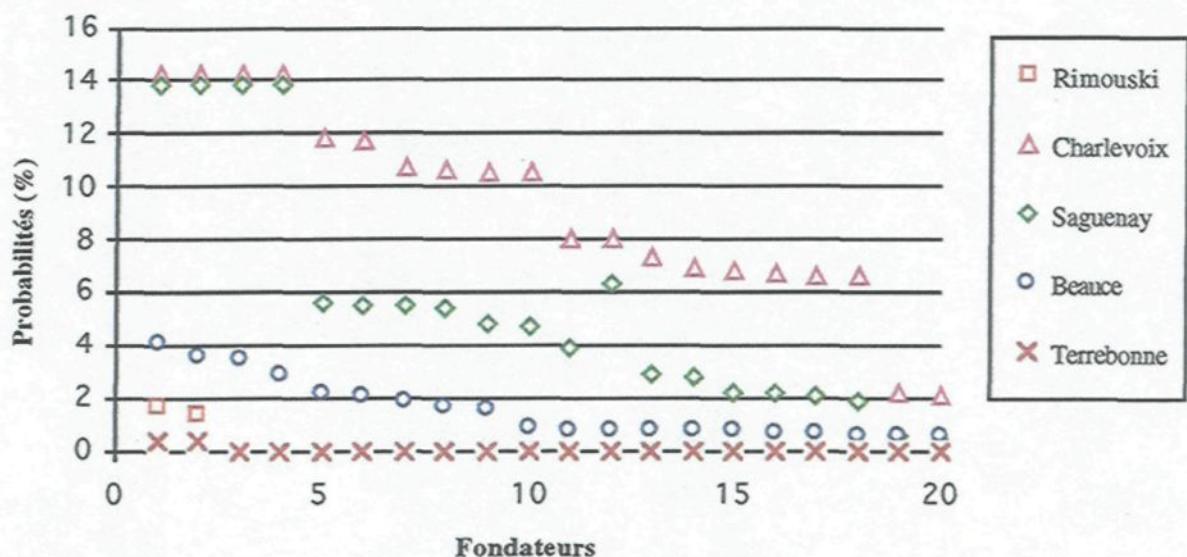
| | Rimouski | | Charlevoix | | Saguenay | | Beauce | | Terrebonne | |
|----|----------|-----------------------|------------|--------|----------|--------|--------|--------|------------|------------------------|
| | fond | prob | fond | prob | fond | prob | fond | prob | fond | prob |
| 1 | 118 | 0,0164 | 316 | 0,1428 | 38 | 0,1388 | 2301 | 0,0409 | 1161 | 0,0034 |
| 2 | 70 | 0,0133 | 239 | 0,1427 | 239 | 0,1387 | 326 | 0,0355 | 1155 | 0,0033 |
| 3 | 110 | 0,0011 | 323 | 0,1427 | 316 | 0,1385 | 295 | 0,0348 | 1406 | 0,0001 |
| 4 | 54 | 0,0008 | 38 | 0,1425 | 323 | 0,1380 | 70 | 0,0289 | 823 | 9,0 x10 ⁻⁵ |
| 5 | 148 | 0,0007 | 217 | 0,1185 | 205 | 0,0558 | 189 | 0,0220 | 1394 | 7,5 x10 ⁻⁵ |
| 6 | 13 | 0,0007 | 7 | 0,1168 | 180 | 0,0551 | 198 | 0,0206 | 791 | 7,1 x10 ⁻⁸ |
| 7 | 211 | 0,0007 | 197 | 0,1072 | 197 | 0,0545 | 571 | 0,0188 | 827 | 5,6 x10 ⁻⁸ |
| 8 | 71 | 0,0007 | 180 | 0,1066 | 206 | 0,0535 | 118 | 0,0162 | 1162 | 2,0 x10 ⁻⁸ |
| 9 | 39 | 0,0006 | 205 | 0,1053 | 217 | 0,0483 | 372 | 0,0157 | 110 | 1,9 x10 ⁻⁸ |
| 10 | 301 | 0,0004 | 206 | 0,1048 | 7 | 0,0470 | 5167 | 0,0082 | 1157 | 7,5 x10 ⁻⁹ |
| 11 | 522 | 0,0003 | 295 | 0,0803 | 295 | 0,0388 | 3574 | 0,0079 | 295 | 6,8 x10 ⁻⁹ |
| 12 | 460 | 0,0002 | 326 | 0,0796 | 326 | 0,0364 | 297 | 0,0077 | 1154 | 6,7 x10 ⁻⁹ |
| 13 | 521 | 0,0002 | 297 | 0,0729 | 297 | 0,0286 | 251 | 0,0077 | 1156 | 5,0 x10 ⁻⁹ |
| 14 | 295 | 0,0002 | 204 | 0,0688 | 196 | 0,0281 | 196 | 0,0076 | 326 | 4,8 x10 ⁻⁹ |
| 15 | 487 | 0,0002 | 196 | 0,0683 | 73 | 0,0218 | 406 | 0,0074 | 1163 | 1,2 x10 ⁻⁹ |
| 16 | 326 | 0,0002 | 8 | 0,0667 | 204 | 0,0213 | 219 | 0,0064 | 2912 | 1,1 x10 ⁻⁹ |
| 17 | 350 | 0,0001 | 74 | 0,0664 | 74 | 0,0207 | 2591 | 0,0063 | 70 | 9,7 x10 ⁻¹⁰ |
| 18 | 338 | 0,0001 | 73 | 0,0660 | 8 | 0,0192 | 482 | 0,0061 | 2953 | 8,3 x10 ⁻¹⁰ |
| 19 | 7611 | 9,4 x10 ⁻⁵ | 33 | 0,0221 | 184 | 0,0055 | 2593 | 0,0060 | 1138 | 7,6 x10 ⁻¹⁰ |
| 20 | 7815 | 8,1 x10 ⁻⁵ | 236 | 0,0207 | 261 | 0,0039 | 433 | 0,0059 | 3102 | 7,6 x10 ⁻¹⁰ |

Source : fichier BALSAC-RÉTRO

En ombragé, fondateurs retrouvés dans plus d'une population, sont exclus les fondateurs communs à Charlevoix et au Saguenay.

Une représentation graphique de la distribution des fondateurs nous permet de voir davantage les écarts entre les probabilités ainsi que les différences ou les ressemblances entre les cinq populations. Le graphique 4.4 nous présente donc la distribution des 20 plus importants fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,035 et 0,05.

Graphique 4.4. Distribution des 20 plus importants fondateurs de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,035 et 0,05



Sur ce graphique, on constate que les distributions des probabilités atteintes sont tout de même assez différentes de celles qui furent calculées pour une fréquence de porteurs située entre 0,01 et 0,05. Les populations de Charlevoix et du Saguenay ont des distributions distinctes, un peu semblables à celles qui avaient été obtenues pour une fréquence de porteurs supérieure ou égale à 0,05.

Ce qui étonne à cette fréquence de porteurs située entre 0,035 et 0,05, c'est le fait que les probabilités atteintes par les quatre principaux fondateurs de Charlevoix et du Saguenay se superposent. Ainsi, les gènes des ces quatre fondateurs ont évolué de façon similaire puisqu'ils ont sensiblement les mêmes chances d'être portés jusqu'à nos jours entre 0,035 et 0,05. Les probabilités des fondateurs suivants de ces populations ne sont pas les mêmes mais elles décroissent de la même façon.

Les probabilités atteintes par les fondateurs de la Beauce sont plutôt faibles, aucune ne dépasse 5%. Pour ce qui est des populations de Rimouski et de Terrebonne, les chances sont encore plus faibles que les gènes de leurs fondateurs soient portés avec une fréquence située entre 0,035 et 0,05 dans les populations contemporaines.

Les fondateurs que nous avons utilisés jusqu'à présent sont les bouts des généalogies. Dans la majorité des cas (environ 85%), nous n'avions pas la date de mariage de ces fondateurs d'où l'incapacité d'établir une distribution des années de mariage des fondateurs de chaque population à moins bien sûr d'en faire une estimation d'après l'année de mariage de leurs enfants. À partir de ce contexte, on peut poser la question suivante: quelle serait l'ampleur des variations dans la distribution des probabilités d'atteindre une fréquence de porteurs cible si on ne considérait que les individus qui sont venus au Québec (les immigrants) et qui sont en réalité, les fondateurs démographiques ?

4.8 Comparaison des deux groupes d'individus pour lesquels les simulations ont été effectuées

Nous avons réalisé une autre série de 50 000 simulations en ne gardant cette fois-ci que les individus qui sont venus au Québec. Afin de mieux saisir en quoi ces simulations diffèrent des précédentes, nous décrivons au tableau 4.9 les deux groupes d'individus sélectionnés, selon divers critères.

Tableau 4.9
Description des deux groupes d'individus

| Critères | Fondateurs série 1 | Immigrants série 2 |
|--|-----------------------|-----------------------|
| Nombre total d'individus pour lesquels les simulations ont été effectuées | 8077 | 5056 |
| Statuts des individus éliminés | aucun | 2 et 50 |
| Proportion des fondateurs ou immigrants avec date de mariage inconnue | 86,9 % | 10,7 % |
| Rapport de masculinité (Rm) | 1,03 | 2,08 |

Source : fichier BALSAC-RÉTRO

Tout d'abord, on remarque que le nombre d'individus pour lesquels nous avons simulé la transmission de deux de leurs gènes diminue dans le second groupe pour atteindre 5 056

individus. En utilisant la variable statut du fichier BALSAC-RETRO, nous avons sélectionné des ancêtres. Ainsi, les individus de statut 2 ou supérieur ou égal à 50 (annexe) ont été écartés pour la seconde série de simulations puisque nous considérions qu'ils n'étaient jamais venus au Québec. Nous désirions avoir que les immigrants qui sont en réalité les fondateurs démographiques. 1 360 individus de la deuxième série (26,9%) ont également été retrouvés dans la première série. Ces individus, des immigrants dont nous n'avions aucune information sur leurs parents, ont été considérés dans la première série comme fondateur. Des écarts importants ont été notés entre les probabilités obtenues par les même individus dans les deux différentes séries de simulations. Dans la seconde série, la proportion des individus dont nous n'avons pu retrouver la date de mariage n'est que de 10,7% alors qu'elle était de 86,9% dans la première série. Cette différence de proportion n'est pas surprenante puisque dans la deuxième série, on ne conserve que les individus venus au Québec alors que dans la première série, la plus grande partie des fondateurs ne sont jamais venus au Québec donc leur date de mariage n'était pas facile à retracer. Finalement, le tableau 4.9 démontre que le rapport de masculinité est deux fois plus élevé dans le second groupe, atteignant deux hommes pour une femme. Ceci est en accord avec ce qui a déjà été démontré pour les immigrants arrivés au Québec au 17e siècle (Charbonneau et alii, 1987).

4.9 Probabilités, pour les immigrants, d'atteindre une fréquence de porteurs située entre 0,035 et 0,05, dans chaque population

Au tableau 4.10, nous reprenons les probabilités obtenues pour les 20 plus importants fondateurs de chaque population d'atteindre une fréquence de porteurs située entre 0,035 et 0,05 (tableau 4.8) et nous comparons ces probabilités avec celles obtenues par les 20 plus importants immigrants de la seconde série. Seule une comparaison avec cette fréquence de porteurs est présentée puisqu'elle nous apparaissait être la plus intéressante. Contrairement aux tableaux précédents, les numéros des fondateurs et des immigrants ne sont pas inscrits puisqu'ils ne réfèrent pas aux mêmes individus. Il est quand même possible de retrouver au total 1 360 immigrants qui sont en même temps des fondateurs mais leur numéro ne correspond pas d'une série à l'autre, conséquence d'une nouvelle numérotation.

Tableau 4.10

Comparaison des probabilités, pour les 20 plus importants fondateurs et immigrants de chaque population, d'atteindre une fréquence de porteurs se situant entre 0,035 et 0,05

| | Rimouski | | Charlevoix | | Saguenay | | Beauce | | Terrebonne | |
|----|----------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----------------------|----------------------|
| | prob. série 1 | prob. série 2 | prob. série 1 | prob. série 2 | prob. série 1 | prob. série 2 | prob. série 1 | prob. série 2 | prob. série 1 | prob. série 2 |
| 1 | 0,0164 | 0,0167 | 0,1428 | 0,2004 | 0,1388 | 0,1961 | 0,0409 | 0,0860 | 0,0034 | 0,0042 |
| 2 | 0,0133 | 0,0145 | 0,1427 | 0,1905 | 0,1387 | 0,1960 | 0,0355 | 0,0487 | 0,0033 | 0,0038 |
| 3 | 0,0011 | 0,0075 | 0,1427 | 0,1897 | 0,1384 | 0,1934 | 0,0348 | 0,0478 | 0,0001 | $3,7 \times 10^{-5}$ |
| 4 | 0,0008 | 0,0069 | 0,1425 | 0,1534 | 0,1380 | 0,1428 | 0,0288 | 0,0470 | $9,0 \times 10^{-5}$ | $3,0 \times 10^{-5}$ |
| 5 | 0,0007 | 0,0066 | 0,1185 | 0,1532 | 0,0558 | 0,1404 | 0,0220 | 0,0463 | $7,5 \times 10^{-5}$ | $7,9 \times 10^{-5}$ |
| 6 | 0,0007 | 0,0064 | 0,1168 | 0,1421 | 0,0551 | 0,1029 | 0,0206 | 0,0441 | $7,1 \times 10^{-8}$ | $7,0 \times 10^{-6}$ |
| 7 | 0,0007 | 0,0041 | 0,1073 | 0,1407 | 0,0545 | 0,0997 | 0,0188 | 0,0416 | $5,6 \times 10^{-8}$ | $5,2 \times 10^{-6}$ |
| 8 | 0,0007 | 0,0034 | 0,1066 | 0,0885 | 0,0535 | 0,0402 | 0,0162 | 0,0410 | $2,0 \times 10^{-8}$ | $3,6 \times 10^{-6}$ |
| 9 | 0,0006 | 0,0021 | 0,1053 | 0,0854 | 0,0483 | 0,0371 | 0,0157 | 0,0390 | $1,9 \times 10^{-8}$ | $8,2 \times 10^{-7}$ |
| 10 | 0,0004 | 0,0020 | 0,1048 | 0,0800 | 0,0470 | 0,0370 | 0,0082 | 0,0378 | $7,5 \times 10^{-9}$ | $6,8 \times 10^{-7}$ |
| 11 | 0,0003 | 0,0018 | 0,0803 | 0,0783 | 0,0388 | 0,0302 | 0,0079 | 0,0365 | $6,8 \times 10^{-9}$ | $5,8 \times 10^{-7}$ |
| 12 | 0,0002 | 0,0018 | 0,0796 | 0,0769 | 0,0364 | 0,0287 | 0,0077 | 0,0363 | $6,7 \times 10^{-9}$ | $5,7 \times 10^{-7}$ |
| 13 | 0,0002 | 0,0018 | 0,0729 | 0,0754 | 0,0286 | 0,0265 | 0,0077 | 0,0350 | $5,0 \times 10^{-9}$ | $5,3 \times 10^{-7}$ |
| 14 | 0,0002 | 0,0018 | 0,0688 | 0,0711 | 0,0281 | 0,0247 | 0,0076 | 0,0327 | $4,8 \times 10^{-9}$ | $3,3 \times 10^{-7}$ |
| 15 | 0,0002 | 0,0013 | 0,0683 | 0,0587 | 0,0218 | 0,0238 | 0,0074 | 0,0286 | $1,2 \times 10^{-9}$ | $2,3 \times 10^{-8}$ |
| 16 | 0,0002 | 0,0013 | 0,0667 | 0,0552 | 0,0213 | 0,0233 | 0,0064 | 0,0217 | $1,1 \times 10^{-9}$ | $1,2 \times 10^{-8}$ |
| 17 | 0,0001 | 0,0012 | 0,0664 | 0,0540 | 0,0207 | 0,0186 | 0,0063 | 0,0198 | $9,7 \times 10^{-10}$ | $9,7 \times 10^{-9}$ |
| 18 | 0,0001 | 0,0008 | 0,0660 | 0,0530 | 0,0192 | 0,0172 | 0,0061 | 0,0188 | $8,3 \times 10^{-10}$ | $6,4 \times 10^{-9}$ |
| 19 | $9,4 \times 10^{-5}$ | 0,0008 | 0,0221 | 0,0526 | 0,0055 | 0,0155 | 0,0060 | 0,0187 | $7,6 \times 10^{-10}$ | $4,6 \times 10^{-9}$ |
| 20 | $8,1 \times 10^{-5}$ | 0,0007 | 0,0207 | 0,0508 | 0,0039 | 0,0109 | 0,0059 | 0,0187 | $7,6 \times 10^{-10}$ | $4,0 \times 10^{-9}$ |

Source : fichier BALSAC-RÉTRO

Dans un premier temps, les probabilités obtenues par les immigrants sont toutes plus élevées que celles obtenues par les fondateurs de la première série. Ce résultat n'est pas surprenant et s'explique par la constitution du fichier de généalogies. Dans ce fichier, on devait retrouver les parents de chaque immigrant qui sont les fondateurs de la série 1. Ainsi, on doit s'attendre à trouver des probabilités deux fois plus élevées pour les immigrants comparativement à celles des fondateurs, à condition que ces fondateurs ne soient parents que d'un seul enfant par généalogie. Toutefois, les parents des immigrants n'ont pas tous été

retracés et on retrouve le plus souvent des probabilités pour les immigrants qui ne sont pas le double de celles de leurs parents. À titre d'exemple, les fondateurs #38 et 316 du Saguenay, parents de l'immigrant #129, ont 13,9% et leur enfant 19,6% de chances d'atteindre une fréquence de porteurs située entre 0,035 et 0,05. On peut supposer que les fondateurs en question, c'est-à-dire les parents des immigrants, ont d'autres enfants parmi les immigrants.

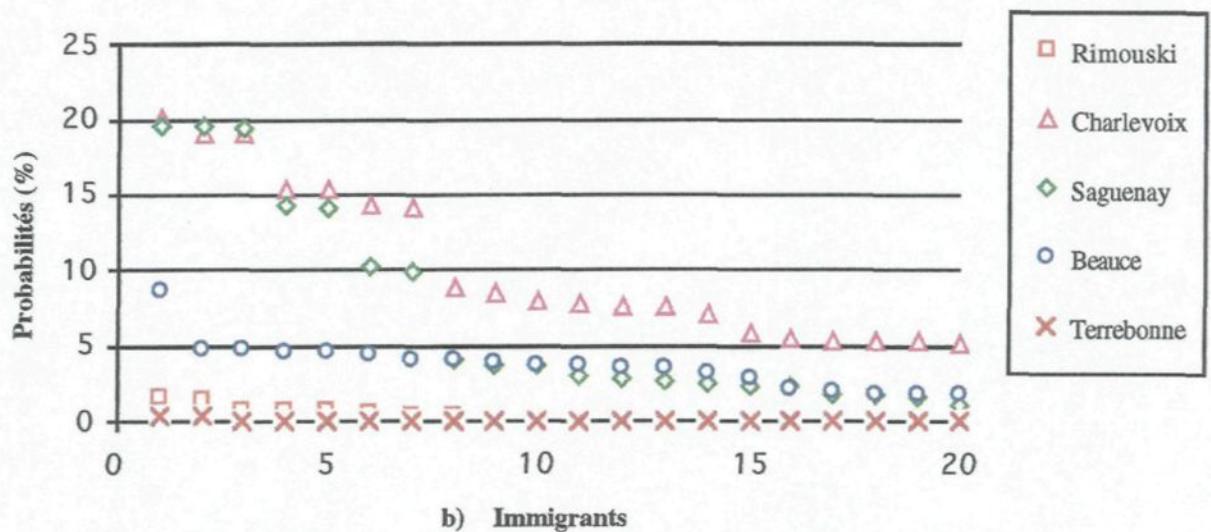
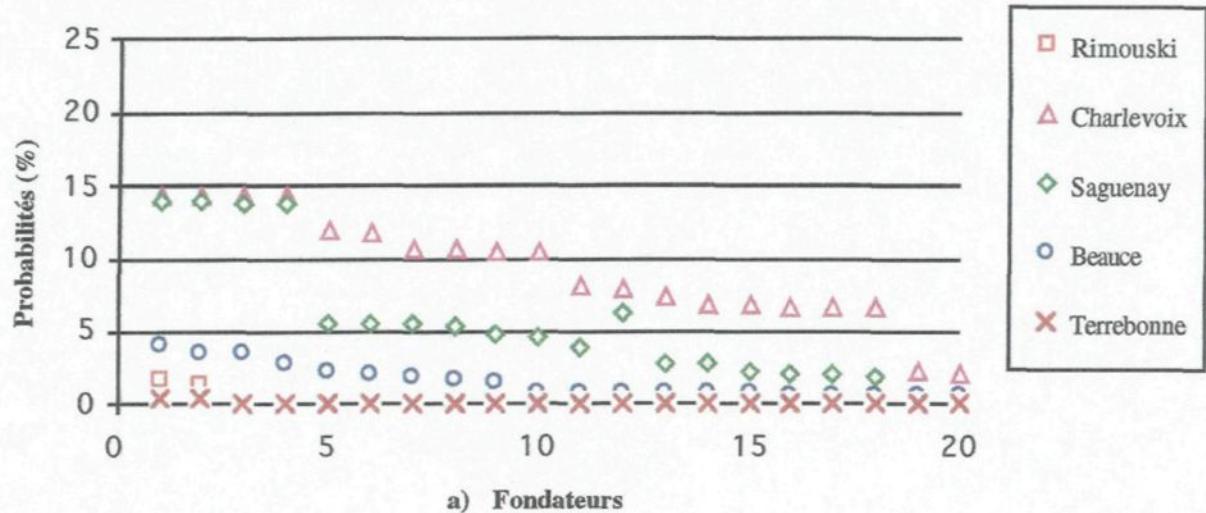
La probabilité la plus élevée pour un immigrant est retrouvée dans la population de Charlevoix (20%). Les probabilités du Saguenay, de la Beauce, de Rimouski et de Terrebonne arrivent par la suite. Le même ordre que celui établi avec les probabilités des fondateurs à la même fréquence est respecté.

Sur le graphique 4.5, on compare les distributions des 20 plus importants fondateurs et immigrants de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,035 et 0,05.

On constate rapidement qu'à une fréquence de porteurs inchangée, les distributions des fondateurs et des immigrants selon leur probabilité sont similaires sur certains points. D'abord l'ordre décroissant des probabilités est le même dans les deux groupes. De plus, l'étalement des fondateurs et des immigrants de la Beauce, de Rimouski et de Terrebonne est sensiblement le même. Par ailleurs, de petites variations sont observées pour les populations de Charlevoix et du Saguenay. Deux immigrants du Saguenay obtiennent des probabilités supérieures à celles de Charlevoix mais par la suite, les probabilités des immigrants de Charlevoix continuent d'être plus élevées.

Il aurait été intéressant de toujours observer des probabilités deux fois plus élevées pour les immigrants de chaque population comparativement aux probabilités des fondateurs. Ce n'est toutefois pas ce qui a été obtenu. Ceci est attribuable à la qualité de l'information qui varie d'un immigrant à un autre. Cet exercice nous aura cependant permis de comprendre le bien fondé de l'utilisation de la plus grande quantité d'informations possible de façon à éviter la redondance d'individus. Par exemple, certains fondateurs peuvent être parents de plus d'un immigrant à la fois et on se retrouve ainsi à attribuer plus d'une paire de gènes à des immigrants qui partagent en réalité une partie du même patrimoine génétique.

Graphique 4.5. Comparaison des distributions des 20 plus importants a) fondateurs et b) immigrants de chaque population selon leur probabilité d'atteindre une fréquence de porteurs située entre 0,035 et 0,05



CONCLUSION GÉNÉRALE

Par cette étude, nous désirions vérifier la présence de différences dans l'évolution des gènes fondateurs de cinq populations du Québec. Des simulations sur des généalogies ascendantes de ces populations nous ont démontré des ressemblances pour les populations de Charlevoix et du Saguenay mais également des différences notables pour Terrebonne et les autres populations.

Tout d'abord, nous avons présenté le nombre de fondateurs distincts par ascendance pour chaque population. Le plus grand nombre de fondateurs distincts par ascendance fut obtenu pour la population de Terrebonne et à l'opposé, le plus petit nombre fut obtenu pour la population de Charlevoix. Même si le nombre de généalogies de départ n'était pas le même pour ces deux populations, ce résultat démontre que la population de Terrebonne est davantage diversifiée par rapport aux autres populations et que Charlevoix l'est moins.

Nous avons par la suite calculé des probabilités d'occurrence pour trois fréquences de porteurs. À une fréquence de porteurs supérieure ou égale à 0,05, Charlevoix (22%) et le Saguenay (10%) obtiennent les probabilités les plus élevées et ce, pour quatre fondateurs communs aux deux populations. Pour les autres populations, il y très peu de chances que les gènes des fondateurs puissent être présents à cette fréquence de porteurs. Une distribution similaire des 20 plus importants fondateurs de Charlevoix et du Saguenay fut observée tandis que pour les trois autres populations, les probabilités atteintes par leurs fondateurs étaient toutes inférieures à 1%.

À une fréquence de porteurs se situant entre 0,01 et 0,05, les probabilités calculées pour l'ensemble des fondateurs de chaque population étaient relativement élevées. Charlevoix et le Saguenay présentaient toujours les probabilités supérieures même si les trois autres populations atteignaient également des probabilités élevées. À l'exception de Terrebonne, les vingt premiers fondateurs de chaque population obtiennent tous des probabilités supérieures à 20% d'atteindre une fréquence de porteurs se situant entre 0,01 et 0,05.

Lorsque nous avons comparé les probabilités obtenues par un même fondateur dans les différentes populations, nous avons constaté que la géographie était un facteur important dans la variation des probabilités. Ainsi, entre les populations plus à l'est du Québec (Rimouski, Charlevoix, Saguenay) et celle plus à l'ouest (Terrebonne), il existe un gradient est-ouest pour

certains fondateurs. Un obstacle naturel tel le fleuve Saint-Laurent a certainement joué un rôle de grande importance dans ce gradient.

La grande majorité des fondateurs de chaque population ont moins d'une chance sur 10 000 d'atteindre une fréquence de porteurs située entre 0,01 et 0,05. Même si c'est Charlevoix qui atteint les probabilités les plus élevées, c'est la population de la Beauce qui compte le plus de fondateurs atteignant des probabilités supérieures à 0,1 ou 10%. Pour l'étalement des probabilités de l'ensemble des fondateurs d'une population, Rimouski et Charlevoix sont les deux populations dont les proportions de fondateurs dans chaque intervalle de probabilités se ressemblent le plus. Ceci indique une évolution similaire des gènes de ces populations. Ainsi, ces deux populations présentent sensiblement les mêmes proportions de fondateurs ayant les mêmes probabilités d'avoir transmis leurs gènes jusqu'à aujourd'hui selon une fréquence se situant entre 0,01 et 0,05.

Pour l'atteinte d'une fréquence de porteurs se situant entre 0,035 et 0,05, Charlevoix et le Saguenay obtiennent toujours les meilleures probabilités pour leurs fondateurs.

Finalement, les simulations effectuées avec les immigrants n'ont pas clairement démontré les résultats anticipés, soit des probabilités deux fois plus élevées pour des immigrants dont les parents étaient considérés comme fondateurs dans les premières simulations. Un tel résultat aurait été obtenu si tous les parents des immigrants avaient pu être retracés et si chaque parent ne présentait qu'un seul enfant parmi les immigrants.

Ce mémoire aura donc servi à faire ressortir des différences notables dans l'évolution des gènes fondateurs pour cinq populations du Québec, différences pouvant se traduire par des incidences de maladies héréditaires plus ou moins élevées dans certaines régions. Même si les simulations ont été effectuées pour des gènes neutres, les résultats n'auraient pas été très différents avec des gènes de maladies récessives puisque ces gènes n'atteignent presque jamais une fréquence assez élevée pour que la sélection agisse (Heyer, 1999).

Pour faire suite à cette étude, il serait intéressant de traiter d'autres régions du Québec dans le but d'avoir une vision de l'évolution génique de cette province dans son intégralité. Il faudra s'assurer que le nombre de généalogies par région soit suffisant et représentatif pour chacune d'elles et qu'une constance soit respectée dans l'attribution des statuts des individus lors des reconstructions généalogiques.

RÉFÉRENCES

Andermann, E., Scriver, C. R., Wolfe, L. S., Dansky, L. & Andermann, F. (1977). Genetic Variants of Tay-Sachs Disease : Tay-Says Disease and Sandhoff's Disease in French Canadians, Juvenile Tay-Sachs Disease in Lebanese Canadians , and a Tay-Sachs Screening Program in the French-Canadian Population. *Program Clinic of Biology Research*, 18 : 161-188.

Bélanger, F., Berberi, S., Breton, J.R., Carrier, D. & Lessard, R. (1990). *La Beauce et les Beaucerons. Portrait d'une région 1737-1987*. Québec: Société du patrimoine des Beaucerons.

Bergeron, M.J. (en préparation). *Étude généalogique de la schizophrénie et de la psychose maniaco-dépressive dans la région de la Beauce*. Mémoire de maîtrise en médecine expérimentale, Université du Québec à Chicoutimi.

Blanchard, R. (1935). *L'Est du Canada français. Tome I*. Paris et Montréal: Masson et Beauchemin.

Bluteau, M.A. & Gauthier, S. (1984). Éléments de recherche en vue d'une étude historique du peuplement de Charlevoix. *Revue de la société d'histoire de Charlevoix*, 2(1), 4-8.

Boleda, M. (1984). Les migrations au Canada sous le régime français (1608-1760). *Cahiers québécois de démographie*, 13-1: 23-29.

Bouchard, G. (1987). *Le Centre SOREP et le fichier-réseau BALSAC Présentation générale*. Document de l'IREP (I-C-64).

Bouchard, G. (1990). *Extension du fichier de population BALSAC à l'ensemble du Québec, du 17e au 20e siècle. Les données du projet*. Document de l'IREP (I-C-90).

Bouchard, G. (1998). *Rapport annuel 1997-1998 du Projet BALSAC*.

Bouchard, G. & De Braekeleer, M., dir. (1990). *Histoire d'un génome. Population et génétique dans l'est du Québec*. Sillery, Québec: Presses de l'Université du Québec.

Bouchard, G. & De Braekeleer, M. (1992). *Pourquoi des maladies héréditaires?* Sillery, Québec: Septentrion.

Bouchard, G., Laberge, C. & Scriver, C.R. (1988). Reproduction démographique et transmission génétique dans le nord-est de la province du Québec (18e-20e siècles). *European Journal of Population*, 4, 39-67.

Bouchard, G., Laberge, C., Scriver, C.R., Glorieux, F., Declos, M., Bergeron, L., Larochelle, J. & Mortezaï, S. (1984). Étude démographique et généalogique de deux maladies héréditaires au Saguenay. *Cahiers québécois de démographie*, 13(1), 117-137.

Bouchard, J.P., Barbeau, A., Bouchard, R., Paquet, M. & Bouchard, R.W. (1979). A Cluster of Friedreich's Ataxia in Rimouski. *Canadian Journal of Neurological Sciences*, 6, 205-208.

Cannings, C., Thompson, E.A. & Skolnick, M.H. (1976). Recursive Derivation of Likelihoods on Pedigrees of Arbitrary Complexity. *Advances in Applied Probability*, 8, 622-625.

Cannings, C., Thompson, E.A. & Skolnick, M.H. (1978). Probability Functions on Complex Pedigrees. *Advances in Applied Probability*, 10, 26-61.

Cavalli-Sforza, L.L. & Zei, G. (1967). Experiments with an Artificial Population. *Proc IIIrd Int. Congr. Hum. Genet.*, Crow, J.F. & Neel, J.V. (Eds.) Plenary Sessions and Symposia (pp. 473-478). Baltimore: Johns Hopkins Press.

Cazes, M.H. & Cazes, P. (1996). Comment mesurer la profondeur généalogique d'une ascendance? *Population*, 1, 117-140.

Charbonneau, H., Desjardins, B., Guillemette, A., Landry, Y., Legaré, J. & Nault, F. (1987). *Naissance d'une population. Les Français établis au Canada au XVIIe siècle*. Paris et Montréal: Institut national d'études démographiques et Presses de l'Université de Montréal.

Commission de toponymie du Québec (1996). *Noms et lieux du Québec. 2e éd.* Québec: Publications du Québec.

Cotterman, C.W. (1940). *A Calculus for Statistico-Genetics*. Dissertation, Ohio State University. Reprinted 1975 in Ballonoff, P. (ed.), *Genetics and Social Structure: Mathematical Structuralism in Population Genetics and Social Theory* (pp.157-272). Stroudsburg, Pennsylvania: Dowden, Hutchinson & Ross.

Cotterman, C.W. (1941). Relatives and Human Genetic Analysis. *Science Monthly*, 53, 227-234.

De Braekeleer, M. & Larochelle, J. (1990). Genetic Epidemiology of Hereditary Tyrosinemia in Quebec and in Saguenay-Lac-St-Jean. *American Journal of Human Genetics*, 47, 302-307.

De Braekeleer, M. & Larochelle, J. (1991). Population Genetics of Vitamin D-Dependant Rickets in Northeastern Quebec. *Annals of Human Genetics*, 55, 283-290.

De Braekeleer, M., Dionne, C., Gagné, C., Julien, P., Brun, D., Ven Murthy, M.R. & Lupien, P.J. (1991). Founder Effect in Familial Hyperchylomicronemia Among French Canadians of Quebec. *Human Heredity*, 41, 168-173.

Dyke, B., Gage, T.B. & Mamelka, P.M. (1986). Genetic Effects of a Maximal Harvest Strategy. *Zoo Biology*, 5, 203-214.

Edwards, A.W.F. (1967). Automatic Construction of Genealogies from Phenotypic Information (AUTOKIN). *Bulletin of European Society of Human Genetics* 1, 42-43.

Edwards, A.W.F. (1968). Simulation Studies of Genealogies (Abstract). *Heredity*, 23, 628.

Edwards, A.W.F. (1969). *Computer Applications in Genetics* (ed. Morton, N.E.), Honolulu: University of Hawaii Press, 81.

Edwards, A.W.F. (1987). *Computers and Genealogies. (Account)*. University of Cambridge.

Edwards, J.H. (1967). Linkage Studies of Whole Populations. *Proc IIIrd Int. Congr. Hum. Genet.*, Crow, J.F. & Neel, J.V. (Eds.) Plenary Sessions and Symposia (pp. 479-482). Baltimore: Johns Hopkins Press.

Gagnon, N. (1998). *Mesure et analyse de l'effet fondateur dans les populations de Charlevoix et du Bas-Saint-Laurent..* Mémoire de maîtrise en médecine expérimentale, Université du Québec à Chicoutimi.

Gareau, C.A. (1927). *Aperçu historique de Terrebonne.* Montréal: Imprimerie des Sourds-Muets.

Gauvreau, D. & Bourque, M. (1988). Mouvements migratoires et familles: le peuplement du Saguenay avant 1911. *Revue d'histoire de l'Amérique française*, 42, 167-192.

Giroux & Joyal (1984). *Géographie du Québec et du Canada.* Outremont : Éditions Lidec.

Guérin, M. (1989). *Peuplement et dynamique démographique de Charlevoix des origines à aujourd'hui.* Mémoire de maîtrise en études régionales, Université du Québec à Chicoutimi.

Guillemette, A. (1996). *Des mères et des enfants. La fécondité et la périnatalité dans Lanaudière de 1979 à 1993.* Régie Régionale de la Santé et des Services Sociaux Lanaudière, Direction de la santé publique.

Guillemette, A. (1997). *Un destin commun. La mortalité selon la cause dans Lanaudière de 1979 à 1993.* Régie Régionale de la Santé et des Services Sociaux Lanaudière, Direction de la santé publique.

Guillemette, A. & St-Cyr, J.P. (1995). *Mosaïque lanaudoise. Profil démographique, socio-économique et sanitaire dans la population lanaudoise.* Régie Régionale de la Santé et des Services Sociaux Lanaudière, Direction de la santé publique.

Hartl, D.L. (1994). *Génétique des populations.* Paris: Flammarion.

Hausfater, G., Saunders, C.D. & Chapman, M. (1981). Some Applications of Computer Models to the Study of Primate Mating and Social Systems. In R.D. Alexander & D.W. Tinkle (Eds), *Natural Selection and Social Behavior* (pp.345-360). Oxford: Blackwell Scientific Publications.

Heyer, E. (1999). One founder-One Gene Hypothesis in a New Expanding Population: The Case of the Saguenay (Québec, Canada). *Human Biology*, 71(1), 99-109.

Hilden, J. (1967). Formalized and Automated Pedigree Processing. *Bulletin of European Society of Human Genetics*, 1, 41.

Hilden, J. (1970). GENEX- An Algebraic Approach to Pedigree Probability Calculus. *Clinical Genetics*, 1, 319-348.

Institut généalogique Drouin. *Répertoire alphabétique des mariages des Canadiens français, 1760-1935*, 64 volumes

Jetté, R. (1983). *Dictionnaire généalogique des familles du Québec: des origines à 1730.* Col. Programme de recherche en démographie historique de l'Université de Montréal, Montréal: Presses de l'Université de Montréal.

Jomphe, M. & Casgrain, B. (1997). *Base de données généalogiques RETRO: structure des données.* Document de l'IREP (III-C-97).

Jomphe, M. & Vigneault, A. (1992). *La reconstruction généalogique à l'aide du fichier RETRO: guide de l'usager SOREP*. Document de l'IREP (III-C-88).

Lavoie, Y. (1972). *L'émigration des Canadiens aux États-Unis avant 1930*. Collection Démographie canadienne, Montréal: Presses de l'Université de Montréal.

Létourneau, E., & Mayer, F.M. (1988). Un modèle d'analyse de généalogie ascendante. *Cahiers québécois de démographie*, 17(2), 213-231.

MacCluer, J.W., VandeBerg, J.L., Read, B. & Ryder, O.A. (1986). Pedigree Analysis by Computer Simulation. *Zoo Biology*, 5, 147-160.

Mayr, E. (1995). *Histoire de la biologie : diversité, évolution et hérédité*. Paris : Librairie générale française.

Pouyez, C., Lavoie, Y., Bouchard, G., Roy, R., Simard, J.P. & St-Hilaire, M. (1983). *Les Saguenayens. Introduction à l'histoire des populations du Saguenay, XVIe-XXe siècles*. Sillery, Québec: Presses de l'Université du Québec.

Roy, R., Bouchard, G. & Declos, M. (1988). La première génération de saguenayens: Provenance, apparentement, enracinement. *Cahier québécois de démographie*, 17, 113-134.

Skolnick, M.H. & Cannings, C. (1973). Simulation of Small Human Populations. *Computer Simulation in Human Population Studies* (eds. Dyke, B. and MacCluer, J.W.), New York: Academic Press, 167-196.

Société de généalogie du Québec (1995). *Répertoire des mariages de la Beauce 1740-1992*: (compilé sous la direction de l'abbé Jean-Guy Roy et Michel Beaulieu). Publication 77.

Tanguay, C. (1975). *Dictionnaire généalogique des familles canadiennes: depuis la fondation de la colonie jusqu'à nos jours*, 7 volumes, Montréal: Éditions Elysée.

Thomas, A. (1990). Comparison of an Exact and a Simulation Method for Calculating Gene Extinction Probabilities in Pedigrees. *Zoo Biology*, 9, 259-274.

Thompson, E.A. (1974). Gene Identities and Multiple Relationships. *Biometrics*, 30, 667-80.

Thompson, E.A. (1978). Ancestral Inference II. The Founders of Tristan da Cunha. *Annals of Human Genetics*, 42, 239-253.

Thompson, E.A. (1979). Ancestral Inference III. The Ancestral Structure of the Population of Tristan da Cunha. *Annals of Human Genetics*, 43, 167-176.

Thompson, E.A. (1981). Pedigree Analysis of Hodgkin's Disease in a Newfoundland Genealogy. *Annals of Human Genetics*, 45, 279-292.

Thompson, E.A. (1986). *Pedigree Analysis in Human Genetics*. Baltimore: Johns Hopkins University Press.

Vogel, F. & Motulsky, A. (1986). *Human Genetics*. Springer-Verlag, Berlin Heidelberg New York Tokyo.

Wright, S. & McPhee, H.C. (1925). An Approximate Method of Calculating Coefficients of Inbreeding and Relationship from Livestock Pedigrees. *Journal of Agricultural Research*, 31, 377-383.

ANNEXE

Liste des codes de statuts de BALSAC-RETRO

| STATUT | DESCRIPTION |
|--------|--|
| 0 | pas de statut (valeur par défaut) |
| 1 | immigrant |
| 2 | jamais venu |
| 3 | interrompu volontairement |
| 4 | origine inconnue |
| 5 | enfant naturel |
| 6 | adoption |
| 7 | amérindien |
| 8 | non marié |
| 9 | fondateur |
| 10 | introuvable |
| 11 | donné par la famille |
| 12 | réhabilitation du mariage |
| 13 | à venir et ou soumis à René Jetté (généalogiste) |
| 14 | mariage précédent introuvable |
| 15 | confession non catholique |
| 16 | émigrant |
| 17 | enfant né hors mariage |
| 18 | arrêt pour cause de veuvage |
| 50 | origine française |
| 51 | origine allemande |
| 52 | origine acadienne |
| 53 | origine américaine |
| 54 | origine anglaise |
| 55 | origine irlandaise |
| 56 | origine écossaise |
| 57 | origine hongroise |
| 58 | origine hollandaise |
| 59 | origine suisse |
| 60 | origine italienne |
| 61 | origine autrichienne |
| 62 | origine étrangère |

