

Université de Montréal

Aux origines du génome québécois :
démographie génétique d'une population nouvelle

par

Alain Gagnon

Département de Démographie
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiae Doctor (Ph. D.) en démographie

mai 2000

© Alain Gagnon



HB
881
U54
2000
v. 007

Université de Montréal

Université de Montréal
Département de génétique et de biologie
démographique

par
Alain Gagnon

Département de Démographie
Faculté des arts et des sciences

8

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de Philosophe Docteur (Ph.D.) en démographie



mai 1999
Alain Gagnon

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Aux origines du génome québécois :
démographie génétique d'une population nouvelle

présentée par :
Alain Gagnon

a été évaluée par un jury composé des personnes suivantes :

Bernard Brais..... Président du jury
Hubert Charbonneau..... Directeur de recherche
Évelyne Heyer..... Codirecteur de recherche
Bertrand Desjardins..... Membre du jury
Hélène Vésina..... Examineur externe

Thèse acceptée le : 21 septembre 2000

SOMMAIRE

Cette étude porte sur la régionalisation de l'effet fondateur québécois au cours des deux premiers siècles du peuplement (1608-1800). Prenant appui sur le Registre de population du Québec ancien (RPQA), elle vise en particulier : (1) à explorer la stratification du bassin génétique québécois à ses origines et (2) à rendre compte de cette stratification à partir de la dynamique démographique et sociale de la population. Bien que son objectif soit double, cette thèse est toutefois divisée en quatre étapes, chacune d'elles constituant le sujet d'un article et poursuivant les deux objectifs stipulés.

Le premier article se penche sur la répartition spatio-temporelle des mariages entre apparentés au Québec ancien et fait ressortir le rôle des différents facteurs qui en modulent la fréquence (isolement géographique, interdits de l'Église, diversité initiale des fondateurs, etc.). Il montre, entre autres, que les différences régionales ne s'expliquent pas en fonction d'une propension plus ou moins grande à chercher son conjoint dans la parenté immédiate, mais à partir de facteurs liés à la disponibilité des conjoints sur le marché matrimonial. À cet égard, comme attendu, les régions de l'est du Québec présentent des coefficients d'apparentement entre conjoints plus élevés que celles de l'ouest.

L'objet du second article est sensiblement le même que celui du premier puisqu'il vise également à décrire la distribution spatio-temporelle des mariages entre apparentés au Québec ancien. Toutefois, les méthodes et les données utilisées pour parvenir à cette fin sont différentes. Il est en effet basé sur des distributions patronymiques plutôt que sur l'armature généalogique de la population. En comparant les résultats générés, il permet ainsi de tester la validité de la méthode patronymique, qui repose sur un certain nombre d'hypothèses rarement vérifiées. Il fournit, par le recours aux numéros de lignée contenus dans le RPQA, un moyen de

tester l'origine monophylétique des patronymes. Il s'agit donc d'une contribution essentiellement méthodologique. Les résultats montrent que si la méthode isonymique permet de mesurer sans trop d'erreurs l'apparentement dans l'ensemble de la colonie, il en va tout autrement à une plus petite échelle.

Le troisième article est consacré à la contribution génétique des fondateurs et aux distances génétiques entre régions. Estimant la part des gènes qui, au sein de chaque région, provient de tel ou tel ancêtre fondateur, il précise les ramifications spatiales de l'effet fondateur québécois au-delà de ce que les enquêtes de biologie moléculaire ont révélé jusqu'ici. Il fournit d'abord une nouvelle mesure permettant d'évaluer l'homogénéité et la concentration d'un bassin génétique : le Nombre Utile de Fondateurs à contribution uniforme (NUF). Il met ensuite en évidence l'existence de trois grands bassins génétiques au Québec ancien (au lieu de deux) et identifie les fondateurs qui expliquent au mieux les distances génétiques entre régions. Les épïcêtres de ces trois bassins sont les principaux ports d'entrée ou pôles d'attraction en Nouvelle-France, soit Montréal, Trois-Rivières et Québec.

Enfin, un dernier article vise à comparer la liaison entre la fécondité utile des parents et de celle de leurs enfants par région au Québec ancien, soit la corrélation intergénérationnelle de la fécondité utile (CIFU), qui constitue un indice synthétique de la transmission verticale de tous les facteurs démographiques. Les résultats montrent que les individus issus de familles nombreuses dans une subdivision territoriale donnée tendent à encourager l'établissement d'un grand nombre de leurs enfants dans la même subdivision. Cette dernière contribution permet ainsi d'explicitier les résultats obtenus au cours des articles précédents.

TABLE DES MATIÈRES

SOMMAIRE.....	1
TABLE DES MATIÈRES.....	3
LISTE DES TABLEAUX.....	7
LISTE DES FIGURES	8
REMERCIEMENTS	10
INTRODUCTION GÉNÉRALE	12
UN TERRITOIRE POUR LA DÉMOGRAPHIE GÉNÉTIQUE.....	12
LA DÉMOGRAPHIE GÉNÉTIQUE AU QUÉBEC : UNE REVUE DE LITTÉRATURE.....	17
<i>L'anthropologie génétique au Québec.....</i>	<i>17</i>
<i>Le Programme de recherches en démographie historique (PRDH).....</i>	<i>25</i>
<i>L'institut interuniversitaire de recherches sur les populations (IREP).....</i>	<i>29</i>
SOURCES ET MÉTHODES.....	38
<i>L'évolution spatio-temporelle de l'apparement des conjoints (Chapitre 1 et 2).....</i>	<i>40</i>
<i>La contribution génétique par région des fondateurs (chapitre 3).....</i>	<i>42</i>
<i>La corrélation intergénérationnelle de la fécondité utile (chapitre 4).....</i>	<i>44</i>
<i>Le découpage territorial.....</i>	<i>46</i>
CHAPITRE 1 - APPAREMENT DES CONJOINTS AU QUÉBEC ANCIEN.....	50
INTRODUCTION.....	50
1.1 DONNÉES ET MÉTHODES.....	52
1.1.1 <i>Données.....</i>	<i>52</i>
1.1.2 <i>Mesures.....</i>	<i>53</i>
1.1.3 <i>Critères présidant au choix des divisions territoriales</i>	<i>55</i>

1.2	RÉSULTATS	56
1.2.1	<i>Portrait d'ensemble de l'évolution de l'apparement entre conjoints</i>	56
1.2.2	<i>Apparement des conjoints selon la région et le temps</i>	59
1.2.3	<i>Poids relatifs des types de mariage sur l'apparement global selon les régions</i>	61
1.2.4	<i>Géographie de l'apparement comparée: quelques études de cas</i>	64
1.2.5	<i>La fiabilité des dispenses</i>	69
	CONCLUSION	71
CHAPITRE 2 - PATRONYMES , NUMÉROS DE LIGNÉE ET GÉNÉALOGIES : ÉVOLUTION		
COMPARÉE DE TROIS INDICES DE L'APPAREMENT AU QUÉBEC ANCIEN (1608-1800).....		
	INTRODUCTION	73
2.1	DONNÉES ET MÉTHODES	75
2.1.1	<i>Les patronymes dans un monde « habituel »</i>	75
2.1.2	<i>Les patronymes dans un monde idéal: les numéros de lignée</i>	76
2.1.3	<i>Patronymes, numéros de lignées et généalogies: trois sources qui se complètent</i>	77
2.2	RÉSULTATS.....	78
2.2.1	<i>Évolution de l'apparement moyen des conjoints dans l'ensemble de la colonie</i>	78
2.2.2	<i>Apparement moyen des conjoints par région.</i>	82
2.2.3	<i>Étude d'un cas particulier : Lotbinière</i>	87
	CONCLUSION	92
CHAPITRE 3 - FRAGMENTATION OF THE QUÉBEC POPULATION GENETIC POOL (CANADA):		
EVIDENCE FROM GENETIC CONTRIBUTION OF FOUNDERS PER REGION IN THE 17TH AND 18TH		
CENTURIES		
	ABSTRACT	94
	INTRODUCTION	95
3.1	DATA AND METHOD.....	97
3.1.1	<i>Historical background and geographical location</i>	97
3.1.2	<i>Database</i>	98

3.1.3	<i>The measure of founders' Genetic Contribution (GC)</i>	100
3.1.4	<i>The Founders' Uniform Contribution Number (FUN)</i>	101
3.1.5	<i>The measure of genetic similarities between regions and their classification</i>	102
3.2	RESULTS	107
3.2.1	<i>Distribution of founders' contribution and the FUN</i>	107
3.2.2	<i>The classification of regions</i>	110
3.2.3	<i>The classification of regions as a function of their homogeneity</i>	115
	DISCUSSION.....	118
	<i>FUN and underlying socio-demographic processes</i>	118
	<i>A three cluster grouping</i>	119
	<i>Past and present classification</i>	121
	<i>Genetic contribution as a proxy for the distribution of inherited disorders</i>	123
	ACKNOWLEDGEMENTS	124
	CHAPITRE 4 - INTERGENERATIONAL CORRELATION OF EFFECTIVE FAMILY SIZE IN EARLY QUÉBEC (CANADA)	125
	ABSTRACT	125
	INTRODUCTION.....	126
4.1	DATA AND METHODS	130
4.1.1	<i>Database and selection of cases</i>	130
4.1.2	<i>The correlations</i>	133
4.1.3	<i>The effect of variance and correlation of EFS</i>	136
4.2	RESULTS	138
4.2.1	<i>Transmission of fertility and of EFS (whole colony)</i>	138
4.2.2	<i>The distribution of EFS</i>	139
4.2.3	<i>Area-based analysis of EFS</i>	140
4.2.4	<i>Region-based analysis of EFS</i>	144

DISCUSSION.....	146
<i>How much do the correlations explain the geographic differentials of the founders' contribution ?</i>	149
ACKNOWLEDGEMENTS	152
CONCLUSION GÉNÉRALE	153
<i>Apparemment et choix du conjoint (Chapitre 1)</i>	153
<i>Apparemment et isonymie (Chapitre 2)</i>	155
<i>Contribution génétique et classification des régions (Chapitre 3)</i>	157
<i>Corrélation intergénérationnelle de la fécondité utile (Chapitre 4)</i>	160
BIBLIOGRAPHIE.....	164
ANNEXE	174

LISTE DES TABLEAUX

Tableau 1.1 - Équivalences entre les relations de parenté et les coefficients de parenté	54
Tableau 1.2 - Proportion des différents types de mariages consanguins selon les régions	62
Tableau 1.3 - Proportion des conjoints ayant en commun au moins un ancêtre (P) et coefficients moyens de parenté (PHI) par région, XVII ^e et XVIII ^e siècles.	66
Figure 1.5 - Coefficient moyen de parenté proche des conjoints (PHI-Proche) selon les régions, 1780-1799.	70
Tableau 2.1 - Proportions attendues (P _A) et observées (P _O) d'unions isonymes à Lotbinière, 1780-1799 (%).....	90
Table 3.1 - Number of spouses and founders involved for each region and for the whole colony.	100
Table 3.2 - Founders' uniform contribution number (FUN) and its ratio to actual number of founders per region.	110
Table 3.3 - Founders' uniform contribution number (FUN) and its ratio to actual number of founders per clusters (FUN/ N _f) in the 1780-99 marriage cohort.	118
Table 4.1 - Number of children born to daughters by the number of children their mothers had (mothers married before 1680).	138
Table 4.2 - Three types of EFS correlations according to three geographic subdivisions (area) in early Quebec	141
Table 4.3 - Percentage of homogeneity increase associated to the Type 4 correlation per area and FUN/N _f ratios calculated from the genetic contribution of founders to the 1780-1799 marriage cohort.	143
Table 4.4 - Correlations of EFS per region in early Quebec.....	145
Table 5. Percentage of homogeneity increase associated to the Type 4 correlation per region and FUN/N _f ratios calculated from the genetic contribution of founders to the 1780-1799 marriage cohort.	146
Tableau A - Effectifs des unions utilisées dans les calculs d'apparentement des conjoints par décennie	175
Tableau B - Effectifs des unions utilisées dans le calcul de l'apparentement des conjoints par région au moyen de la méthode généalogique.....	176
Tableau C - Effectifs des unions utilisés dans le calcul de l'apparentement des conjoints par région au moyen de la méthode isonymique (patronymes et numéros de lignée).....	177
Tableau D. Les régions et les paroisses.....	178

LISTE DES FIGURES

Carte 1 – La Vallée du Saint Laurent à l’époque de la Nouvelle-France.....	46
Figure 1.1- Évolution de l’apparementement entre conjoints, ensemble de la colonie, XVII ^e et XVIII ^e siècles.	57
Figure 1.2 - Coefficient moyen de parenté des conjoints (PHI) selon les régions, XVIII ^e siècle.	60
Figure 1.3 - Proportion de mariages de degré 3:3 selon les régions, XVIII ^e siècle.....	63
Figure 1.4 - Coefficient moyen de parenté des conjoints calculé à partir des dispenses (PHI-Dispenses) selon les régions, 1780-1799.	69
Figure 2.1 Évolution de l’apparementement des conjoints au Québec ancien.	78
Figure 2.2. Coefficient de parenté des conjoints par région calculé à partir des généalogies (Φ_G).....	83
Figure 2.3. Coefficient de parenté des conjoints par région calculé à partir des patronymes (Φ_P).....	83
Figure 2.4 - Coefficient de parenté des conjoints par région, calculé à partir des numéros de lignée (Φ_L)	84
Carte 2 – Le peuplement de Lotbinière	88
Figure 3.2 - Distribution of founders’ contribution to the 1780-99 marriage cohort.	108
Figure 3.3 - Plot of the first three principal components (PC) of founders’ contribution per region (1780-99 marriage cohort). The three PC account for 56.0% of the variation. Axes have been rotated by varimax method.	111
Figure 3.4 - Cluster analysis of regions according to GC of founders to the 1780-99 marriage cohort. Measure of similarity: standardized covariation about total GC. 3a) $N_f = 9659$ 3b) $N_f = 300$. Grouping made by UPGMA.....	112
Figure 3.5 - Variance and overall genetic contribution to the three clusters of the 9659 founders.....	114
Figure 3.6 - Cluster analysis of Regions according to GC of founders to the 1780-99 marriage cohort. Measure of similarity: probability of choosing two genes that come from the same founder. $N_f = 9659$. Grouping made by UPGMA.	116
Figure 4.1 - Schematic representation of some families’ pedigrees showing which individuals are included in the analysis and to which generation they belong.	132
Figure 4.2 - Distribution of the year of marriage of the grand sons whose paternal grandfather was married between 1650 and 1680. (Number of grandfathers: 3110; number of grandsons: 16309).....	133
Figure 3. Distribution of EFS of women married before 1730, per area	140
Figure 4.4 - Extract of a pedigree representing a high in situ transmission of EFS, but a low overall transmission of EFS.	142

À tous les déracinés de la ville de Gagnon

REMERCIEMENTS

Je tiens à remercier tous ceux qui, à un moment ou à un autre, m'ont aidé dans la réalisation de cette recherche. En premier lieu, je voudrais remercier mon directeur de thèse, Hubert Charbonneau, qui a su m'initier à cette belle discipline qu'est la démographie historique. Il m'a également prodigué de nombreux et judicieux conseils, tout en accordant un intérêt soutenu à mes travaux. Que soient remerciés par la même occasion ses collègues de toujours. Visionnaires, les membres du PRDH ont su déceler et mettre à profit l'immense potentiel qu'offrait la reconstitution de la population québécoise à ses origines. Sans eux, l'idée même de cette recherche n'aurait jamais vu le jour. Merci à Bertrand Desjardins pour avoir su traduire en langage clair les requêtes que je lui ai adressées. Ses commentaires sur mes articles m'auront été par ailleurs fort utiles. D'autres ont aujourd'hui quitté le PRDH, je pense ici à Jean Daviault et Josée Tassé, mais ils m'auront grandement aidé lorsqu'ils étaient encore présents.

Je tiens également à exprimer ma profonde gratitude envers Évelyne Heyer, ma co-directrice qui, malgré l'éloignement, a pu assurer un suivi constant au déroulement de ma recherche, tout en l'alimentant de nombreuses hypothèses. Les innombrables messages électroniques qui remplissent aujourd'hui ma boîte aux lettres attestent de notre correspondance ininterrompue depuis bientôt trois ans. J'aurai eu l'heur d'avoir été accueilli à deux reprises au sein de sa belle équipe du Musée de l'Homme, dont le dynamisme et l'expertise m'ont convaincu d'y entreprendre une recherche post-doctorale.

Merci également à Bernard Brais pour m'avoir offert de participer à un stage d'initiation à la biologie moléculaire dans son laboratoire. J'ai pu ainsi profiter de sa grande connaissance en la matière, tout en y côtoyant de jeunes chercheurs qui font déjà figure de professionnels, dont

Hugo Lavoie. Grâce à eux, je comprends mieux ces aspects de la biologie moléculaire qui m'apparaissaient au départ relativement flous.

Enfin, merci à Mylène qui partage mes jours, mes joies, mes épreuves et ma passion pour la recherche. Cette thèse doit beaucoup à son soutien et sa confiance, mais aussi à son sens critique et sa rigueur.

INTRODUCTION GÉNÉRALE

UN TERRITOIRE POUR LA DÉMOGRAPHIE GÉNÉTIQUE

Peu de populations humaines ont une histoire démographique et biologique aussi captivante que celle du Québec. Fondée en 1608 aux abords d'un grand fleuve, elle connaîtra un développement formidable, décuplant sa mise initiale en passant de quelques milliers d'individus à quelques millions en l'espace de quatre cents ans. Pourtant, au cours de cette expansion, l'immigration en provenance de la mère patrie n'aura eu qu'un rôle secondaire. Au moment où s'amorce le peuplement de l'Amérique, la France se croit aux prises avec un grave problème de dépopulation. Elle hésite à envoyer ses habitants de l'autre côté de l'Atlantique pour peupler des contrées peu hospitalières, où d'ailleurs l'or et les métaux précieux, promis un siècle plus tôt par Jacques Cartier, brillent encore par leur absence. (Charbonneau et al., 1987).

Néanmoins, lorsque Louis XIV prend en charge la colonie en 1663, on observe une intensification des efforts de peuplement. Des contingents plus appréciables font voile vers le Canada et viennent alimenter la colonie qui a du mal à prendre son envol. C'est l'époque des Filles du Roi, envoyées pour rétablir l'équilibre des sexes, et du Régiment Carignan, dépêché sur le territoire pour contrer la menace iroquoise. Mais cet effort ne durera au plus qu'une quinzaine d'années. D'ailleurs, déjà en 1666, le ministre Colbert rappelait à l'intendant Talon qu'

« il ne seroit pas de la prudence [du Roi] de dépeupler son Royaume comme il faudroit faire pour peupler le Canada »¹, accordant ainsi au mouvement naturel la primauté dans le développement de la colonie. La Conquête anglaise de 1760 allait subrepticement confirmer ses visées puisque à partir de ce moment, les entrées se firent extrêmement rares, pendant que la fécondité atteignait des sommets rarement égalés dans les populations humaines (Charbonneau et al., 1987, Charbonneau, 1984; Henripin 1954).

Sur des terres, qui à tout prendre se révélèrent très fertiles, la jeune population connaît ainsi une expansion inespérée. Et quelle expansion! De Québec à Montréal, elle se répand d'abord le long d'une vallée immense où, de proche en proche, de nouveaux fronts pionniers s'ouvrent dans l'arrière pays, à mesure que les vieilles terres qui jonchent le fleuve parviennent à l'état de saturation. Elle colonise ensuite de nouvelles régions, s'étendant en aval vers le Bas-Saint-Laurent sur la rive sud, et vers Charlevoix sur l'autre rive. Au XIXe siècle, le Richelieu, la Yamaska et le Saint-François conduisent les colons jusqu'aux Cantons de l'Est. À la même époque, d'autres colons quittent les côtes de Charlevoix pour rejoindre les rives du Saguenay. Les descendants de ces pionniers encercleront un peu plus tard le Lac Saint-Jean. Tout en gardant ses positions, la population déborde au même moment vers l'ouest et déverse son trop plein vers les États-Unis. Alimentées par une fécondité toujours exceptionnelle, d'autres régions seront encore colonisées au XXe siècle, dont l'Abitibi, la Côte-Nord et le Nouveau-Québec. Jean Benoist (1977) résume d'une façon merveilleuse ce que le généticien des populations doit en conclure : « Un patrimoine héréditaire, commun initialement à un groupe restreint immigré dans un milieu nouveau, devient celui d'un peuple dispersé sur de vastes espaces ».

¹ *Rapport de l'Archiviste de la Province de Québec (RAPQ)*, 1931, p.41 (cité dans Charbonneau et al., 1987).

Bien sûr, au cours des siècles, la population se sera enrichie de nouveaux apports migratoires de provenances diverses. Mais la contribution des nouveaux arrivants au patrimoine génétique canadien-français sera d'autant moins considérable qu'ils seront peu nombreux, et que la population issue du noyau initial n'aura cessé de croître rapidement. Pour cette raison, et à l'exemple de Bouchard et De Braekeleer (1991a), on peut aborder le profil épidémio-génétique de la population québécoise au moyen du modèle de « l'effet fondateur ». Un tel phénomène est en effet observé lorsqu'une population est fondée par un groupe restreint d'individus et qu'elle demeure isolée par la suite (Mayr, 1963). Dans de telles circonstances, les gènes portés par les premiers arrivants constituent un échantillon non représentatif et, en quelque manière, appauvri, de la population mère. Par la suite, sous l'effet de la dérive génétique, certains gènes se diffusent en grand nombre, alors que d'autres disparaissent. La série d'échantillonnages consécutifs qu'amène la succession des générations peut ainsi faire en sorte qu'un gène délétère, peu fréquent dans la population mère, soit à terme très fréquent dans la population qui en est issue².

Devant tout ceci, le démographe historien, comme le généticien des populations, s'interroge : à partir de quel moment la consanguinité fait-elle son apparition ? Quelle en est l'évolution et la distribution spatiale ? Quelle est la contribution génétique par région des fondateurs ? Comment leur fécondité et celle de leurs descendants infléchissent-elle la fréquence actuelle des divers gènes qu'ils portaient au départ ? Comment, en se dispersant et en se cloisonnant, les premiers Canadiens se différencient-ils entre eux ? Quelles conclusions

² À ce titre, le Québec peut être rangé parmi d'autres populations à « effet fondateur », comme celles de la Finlande (O'Brien et al. 1994), l'Islande (Stefansson, 1999) ou encore certains isolats religieux comme les Huttérites (Klinger et al, 1990).

peut-on tirer de tout ceci sur la distribution actuelle des maladies héréditaires dans la population?

À l'endroit de la quasi-totalité des populations humaines, ces questions apparaissent d'emblée utopiques. Mais au Québec, grâce à la richesse des archives et à l'existence d'une grande banque de données informatisée (dont il sera fait état plus loin), il est possible de reconstituer pièce par pièce l'histoire démographique de l'ensemble de la population, au cours des deux premiers siècles du peuplement. On peut ainsi apporter un éclairage nouveau sur la manière avec laquelle le patrimoine génétique s'est mis en place et s'est redistribué de génération en génération pendant cette période.

C'est là notre objectif. Prenant la population « à bras le corps » à ses tout débuts, nous voulons suivre sa marche au cours des deux siècles suivants, de manière à rendre compte, du moins en partie, de l'évolution de sa structure génétique. L'étude comporte quatre chapitres qui se présentent sous la forme de quatre articles. Les deux premiers chapitres forment un tout : ils se penchent tous deux sur l'apparition et l'évolution spatio-temporelle de l'appareil des conjoints dans la vallée laurentienne. Le troisième chapitre aborde la stratification du bassin génétique québécois à ses origines au moyen de l'analyse de la contribution génétique des fondateurs par région. Enfin, le quatrième et dernier chapitre se penche sur la transmission intergénérationnelle de la fécondité utile. Mais avant d'aborder plus en détails les données et méthodes dont nous ferons usage, ainsi que le corps de la thèse, nous avons cru utile de fournir

un bref panorama de la démographie génétique québécoise³. Celui-ci permettra de situer la problématique dans un contexte plus large et d'en solidifier les assises.

Quelques précisions d'abord. Cette revue ne prétend nullement à l'exhaustivité. Nous avons simplement essayé de relater les origines et les « moments forts » de la discipline au Québec, tout en développant plus avant, çà et là, les aspects qui se rapprochent le plus de la problématique qui nous préoccupe. Ainsi, lorsque nous abordons le travail des anthropologues, nous nous penchons surtout sur les aspects qui se rapportent au « choix du conjoint », sujet des deux premiers articles, et sur l'articulation entre la fécondité et les migrations, sujet de notre quatrième article. De même, lorsque nous citons des études de biologie moléculaire, nous ne cherchons qu'à en extraire les aspects qui touchent spécifiquement à la stratification du bassin génétique québécois, un thème qui sous-tend l'ensemble de notre étude, mais qui est plus amplement abordé dans notre troisième article. Ainsi, nous ne précisons par exemple ni la localisation des mutations sur les chromosomes, ni la nature et la taille des haplotypes ancestraux, ni le caractère récessif ou dominant des maladies citées, etc. Tous ces aspects sont bien évidemment d'une importance capitale. Mais ils dépassent les visées et les moyens de ce travail. Pour les aspects se rapportant au contexte médical et épidémiologique, on consultera l'excellente revue de littérature qu'en a fait H. Vézina (1996) au Québec. D'autre part, j'essaierai de relever non seulement les études qui ont été menées au Saguenay-Lac-Saint-Jean (SLSJ) et à Charlevoix, mais aussi celles qui se rapportent à l'ensemble de la population québécoise. Quant à la démographie génétique à l'extérieur du Québec, elle a déjà fait l'objet d'un examen de synthèse (Gagnon, 1998). Il n'en sera donc point fait état ici.

³ On peut définir la démographie génétique comme étant l'étude des rapports entre la composition et les comportements démographiques d'une population d'une part, et la structure et l'évolution de son bassin génétique d'autre part (Vézina, 1996; Bouchard, 1993; pour une définition plus détaillée : Johnston et Albers, 1973: 202-203; Adams, 1990).

LA DÉMOGRAPHIE GÉNÉTIQUE AU QUÉBEC : UNE REVUE DE LITTÉRATURE

Si le Québec représente un véritable laboratoire pour l'étude de la génétique des populations (Laberge, 1969; 1976), il le doit non seulement à l'histoire particulière de son peuplement mais également au fait que la grande majorité de ses habitants sont catholiques romains. Depuis le début de la Nouvelle-France, tous les événements démographiques qui les concernent (baptêmes, mariages, décès) ont été enregistrés dans les livres des paroisses. Il ne serait donc pas exagéré de dire que l'origine de la démographie génétique québécoise remonte au XVI^e siècle, à l'époque du Concile de Trente (1563). En effet, celui-ci étendit à l'Église universelle l'obligation d'enregistrer les mariages et les baptêmes. Le *Rituale Romanum*, promulgué par Paul V en 1614, exigera ensuite que les décès soient également enregistrés (LaRose, 1976) Ainsi, au moment où débute le peuplement de la Nouvelle-France, les prêtres sont tenus de mettre scrupuleusement par écrit les moments clés, les temps forts, qui organisent, par leur interaction, le devenir biologique d'une population.

C'est de ce zèle remarquable de la part des autorités religieuses dont profitera la démographie historique et génétique. À notre avis, trois institutions ont permis l'éclosion de la discipline, et pavé la voie à cette thèse. Nous allons les présenter ici dans un ordre à peu près chronologique, sans préjuger de leur importance. Il s'agit du Département d'anthropologie de l'Université de Montréal, du Programme de recherche en démographie historique et de l'Institut interuniversitaire de recherches sur les populations.

L'anthropologie génétique au Québec

Sous l'égide du Département d'anthropologie de l'Université de Montréal, la génétique des populations connaît un essor remarquable. De la fin des années 60 à la fin des années 70, une équipe de chercheurs met en place un programme d'anthropologie génétique consacré à plusieurs populations, mais également aux « isolats » québécois. En effet, l'isolat apparaît alors

comme le cadre privilégié pour observer l'évolution biologique d'une population humaine. Il répond à la définition de la population panmictique, c'est-à-dire celle dans laquelle les membres d'un sexe ont des chances égales de se croiser avec n'importe quel autre membre de l'autre sexe. Cela constitue un avantage dans la mesure où les modèles prévoyant l'évolution des fréquences géniques sont basés sur le concept de panmixie. D'autre part, la reconstitution des généalogies est plus aisée dans une population isolée, généralement de petite taille, ce qui facilite du même coup l'identification de maladies rares causées par des gènes récessifs à l'état homozygote (Mayer, 1981).

Dès le départ, l'équipe met explicitement l'emphase sur la nécessité d'aborder les populations d'un point de vue pluridisciplinaire, en accordant une large part à la démographie : « Lorsqu'en 1966 fut entreprise l'étude systématique d'un certain nombre de communautés rurales isolées canadiennes-françaises, notre premier souci a été d'élaborer une méthode aussi exhaustive et poussée que possible, permettant de rendre compte de leur évolution socio-démographique, avant que soit abordée leur étude biologique » (Gomila et Guyon, 1969). En témoignent manifestement les études recensées, qu'il s'agisse d'études à caractère anthropométrique (Benoist et Auger, 1965; Auger, 1966), ou d'études portant sur les groupes sanguins (Magnan et Benoist, 1969). Évidemment, l'emphase est encore plus grande dans les études spécifiquement démo-génétiques. Nous en avons retenu quelques-unes.

Basée à la fois sur le dépouillement d'archives paroissiales et sur des enquêtes orales, l'étude de la communauté de « Bois-Vert » (Gomila et Guyon, 1969) se propose de cerner de près les facteurs qui influencent l'évolution de la consanguinité. En suivant pas à pas cette paroisse de colonisation depuis sa fondation en 1863, elle met d'abord en évidence le phénomène de la contribution inégale des fondateurs aux générations futures. Par le jeu de facteurs comme les échanges migratoires et la fécondité différentielle, quatre couples, arrivés au

tout début, représentaient encore 78% de la contribution démographique au moment de l'enquête⁴. Les auteurs font ensuite ressortir l'importance de la consanguinité « éloignée ». Ils démontrent que s'ils n'avaient pas pris en compte les liens remontant à des ancêtres communs arrivés au tout début, ils auraient décelé très peu d'unions consanguines. Toutefois, la consanguinité, tant éloignée que proche, n'aura pas connu une croissance linéaire dans la population. À partir de 1915, on assiste en effet à une réduction drastique des coefficients. D'une valeur maximale de 0,012 à cette date, le coefficient moyen, calculé avec la totalité de l'information généalogique disponible, chutera à une valeur minimale de 0,001 en 1945, pour remonter à 0,005 en 1965. Les auteurs semblent indiquer (ils ne sont pas très explicites à cet égard) que la réduction est due, en dernière instance, aux fluctuations du rapport de masculinité. Dans une petite paroisse, elles sont appréciables et peuvent amener une réduction drastique du nombre de conjoints potentiels sur le marché matrimonial. Ceci peut inciter les habitants à aller chercher alliance à l'extérieur de la communauté. De fait, on s'aperçoit que les mariages exogames croissent parallèlement à la totalité des mariages célébrés et sont presque deux fois plus nombreux que les unions endogames. Quant à ces derniers, ils sont systématiquement consanguins. Or, il se peut bien que les variations du rapport de masculinité, bien qu'elles soient importantes, n'aient ici qu'un rôle secondaire. La consanguinité, lorsqu'elle est excessive, contient en elle le germe de sa propre annihilation. Pour ne pas contrevenir aux règles édictées par les autorités religieuses, ceux qui sont le plus apparentés à leurs conjoints potentiels sont littéralement forcés de contracter une union exogame. Dans son étude sur l'île de Saint-Barthélemy, Leslie (1980) ne dit pas autre chose. Il montre en outre que les célibataires les plus apparentés à leur communauté entraîne une réduction de la consanguinité

⁴ La contribution *démographique* d'un fondateur est égale à $\sum_g (1/2)^g \cdot N_g$, où N_g est le nombre de ses descendants séparés de lui par g générations (Bouchard et al., 1995).

et décroît le rythme de la divergence génétique entre les populations qui échangent ainsi leurs conjoints potentiels.

L'Isle-aux-Coudres a fait l'objet de nombreuses études de la part des anthropologues (entre autres : Philippe et Gomila, 1971; 1972; Philippe, 1972; Boisvert et Mayer, 1994; voir aussi Laberge, 1976). Toutes soulignent son caractère hautement consanguin, et pour cause, plus de 90% des mariages qui y ont été célébrés depuis 1939 constituent des unions entre apparentés (parentés proches et éloignés). On remarque par ailleurs que les taux de mortalité infantile y ont toujours été très élevés. Encore aujourd'hui, on observe une sur-hospitalisation des enfants de 0 à 4 ans. La composante génétique de ces anomalies congénitales est indéniable (Boisvert et Mayer, 1994). On conçoit donc aisément que l'île ait suscité un certain intérêt. Évidemment, on peut invoquer son isolement relatif pour expliquer son profil épidémiogénétique particulier. Toutefois, on ne saurait s'en tenir à ce seul facteur, et l'analyse gagne à mettre en relation l'ensemble des paramètres démographiques qui concourent à l'évolution génétique d'une population. L'étude de P. Philippe et de J. Gomila (1971) est à cet égard exemplaire. En sus des résultats, elle fournit en outre une intéressante contribution méthodologique. Les auteurs développent plus avant les concepts « d'enfants génétiquement utiles » et « d'indice de variation ». Le premier concept, qui fera d'ailleurs l'objet de notre quatrième article, se réfère au nombre d'enfants qui, à l'âge adulte, participent à la procréation. Quant au second, couramment utilisé en statistique, il désigne le ratio de la variance à la moyenne. Autrement dit, $I = S_x^2/M_x$, où S_x^2 et M_x sont, respectivement, la variance et la moyenne.

En une première phase, les auteurs limitent leurs observations aux unions endogames, c'est-à-dire celles composées de deux conjoints originaires de l'île. Ils remarquent ainsi que, de 1770 à 1960, seule la moitié des individus qui y naissent y fondent un foyer. Les autres

émigrent, décèdent avant l'âge du mariage ou restent célibataires. En d'autres termes, pour toutes les périodes considérées, le nombre de naissances est deux fois supérieur à la taille de la famille « utile ». Toutefois, les *indices de variation* de la natalité « brute » et de la famille utile se suivent de très près et ne s'éloignent guère de 1. Cela signifie que la distribution de la fécondité des familles endogames peut être approchée par une loi de Poisson. On pourrait penser que cela donne la possibilité de modéliser le devenir des fréquences géniques au moyen des modèles simples de la génétique des populations, ceux-ci faisant en effet largement usage de la loi de Poisson. Or, en prenant en compte l'ensemble des unions (endogames et exogames célébrées dans l'île), on s'aperçoit que les indices de variation sont beaucoup plus élevés (environ deux fois plus). En effet, ces unions comprennent des mariages contractés avec un conjoint étranger à l'île; ceci ajoute chaque fois, une famille de taille 1 au calcul de la moyenne et de la variance (car le conjoint étranger est le plus souvent le seul représentant de sa famille dans la communauté). Il s'ensuit que la taille moyenne de la famille utile est diminuée et sa variance accrue.

Dans la dernière partie de leur étude, les auteurs utilisent une formule établie par Jacquard et Nadot (1968) pour prévoir la fréquence des mariages consanguins du premier degré (cousins germains) et du second degré (cousins issus de cousins germains). Cette formule permet d'établir la fréquence attendue (au hasard) de telles unions, en fonction de la moyenne et de la variance de la fécondité utile. Ils remarquent ainsi qu'il y a moins de mariages entre cousins germains que si les unions avaient lieu au hasard, et que malgré cet évitement, « tout se passe comme si les issus de germains ne pouvaient échapper intégralement à l'indice de structure, ce qui montre l'existence d'un frein, à ce niveau, au libre choix du conjoint et la présence d'un déterminisme structurel démographique » (p.716). Or, ce modèle ne tient pas compte de la corrélation intergénérationnelle de la fécondité utile, que nous présenterons plus loin. Celle-ci a pour effet d'augmenter encore la probabilité d'occurrence des unions

consanguines (Voir MacCluer et al., 1971). Étant donné cette corrélation, il se peut que les valeurs attendues au hasard soient encore plus élevés. Cela voudrait dire que l'évitement est plus prononcé que ne le donne à penser la seule prise en compte de la variance et de la moyenne de la fécondité utile.

Nous l'avons vu à quelques reprises, les migrations ont un effet considérable sur les différents indicateurs de structure génétique. La prochaine étude que nous allons présenter aborde le problème de front.

Bien qu'isolée géographiquement, la paroisse de « La Serpentine » (Cantons de l'Est) met à mal le concept d'isolat. F. Mayer (1981) nous montre en effet qu'elle est constamment traversée par des échanges migratoires. À un point tel que « la contribution respective des événements démographiques aux effectifs de la population donne un total de 54% pour la migration » (Mayer, 1981 : 42). Pour cette raison, on constate que l'arrivée d'immigrants fondateurs⁵ se distribue tout au long de l'histoire de la communauté. Leur succès en terme de contribution aux générations futures sera très variable. Car, à la suite d'échanges migratoires, la naissance d'une nouvelle lignée s'accompagne souvent de l'extinction d'une autre. Les chiffres sont à cet égard éloquentes. Pour l'ensemble de l'histoire génétique de la population, 833 couples fondateurs auront participé aux différents pools géniques. Mais parmi eux, seulement 204 seront encore représentés dans le pool génique de 1968. De plus, les 1068 individus qui composent la population à ce moment-là descendent de 306 immigrants féconds dont 115 sont toujours vivants. Il y a là un contraste frappant avec ce qui a été remarqué plus haut à l'endroit de la

⁵ En génétique des populations, on désigne par « fondateurs » les ancêtres dont on ne connaît pas l'ascendance, qu'ils aient appartenu au noyau initial de peuplement, ou qu'ils s'y soient greffé par la suite (ibid.: 56).

communauté de Bois-Vert, où les quatre couples arrivés en 1868 représentaient encore 78% de la contribution démographique de la population en 1967.

Toutefois, deux couples, arrivés au tout début, auront maintenu leur importance dans la composition de la population de la Serpentine, puisqu'en 1968, environ 40% de ses habitants pouvaient se tracer une voie généalogique jusqu'à l'un et l'autre de ces deux couples. Fait intéressant, c'est surtout par les hommes qu'est maintenue l'intégrité génétique de la population. À toutes les époques, plus d'hommes que de femmes se rattachent à un ancêtre fondateur éloigné, ce qui atteste le phénomène, plusieurs fois observé, de la stabilité des premiers dans la communauté par rapport aux secondes. Ainsi, tout au moins à l'échelle locale, les gènes suivent-ils surtout le tracé des lignées masculines.

Enfin, la liaison structurelle entre le phénomène migratoire et la nuptialité nécessite la création d'une typologie très fouillée des unions. On observe, bien entendu, celles qui mettent en relation deux individus nés et résidant dans la communauté (union endogames). Mais, pour qui veut saisir jusqu'au bout la dynamique bio-sociale d'une communauté, une kyrielle d'autres doivent encore être considérées. Par exemple, dans une union dite endogame, l'un des conjoints, résidant dans la communauté, peut bien être né à l'extérieur. Une union exogame peut signifier le gain ou la perte d'un individu pour la communauté, etc. Si bien qu'au total, la typologie nécessite pas moins de 17 catégories d'unions différentes.

On le voit, cette étude souligne la difficulté que pose le phénomène des migrations, et met en relief l'impasse à laquelle peut conduire le concept d'isolat. Celui-ci n'est par ailleurs en aucun cas la règle des populations humaines; il en est même de plus en plus l'exception. C'est justement ce que l'auteur voulait souligner et sa démonstration est tout à fait claire à cet égard.

Pour notre part, nous croyons utile d'ajouter en dernier lieu que l'étude des registres d'une seule paroisse ne permet pas de saisir et d'enregistrer les « migrations de retour ». D'un point de vue génétique, celles-ci ont en effet un rôle déterminant. Lorsque les individus émigrent, les gènes qu'ils portent ne disparaissent pas obligatoirement de leur communauté d'origine : leurs descendants peuvent très bien revenir s'y établir une ou deux générations plus tard. Ainsi risque-t-on, en limitant l'analyse à l'échelle locale, de sous-estimer la consanguinité, ou encore d'accorder une importance trop faible à un fondateur dont la descendance aurait connu de tels aller-retour.

Nous venons d'insister sur les impairs du concept d'isolat. En fait, l'emploi qu'en auront fait les anthropologues de l'Université de Montréal leur aura permis d'apporter une contribution majeure à la démographie génétique québécoise. Nous leur devons d'avoir fourni une connaissance très minutieuse des mécanismes socio-démographiques qui façonnent le devenir biologique des populations. Cette connaissance n'a de cesse de générer, encore aujourd'hui, une foule d'hypothèses, et rien n'empêche que ces hypothèses soient aussi mises à profit lorsqu'on a accès à des données sur une plus large échelle.

La génétique des populations humaines occupe une place à part dans les sciences du vivant, dans la mesure où elle peut recourir à la mémoire collective, qu'elle soit écrite, orale ou institutionnelle. Au Québec, nous sommes doublement choyés puisqu'une bonne partie de cette mémoire collective, celle se rapportant aux événements démographiques, a été fixée sur support informatique. Dans les trente dernières années, deux projets de grande envergure ont vu le jour.

Le Programme de recherches en démographie historique (PRDH)

En 1966, des démographes de l'Université de Montréal se donnent comme objectif de reconstituer l'ensemble de la population québécoise à ses origines. Pionniers d'une ère nouvelle, ils mettent sur pied une grande base de données démographiques qui impressionne tant par son caractère exhaustif que par les possibilités qu'elle offre. Formés à l'école de Louis Henry, père de la démographie historique, ils réalisent très tôt le potentiel qu'offre la population québécoise pour l'application des méthodes de la reconstitution des familles. Celles-ci avaient déjà fait leurs preuves et jeté une lumière nouvelle sur l'histoire démographique et sociale des populations anciennes (entre autres : Henry, 1956; Deniel et Henry, 1965; Charbonneau, 1970). Mais, centrées qu'elles étaient sur l'observation suivie de petites communautés rurales, elles se heurtaient au problème des « sorties d'observation », que provoquent inévitablement les migrations (Charbonneau et al., 1987; Desjardins, 1991). Le même obstacle, qui empêchait la pleine réalisation de l'anthropologie génétique, se dressait également devant les démographes historiens. Or, la population canadienne à ses origines pouvait offrir une voie de contournement. Elle constitue en quelque sorte un ensemble semi-fermé, de taille relativement réduite. On pouvait donc envisager de la reconstituer en entier. Ainsi, les jeunes célibataires partis chercher épouse dans la paroisse voisine, ou les familles parties s'établir sur d'autres terres ne peuvent échapper au crible qu'offre une observation exhaustive de l'ensemble de la population. Toutefois, malgré la petitesse relative de la population, les nombres en présence demeuraient considérables. Pour reconstituer l'ensemble de la population, encore fallait-il être outillé adéquatement. C'est là la seconde grande innovation du PRDH, l'établissement d'un protocole informatisé de reconstitution *automatique* des familles. (Légaré et al. 1973; Desjardins et al. 1977; Beauchamp et al. 1977; Légaré, 1981; Desjardins, 1991).

Dès le départ, les artisans de ce projet sont aussi conscients de l'intérêt que la population canadienne représentait pour les études de génétique des populations. Rappelons-

nous qu'à la même époque, l'épidémiologie et l'anthropologie génétique faisaient leur apparition au Québec. Ils prendront donc soin d'inclure dans leur protocole les moyens permettant de reconstituer intégralement l'armature généalogique de la population (Légaré, communication personnelle). Ainsi, une foule de problématiques bio-démographiques deviendront envisageables, en particulier l'héritabilité de la longévité (Charbonneau et Desjardins, 1990), la transmission de la fécondabilité des mères à leurs filles (Desjardins et al., 1991) ou le rapport entre la fécondité précoce et la longévité (Le Bourg et al., 1993). Malheureusement (ou heureusement!), ces études soulignent la difficulté de dissocier l'hérédité des composantes sociales et environnementales. D'autres études aborderont le problème dans l'autre sens et mettront en relief l'influence des facteurs environnementaux (notamment le climat) sur le rapport de masculinité à la naissance (Nonaka et al., 1999). Enfin, soulignons également quelques travaux d'intérêt biologique, dont un article portant sur les âges limites à la procréation (Desjardins et al. 1994), une série d'autres sur les jumeaux (Philippe, 1990; 1991a; 1991b) et un autre encore traitant des composantes de la parenté biologique dans un régime de fécondité naturelle (Naud et al., 1998).

Ce dernier article démontre que le nombre de « parents proches » est facilement prévisible dans une population relativement stable. D'abord puisque ce qui compose la parenté est sujet à de multiples événements « aléatoires », sa distribution doit en principe se rapprocher d'une loi normale, ce qui est effectivement constaté. Un modèle simple permet ensuite d'évaluer le nombre moyen de « parents proches » qu'un individu aura été susceptible de côtoyer durant sa vie. Par exemple, selon le modèle, le nombre moyen de parents des premiers et deuxièmes degrés des enfants nés de la première génération de Canadiens devrait être de $231 (= 4x^2 + x + 3)$, où x est la taille moyenne des fratries), ce qui est relativement près du chiffre observé de 220 parents. Un autre résultat attire notre attention. Les auteurs remarquent en effet qu'en moyenne, les enfants de familles nombreuses ont tendance à avoir eux-mêmes beaucoup d'enfants. Cela

fait en sorte que la distribution du nombre moyen de petits-enfants attendus diverge sensiblement de ce qu'on attend en l'absence de ce lien intergénérationnel.

Nous avons dit plus haut que la naissance du Québec se caractérisait par un « effet fondateur » et que celui-ci appauvrissait, par le biais d'échantillonnage qu'il implique nécessairement, la diversité génétique d'une population à naître. Or, un autre phénomène, qui n'a pas partie liée avec l'effet fondateur comme tel, conduit également à l'appauvrissement du pool génique d'une population : l'inégale contribution des fondateurs qui lui ont donné naissance. Il a été possible d'en apprécier l'ampleur pour le Québec grâce au Registre de Population du Québec ancien (RPQA), qui a ouvert à cet égard une voie fort prometteuse.

L'utilisation du RPQA permet d'abord une bien meilleure estimation du nombre de pionniers ayant fait souche dans la population que celle qu'avait fournie le généalogiste Archange Godbout (1946). On apprend par exemple qu'avant 1760, entre 8000 et 10000 individus font souche dans la Vallée du Saint-Laurent. Ces chiffres représentent environ le tiers des immigrants qui auront fait voile vers le Canada aux mêmes époques (Boleda, 1984). Le gros de la vague de l'immigration pionnière aura eu lieu entre 1663 et 1680. Après 1680, le mouvement naturel de la population prend définitivement le pas sur l'accroissement dû à l'apport migratoire. C'est pourquoi on peut considérer les 3380 immigrants établis avant cette date comme les véritables pionniers de la population. En 1987, ils feront l'objet d'une publication majeure en démographie québécoise : *Naissance d'une population* (Charbonneau et al., 1987). En plus d'étudier en profondeur leur nuptialité, leur fécondité et leur mortalité, cet ouvrage précise l'année d'établissement et le lieu de provenance des pionniers. Il estime en outre leur

descendance et leur contribution génétique⁶ par période, de même que la part des provinces françaises dans la formation du patrimoine héréditaire québécois.

Voyons quelques résultats qui nous intéressent au premier chef. De 1680 à 1730, la proportion de pionniers dont la contribution est faible augmente constamment, alors qu'inversement, celle dont la contribution est grande diminue sans cesse. Ainsi, en 1680, près de la moitié des fondateurs ne représentent qu'environ 22% du total de leur contribution, alors que seulement 17% représentent près de 40% de ce total. En 1730, la contribution de la moitié des fondateurs s'abaissera à 10%, alors que 17% des fondateurs suffiront à rendre compte de la moitié du total. Quant aux fondateurs arrivés après 1680, après la Conquête, ou encore plus tard, ils n'ont pu entamer l'avance qu'avaient pris leurs prédécesseurs. On arrive ainsi à estimer qu'en définitive, les deux tiers du génome des Canadiens français actuels provient des quelques 2600 pionniers qui ont fait souche dans la population avant 1680. Dans notre troisième article, nous aborderons également cette question de la « concentration » de la contribution des fondateurs.

En limitant l'analyse aux principales régions d'origine (dont la Normandie, le Perche, la région parisienne et le Centre-Ouest), les auteurs démontrent que seule la Normandie présente une contribution significativement supérieure à ce qui était attendu. Cette région a en effet fourni 23% des pionniers alors qu'elle compte pour environ 27% de la contribution génétique en 1730. Parmi eux, ceux venus du Perche (4,7% des pionniers) représenteront 7,6% de la contribution à la même date. C'est l'ancienneté de l'établissement des pionniers originaires de cette région qui

⁶ Dans *Naissance d'une population*, ce qui est appelé « contribution génétique » réfère en fait à la « contribution démographique », telle que définie plus haut. Toutefois, les deux mesures se recoupent à peu près (Bouchard et al., 1995).

leur aura conféré un tel avantage. Deux autres études viendront plus tard étoffer ces résultats (Charbonneau, 1990; Desjardins, 1990). Cette fois, l'analyse aura porté sur l'ensemble des fondateurs installés dans la Vallée laurentienne avant 1765. Plus récemment encore (Naud, 1997), un mémoire de maîtrise a développé ces questions en croisant le lieu de provenance des immigrants et leur lieu d'établissement. De plus, dans ce mémoire, un modèle simple de contribution génétique, qui toutefois ne prend pas en compte les migrations internes, tente d'estimer la part qu'occupe chaque région de provenance française dans chaque région laurentienne au moment de la Conquête.

Ce qui précède montre clairement la pertinence de faire appel au RPQA pour l'étude de l'impact des facteurs socio-démographiques sur la structure génétique de la population. Toutefois, suivant en cela son mandat initial, le PRDH aura surtout contribué à accroître nos connaissances démographiques de la population québécoise à ses origines. En réalité, le groupe de recherche qui aura peut-être le plus contribué à l'essor de la démographie génétique québécoise est l'*Institut interuniversitaire de recherches sur les populations*, que nous allons présenter à l'instant.

L'institut interuniversitaire de recherches sur les populations (IREP)

Si les membres du PRDH ont eu le mérite d'être les premiers à avoir informatisé et automatisé la reconstitution des familles, l'approche combinant l'histoire, la démographie et la généalogie à la génétique a été le mieux incarnée au sein de l'IREP. Créée à l'initiative de G. Bouchard dans les années 70, la base de données démographiques et généalogiques élaborée par cet organisme couvre aujourd'hui l'ensemble des régions du Saguenay-Lac-Saint-Jean et de Charlevoix, des origines à nos jours (fichier BALSAC). Elle comprend en outre un échantillon de généalogies d'individus répartis dans diverses régions du Québec (fichier RETRO). En plus d'avoir permis une connaissance accrue des mécanismes socio-démographiques qui influent sur

la structure génétique d'une population, la base de données a notamment donné lieu à plusieurs enquêtes à caractère épidémiogénétique. Ici, nous nous contenterons de relever ce qui semble être le plus pertinent pour notre travail.

Mentionnons tout d'abord un ouvrage fort important, publié au début des années 90, sous la direction de G. Bouchard et M. De Braekeleer : *Histoire d'un génome* (1991). Nous avons retenu ici deux articles de ce collectif (bien que d'autres sont cités dans le corps de la thèse).

Le premier se penche sur l'articulation entre « mouvements migratoires, effets fondateurs et homogénéisation génétique » (Bouchard et De Braekeleer, 1991b). Reprenant quelques résultats exposés un an plus tôt par M. De Braekeleer (1990), il fournit une explication originale au « paradoxe de l'homogénéisation de la population canadienne-française ». L'étude de De Braekeleer (1990) avait en effet mis en évidence une diversité génétique « remarquable » dans la population québécoise (diversité allélique et haplotypique comparable à celle relevée en France). Ceci pouvait représenter un véritable puzzle, dans la mesure où la population a été fondée par un noyau restreint d'individus. Bien sûr, on pouvait soupçonner que la diversité des origines des fondateurs y était pour quelque chose. Pour la plus grande part d'ailleurs, l'immigration pionnière en Nouvelle-France était de type individuel (Charbonneau et al. 1987), contrastant avec une immigration à caractère familial comme ce fut le cas dans le transfert migratoire Charlevoix-Saguenay. Mais une pièce manquait encore pour résoudre le paradoxe.

Il faut, nous disent les auteurs, voir dans la France d'ancien régime une mosaïque de sous-populations relativement cloisonnées et, de ce fait, relativement isolées génétiquement les unes des autres. Supposons qu'on puisse faire l'étude de la structure de la population française à cette époque. Si l'unité d'analyse est la sous-population, on en conclura que la population

dans son ensemble était très hétérogène : relativement isolées et ainsi en partie soumises à la dérive génétique, les sous-populations ont en effet divergé les unes des autres pendant quelques siècles. Si toutefois l'analyse porte sur les individus, alors on ne manquera pas de relever l'homogénéité que leur aura conféré cet isolement relatif. Or, la population québécoise a « puisé » un peu partout dans cette mosaïque. Puis, assez paradoxalement, par l'action du processus d'homogénéisation *sociale* s'étant mis en place dès le début du peuplement (notamment l'uniformisation par la langue française, les échanges matrimoniaux et migratoires, etc.), les différents segments de la diversité française inter-régionale auraient « fusionné » au sein des génomes individuels canadiens-français: « En somme, des éléments géniques qui n'étaient que juxtaposés en France se seraient trouvés combinés au Québec » (p.307).

Le deuxième article que nous avons retenu porte sur la reproduction familiale et ce que G. Bouchard appelle « l'effet multiplicateur » (Bouchard, 1991; voir aussi Bouchard et Roy, 1990 et Bouchard, 1996). Cet effet multiplicateur désigne l'ensemble des facteurs sociaux qui prennent en quelque sorte le relais de l'effet fondateur et favorisent « l'expansion d'un segment de la population au détriment des autres » (Bouchard, 1991 : 244). En l'occurrence, la part des gènes saguenayens provenant de la souche initiale de Charlevoix a été augmentée par l'action de ces facteurs. On remarque par exemple que parmi les premières générations de Saguenayens, ceux qui étaient issus de Charlevoix se sont davantage enracinés dans la région. Un plus grand nombre de leurs enfants s'y sont également installés. À cela s'ajoute un jeu d'alliances préférentielles mettant en relation, encore une fois, les descendants de Charlevoix. Tout ceci est lié au mode d'occupation du sol en un « terroir neuf » et au système de reproduction familiale qui en découle. Ainsi, les premiers arrivants ont pu accaparer plus de terre et établir à leur proximité un plus grand nombre de leurs descendants. Dans un travail ultérieur, Bouchard (1996) développera en détail le mécanisme de cette reproduction familiale. Il y fera en outre l'hypothèse qu'un tel mécanisme a pu également sous-tendre le développement de la

colonie à l'époque de la Nouvelle-France. Nous verrons que le quatrième chapitre de cette recherche tend à lui donner raison.

Le phénomène de l'effet multiplicateur peut également être associé à une corrélation intergénérationnelle de la fécondité utile. Voyons comment.

Dans un article consacré à la variabilité de la contribution génétique des fondateurs, E. Heyer et M. Tremblay (1995) se penchent sur les facteurs démographiques qui peuvent expliquer la fréquence relativement élevée de certaines mutations au SLSJ. L'un de ces facteurs, nous en avons déjà parlé, est la fécondité, dont la variation peut être très élevée dans une population en croissance. En effet, puisque dans une telle population la taille moyenne des familles est grande, la variance l'est également. Il en résulte qu'au bout de quelques générations, les contributions génétiques des fondateurs deviennent très concentrées : un petit nombre compte pour beaucoup alors qu'un grand nombre compte pour très peu. Toutefois, ce facteur à lui seul ne peut expliquer des taux de porteurs pouvant atteindre 1/20. Il donc faut faire appel à d'autres hypothèses.

Sans aborder explicitement cette question, M. Tremblay (1997) apporte des éléments de réponse. En effet, il montre « qu'en moyenne, les enfants des mères qui ont eu beaucoup d'enfants *mariés* au Saguenay ont eu plus d'enfants que ceux qui sont issus de familles moins nombreuses » (p.135)⁷. Il faut bien noter ici que la comparaison porte sur le nombre d'enfants *mariés* des mères. Car la liaison entre nombre d'enfants des enfants mariés et le nombre « brut » d'enfants des mères est beaucoup plus ténue. Comme l'auteur le mentionne, cela n'a

⁷ La sélection est faite sur 3290 mères dont les filles sont nées (entre 1850 et 1880) et se sont mariées au Saguenay.

rien de surprenant. D'une part, et cela a été soulevé dans maints travaux de démographie historique, la mortalité infantile est plus élevée dans les familles nombreuses. Pour cette raison, les enfants issus d'une grande famille sont, en comparaison de leur effectif, moins nombreux à parvenir à l'âge de la reproduction que ceux provenant d'une famille moins féconde. Cela a pour effet de diminuer la liaison moyenne entre la fécondité « brute » des parents et celles de leurs enfants mariés. D'autre part, et ce second facteur est intimement lié aux effets multiplicateurs, on peut supposer que les familles dont plusieurs des enfants auraient quitté la région ont en quelque sorte transmis leur « instabilité » à leurs enfants restés dans la région : « on pourrait alors parler de transmission intergénérationnelle de la mobilité » (p.135).

Prenant le relais de cette étude de nature plutôt démographique afin d'en tirer les conséquences biologiques, Austerlitz et Heyer (1998) montrent qu'il ne suffit pas de faire appel à la variance de la fécondité utile pour expliquer la fréquence actuelle de certains allèles mutants au SLSJ . Il faut invoquer en plus la corrélation intergénérationnelle de la fécondité utile (CIFU). Autrement dit, il ne suffit pas que certains aient beaucoup d'enfants « utiles » et d'autres très peu (grande variance de la fécondité utile); encore faut-il que cette disposition se transmette de génération en génération. À l'instar de la « contribution génétique », la CIFU constitue en quelque sorte un sommaire de la transmission verticale de tous les facteurs démographiques (fécondité, mortalité, nuptialité et migration). Les études de la transmission de l'un ou l'autre des facteurs démographiques (pris séparément) n'ont fait état jusqu'à maintenant que de liens intergénérationnels plutôt ténus (Desjardins et Charbonneau, 1990; Desjardins et al., 1991). Or, lorsqu'on considère d'un seul tenant tous ces facteurs, la corrélation d'une génération à l'autre apparaît (Austerlitz et Heyer, *ibid.*). Elle constitue ainsi un indice synthétique des mécanismes socio-démographiques qui modulent la fréquence des allèles mutants.

Pour notre part, nous pouvons ajouter encore une fois que cette transmission du comportement reproducteur renvoie aux effets multiplicateurs. En effet, elle dénote la présence de mécanismes de reproduction sociale favorisant l'implantation sur le territoire de certains (en l'occurrence, les descendants de Charlevoix) au détriment des autres. Autrement dit, si certaines mutations ont atteint une telle fréquence dans la population, c'est en grande partie parce qu'elles ont été en quelque sorte « portées » de génération en génération par la vigueur de l'enracinement des familles qui les ont introduites. En favorisant l'établissement des leurs, ces familles ont assuré une large diffusion aux mutations qu'elles portaient. Il semble en réalité, comme la prochaine étude va le montrer, que cette tendance se soit manifestée dès l'aube du peuplement de la Nouvelle-France.

Dans une étude mettant à contribution à la fois les données de l'IREP et celles du PRDH, Bouchard et al. (1995) se penchent sur le rapport entre la mobilité géographique et la stratification du bassin génétique canadien-français à ses débuts⁸. Après avoir mis en évidence la dichotomie maintes fois observée entre les pools géniques de l'Est et de l'Ouest de la province (entre autres : Bouchard et de Braekeleer, 1991a; Tremblay et al., 1998), les auteurs s'interrogent plus spécifiquement sur les comportements démographiques des fondateurs à l'origine du pool génétique de la population saguenayenne, en particulier, les fondateurs qui y auraient vraisemblablement introduit les pathogénies (ils figurent dans les têtes de lignée de 706 proposants, atteints d'une maladie héréditaire). Deux hypothèses sont proposées:

- (1) Les fondateurs importants pour le pool génique des proposants se seraient d'abord installés dans les régions de Québec et de Charlevoix, et une grande partie de leurs descendants se seraient ensuite déplacés vers le Saguenay;

⁸ Cette étude a d'ailleurs servi de point de départ à la présente thèse.

(2) Ces fondateurs et leur descendance se seraient au contraire répartis dans l'ensemble du Québec et, par conséquent, les maladies héréditaires observées au Saguenay résulteraient de conditions spécifiques à cette région (isolement relatif, faible diversité du pool génique, dérive génétique, etc.).

Le relevé des lieux d'établissement des fondateurs importants et de ceux de leur descendance utile sur deux générations permet de montrer que la première hypothèse est la bonne. En moyenne, les fondateurs dont la contribution au pool génique des proposants est importante conservent neuf de leurs enfants sur dix et quatre de leurs petits enfants sur cinq dans le Gouvernement de Québec. Par contre, les fondateurs qui comptent pour peu dans l'ascendance des proposants ne se sont établis dans ce Gouvernement qu'une fois sur deux et n'y ont « retenu » que le tiers de leurs petits enfants. Ainsi est mis en évidence le caractère particulier de la souche qui a donné naissance aux régions de Charlevoix et du SLSJ, ce qui confirme l'origine lointaine des processus qui ont donné lieu à cette fragmentation spatiale du bassin génétique québécois.

Comme nous venons de le voir, la recherche en démographie génétique québécoise a tout intérêt à étendre l'observation à l'ensemble du territoire. La dernière série de travaux que nous allons présenter en montre encore une fois la pertinence. En particulier, ces travaux s'interrogent sur la possibilité de retracer l'origine des fondateurs qui ont introduit les génopathies dans la population québécoise. Pour cette raison, les données qu'ils exploitent se déploient en amont et en aval de l'horizon de fondation de la population.

Au moyen d'une simulation de Monte Carlo, De Braekeleer et Dao (1994a, 1994b) montrent que les Percherons ont une contribution génétique significativement supérieure à la part qu'ils occupent dans l'ensemble des fondateurs figurant en tête de lignée d'un ensemble de proposants atteints de l'une ou l'autre des 28 maladies recensées. Cela les amène à dire que :

« The results presented here strongly suggest that several disorders were brought to Nouvelle-France by settlers coming from Perche » (p. 220). Or, une étude menée conjointement par des membres de l'IREP et du PRDH (Heyer et al., 1997) est beaucoup moins catégorique à ce compte. Cette étude est basée sur l'analyse de l'origine génétique de 673 proposants (atteints de l'une ou l'autre des 6 maladies recensées) et de 99 témoins nés au Saguenay. En analysant la contribution des fondateurs arrivés avant le XVIIIe siècle à ces individus, les auteurs montrent qu'il est difficile, voire impossible, d'identifier le lieu de provenance des mutations sur le territoire français ou européen. En effet, en prenant en compte un ensemble de facteurs (la contribution génétique, le niveau d'apparentement, le sexe, la date de naissance, etc.), ils trouvent plusieurs candidats possibles, qui conduisent à différentes régions d'origine en France (Aunis, Maine, Normandie, Orléanais, Perche, et autres) ou à l'extérieur (Grande-Bretagne et autres pays européens). De plus, les fondateurs qui sont communs à 95% des probants pour une maladie donnée sont aussi communs à 95% des probants pour au moins une autre maladie. Et même, parmi les fondateurs, 29 sont communs à 95% des probants de chacun des groupes (incluant le groupe témoin).

On le voit, cette difficulté d'identifier les ancêtres « introducteurs » pour la population saguenayenne tient largement au fait qu'elle n'est pas « stratifiée ». D'ailleurs, une étude, déjà citée (Heyer et Tremblay, 1995) montre également qu'aucun sous-groupe de fondateurs n'est spécifique à une maladie particulière. En termes épidémiogénétiques, la population forme par ailleurs un tout relativement homogène sur le plan spatial, telle ou telle maladie n'étant pas nécessairement plus fréquente dans telle subdivision que dans telle autre. Gradie et al. (1988; 1991) l'ont bien démontré au moyen d'un indice simple, soit le rapport de la proportion des cas d'une maladie dans une subdivision sur la part de la subdivision dans la population totale. On constate en effet que pour trois maladies (ataxie spastique, polyneuropathie et Tyrosinémie), le

rapport est souvent voisin à 1, ce qui signifie que chacune de ces maladies n'est pas plus fréquente dans une subdivision de la région que dans une autre.

Mais, si elle s'avère très difficile pour le Saguenay, cela ne signifie pas que la tentative soit vouée à l'échec à l'endroit de l'ensemble de la population québécoise. Une étude mettant à contribution, encore une fois, à la fois l'expertise de l'IREP et celle du PRDH (Brais et al., 1997), prouve que c'est possible dans certains cas. Dès les années 60, on avait remarqué que la dystrophie oculopharyngée était très fréquente dans les comtés de Montmagny et de l'Islet. La généalogie ascendante de 89 patients atteints de la région convergeait invariablement vers un couple fondateur, établi en 1648 (Barbeau, 1966), ce qui laissait croire que l'un ou l'autre des membres de ce couple avait introduit la mutation dans la population. Or, en remontant les généalogies d'autres individus atteints, mais vivant dans d'autres régions du Québec, ou même aux États-Unis, B. Brais et ses collègues se sont aperçu qu'en fait, ce couple n'y était pour rien; étant donné sa très grande descendance, il figure dans la généalogie de la plupart des Québécois, dans celle des individus atteints comme dans celle des témoins. En réalité, le couple « spécifique » à l'ensemble des malades n'est jamais venu au Canada, mais trois de leurs filles y ont immigré et y ont fait souche (Brais et al., *ibid.*; Desjardins, communication personnelle). Par ailleurs, la distribution de la descendance, sur trois générations, de ces trois sœurs est à cet égard éloquent: elle recoupe largement la distribution actuelle des proposants atteints. L'étude prouve donc, si besoin était, toute l'importance des comportements démographiques des premiers habitants de la Nouvelle-France. En suivant ces pionniers et leur descendance de près, on peut prévoir en partie la distribution spatiale des maladies héréditaires.

SOURCES ET MÉTHODES

La totalité des données de la présente étude a été puisée dans le *Registre de population du Québec ancien* (RPQA). Nous ne décrivons pas celui-ci en détails; d'autres l'ont déjà fait (Légaré, 1981; Charbonneau et al., 1987; Desjardins, 1991), beaucoup mieux que nous ne saurions le faire. Il s'agit d'une base de données constituée d'informations provenant des registres paroissiaux et d'autres sources comme les recensements ou les actes notariés. Couvrant la période allant du début de la colonisation jusqu'à 1800, elle permet de connaître une foule de renseignements sur les individus qui ont laissé des traces dans les registres de paroisse, comme les dates et lieux de leur naissance, mariage(s) et décès, de même que les liens filiaux et conjugaux qu'ils entretenaient avec d'autres individus. De plus, l'armature généalogique de la population est solidement ancrée jusqu'aux fondateurs de la population, dont on connaît, à proportion d'environ 95%, le lieu d'origine en Europe. Au total, la base de données comprend environ 712000 actes (Desjardins, 1991; voir aussi le site Internet du PRDH : <http://www.genealogie.umontreal.ca>).

Étant donné les objectifs de notre étude, nous avons cru opportun d'avoir recours aux seuls actes d'union et de mariage contenus dans cette base de données. Exception faite du quatrième chapitre (voir plus bas), aucune donnée relative à la mortalité ou à la fécondité des premiers Canadiens n'est employée. La raison en est très simple, tous les mariages ont été traités et leurs conjoints reliés à leurs parents jusqu'en 1800, alors que le couplage systématique des naissances et des décès ne dépasse pas 1770. Comme nous le mentionnons au troisième chapitre, l'utilisation des unions permet ainsi d'avoir prise sur la population jusqu'à une époque plus récente. D'autre part, et cela est également mentionné au chapitre 3, les unions constituent en quelque sorte les vecteurs de la transmission des gènes d'une génération à l'autre. Cela n'est plus vrai aujourd'hui, mais à l'époque, c'est par le mariage qu'un individu accédait à la

reproduction. Néanmoins, en observant, comme cela est fait au chapitre 4, la descendance mariée des premiers Canadiens sur trois générations, nous avons pris sur leur fécondité utile, dont nous avons dit qu'elle constituait le cadre privilégié d'observation de la reproduction biologique humaine.

Voyons maintenant les données et méthodes mises à contribution dans nos quatre chapitres. Afin de ne pas répéter ce qui est amplement exposé dans le corps de cette thèse, nous allons nous en tenir ici à l'essentiel.

L'évolution spatio-temporelle de l'apparentement des conjoints (Chapitre 1 et 2)⁹

Les deux premiers chapitres, dont le but est d'étudier l'évolution spatio-temporelle de l'apparentement au moyen de divers indicateurs, sont essentiellement basés sur les mêmes sources. Pour mesurer l'apparentement par la méthode généalogique, nous avons eu recours à l'ensemble des unions contenues dans le Registre (au nombre de 70869). Cette information provient essentiellement du relevé des actes de mariage ou notariés, mais est aussi déduite de la reconstitution des familles opérée par le PRDH, dans le cas où, pour une raison ou une autre, l'acte confirmant l'union entre deux conjoints est manquant. Si ces derniers n'ont pas été que de passage dans la colonie, on ne manquera pas, dans la quasi-totalité des cas, de retrouver une mention (naissance d'un enfant, décès, etc.) qui atteste leur union. Si par contre les conjoints en question n'ont pas fait souche dans la population, alors il importe peu d'en avoir perdu la trace, puisqu'ils n'auront pas participé à son renouvellement biologique. Il reste toutefois qu'une étude dont l'objectif serait d'analyser en détail le marché matrimonial (nombre de conjoints potentiels, etc.) de la Nouvelle-France ne pourrait faire abstraction de ces individus « fugaces » qui échappent au crible de l'observation (les individus qui n'ont pas fait souche ont pu, à un moment ou à un autre, être présents sur le marché matrimonial, et faire partie des conjoints potentiels).

Souvent, lorsque les données font défaut, les généticiens des populations sont amenés à fournir des estimations de l'apparentement basées sur les relevés des dispenses de consanguinité ou sur la fréquence des unions mettant en relation des conjoints qui ont le même patronyme (isonymie). Dans ces cas, ils ne peuvent bénéficier de l'ajout d'information qu'apporte la reconstitution des familles. Ainsi, afin de tester la validité de ces méthodes, nous n'avons délibérément eu recours qu'aux seuls actes de mariages contenus dans la banque de données

⁹ Pour le détail des effectifs en présence, pour chaque région, voir annexe, tableaux A, B et C.

(soit 67777 actes). En comparant les résultats obtenus à partir des actes de mariage à ceux obtenus à partir des unions *connues*, nous avons pu mettre en évidence certains biais, et en retracer la source. Par exemple, l'une des hypothèses de la méthode isonymique est que les patronymes doivent avoir une origine monophylétique, c'est-à-dire qu'ils doivent provenir d'une souche unique. Évidemment, puisque les individus sont parfois nommés en vertu de la profession qu'ils exercent, il arrive que des souches distinctes, ayant fait leur apparition dans des endroits différents, soient porteuses d'un même patronyme. Pour cette raison, la mesure de la consanguinité effectuée à partir du relevé des unions isonymes est souvent surévaluée. Or, dans le RPQA, chaque individu est identifié par un « numéro de lignée » qui le rattache, en ligne paternelle, jusqu'à son ancêtre le plus loin rencontré dans sa généalogie. Ce numéro de lignée peut servir de « patronyme » dont l'origine est nécessairement monophylétique. Nous disposons donc d'une estimation totalement exempte du biais qu'entraîne l'origine polyphylétique.

Enfin, une précision terminologique : normalement, les auteurs appellent "F" le coefficient Φ (ou PHI), sous-entendant par là que l'apparement moyen des conjoints correspond exactement à la consanguinité moyenne de leurs descendants. Or, on ne saurait être assuré de ce qu'il en aille toujours ainsi. Souvent, on constate que la fécondité utile tend à avantager, sur un territoire donné, les unions entre apparentés. Les effets multiplicateurs qui ont eu cours dans la population du Saguenay en témoignent clairement (voir plus haut). En effet, non seulement les descendants de Charlevoix se mariaient préférentiellement entre eux au Saguenay, mais en plus, ils parvenaient à y établir un plus grand nombre d'enfants. En conséquence, la consanguinité moyenne des enfants était plus élevée que l'apparement moyen des parents. Nous préférons donc, pour plus de rigueur, nous en tenir au coefficient de parenté (PHI dans le premier chapitre et Φ dans le second) des conjoints plutôt qu'au coefficient de consanguinité F pour caractériser l'évolution de l'identité des gènes dans la population.

La contribution génétique par région des fondateurs (chapitre 3)¹⁰

Le troisième chapitre se penche sur la contribution génétique par région des fondateurs au Québec ancien. Précisons d'abord que ce qui est entendu ici par fondateur diffère quelque peu de la définition souvent adoptée dans les études d'épidémiologie génétique. Souvent, dans le jargon des médecins généticiens, est fondateur l'ancêtre qui a introduit un gène mutant dans la population. On dira d'un tel individu qu'il est fondateur pour telle ou telle mutation. Or, en démographie historique et génétique, un fondateur est un immigrant qui a eu une descendance dans la population. Pour la cohorte des unions célébrées entre 1780 et 1800, qui compte 44510 conjoints, nous avons pu retracer un nombre de 9659 fondateurs ainsi définis. Nous avons ensuite estimé la contribution génétique des seconds aux premiers. En termes simples, la contribution génétique d'un fondateur est la part qu'il occupe dans le génome d'une population (Roberts, 1968; Heyer, 1995a; 1995b). Par exemple, la contribution d'un grand-père à son petit-fils est de 1/4. Cela se conçoit aisément si l'on considère que le petit-fils doit son existence à quatre grands-parents. Pour évaluer la contribution d'un fondateur à l'ensemble d'une région donnée, on additionne ses contributions à chacun de ses descendants et on divise le résultat par l'effectif total de la région. Elle dépend alors de sa fréquence d'apparition dans les généalogies, pondérée (lors de chaque occurrence) en fonction du nombre de générations qui le séparent de chacun de ses descendants. Puisque le fondateur est par définition un immigrant, ce peut être également un enfant d'un autre fondateur, auquel cas il est qualifié de « semi-fondateur ». Dans le calcul de la contribution génétique, ce dernier compte alors pour moitié.

La contribution génétique d'un fondateur peut également être conçue comme la probabilité qu'un gène, choisi au hasard chez les descendants, provienne de ce fondateur. C'est

¹⁰ Les effectifs en présence figurent dans le corps de l'article, au tableau 3.1.

la conception en terme de « probabilité d'origine des gènes » (Jacquard, 1974a; 1974b; Cazes, 1986; Heyer, 1995a; 1995b). Cette conception nous a permis de développer un outil permettant de comparer les populations sur le plan de l'homogénéité et de la concentration de leur bassin génétique. Il s'agit du « nombre utile de fondateurs à contribution égale » ou NUF (en anglais : *Founders' Uniform contribution Number*, FUN). Le NUF se définit comme l'inverse de la probabilité que deux gènes, choisis au hasard dans la population, viennent du même fondateur et s'exprime par :

$$NUF_x = 1/\sum x_i^2$$

où x_i est la contribution génétique du i^e fondateur. Plus ce nombre est réduit, plus la population est homogène. L'intérêt du NUF vient surtout du fait qu'il peut être rapporté au nombre de fondateurs *observés* (N_f), ce qui donne une indication chiffrée de la *concentration* de la distribution de la contribution génétique. Plus le rapport NUF/N_f est petit, plus la distribution de la contribution génétique est concentrée. Cela veut dire en effet que, pour un nombre de fondateurs donné, la plupart ont eu très peu de descendants, alors qu'une infime minorité d'entre eux en ont eu beaucoup. C'est une distribution qui ressemble à s'y méprendre à celle des richesses : très peu en ont beaucoup alors que la majorité en ont très peu. Ainsi, les fondateurs n'ont pas tous le même poids, et leur nombre « brut » peut induire en erreur. Imaginons par exemple deux populations. La première doit son existence à 100 pionniers. Ces derniers ont tous, après quelques générations, à peu près le même nombre de descendants. Dans la seconde population, on retrace 1000 pionniers. Toutefois, 10 d'entre eux, très prolifiques, sont responsables de la presque totalité de la descendance. Laquelle de ces deux populations est la plus homogène ? Si l'on ne considère que le nombre brut de fondateurs, nous en concluons que c'est la première, puisqu'elle a été fondée par un plus petit nombre d'individus... Mais en réalité, la seconde est sans doute plus homogène puisque parmi ses 1000 fondateurs, seulement 10 ont vraiment compté ; c'est bien moins que 100... L'utilisation de la

notion de nombre utile de fondateurs permet donc d'avoir une idée plus juste du nombre de fondateurs à l'origine d'une population.

Dans la deuxième partie du troisième chapitre, nous utilisons quelques méthodes de statistiques descriptives ou exploratoires pour *classifier* les régions les unes par rapport aux autres sur le plan de la contribution différentielle des fondateurs à chacune d'elles. Ces méthodes ont généralement pour objectif de convertir un gros tableau de données en les projetant dans un espace à deux ou trois dimensions, tout en conservant le maximum d'information (Jacquard, 1974; Pontier et al. 1990; Bry, 1995). Ici, nous avons affaire à un tableau de 9659 lignes (le nombre de fondateurs) par 10 colonnes (le nombre de régions). Nous voulons voir le positionnement des régions dans un espace à deux ou trois dimensions qui reflète au mieux l'information contenue au départ. À cette fin, nous utilisons d'abord l'Analyse en Composantes Principales (ACP) (Bry, 1995). Pour valider les résultats obtenus, nous procédons ensuite à une analyse des regroupements, ou analyse en cluster (Jacquard, 1974).

La corrélation intergénérationnelle de la fécondité utile (chapitre 4)

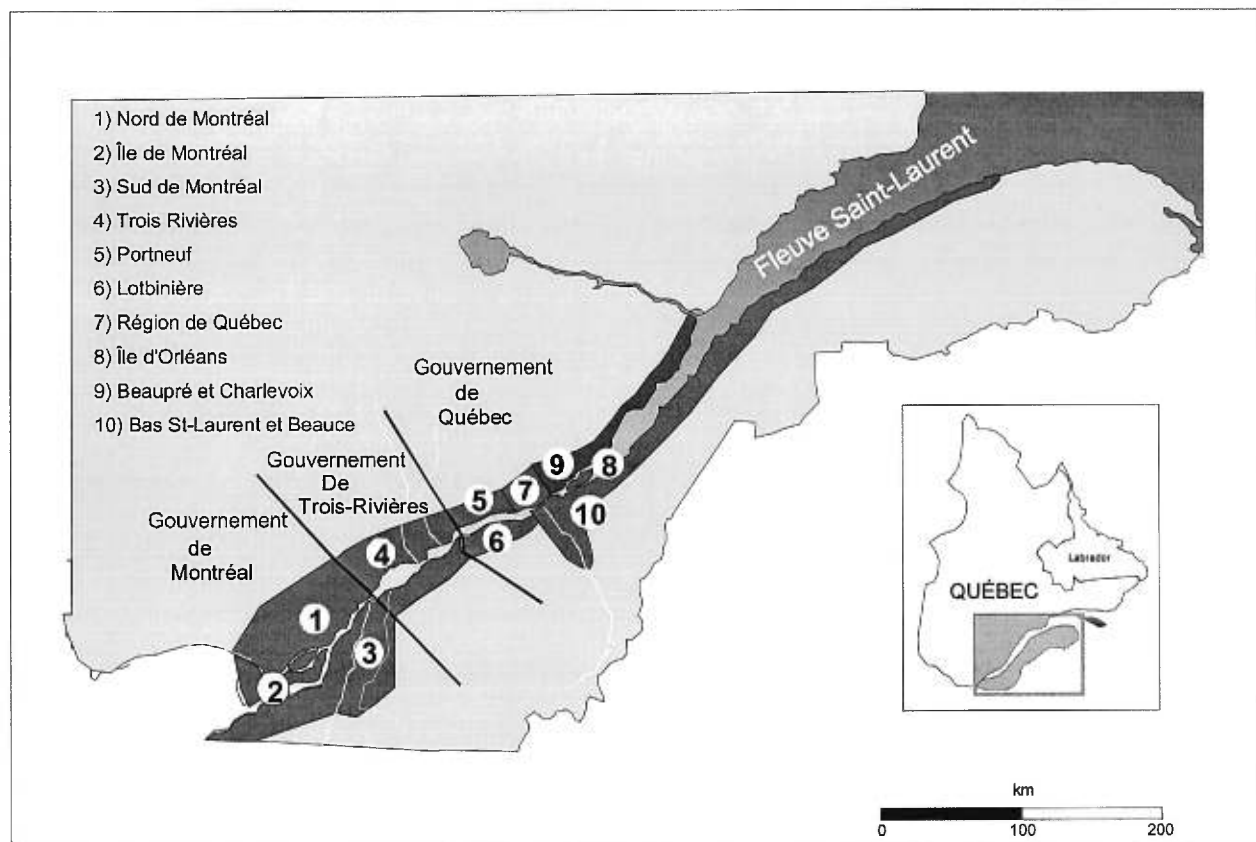
Le quatrième article accorde une attention particulière aux processus socio-démographiques qui modulent la fréquence des gènes dans une population. En particulier, il tente de vérifier s'il y a une corrélation intergénérationnelle de la fécondité utile au Québec ancien et si oui, quel en est l'effet sur l'homogénéisation de la population. L'article s'inspire donc largement des contributions de M. Tremblay (1997) et de F. Austerlitz et E. Heyer (1998), présentées plus haut. Étant donné que le RPQA porte sur l'ensemble du territoire du Québec ancien, nous pouvons mesurer la variation régionale de cet indicateur. Plus précisément, pour chaque unité territoriale, nous comparons le nombre d'enfants utiles des parents établis avant 1680 à celui de leurs enfants. Enfin, comme nous le disions plus haut, nous utilisons également les données relatives à la descendance finale des familles comprises dans l'échantillon. Cela

nous a permis de vérifier si la transmission de la fécondité utile est plus importante que la transmission de la fécondité « brute ».

Le découpage territorial

Étant donné que cette étude se penche sur la stratification du bassin génétique québécois à ces origines, il nous faut glisser ici quelques mots sur les choix qui nous ont guidé dans son découpage territorial (voir Carte 1).

Carte 1 – La Vallée du Saint Laurent à l’époque de la Nouvelle-France¹¹



PRDH - S. Gilbert et A. Gagnon

¹¹ Les divisions territoriales de cette carte ont été agrandies pour en faciliter la lecture; à l’époque, la colonie est essentiellement constituée de petits îlots de peuplement qui s’éloignent peu des berges du Saint-Laurent et de ses principaux affluents.

Précisons d'entrée de jeu que pour assurer la cohérence de l'ensemble de la thèse, nous avons conservé le même découpage territorial d'un article à l'autre. Il aurait peut-être été préférable d'opérer, dans certains cas, des coupes plus fines, afin d'être en mesure de relever des processus s'établissant à des échelles plus petites. Des analyses par paroisse auraient sans doute révélé des divergences par rapport à ce qui se passe à plus grande échelle. Du même coup, les résultats obtenus auraient pu être comparés à ceux qu'ont obtenus les anthropologues de l'Université de Montréal à l'endroit des « isolats » québécois. Mais, notre premier coup de sonde dans la stratification du bassin génétique québécois à ses origines se devait de porter sur des ensembles plus grands, afin d'aller droit à l'essentiel.

En premier lieu, nous avons voulu rester fidèle aux divisions administratives de la Nouvelle-France. Celle-ci était divisée en trois « Gouvernements » qui constituaient les trois principaux pôles d'attraction : Montréal, Trois-Rivières et Québec. Les régions que nous avons retenues sont entièrement circonscrites à l'intérieur de l'un ou l'autre de ces Gouvernements¹². Celui de Montréal est de loin le plus peuplé, du moins à partir du XVIIIe siècle. Nous l'avons divisé en trois régions, (1) Nord de Montréal, (2) Île de Montréal et (3) Sud de Montréal. Ensuite, étant donné sa taille relativement réduite, nous n'avons retenu qu'une seule région (4) pour tout le Gouvernement de Trois-Rivières. Enfin, nous avons divisé le Gouvernement de Québec en six régions : (5) Portneuf, (6) Lotbinière, (7) Ville de Québec, (8) Île d'Orléans, (9) Beaupré et Charlevoix et (10) Bas-Saint-Laurent et Beauce. La raison pour laquelle nous avons ainsi subdivisé ce dernier Gouvernement en autant de régions est qu'il présente des contrastes beaucoup plus marqués sur le plan de l'apparement des conjoints que les deux autres. Il est aussi beaucoup moins homogène sur le plan géographique. Le relief y est parfois très accidenté

¹² Voir en annexe, Tableau D, pour le détail des paroisses constituant chacune des régions.

(p.ex. Beaurpré et Charlevoix) ou bien comparable à celui qui caractérise le reste de la vallée laurentienne (p.ex. Le Bas-Saint-Laurent).

Ainsi, le découpage adopté fait en sorte que les subdivisions (les régions) n'ont pas toutes la même taille. En l'occurrence, les régions de Portneuf (5), Lotbinière (6) et de l'Île d'Orléans (8) sont beaucoup moins peuplées que les autres¹³. En fait, le premier critère qui a présidé au découpage n'est pas celui d'une taille uniforme des régions. Il s'agissait plutôt de faire ressortir au mieux les différences et les contrastes. Par exemple, suivant Jean-François Naud (1997) nous avons cru bon au départ de regrouper Portneuf et Lotbinière d'une part, et d'autre part, toute la région située en aval de Québec sur la rive nord (Beaurpré, Charlevoix et Île d'Orléans). Or, comme les deux premiers articles le montrent, l'examen des coefficients de parenté des conjoints a révélé qu'il valait mieux distinguer Portneuf de Lotbinière d'une part, et l'île d'Orléans de Beaurpré et Charlevoix d'autre part.

On peut se demander si l'analyse de la contribution génétique des fondateurs par région (troisième article) ne justifiait pas d'autres découpages. Une bonne partie de cet article est consacrée à la classification des régions en fonction de la contribution différentielle des fondateurs à chacune d'elles. Comme le lecteur sera à même de le constater, à l'aube du XIXe siècle, trois regroupements se dégagent de l'analyse en composantes principales comme de l'analyse en *cluster* : l'Ouest, le Centre et l'Est. Or, les résultats d'une classification peuvent varier selon le découpage adopté. Pour nous assurer de leur validité, nous avons refait plusieurs analyses basées sur des découpages comprenant 12 ou 15 régions plutôt que 10. Ces analyses ne sont pas présentées dans l'article en question, faute d'espace (les revues spécialisées ont

¹³ Voir en annexe les différents tableaux présentant les effectifs utilisés pour chaque article.

des exigences assez strictes à cet égard). Mais il convient d'indiquer tout de même que toutes font état d'une division tripartite, ce qui confirme la validité des analyses menées avec 10 régions.

Les bases de cette étude étant précisées, nous pouvons entrer dans le vif du sujet avec le premier chapitre que voici.

CHAPITRE 1 - APPARENTEMENT DES CONJOINTS AU QUÉBEC

ANCIEN

Avec la collaboration de H. Charbonneau et de B. Desjardins. Paru dans : Le choix du conjoint, sous la direction de G. Brunet, A. Fauve-Chamoux et M. Oris , Les chemins de la recherche, no 43, 1998, Lyon, BOSC: 133-150. Cette étude a été financée en partie par le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR) du Québec et par le Conseil de Recherches en Sciences Humaines du Canada (CRSH). Je tiens à remercier Jean Daviault et Josée Tassée pour leur assistance technique.

INTRODUCTION

L'incidence des génopathies au sein d'une population relève en grande partie de facteurs socio-culturels, ceux-ci façonnant à long terme le pool génique de celle-là. Le mariage au près (entre cousins germains par exemple), de loin le plus souvent évoqué, constitue l'un de ces facteurs. Le raisonnement est classique: la progéniture issue d'un mariage entre proches parents risque fort de recevoir, à un locus donné, une double copie d'un gène délétère récessif et donc d'être atteinte d'une maladie héréditaire. Or cette disposition en matière de choix nuptial n'est pas la seule à contribuer à l'augmentation de l'homozygotie et, *a fortiori*, à celle des génopathies autosomales récessives. Les conditions initiales y sont également pour beaucoup. Une population constituée au départ d'une poignée d'immigrants fondateurs et qui par la suite se retrouve quasiment isolée, devant dès lors compter sur son propre mouvement naturel pour assurer son renouvellement, contient en germe, couve en son sein les conditions optimales d'expression des tares héréditaires de fréquence rare. Après plusieurs générations, au gré d'une inévitable augmentation de l'apparement entre conjoints, la plupart des descendants en viennent à avoir un ou plusieurs ancêtres en commun, desquels ils auront reçu copies des

mêmes gènes, et alors peuvent apparaître les maladies sans qu'il y ait eu nécessairement une fréquence élevée de mariages entre proches.

Ce modèle de *l'effet fondateur* correspond largement aux paramètres de la formation et de l'évolution de la population francophone du Québec (Bouchard et De Braekeleer, 1991a). Au lendemain de la conquête anglaise en 1760, il est encore bien peu de colons installés dans la vallée du Saint-Laurent qui n'aient pas mis les voiles depuis la France. Et après cette date, l'immigration en provenance de celle-ci s'interrompt brusquement. Le noyau de peuplement initial ainsi constitué était entouré d'une membrane pratiquement hermétique. Amérindiens ou Anglais, rares ont été ceux qui parmi eux s'y sont intégrés. On le devine, les mariages mixtes n'avaient pas la faveur du clergé québécois. Dès lors, l'accroissement de la population ne pouvait être que le fait du mouvement naturel - qui bénéficia d'ailleurs d'une fécondité très élevée. À cet égard, on a pu établir qu'une minorité de fondateurs de la première heure, quelques 2600 hommes et femmes arrivées avant 1680, est responsable des deux tiers des gènes de la population canadienne-française actuelle (Charbonneau et al., 1987).

Il n'est guère étonnant dans ces conditions que l'homogénéité génétique de cette population ait souvent été proclamée sans retenue. Il faut dire que ce préjugé vétuste, mais persistant, a été renforcé par les premiers travaux de génétique médicale au Québec. En effet, nous disent G. Bouchard et M. De Braekeleer (1991b), ces travaux mettaient en relief et portaient au devant de la scène les maladies héréditaires propres aux Canadiens français. Or, les conclusions tirées des récentes enquêtes de génétique moléculaire paraissent sans appel, réhabilitatrices: le génome canadien-français est remarquablement diversifié (ibid.; De Braekeleer, 1990). En regard de ces travaux, la population québécoise est loin de se présenter comme une mosaïque uniforme. Les processus d'homogénéisation semblent avoir été confinés à certaines régions particulières, celles de l'est, et les maladies qui y ont été dépistées n'ont que très peu

d'incidence dans celles du centre et de l'ouest, ces dernières étant plutôt caractérisées par des maladies souvent répertoriées dans les populations caucasiennes en général (Bouchard et al., 1995).

C'est dans ce cadre que nous nous proposons d'apporter notre contribution. Pour ce, nous entendons étudier ici l'évolution et la répartition géographique des mariages entre apparentés au Québec ancien. Cela permettra de rendre compte, en partie, du profil épidémiogénétique du Québec et d'en préciser les contours pour les études à venir.

1.1 DONNÉES ET MÉTHODES

1.1.1 Données

Le Registre de population du Québec ancien (RPQA) élaboré par le Programme de recherches en démographie historique (PRDH) est tout désigné à cette fin. C'est une immense banque de données sur support informatique (il contient environ 800 000 actes de baptême, mariage et sépulture) qui concerne une population pratiquement fermée, "en marche" depuis deux siècles pour ainsi dire (entre 1608 et 1800), et qui de ce fait se laisse aisément appréhender par l'appareillage conceptuel et méthodologique de la génétique des populations.

Pour ce qui nous intéresse plus particulièrement ici, le registre contient 70869 actes ou contrats de mariage célébrés entre 1616 et 1799. L'ensemble des généalogies des conjoints impliqués y est reconstitué, de telle manière que les éventuels liens de parenté pouvant exister entre ceux-ci sont aisément décelés. En outre, la connaissance exacte du lieu du mariage est assurée dans plus de 95% (= 67570/ 70869) des cas. À propos des conjoints dont on ne connaît pas le lieu du mariage, nous avons eu recours au palliatif que constitue le lieu de naissance du premier enfant. Au total, il ne subsiste que 997 actes (1,4%) pour lesquels ni l'une ni l'autre de

ces informations ne sont disponibles. Mais c'est bien peu en regard de la masse de données accessibles (69872 actes).

1.1.2 Mesures

Les façons de mesurer l'apparentement sont aussi diverses que multiples. On a souvent recours aux *codes de consanguinité* dont l'Église, dans les évaluations nécessaires à l'octroi des dispenses, a fait grand usage au cours des siècles. Puisque ceux-ci se conçoivent d'une manière pratiquement intuitive - ils mesurent simplement la distance généalogique des conjoints à l'ancêtre commun le plus près - nous avons jugé bon d'y recourir. Nous calculerons à cet effet les différentes proportions de mariage de degré 2:2 (cousins germains), de degré 3:2, de degré 3:3 (cousins issus de cousins germains), de degré 4:3, ainsi que de degré 4:4. Nous considérerons également la *proportion* des conjoints dont l'apparentement est supérieur ou égal à celui de degré 4:4, que nous appelons « apparentement proche » et que nous désignerons par le symbole "P≥4" dans les tableaux et figures présentés ci-après¹⁴. En outre, nous utiliserons une troisième mesure de proportion, assez grossière, mais fort instructive lorsqu'elle est utilisée en parallèle avec d'autres: la proportion de conjoints ayant *au moins un* ancêtre commun - celui-ci pouvant se situer à une quelconque distance généalogique -, soit "P".

¹⁴ Deux clarifications s'imposent ici. D'abord, le terme « supérieur » dénote, *stricto sensu*, l'apparentement et non le chiffre représentant le degré. Ainsi, le degré de parenté 3:3 est supérieur au degré 4:4. Ensuite, le choix du degré 4:4 comme limite inférieure repose en quelque sorte sur un critère de nature purement historique; il s'agit du degré d'apparentement en deçà duquel il n'était pas nécessaire d'obtenir une dispense pour procéder au mariage. Cela ne signifie en aucune façon que l'apparentement soit nul en dessous de cette limite conventionnelle.

Les données dont nous disposons nous interdisent toutefois d'en rester là. Aussi, il est justifié de faire appel à une mesure plus synthétique, qui tienne compte de l'accumulation de l'apparemment le long des chemins généalogiques reliant les conjoints par le biais de chacun des ancêtres qu'ils ont en commun, soit le *coefficient de parenté*. Proposé par Malécot en 1948 et depuis lors couramment utilisé, le coefficient de parenté φ de deux individus est la probabilité qu'à un locus donné, des gènes désignés au hasard chez l'un et l'autre soient identiques (c'est-à-dire la probabilité que ces gènes soient la copie d'un même gène porté par un ancêtre commun)¹⁵. Étendu à l'ensemble d'une population, il s'exprime par une moyenne pondérée. Plus précisément, nous désignerons par "PHI" le coefficient de parenté moyen des conjoints et nous le calculerons comme suit:

$PHI = \sum p_i \varphi_i$, où les p_i représentent la proportion des couples ou paires de conjoints ayant un coefficient de parenté φ_i .

Tableau 1.1 - Équivalences entre les relations de parenté et les coefficients de parenté

Relation de parenté	φ (coefficient de parenté)
1:2	0,12500000
2:2	0,06250000
2:3	0,03125000
3:3	0,01562500
3:4	0,00781250
4:4	0,00390625
4:5	0,00195315
5:5	0,00097650
aucune	0

¹⁵ Pour deux individus A et B: $\varphi_{AB} = \sum (1/2)^{n_i + p_i + 1} (1 + f_{C_i})$, où n_i = nb d'ancêtres séparant A de l'ancêtre commun C_i , p_i = nb d'ancêtres séparant B de l'ancêtre C_i et où f_{C_i} = le coefficient de *consanguinité* de l'ancêtre C_i ; la sommation est ici étendue à tous les chemins généalogiques reliant A à B par l'intermédiaire d'un ancêtre commun C_i (Jacquard, 1974a; 1974b).

On peut reprocher au coefficient de parenté de ne pas être intelligible au premier coup d'œil. Que peut bien signifier un coefficient de 0,0078125 par exemple ? Pour aider à la compréhension, on peut toutefois traduire ces coefficients en relations de parenté (tableau 1.1). Par exemple, deux cousins issus de germains (degré 3:3) ont un coefficient de parenté égal à 0,015625.¹⁶ Or, des individus qui sont *deux fois* parents de degré 3:4 (pour chacune des deux paires d'ancêtres qu'ils ont en communs) sont également caractérisés par un tel coefficient¹⁷. Ainsi, on peut parler d'une véritable accumulation de l'apparentement à mesure que se multiplie le nombre d'ancêtres communs. Dans cet esprit, nous avons introduit une nouvelle mesure, forme hybride des précédentes: la proportion des conjoints dont l'apparentement *cumulé* équivaut à un degré supérieur ou égal au degré 4:4 (ou "P'≥4")¹⁸.

Enfin, il est fort intéressant de vérifier la fiabilité des données concernant la parenté déclarée, c'est-à-dire celles obtenues à partir d'un relevé des dispenses de consanguinité. Pour ce, nous examinerons les dispenses ayant été accordées entre 1780 et 1799, époque où se concentre la grande majorité des cas.

1.1.3 Critères présidant au choix des divisions territoriales

Le premier critère que nous avons retenu pour la subdivision des régions est évidemment celui de la proximité géographique. Le deuxième est celui de l'homogénéité: deux sous-régions proches seront regroupées en une seule si elles sont caractérisées par des niveaux d'apparentement semblables. Le troisième est celui de la nécessité d'avoir un effectif

¹⁶ En effet, $\varphi = (1/2)^{3+3+1} + (1/2)^{3+3+1} = 0,015625$; il y a deux termes à additionner car deux ancêtre communs sont invoqués, soient les deux arrières grand parents.

¹⁷ En effet, $\varphi = 4.(1/2)^{3+4+1} = 0,015625$; il y a 4 termes puisque 4 ancêtres communs sont invoqués.

¹⁸ D'ailleurs, le fond probabiliste qui sous-tend la mesure de la parenté nous y autorise. Sur le plan proprement dit de la notion d'apparentement, la probabilité que chaque paire des conjoints présentés ici, prise isolément, ait en un locus donné des gènes identiques est la même.

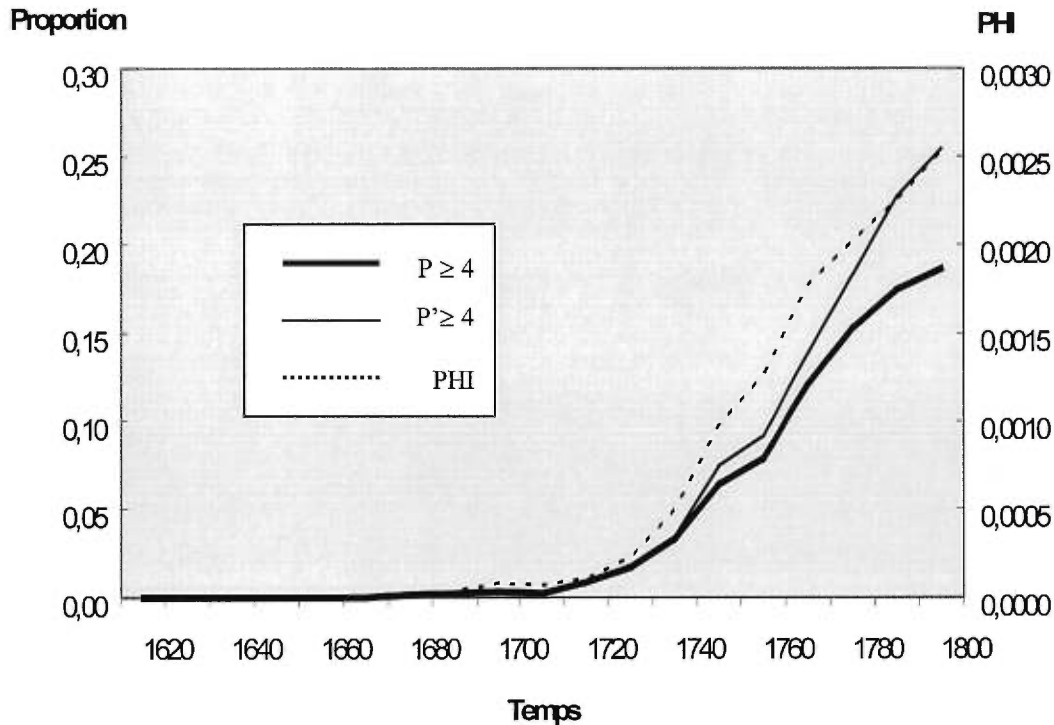
suffisant. Ainsi, les régions ne doivent donc pas être considérées ici comme des entités en soi, ni d'ailleurs comme des répliques à l'identique des circonscriptions administratives; elles ont plutôt été retenues à la faveur de nombreux tests, par tâtonnement pour ainsi dire. Nous avons ainsi obtenu 10 régions (voir Carte 1, Introduction):

1.2 RÉSULTATS

1.2.1 Portrait d'ensemble de l'évolution de l'apparentement entre conjoints

Avant d'aborder les questions relatives à la répartition territoriale des mariages entre apparentés, voyons d'abord l'évolution de l'apparentement des conjoints dans l'ensemble de la colonie aux XVII^e et XVIII^e siècles. C'est une occasion on ne peut plus rare qu'est celle de pouvoir visualiser l'apparition et la progression de l'apparentement au sein d'une population en plein essor. Reportons-nous à la figure 1.1 montrant, premièrement, en trait gras, l'évolution de la proportion de conjoints apparentés à un degré supérieur ou égal à 4:4 ($P \geq 4$), deuxièmement, en trait fin, celle de la proportion des mariages entre conjoints dont l'apparentement cumulé équivaut à un degré supérieur ou égal à 4:4 ($P' \geq 4$) et troisièmement, en trait pointillé, celle du coefficient de parenté moyen des conjoints (PHI). L'image qui en ressort est très nette. Jusqu'à 1720, l'apparentement n'a pratiquement pas évolué et n'a connu que des fluctuations de faible amplitude: aucune tendance ne s'y est encore généralisée. Après, l'augmentation est progressive, quasi-linéaire et très marquée. Elle se poursuivra jusqu'à la toute fin de la période prise en compte.

Figure 1.1- Évolution de l'apparement entre conjoints, ensemble de la colonie, XVII^e et XVIII^e siècles.



Les mesures ici considérées sont très sensibles à l'immigration, cette dernière fournissant en effet de nouveaux contingents non apparementés aux colons en place. Avant 1680, les conjoints immigrants sont encore majoritaires. Ensuite, ils deviennent rapidement minoritaires, si bien qu'ils ne représentent plus que le cinquième ou même le dixième environ de l'ensemble des conjoints au début du XVIII^e siècle. Or, il faut au moins un demi siècle (ou deux générations) avant que n'apparaisse, *ex nihilo*, la parenté entre les descendants d'une population fondatrice semi-fermée (où la part des immigrants est très faible). Cela explique l'évolution à la hausse des courbes à partir de 1720 et surtout de 1740; les conjoints ont dès lors plus de chance qu'auparavant d'être apparementés de quelque manière.

Que dire du niveau de l'apparement des conjoints en fin de période ? Entre 1790 et 1799, le coefficient moyen de parenté atteint 0,0025 (figure 1.1). En se reportant au tableau des équivalences (tableau 1.1) on en déduit que les conjoints sont alors, en moyenne, apparementés à un degré de parenté cumulé légèrement supérieur au degré 4:5 (pour lequel $\varphi \cong 0,002$). Le niveau de leur apparement se situe donc à la lisière de l'apparement proche (degré 4:4 ou plus). Sur le plan génétique, le coefficient obtenu signifie que le patrimoine génétique de la population québécoise s'est appauvri d'environ un quart de 1% entre l'époque de la fondation de la Nouvelle-France et la fin du XVIII^e siècle. Ce qui veut dire que la probabilité qu'un individu choisi au hasard parmi la progéniture des conjoints s'étant marié entre 1790 et 1800 soit hétérozygote pour un locus donné est inférieure d'environ 0,25% à ce qu'elle était chez les fondateurs¹⁹ (voir à ce sujet Chapman et Jacquard, 1971; Gauvreau et Jetté, 1992). Cela paraît assez faible. Or, il faut savoir que les coefficients de parenté ou de consanguinité observés jusqu'à maintenant dans les populations humaines dépassent rarement 0,01 et que le chiffre obtenu ici cache une variation régionale importante²⁰.

¹⁹ Cela, si l'on suppose un apparement nul entre les fondateurs et une absence d'association entre la fécondité et l'apparement des mariés de la fin du XVIII^e siècle.

²⁰ Par exemple, Chapman et Jacquard (opt.cit.) ont obtenu, sur cinq générations d'une population hautement endogame et de taille très réduite (les Jicaques du Honduras), un coefficient de consanguinité d'environ 8,8%.

Par ailleurs, la *proportion* de conjoints dont *l'apparentement cumulé* est égal ou supérieur au degré 4:4 ($P' \geq 4$) est alors d'un peu plus de 25%, résultat près d'une fois et demi supérieur à celui obtenu lorsqu'on ne tient compte que du premier ancêtre rencontré (soit $P \geq 4$, qui ne dépasse pas 18%). On remarque à cet égard que l'écart entre ces deux mesures, apparu vers 1740, n'aura cessé d'augmenter jusqu'à la fin du XVIIIe siècle . Et il y a lieu de supposer que cette tendance se sera maintenue jusqu'à nos jours.

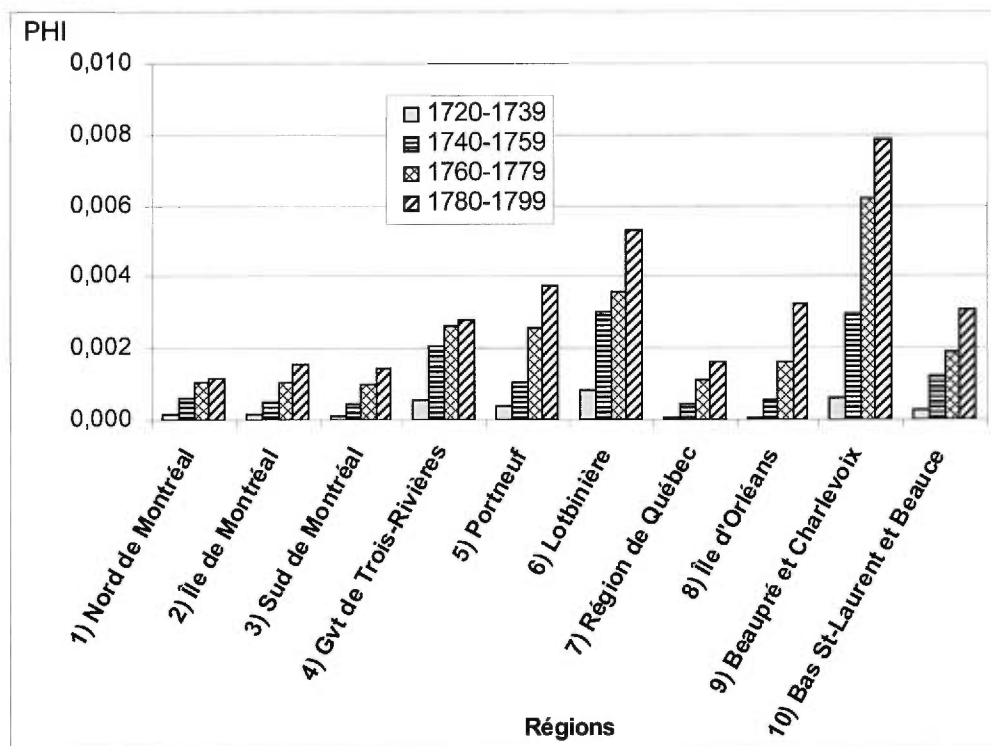
C'est dire l'importance de la parenté ambiante, celle qui relève non pas d'une pratique sociale encourageant le mariage au près, mais de l'accumulation de la parenté éloignée. Elle aura contribué à accroître le potentiel d'expression des tares héréditaires au sein de la population québécoise, sans toutefois homogénéiser substantiellement son bassin génétique²¹. Cependant, nous l'avons mentionné au tout début, les régions n'ont pas toutes connu le même destin épidémiologique. Certaines ont été plus touchées que d'autres. Voyons comment, sur le plan régional, les dispositions différentielles en matière de choix du conjoint - ou plutôt de disponibilité - peuvent nous éclairer à ce compte.

1.2.2 Apparentement des conjoints selon la région et le temps

Il convient d'abord de caractériser l'évolution spatio-temporelle de la parenté à l'aune du coefficient moyen de parenté PHI. C'est la mesure la plus synthétique et aussi la plus valable. Sur la figure 1.2, on voit aisément se dessiner la hiérarchie des régions.

²¹ À ce sujet, on consultera G. Bouchard et M. De Braekeleer: « En résumé, une forte consanguinité éloignée a causé la fréquence élevée de certaines génopathies, mais la consanguinité proche fut en général trop faible pour homogénéiser l'ensemble du pool génique » (1991a: 310).

Figure 1.2 - Coefficient moyen de parenté des conjoints (PHI) selon les régions, XVIII^e siècle.



La région de Beaupré et Charlevoix, cela est attendu, arrive bonne première, le coefficient moyen de parenté y culminant à 0,0078485 au cours de la période 1780-99, ce qui correspond approximativement à un apparentement moyen de degré 3:4 ($\varphi = 0,0078125$; voir tableau 1.1 des équivalences plus haut). Cela n'est guère surprenant puisque qu'on a dépisté dans cette région - de même que dans celle du Saguenay-Lac-Saint-Jean, qui en est issue - de nombreuses maladies héréditaires (Bouchard, De Braekeleer et al., 1991a). La deuxième région, et c'est là un résultat qu'on n'eût pas escompté au départ, est celle de Lotbinière, c'est-à-dire la partie du gouvernement de Québec située en amont de la ville, du côté sud du fleuve²². À notre connaissance, aucune étude ne l'a jusqu'à maintenant ciblée en vue d'y repérer des

²² Claude Laberge, dans son étude pionnière sur la consanguinité des Canadiens Français (1969), n'y a pas relevé un coefficient de consanguinité important. Or, ses résultats, qui portent sur le tournant du XXe siècle, ont été établis à partir des dispenses de consanguinité; ainsi toute la parenté accumulée les deux siècles précédents n'y est pas prise en compte.

généopathies. Viennent ensuite, en ordre d'importance décroissant, les régions de Portneuf (5), de l'Île d'Orléans (8), du Bas Saint-Laurent/Beauce (10) et de Trois-Rivières (4). Enfin, la région de Québec (7) et tout le gouvernement de Montréal (1,2 et 3) ferment la marche.

Il convient de remarquer ici que l'écart entre la région de Beauré/Charlevoix et les autres eût été bien plus grand si l'on avait tenu compte de l'apparentement entre les fondateurs de la Nouvelle-France (que l'on a supposé ici égal à zéro). Les généalogies des habitants de ladite région convergent souvent vers une contrée particulière en France, le Perche, alors que les généalogies des proposants des autres régions du Québec (surtout celles de l'ouest) sont en général sur ce plan beaucoup plus diversifiées (voir à ce sujet, Naud, 1997).

1.2.3 Poids relatifs des types de mariage sur l'apparentement global selon les régions

Quels sont les mariages qui pèsent le plus sur le niveau de l'apparentement global ? D'un point de vue génétique, une telle question peut sembler dénuée de sens: n'avons-nous pas insisté sur le caractère cumulatif de la parenté, sur le fait, par exemple, que deux conjoints qui ne se reconnaissent aucun lien peuvent être, au contraire, parents d'assez près? Toutefois, sur le plan social, la question n'est pas sans intérêt. À cet effet, le tableau 1.2 présente, pour chaque région et pour l'ensemble des XVII^e et XVIII^e siècles, le nombre et la proportion des mariages de degré 2:1, 2:2, 3:2, 3:3, 4:3, 4:4 et inférieur à 4:4 ($P < 4$) respectivement, ainsi que le coefficient moyen de parenté des conjoints PHI.

Tableau 1.2 - Proportion des différents types de mariages “consanguins” selon les régions, ensemble des XVIIe et XVIIIe siècles²³

Région	N	Degré de parenté							PHI (X 10 ⁻⁴)
		2:1	2:2	3:2	3:3	4:3	4:4	< 4:4	
1) Nord de Montréal	9214	0	13 (0,1)	35 (0,4)	191 (2,1)	178 (1,9)	330 (3,6)	628 (6,8)	9,9
2) Île de Montréal	11499	0	22 (0,2)	61 (0,5)	208 (1,8)	161 (1,4)	235 (2,0)	329 (2,9)	8,7
3) Sud de Montréal	12907	0	14 (0,1)	64 (0,5)	268 (2,1)	281 (2,2)	474 (3,7)	581 (4,5)	10,1
4) Trois Rivières	6289	1	42 (0,7)	87 (1,4)	244 (3,9)	172 (2,7)	197 (3,1)	284 (4,5)	21,5
5) Portneuf	2520	0	9 (0,4)	23 (0,9)	118 (4,7)	87 (3,5)	118 (4,7)	108 (4,3)	20,3
6) Lotbinière	1462	0	7 (0,5)	27 (1,8)	113 (7,7)	74 (5,1)	101 (6,9)	125 (8,5)	35,2
7) Région de Québec	8672	0	21 (0,2)	27 (0,3)	94 (1,1)	67 (0,8)	144 (1,7)	235 (2,7)	6,2
8) Île d'Orléans	2482	0	1 (0)	4 (0,2)	52 (2,1)	56 (2,3)	121 (4,9)	185 (7,5)	10,0
9) Beauré et Charlevoix	4098	0	6 (0,1)	36 (0,9)	366 (8,9)	294 (7,2)	383 (9,3)	377 (9,2)	40,9
10) Bas St-Laurent et Beauce	10729	0	10 (0,1)	53 (0,5)	447 (4,2)	490 (4,6)	809 (7,5)	138 (12,9)	20,8
Total:	69872	1	145	417	2101	1860	2912	4238	
En termes relatif:	100%	0%	0,2%	0,6%	3,0%	2,7%	4,2%	6,1%	

On remarque tout de suite que dans l'ensemble, les mariages entre très proches parents (oncle-nièce ou cousins germains) constituent l'exception plutôt que la règle: ils ne comptent en effet que pour 0,2% du total. La hiérarchie des régions suivant le coefficient de parenté moyen n'est donc pas tributaire de l'importance différentielle de ces mariages. Force est d'admettre que les pressions exercées par les autorités ecclésiastiques pour dissuader leurs fidèles de chercher alliance dans la parenté immédiate ont porté fruit, particulièrement à Beauré et Charlevoix, où les effectifs non apparentés sur le marché matrimonial n'étaient pas monnaie courante (Morissette, 1991). En tout état de cause, si les conditions de la panmixie²⁴ avaient prévalu, les mariages entre très proches parents y auraient certes été plus fréquents car les fondateurs qui

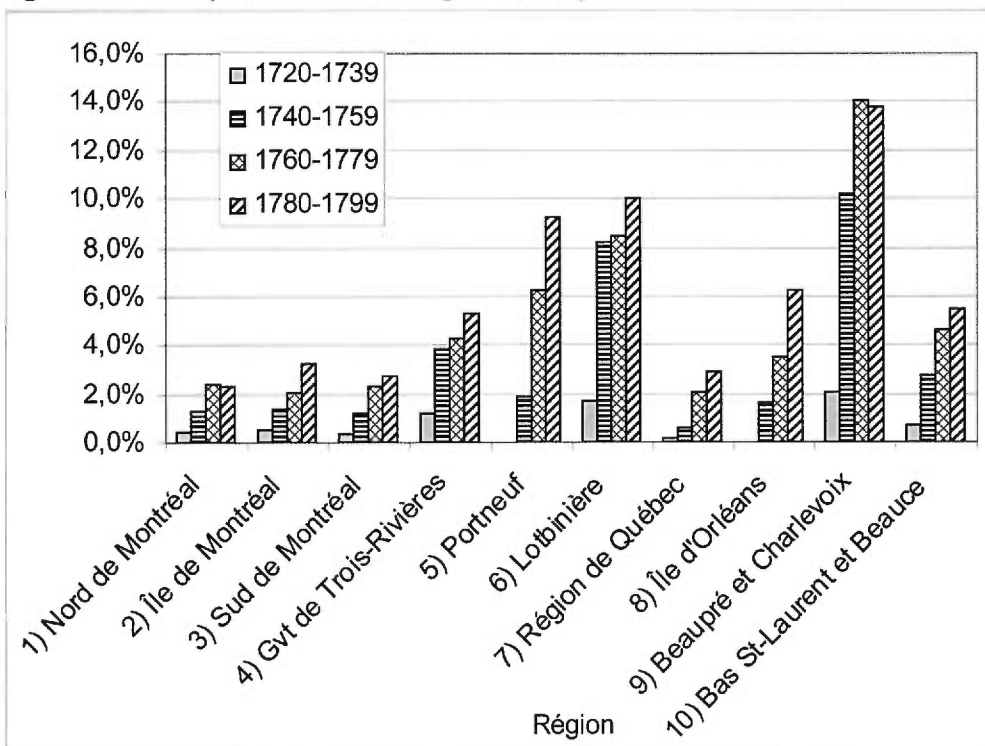
²³ Les chiffres entre parenthèses donnent les pourcentages respectifs.

²⁴ Il y a panmixie lorsque le choix du conjoint se fait au hasard.

sont à l'origine de cette région sont peu nombreux et souvent apparentés²⁵. On est ainsi amené à supposer une sorte de surcompensation, un zèle de la part des candidats au mariage à éviter leurs proches, là même où ils ont le plus de chance de se rencontrer.

Ce sont surtout les mariages entre cousins issus de germains (degré 3:3) qui pèsent sur le coefficient moyen de parenté des conjoints. L'importance relative de ceux-ci est en effet systématiquement associée à la mesure de celui-là. On peut d'ailleurs s'en rendre aisément compte en observant la figure suivante, montrant la proportion de mariages de degré 3:3 selon les régions (figure 1.3): il est, sauf exception, pratiquement superposable à la figure 1.2, où est représenté le coefficient de parenté moyen des conjoints:

Figure 1.3 - Proportion de mariages de degré 3:3 selon les régions, XVIII^e siècle.



²⁵ Jetté, Gauvreau et Guérin (1991: 96) ont estimé que dans Charlevoix, 38% seulement des fondateurs n'étaient apparentés à aucun autre fondateur et que la proportion de frères et de soeurs atteignait 49%.

Bien sûr, les mariages de type 4:3 et 4:4 n'ont pas un effet négligeable sur le coefficient moyen de parenté, mais cet effet est déjà nettement moins affirmé, car bien qu'ils soient à peu près aussi fréquents que les mariages de type 3:3, leurs poids respectifs sont moins importants. N'oublions pas toutefois que tout ceci n'a de sens que si l'on adopte une vision pour ainsi dire tronquée de ce qu'est la parenté. En fait, dans Beauré et Charlevoix, par exemple, plusieurs relations de parenté que l'on dit être de degré 3:3 sont en fait beaucoup plus complexes; un tel se marie avec sa petite cousine il est vrai, mais celle-ci lui est en outre apparentée de plusieurs autres façons.

1.2.4 Géographie de l'apparentement comparée: quelques études de cas

Les résultats qui précèdent sont en partie conformes au schéma dégagé depuis peu par la recherche épidémiologique québécoise (entre autres : Bouchard et De Braekeleer, 1991a). Il semble à cet égard que les conditions qui prévalaient avant le XIX^e siècle soient directement à l'origine de la situation d'aujourd'hui puisque, comme prévu, les coefficients moyens de parenté sont en général plus élevés dans le Gouvernement de Québec (« Est ») que dans ceux de Trois-Rivières et de Montréal (« Ouest »)²⁶. Or, à la différence de l'Ouest, l'Est de la colonie présente plusieurs contrastes à cet égard, les coefficients de parenté ne s'y agencant aucunement en un gradient continu d'une région à l'autre. Aussi l'analyse qui suit accordera-t-elle plus d'attention au second qu'au premier de ces regroupements territoriaux.

²⁶ Bien entendu, la région mère de Québec fait exception. C'est une agglomération urbaine: elle reçoit plus d'immigrants et la mobilité des familles y est plus grande qu'à la campagne.

S'il est vrai que la faiblesse relative du nombre de fondateurs prédispose parfois à ce que l'apparemment global soit élevé (comme cela est manifeste dans Beauré et Charlevoix), ce n'est pas toujours le cas: on peut imaginer une région où les liens de parenté entre les conjoints sont moins serrés que ne le laisserait supposer la seule prise en compte du nombre réduit de fondateurs qui lui a donné naissance²⁷. Il nous semble que la région "8" (comprenant la Beauce et le Bas Saint-Laurent) était précisément dans cette situation. Nous avons remarqué que sur le plan de la *proportion* de conjoints ayant en commun *au moins un* ancêtre (soit "P"), cette région arrive tout juste derrière celle de Beauré et Charlevoix (9), si l'on considère l'ensemble des XVIIe et XVIIIe siècles: environ trois couples sur dix y sont constitués de conjoints qui peuvent se trouver au moins un ancêtre commun. Pourtant, si l'on se reporte au coefficient moyen de parenté (PHI), on voit que cette région se place loin derrière Beauré et Charlevoix (voir tableau 1.3).

²⁷On peut même, à l'inverse, imaginer une région fondée par un grand nombre d'immigrants mais où, pour des raisons sociales (qui relèvent de la nécessité de remembrer les patrimoines familiaux par exemple), les mariages entre parents très proches seraient très fréquents.

Tableau 1.3 - Proportion des conjoints ayant en commun au moins un ancêtre (P) et coefficients moyens de parenté (PHI) par région, XVII^e et XVIII^e siècles.

Régions	N	P	PHI (x 10 exp-4)
1) Nord de Montréal	9214	1375 (14,9%)	9,9
2) Île de Montréal	11499	1016 (8,8%)	8,7
3) Sud de Montréal	12907	1682 (13%)	10,1
4) Trois Rivières	6289	1026 (16,3%)	21,5
5) Portneuf	2520	463 (18,4%)	20,3
6) Lotbinière	1462	447 (30,6%)	35,2
7) Région de Québec	8672	588 (6,8%)	6,2
8) Île d'Orléans	2482	419 (16,9%)	10,0
9) Beaupré et Charlevoix	4098	1462 (35,7%)	40,9
10) Bas St-Laurent et Beauce	10729	3195 (29,8%)	20,8
Total:	69872	11673	

De toute évidence, les ancêtres communs des conjoints de la région "10" sont d'une part, moins nombreux et d'autre part, à une plus grande distance généalogique que ceux des conjoints de la région "9". En effet, pour chaque couple apparenté, on dénombre en moyenne dans la première de ces régions 3,1 ancêtres communs alors que dans la seconde, on en dénombre 4,0. De plus, un simple coup d'œil au tableau 1.2 révèle que les mariages se font plus souvent dans la parenté plus éloignée dans la première que dans la seconde. On remarquera qu'il en va de même, quoique dans une moindre mesure, lorsqu'on compare la région du Nord de Montréal à celle de l'Île de Montréal, sauf que cette fois, ce sont les proportions de conjoints ayant au moins un ancêtre commun qui diffèrent pour un coefficient de parenté équivalent.

Il faudrait, pour y voir plus clair, étudier en profondeur les modalités de l'interaction entre le cloisonnement géographique, le nombre de fondateurs, le niveau d'apparentement entre ceux-ci, leur descendance utile, etc. Mais on peut déjà supposer que la différence tient à ce que les aires de mariages sont beaucoup plus réduites dans un cas (9; Beaupré et Charlevoix) que dans l'autre (10; Bas Saint-Laurent et Beauce). Sur la Rive Nord, on trouve de petites enclaves

recroquevillées en de véritables cellules auto-reproductives, où donc se multiplient les chemins généalogiques reliant l'un à l'autre les conjoints potentiels. De l'autre côté, sur la rive sud, c'est une tout autre histoire: étant donné une plus grande mobilité familiale, le brassage des effectifs mariables de provenances diverses y est constant et de fait, les liens généalogiques y sont plus ténus.

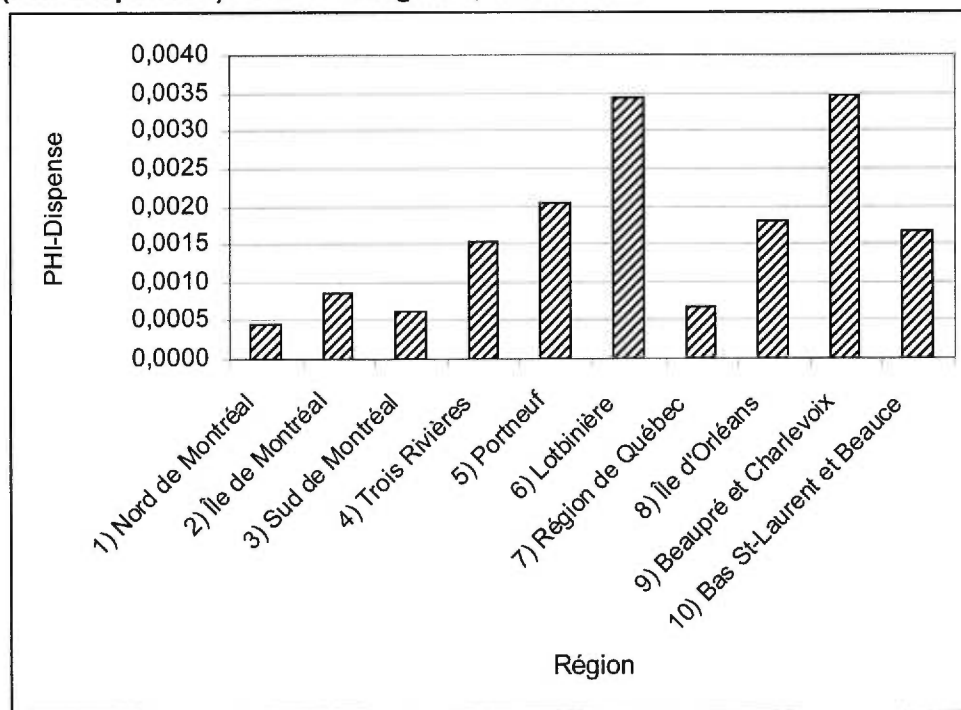
Nous n'avons d'autre choix que de supposer un certain déterminisme géographique pour expliquer la grande différence qui se pose entre ces deux régions. Voyons Beupré et Charlevoix d'abord. Coincés entre la montagne et le fleuve, ses habitants ne disposent que d'une mince frange de terres labourables pour cultiver (Blanchard, 1935). Les contacts entre paroisses ne sont pas chose facile; il faut traverser de hauts cols, souvent tard enneigés, pour se rendre au village le plus près. On s'y reproduit donc entre soi. À la différence de ce qui se passe dans les contrées montagneuses du vieux continent (Pyrénées, Alpes par exemple), la nuptialité et la fécondité n'accusent aucun recul devant l'accroissement de la population. Ainsi, lorsque la pression démographique fait mine d'écraser les habitants contre les parois des montagnes, l'enceinte cloisonnée éclate. Dans la première moitié du XIXe siècle, des contingents quittent leurs communautés respectives et fondent le Saguenay-Lac-Saint-Jean, d'où l'expression « colonisation par bonds successifs » de G. Bouchard (1996) pour rendre compte de l'évolution de l'occupation de la rive Nord du Saint-Laurent à l'est de Québec. Bien sûr, sur la rive sud du fleuve (région 10), chaque village constitue également le relais pour le peuplement de nouvelles terres, mais le processus y est plus graduel. La plaine du Saint-Laurent s'y fait plus large et s'y prolonge d'autre part beaucoup plus à l'est. Ce n'est qu'au Bic que les Appalaches rappellent leur présence, avec éloquence il faut dire. De même, le relief beauceron n'est guère plus accidenté. La mobilité ainsi facilitée est un vecteur d'hétérogénéité génétique qui agit en continu, compte tenu des conditions de peuplement qui y prévalaient au départ, s'entend.

Les facteurs géographiques, toutefois, n'épuisent pas à eux seuls l'ensemble des déterminants. Et même, ils semblent parfois ne jouer aucun rôle. Voyons l'Île d'Orléans (région 8), un cas fort intéressant. On y eût imaginé un apparentement entre conjoints beaucoup plus important (voir figure 1.3). C'est une île, soit le prototype classique de l'isolat géographique dans la plupart des études qui portent sur la consanguinité. Qu'on pense par exemple à l'Île-aux-Coudres. De plus, elle se situe dans la périphérie immédiate de la région de Beaupré et Charlevoix, précisément celle où l'apparentement global est à son plus haut niveau. À peine 1km la sépare de la côte qu'elle longe scrupuleusement, comme si elle venait tout juste, à l'échelle du temps géologique, d'en être détachée. Or cela s'explique: d'une part, cette île n'était pas très isolée et d'autre part, elle a été peuplée très tôt « d'un seul jet » par des fondateurs de provenances diverses (en somme, des immigrants fraîchement débarqués dans la colonie). Ce n'est pas que les mariages dits consanguins par l'Église (de degré supérieur ou égal à 4:4) n'y soient pas courants (voir tableau 1.2). À la fin du XVIII^e en tout cas, ils le sont autant que dans Lotbinière, qui pourtant se caractérise alors par un coefficient moyen de parenté significativement plus élevé (figure 1.2). Ce sont en fait, parmi les mariages consanguins, les mariages de type 3:3 qui y font défaut (rappelons-nous que ce sont eux qui contribuent le plus à l'apparentement global). Ainsi, tout se passe comme s'il y avait plus souvent à l'Île d'Orléans que dans Lotbinière un conjoint disponible de degré 4:4. À l'inverse, il semble que les habitants de Lotbinière soient, plus souvent qu'à l'Île d'Orléans, forcés de chercher leur conjoints dans la parenté plus proche (degré 3:3), parce que les fondateurs de la première heure, moins nombreux et moins diversifiés, y auraient laissé une descendance globalement plus apparentée. À cet égard, il serait intéressant d'évaluer la proportion de frères et sœurs qui ont fondé cette région: sans doute est-elle assez importante.

1.2.5 La fiabilité des dispenses

Pour terminer, il peut être utile de vérifier si le relevé des dispenses constitue un moyen sûr de caractériser l'apparentement entre conjoints. On le sait, la mesure qu'on en tire sous-évalue nettement l'apparentement puisqu'elle ne tient compte, la plupart du temps, que des deux ancêtres communs les plus près des conjoints (p.ex. les arrière-grands-parents) et ne va pas au delà de quatre générations (degré 4:4). Elle ne concerne que la parenté proche. Mais, et cela est ici d'une importance capitale, risque-t-elle de ne pas bien reproduire la hiérarchie des régions ? À en juger par la figure 1.4, où sont calculés les coefficients de parenté à partir des dispenses (PHI-Dispenses) pour chaque région et pour la période 1780-1799, cela n'est pas certain.

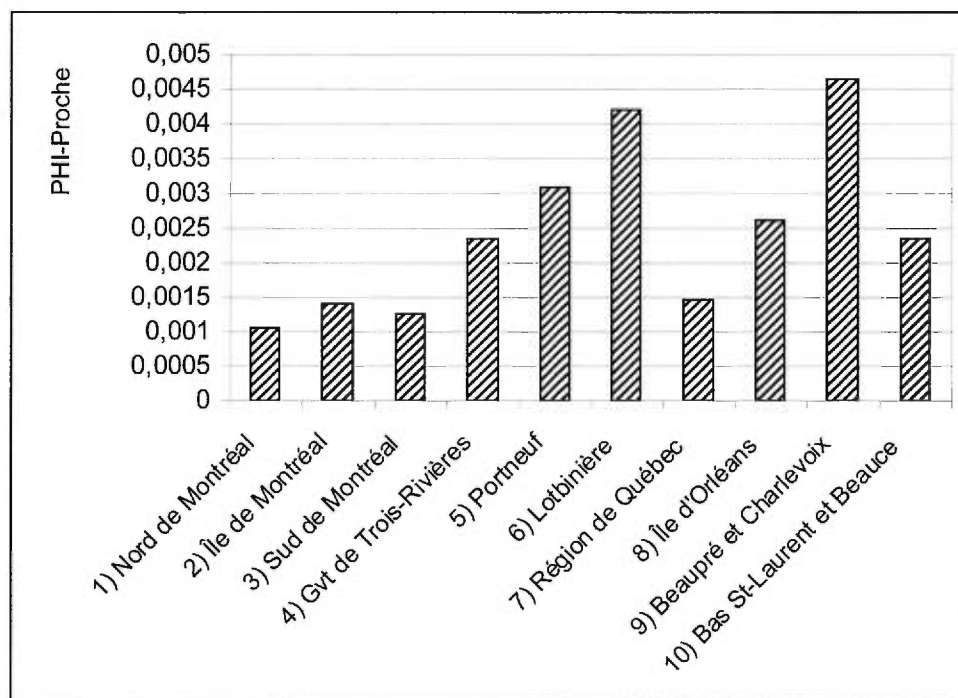
Figure 1.4 - Coefficient moyen de parenté des conjoints calculé à partir des dispenses (PHI-Dispenses) selon les régions, 1780-1799.



On constate en effet que, tout au moins à la fin du XVIII^e siècle, Lotbinière y rejoint presque Beauré et Charlevoix. La vigilance des autorités ecclésiastiques aurait-elle été à ce point plus grande dans la première de ces régions que dans la seconde ? C'est peu vraisemblable. On peut s'en assurer en calculant un coefficient de parenté qui ne tient compte que des deux

ancêtres communs les plus près et dont la distance généalogique n'excède pas 4 unités, mais cette fois à partir de l'armature généalogique du RPQA, soit le PHI-Proche. On trouve en effet le même classement: Lotbinière et Beauré/Charlevoix arrivent presque nez à nez (figure 1.5).

Figure 1.5 - Coefficient moyen de parenté proche des conjoints (PHI-Proche) selon les régions, 1780-1799.



On peut donc dire que le relevé des dispenses permet de dégager des tendances générales, de comparer sommairement les régions, mais qu'il ne permet pas toujours d'évaluer dans quelle mesure ces régions diffèrent les unes des autres, du moins est-il difficile d'en être assuré. En effet, si les dispenses avaient constitué notre seule source d'information, nous aurions rangé Beauré/Charlevoix et Lotbinière côte à côte au premier rang. Remarquons d'autre part qu'à ne s'en tenir qu'au calcul du PHI-Proche, on risque encore de ne pas bien mettre en lumière l'évolution globale de la parenté. À la figure 1.2 (PHI), la première se démarque nettement de la seconde alors qu'il en va tout autrement ici. L'avance qu'aura prise la région de Beauré et Charlevoix au cours des périodes précédant celle de 1780-1799 ne s'y répercute pas. D'ailleurs,

dans l'ensemble, les écarts relatifs entre les régions sont ici moins importants que ceux relevés à la figure 1.2. Mais cette diminution des écarts relatifs s'explique surtout par la baisse qu'accusent les régions les plus consanguines, alors que les autres conservent à peu près le même résultat.

CONCLUSION

Effets fondateurs « amplifiés » et effets fondateurs « neutralisés »: tels sont les axes majeurs de l'évolution et de la répartition régionale de l'apparentement entre conjoints au Québec ancien. Certes la colonie dans son ensemble est peuplée par un nombre restreint d'immigrants, des immigrants qui pour la plupart arrivèrent au XVII^e siècles et auxquels nous devons l'essentiel de la population actuelle. Certes le niveau de l'apparentement global, pour cette raison, n'a de cesse d'y augmenter au cours du XVIII^e siècle. Mais en définitive, le filet de l'homogénéisation génétique ne recouvre pas uniformément l'ensemble du territoire. Dans certaines régions, le noyau du peuplement originel est si compact et si peu diversifié qu'il n'est point aisé, pour les descendants qui en sont issus, de rencontrer un conjoint qui ne leur soit pas apparenté. Ce n'est pas que l'on y choisisse délibérément un parent. C'est plutôt que le conjoint étranger se fait rare. Ainsi dans ces contrées les enjeux du marché matrimonial sont davantage déterminés par la disponibilité que par le choix du conjoint. Ailleurs, le peuplement est plus hétérogène. L'apport de nouveaux immigrants - qui s'estompe toutefois après la Conquête -, de même que la mobilité inter-paroissiale, alimentent en continu le marché matrimonial de sang nouveau et contribuent en conséquence à ce que, la plupart du temps, aucun ordre de parenté autre que celui de l'alliance ne lie les conjoints entre eux.

Ce travail constitue un premier jalon dans l'étude de la stratification du bassin génétique québécois à ses origines. Il s'agissait de suivre de près cette trame ininterrompue de la

multiplication et de l'enchevêtrement des liens de parenté au sein de la colonie afin d'en faire ressortir les paramètres socio-démographiques, pouvant éclairer les données récentes de la biologie moléculaire. Il convient maintenant, et nous y viendrons dans un prochain travail, de mesurer la contribution génétique des fondateurs de la Nouvelle-France et d'évaluer les distances génétiques entre régions. Cela nous permettra d'acquérir une meilleure connaissance de la structure génétique de la population du Québec et de clarifier le rôle d'autres facteurs démographiques non abordés ici, dont, entre autres, les migrations et la fécondité.

CHAPITRE 2 - PATRONYMES , NUMÉROS DE LIGNÉE ET GÉNÉALOGIES : ÉVOLUTION COMPARÉE DE TROIS INDICES DE L'APPARENTEMENT AU QUÉBEC ANCIEN (1608-1800)

À paraître dans : Les chemins de la recherche, no 44, sous la direction de A. Bideau, G. Brunet, P. Darlu, B. Desjardins et G. Zei. Cette étude a été financée en partie par le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR) du Québec et par le Conseil de Recherches en Sciences Humaines du Canada (CRSH). Je tiens à remercier messieurs Hubert Charbonneau, Bertrand Desjardins, Jean Daviault et Stéphane Gilbert pour leur aide précieuse.

INTRODUCTION

Par une heureuse disposition, la fréquence des mariages impliquant des individus dont le patronyme est le même nous donne une estimation de l'apparement des conjoints. Le principe est fort simple. En général, un oncle et une nièce (ou bien une tante et son neveu) partagent la moitié du temps le même patronyme alors que leur coefficient de parenté est égal à $1/8$. De même, les cousins sont isonymes une fois sur quatre (pour un coefficient de parenté de $1/16$), les petits cousins une fois sur seize (pour un coefficient de parenté de $1/64$) et ainsi de suite. Ainsi, pour deux individus susceptibles de contracter une alliance, la probabilité qu'ils aient le même patronyme est quatre fois supérieure à leur coefficient de parenté. Il suffit donc de diviser la fréquence des mariages isonymes par quatre pour obtenir une estimation du coefficient moyen de parenté des conjoints.

La méthode paraît bien simple et très prometteuse. De fait, au lendemain de sa publication (Crow et Mange, 1965), elle connut un franc succès. Toutefois, elle nécessite certaines hypothèses qui s'avèrent souvent difficiles à vérifier (Ellis et Startmer, 1978; Crow, 1980; 1983; Tay et Yip, 1984; Lasker, 1985; 1989). Voici les principales: (1) d'abord, et c'est la plus importante, chacun des patronymes doit avoir une origine monophylétique, c'est-à-dire qu'il doit provenir d'une souche unique; (2) ensuite, la transmission des patronymes doit être strictement corrélée à la transmission biologique. Ici sont en cause les changements de noms, de leur épellation, les adoptions, l'illégitimité et les simples erreurs de transcription; (3) puis, l'écart à la panmixie doit être symétrique en regard du sexe. Par exemple, les cousins issus de frères doivent se marier dans la même proportion que ceux issus de sœurs. (4) Enfin, parmi les migrants, les hommes et les femmes doivent être également représentés.

Bien que l'on attire souvent l'attention sur ces hypothèses, il est rare qu'on soit à même d'en tester la validité. Or, la population du Québec ancien se prête remarquablement bien à l'exercice. L'essentiel des événements démographiques qui la caractérise depuis ses origines (en 1608) jusqu'à 1800 est en effet reconstitué dans une base de données informatisée contenant plus de 700000 actes de baptême, mariage et sépulture (Le RPQA). Tous les mariages y sont jumelés et les liens qui rattachent les conjoints jusqu'aux fondateurs sont connus. Cette connaissance des conditions initiales constitue un atout majeur que nous tenterons de mettre à profit.

Dans un article antérieur (Gagnon et al., 1998), le registre a fait l'objet d'une mesure de l'apparentement des conjoints par région à partir de la méthode généalogique. Il s'agira maintenant de comparer cette mesure à celle qui peut être obtenue au moyen des patronymes. De plus, nous proposerons une nouvelle manière d'estimer l'apparentement et qui fait appel au « numéro de lignée paternelle », un « patronyme idéal », dont nous ferons état ici. Chaque

mesure sera comparée tour à tour avec chacune des deux autres, selon le même découpage temporel et géographique qui a été adopté dans l'étude susmentionnée (voir Carte 1; introduction).

2.1 DONNÉES ET MÉTHODES

2.1.1 Les patronymes dans un monde « habituel »

Pour tester la validité de la méthode isonymique, nous tenterons de nous rapprocher le plus possible des conditions qui prévalent en général lorsqu'elle est employée. Autrement dit, nous procéderons comme si nous ne disposions que d'un répertoire d'actes de mariage non-jumelés, soit l'index des 67777 actes de mariages du RPQA. Les unions connues, c'est-à-dire aussi bien celles confirmées par un acte de mariage que celles déduites de la reconstitution des familles, sont au nombre de 70869. Toutefois, puisque nous voulons être au diapason des études isonymiques, nous utiliserons les patronymes *standardisés* auxquels sont rattachées les mentions qui apparaissent dans ces actes. En effet, un vaste travail de standardisation orthographique précède toujours l'analyse proprement dite (voir à ce sujet Weiss et al., 1983). Dans le cas présent, la standardisation est faite à partir de l'ensemble des mentions (qu'il s'agisse d'acte de mariage, de baptême ou de sépulture) où apparaissent les variantes d'une forme orthographique de base. Le critère qui préside à l'attribution d'une forme définitive est généralement celui de la fréquence; si une « manière d'écrire » revient plus souvent que les autres, elle constitue alors la « vedette », et est retenue à titre de patronyme standardisé.

Sur la base de cet index des patronymes standardisés, nous estimerons le *coefficient moyen de parenté des conjoints* de manière telle qu'indiquée en introduction. En d'autres termes, nous estimerons $\Phi_p = P_p / 4$, où P_p = proportion de mariages impliquant des conjoints dont le patronyme est identique. Comme on peut déjà s'y attendre toutefois, cette estimation

sera biaisée car les quatre hypothèses dont nous avons fait état (origine monophylétique, transmission intégrale des patronymes, écart symétrique à la panmixie en regard du sexe et immigration égale des sexes) ne sont certainement pas entièrement respectées. Si l'étude s'arrêtait là, on n'aurait aucun moyen d'évaluer la part respective des erreurs qu'entraînent l'un ou l'autre des écarts à ces hypothèses. Toutefois, comme nous l'avons mentionné dans l'introduction, nous disposons de données qui nous permettront d'y voir plus clair.

2.1.2 Les patronymes dans un monde idéal: les numéros de lignée

Dans un monde idéal pour le généticien, tous les noms de famille ont une souche unique. Chaque patronyme introduit par un immigrant ne l'est qu'une seule fois et demeure invariable dans le temps. Il ne change pas d'orthographe d'une génération à l'autre de manière à en engendrer de nouveaux, ou à s'agglutiner à d'autres dont la calligraphie est semblable. Aucun père ne transmet son prénom ou son surnom à ses enfants. Aucun individu n'est nommé en vertu de la profession qu'il exerce. En remontant une généalogie en ligne paternelle, on rencontre invariablement le même patronyme (dans son orthographe intégrale) et l'on se rend sans ambages à l'immigrant qui l'a introduit.

Le *Registre de population du Québec ancien* nous permet, de manière détournée, de se rapprocher très près de ces conditions pour le moins utopiques. En effet, à chaque individu qui figure dans ce registre est associé un numéro de lignée qui le rattache, en ligne paternelle, à l'ancêtre le plus loin rencontré dans sa généalogie, sa tête de lignée. Ce peut être un immigrant ou encore le père de cet immigrant, puisque ce dernier, ainsi que sa femme, est souvent nommé dans l'acte de mariage dudit immigrant. La plupart du temps, le patronyme et le numéro de lignée correspondent. Par exemple, tous les individus qui portent le nom « Tremblay » (dans son orthographe standardisée) ont le même numéro de lignée. Mais souvent, la correspondance est

très ténue. Ainsi en va-t-il des dénommés « Roy ». À ce seul nom correspondent au moins 29 lignées différentes!

Nous disposons donc d'une série de patronymes « idéaux », à partir desquels nous pouvons estimer l'apparentement des conjoints comme suit: $\Phi_1 = P_1 / 4$ où P_1 est la proportion de conjoints qui partagent un *numéro de lignée* identique, c'est-à-dire la proportion de ceux qui sont rattachés en ligne paternelle au même ancêtre, qui constitue la tête de lignée.

2.1.3 Patronymes, numéros de lignées et généalogies: trois sources qui se complètent

Comme nous l'avons mentionné en introduction, nous pouvons enfin faire appel au coefficient de parenté Φ_g estimé à partir de la méthode généalogique. Ceci porte à trois le nombre de mesures que nous allons comparer. Il est important de souligner d'entrée qu'aucune d'elles n'a de valeur absolue. De plus, ni l'une ni l'autre ne peut constituer un étalon définitif à partir duquel on tenterait d'évaluer les biais que contiennent les deux autres. Il s'agit plutôt d'obtenir de leur conjugaison une meilleure idée de l'apparentement des conjoints. Certes, la méthode généalogique, particulièrement lorsque tous les liens sont connus jusqu'aux fondateurs de la population (et c'est le cas ici), est la plus appropriée. Mais, elle suppose que ces fondateurs ne sont pas apparentés entre eux - ce qui, on le sait n'est vrai qu'en partie. Par contre, la méthode isonymique permet de prendre en compte l'apparentement des fondateurs, sous réserve que les patronymes ont une souche unique (Crow, 1980). Si la mesure qu'on en tire dépasse de loin celle obtenue au moyen des généalogies, alors le dilemme suivant se pose: est-on en présence de souches multiples? Ou bien les fondateurs étaient-ils effectivement apparentés au départ? Ou bien encore ces deux effets se conjuguent? Comment lever l'incertitude? Par le recours aux numéros de lignée. De cette manière en effet, on élimine les cas d'homonymie et on s'assure que chaque souche est unique, du moins depuis son introduction dans la population canadienne! (Car deux numéros de lignée distincts peuvent encore

correspondre à un patronyme monophylétique qui témoignerait d'une parenté lointaine dans la population d'origine). De cette manière également, nous pouvons apprécier l'effet des processus socio-démographiques qui entraînent des écarts par rapport aux hypothèses qui justifient la méthode isonymique.

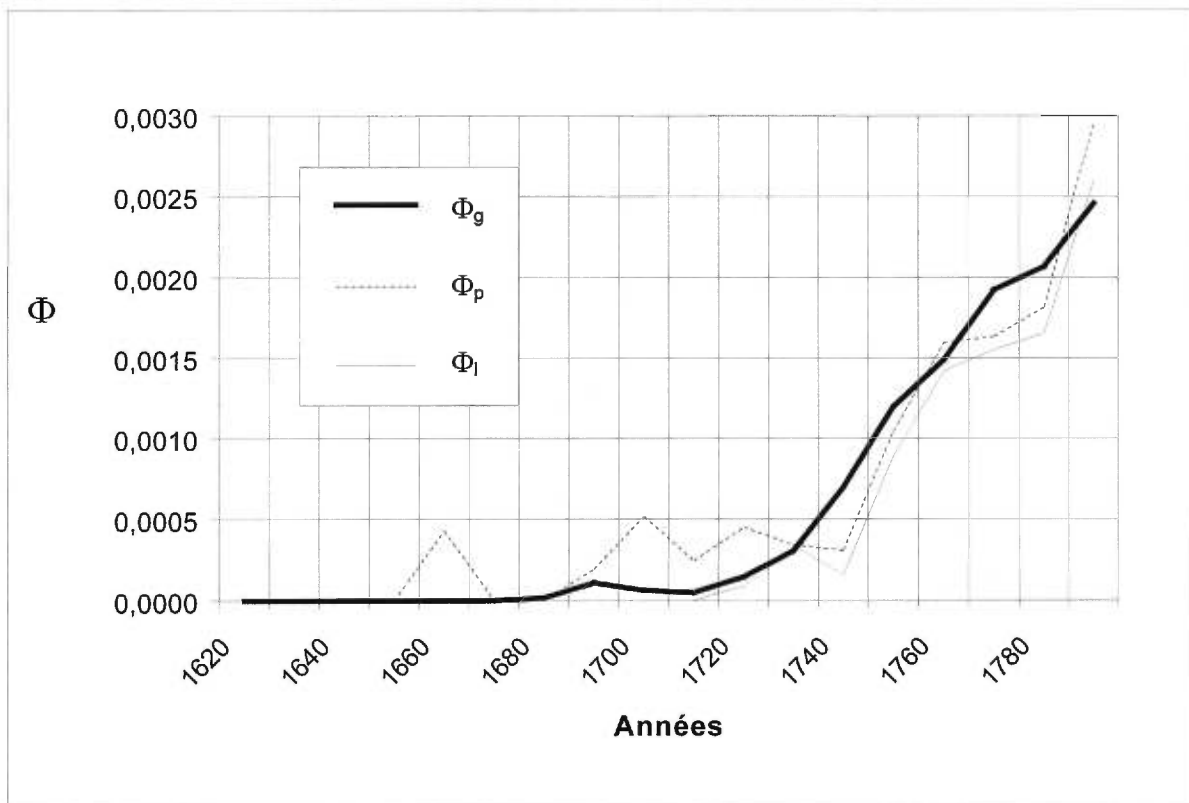
2.2 RÉSULTATS

2.2.1 Évolution de l'apparement moyen des conjoints dans l'ensemble de la colonie

À l'instar de ce qui a été fait dans l'étude généalogique, nous allons d'abord traiter de l'évolution de l'apparement des conjoints dans l'ensemble de la colonie, entre 1620 et 1799.

Les résultats des trois coefficients Φ_g , Φ_p et Φ_l sont représentés sur la figure 2.1 qui suit:

Figure 2.1 Évolution de l'apparement des conjoints au Québec ancien.



Remarquons tout d'abord que les deux courbes isonymiques (Φ_p et Φ_l) sont beaucoup plus accidentées que la courbe généalogique (Φ_g). Cela est attendu, compte tenu du fait que les

deux premières sont beaucoup plus sujettes aux fluctuations aléatoires (voir à ce sujet Lasker, 1988). Un mariage est ou n'est pas isonyme; dans le premier cas, sa contribution, dans le calcul du coefficient moyen de parenté, est de 1/4 dans le second, elle est de zéro. Il en va autrement de la contribution d'un mariage consanguin à la mesure de Φ_g : cette contribution constitue une variable pratiquement continue (d'autant plus continue en fait que la profondeur généalogique est grande). Or, les fluctuations deviennent de moins en moins importantes au fil du temps, à mesure que l'augmentation des effectifs accroît du même coup la probabilité des « rencontres » isonymiques.

On note également que les trois courbes n'ont pas le même « point de départ ». Le coefficient calculé au moyen des patronymes standardisés (en pointillé) connaît un départ rapide (dès la décennie 1660-1670) dont on ne saurait se surprendre, compte tenu des nombreux cas d'homonymie qui figurent parmi les fondateurs de cette période (les « Roy » en forment un exemple probant). D'ailleurs, l'apparement tel que déduit à partir des numéros de lignées patronymiques (Φ_i) n'apparaît pas avant la décennie 1720-1730. Il s'agit là d'un « départ » très tardif qui pourrait sembler difficile à expliquer. En effet, l'apparement des conjoints, mesurée au moyen de l'armature généalogique (en trait gras) le précède d'environ 40 ans.

Pour expliquer ce « retard », il faut considérer un autre aspect, qui tient à la nature du peuplement du Québec ancien. Cette population ne s'est pas constituée d'un seul jet, à la faveur d'une immigration instantanée d'un nombre égal d'hommes et de femmes. Les travaux du PRDH (voir en particulier Charbonneau et al., 1987) nous apprennent que dans la première moitié du XVIIe siècle, l'immigration pionnière (qui débuta en 1608) était fort ténue et qu'elle avantageait nettement les hommes (l'immigration pionnière est constituée exclusivement de la portion de l'immigration qui s'est fixée définitivement dans la colonie en s'y établissant en famille avant 1680; *ibid.*: 14). Avant 1663 en effet, on ne compte encore qu'un peu plus de 1400 pionniers

dont le rapport de masculinité est de 169 hommes pour 100 femmes. De 1663 jusqu'à 1680, les entrées s'intensifient et le déséquilibre des sexes s'atténue: la Nouvelle-France reçoit alors près de 2000 nouveaux pionniers dont le rapport de masculinité est cette fois de 118 hommes pour 100 femmes. Par la suite, le mouvement naturel prend définitivement le pas sur le mouvement migratoire et c'est pour cette raison qu'on peut considérer les nouveaux arrivants comme des immigrants plutôt que comme des pionniers. Cette immigration est surtout le fait d'engagés et de militaires et est donc essentiellement composée d'hommes (ibid.).

Ces conditions initiales du peuplement ont une influence déterminante sur le profil des courbes isonymiques. Elles entraînent une sous-estimation. En effet, *le déséquilibre des sexes des pionniers et immigrants a eu pour effet d'introduire une diversité patronymique plus grande que si le mouvement migratoire avait été sous ce rapport équilibré*, ce qui contredit les hypothèses de base de la méthode²⁸. Imaginons le scénario suivant : Un immigrant épouse une Canadienne. Ses enfants reçoivent alors un patronyme qui est par définition absent de la population. Lorsqu'à leur tour ces enfants entrent sur le marché matrimonial, il est possible que l'un d'entre eux, disons un garçon, épouse une parente en ligne maternelle. Or, la mesure isonymique ne pourra prendre en compte la contribution de ce mariage pourtant consanguin. En effet, celui-ci mettra en relation des individus qui, nécessairement, n'ont pas le même patronyme. Il faut donc interpréter avec prudence les résultats que révèle la méthode pour les périodes anciennes de la colonie. Mais que peut-on dire à cet égard des périodes plus récentes?

²⁸ Si l'immigration avait avantage les femmes, alors on aurait surestimé l'apparement. En fait, pour plus de généralité, on peut raisonner en terme de bilan migratoire et prendre en compte également l'*émigration*. Par exemple, une proportion d'émigrants plus grande chez les femmes que chez les hommes conduirait également à une sous-estimation de l'apparement et inversement (voir à ce sujet Ellis et Startmer, 1978).

Considérons les valeurs vers lesquelles s'acheminent nos trois courbes en fin de période. Pour la décennie 1790-1799, $\Phi_p = 0,00296$, $\Phi_l = 0,00259$ et $\Phi_g = 0,00245$. Contrairement à ce qui est souvent constaté, les valeurs calculées au moyen de l'isonymie y sont relativement près de celles calculées au moyen de l'armature généalogique de la population, et en constituent donc une très bonne approximation. Certains chercheurs font état d'une valeur pouvant être 3 ou 4 fois supérieure à la valeur généalogique (Crow et Mange, 1965; Ellis et Friel, 1976, Prost, 1998), surestimation qui est à mettre sur le compte de l'action de deux causes qui peuvent s'additionner : l'origine polyphylétique des patronymes et l'apparentement qui remonte au delà de la profondeur généalogique connue. Ici, ni l'origine polyphylétique, ni l'apparentement des fondateurs, ne semblent jouer un grand rôle. On pourrait donc conclure que la méthode isonymique est dans ce cas-ci tout à fait appropriée pour estimer l'apparentement des conjoints dans la population.

Toutefois, on ne sait pas dans quelle mesure l'effet de l'immigration, toujours à forte composition masculine, se répercute dans le temps. Théoriquement, à mesure que le poids de l'immigration diminue par rapport à celui du mouvement naturel, le biais devient de moins en moins important et, au fil des générations, il tend à s'effacer²⁹. Au Québec ancien, la proportion d'immigrants par rapport à la population totale passe sous les 10% dès 1680 et sous les 3% à partir de 1760 lors de la Conquête anglaise (Charbonneau, 1984). Quels auront été les effets de ces mouvements sur notre estimation isonymique en fin de période? Cela reste à déterminer de manière précise.

²⁹ Le biais ne sera jamais nul car au bout de n générations, quel que soit n, les conjoints sur le marché matrimonial partageront toujours un stock patronymique plus diversifié que si les sexes avaient été également représentés parmi les fondateurs et les immigrants. Un travail est présentement en cours afin d'évaluer l'importance de ce biais.

Voilà donc pour l'évolution de l'apparement des conjoints dans l'ensemble de la colonie. Voyons maintenant si ce portrait d'ensemble cache des variations régionales importantes. Nous aurons alors l'occasion d'observer des effectifs plus réduits, où donc les fluctuations aléatoires peuvent jouer un grand rôle.

2.2.2 Apparement moyen des conjoints par région.

Afin de faciliter la comparaison, nous utiliserons le découpage qui a été adopté dans un précédent article (Gagnon et al., 1998). Nous avons alors subdivisé le Québec ancien en 10 régions et exprimé l'évolution de l'apparement des conjoints par tranches de 20 ans, de 1720 à 1800. Comparons d'abord les coefficients Φ_g et Φ_p (figure 2.2 et figure 2.3)³⁰:

³⁰ Pour les trois prochaines figures, les effectifs en présence (par région et par période) sont présentés en annexe.

Figure 2.2. Coefficient de parenté des conjoints par région calculé à partir des généalogies (Φ_G)

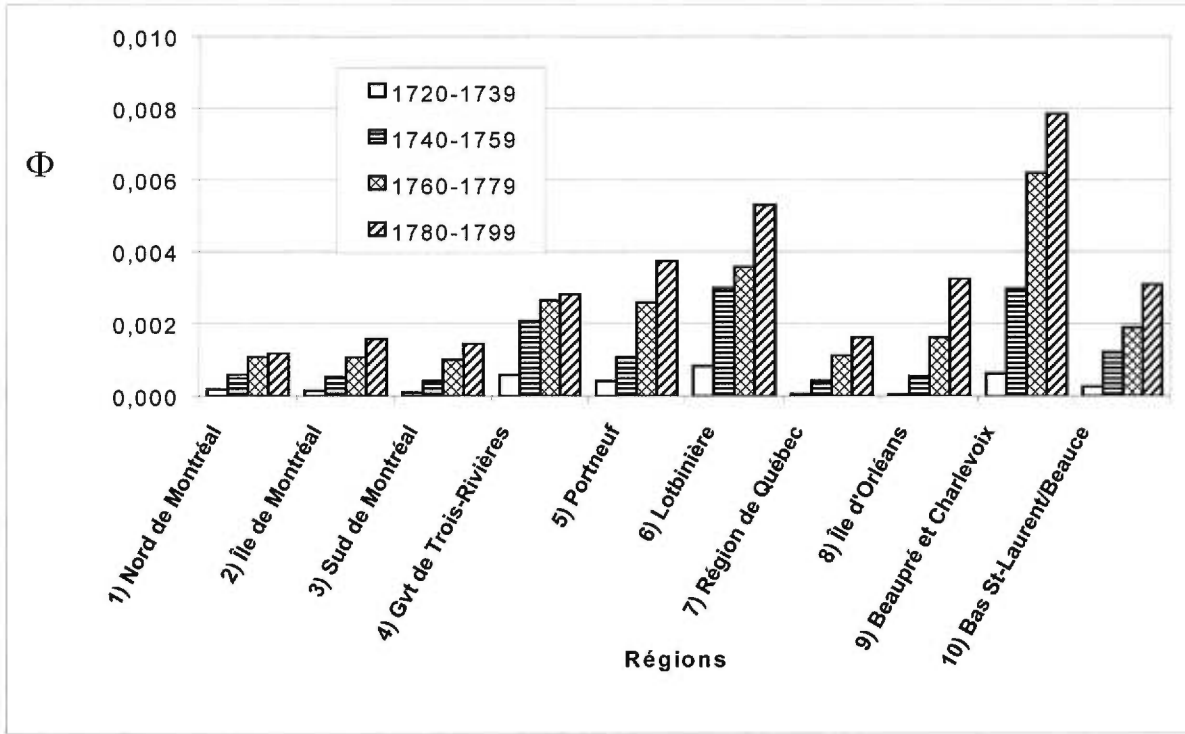
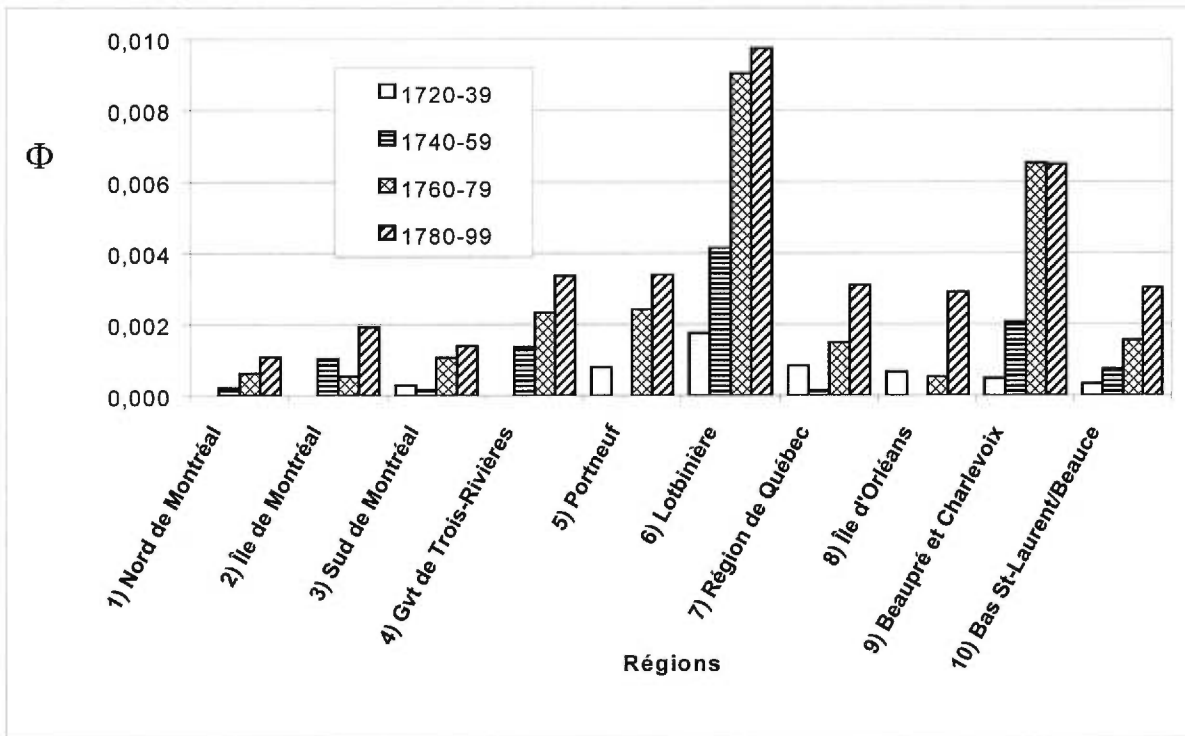


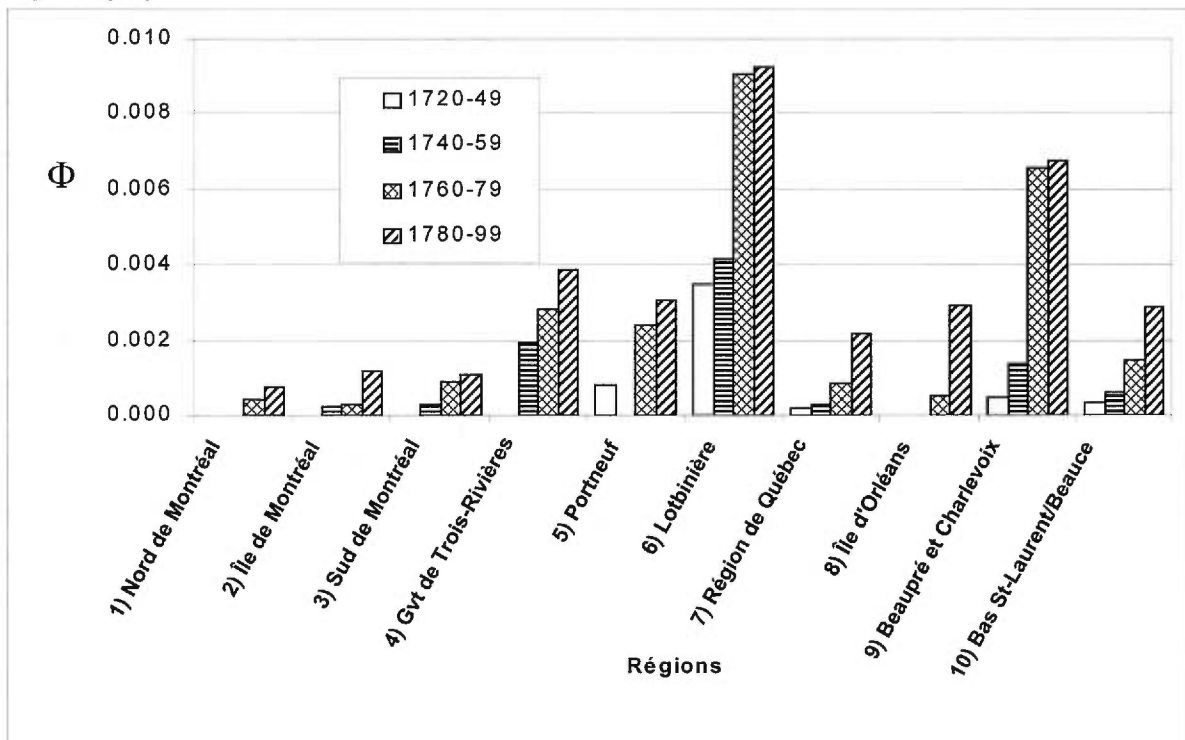
Figure 2.3. Coefficient de parenté des conjoints par région calculé à partir des patronymes (Φ_P)



Ce qui frappe en premier lieu, ce sont les résultats obtenus à l'endroit de la région de Lotbinière: pour chaque période considérée, le coefficient Φ_p (figure 2.3) y est environ de deux fois supérieur au coefficient Φ_g (figure 2.2). Si l'on s'en remettait à la seule estimation isonymique, on conclurait que l'apparement des conjoints est à son plus haut niveau dans cette région. Or, si l'on se réfère encore à l'estimation généalogique (figure 2.2), on voit que c'est plutôt celle de Beauré/Charlevoix qui occupe le premier rang. De plus, dans Beauré et Charlevoix, l'estimation isonymique (figure 2.3) est même inférieure à l'estimation généalogique (figure 2.2). Mais, pour les autres régions, les résultats sont très similaires (tout au moins de 1760 à 1799).

Voyons maintenant ce qu'il advient de l'apparement lorsqu'il est calculé à partir des numéros de lignée (figure 2.4), dont nous avons dit qu'ils constituaient des « patronymes idéaux ».

Figure 2.4 - Coefficient de parenté des conjoints par région, calculé à partir des numéros de lignée (Φ_L)



Ces nouveaux résultats (figure 2.4) sont très similaires aux précédents (figure 2.3). Pour la période 1780-99 par exemple, l'apparement moyen des conjoints de Lotbinière est ici estimé à $\Phi_1 = 0,00924$ ($=1/4.18/487$) alors qu'il était précédemment estimé à $\Phi_p = 0,00976$ ($=1/4.19/487$). Cela indique que l'origine polyphylétique ne joue ici aucun rôle et que nous ne pouvons la tenir pour responsable de l'écart constaté entre les mesures généalogiques et patronymiques. D'autre part, la région de Lotbinière devance encore celle de Beupré et Charlevoix. Il en va de même des résultats qui concernent les régions 5, 7 et 10 (respectivement, Portneuf, Île d'Orléans et Bas-Saint-Laurent/Beauce); ils sont sensiblement les mêmes sur l'un et l'autre graphique.

Il y a toutefois quelques différences. Pour la région de Québec (6) par exemple, il semble y avoir une « correction » à la baisse : le coefficient Φ_1 , et ce dans toutes les périodes considérées, y est ramené à une valeur assez près du coefficient Φ_g (figure 2.2). Même son de cloche pour l'Île de Montréal (2), sauf que cette fois, il s'agit plutôt d'une « sur-correction » à la baisse. Dans ces deux régions, plusieurs patronymes homonymes proviennent de souches distinctes: cela explique l'écart entre les coefficients Φ_1 et Φ_p . Pour la région de Trois-Rivières par contre, c'est le phénomène inverse: Φ_1 (figure 2.4) y est légèrement plus élevé que Φ_p (figure 2.3), différence qui doit être imputée cette fois aux fissions de patronymes (séparation d'une souche patronymique en deux ou plusieurs lignées distinctes).

Nous constatons donc deux cas de figure : (1) surestimation de l'apparement due essentiellement à l'origine polyphylétique et (2) sous-estimation due aux fissions patronymiques. Il y en a un troisième : la compensation entre ces deux biais. Bien que nous n'ayons pas engagé ici de procédures systématiques en ce sens, nous ne pouvons manquer de supposer que leurs effets respectifs, qui s'opposent, s'annulent dans certaines régions. C'est probablement le cas

dans la région 10 (Bas-Saint-Laurent/Beauce). Nous avons constaté plus haut que les deux mesures isonymiques (Φ_1 et Φ_p) y étaient pratiquement identiques. Or, dans cette région, on trouve en grande proportion le patronyme « Roy », dont on sait qu'il se compose d'environ 29 souches différentes³¹. Par ailleurs, plusieurs souches s'y émettent en des patronymes distincts, ce qui compense pour l'origine polyphylétique.

Pour conclure cette section, résumons-nous. D'abord, si l'utilisation de la méthode isonymique s'avère assez fiable à l'endroit de l'ensemble de la colonie, elle l'est peut-être moins lorsqu'on considère séparément les régions. Nous avons vu par exemple que pour la région de Lotbinière, la mesure isonymique est environ deux fois supérieure à la mesure généalogique. Ensuite, l'emploi des numéros de lignées fournit des résultats assez près de ceux obtenus à partir des patronymes. Toutefois, cette correspondance masque parfois certains écarts, étant donné la compensation qui peut s'établir entre des phénomènes dont les effets s'annulent. Maintenant, il convient de s'attarder à une région particulière afin de bien circonscrire les mécanismes socio-démographiques qui, à cette échelle, peuvent engendrer des écarts importants par rapport aux hypothèses qui sous-tendent la méthode isonymique. La région de Lotbinière paraît à cet égard toute choisie.

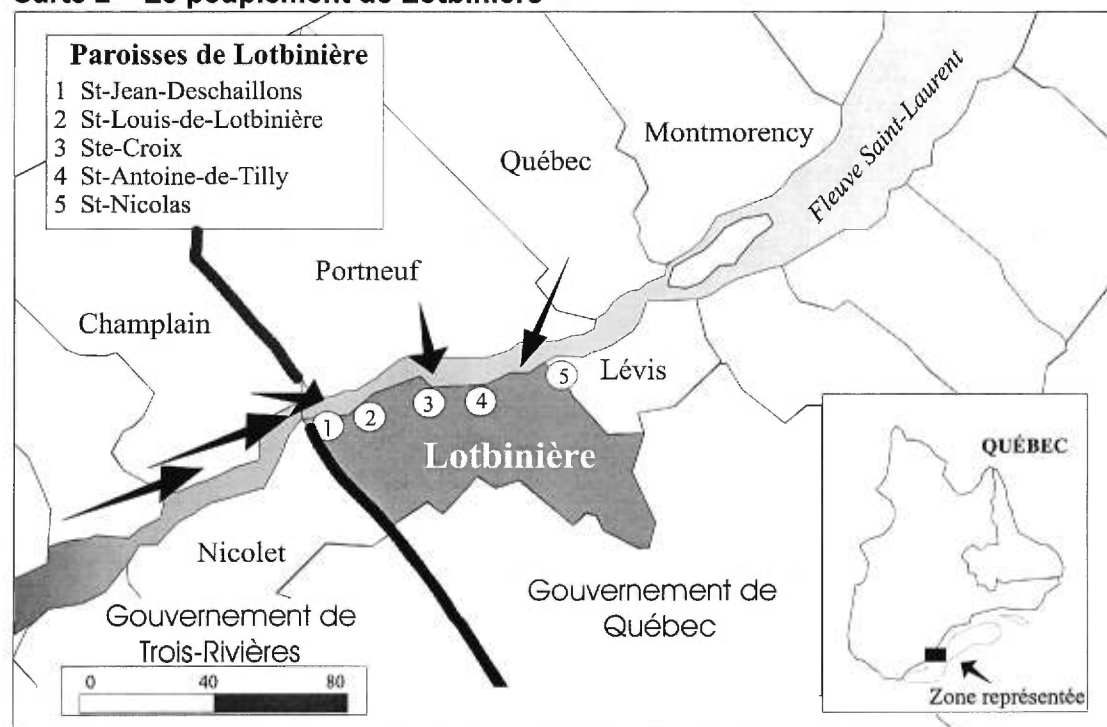
³¹ En fait, parmi ceux qui entrent en union entre 1780 et 1800 et qui portent ce nom, près de la moitié (46%) se trouvent dans cette région.

2.2.3 Étude d'un cas particulier : Lotbinière

L'utilisation des numéros de lignée nous a permis de lever une partie de l'ambiguïté quant aux résultats obtenus pour Lotbinière: la surestimation de l'apparentement ne relève pas de l'origine polyphylétique des patronymes puisque les deux mesures isonymiques (Φ_p et Φ_l) sont à peu près équivalentes. Comment alors expliquer la surestimation? Il faut d'abord, pour s'y employer, faire intervenir l'histoire du peuplement de cette région. Grâce à la richesse des données, nous pourrons ainsi mettre en lumière le rôle des conditions initiales, et les saisir dans le vif de leur manifestation empirique. Ces conditions sont ici déterminantes puisque l'écart constaté entre les mesures apparaît dès la période 1720-1740, qui recoupe en partie celle de l'entrée sur le marché matrimonial des premières générations nées dans la région. Toutefois, nous ne proposerons ici qu'une analyse partielle qui demandera à être complétée par d'autres études plus poussées.

Le peuplement de Lotbinière (voir carte 2), à l'instar de celui du reste de la colonie d'ailleurs, s'est fait avec une extrême lenteur. Bien que des fiefs y aient été créés dans la première moitié du XVIIe siècle, il faut attendre les années 1670 avant d'y trouver quelques familles établies (Samson et Héroux, 1996). À l'époque, la menace iroquoise pèse toujours sur la rive sud du Saint-Laurent! De quoi est composé ce front pionnier? Essentiellement d'hommes qui cherchent à s'établir avec leurs fils, car les meilleurs emplacements de la rive nord, ceux qui longent le Fleuve, sont alors tous occupés. Des cousins viennent ensuite rejoindre les premiers arrivants.

Carte 2 – Le peuplement de Lotbinière



PRDH - S. Gilbert et A. Gagnon

Ces chaînes migratoires conduisent ainsi les pionniers à s'établir près des leurs, dans la même paroisse, sinon dans celle avoisinante (ibid.). Or, au recensement de 1681, on ne compte encore à Saint-Louis-de-Lotbinière, chef lieu de la région, que onze ménages, six hommes célibataires entre 20 et 35 ans, et une seule fille en âge d'être épousée³². La petite population se trouve alors dans la nécessité d'importer des filles de l'extérieur, ce à quoi elle parvient, puisque à l'aube du XVIIIe siècle, la rapport des sexes est plus favorable à son expansion. Or, étant donné qu'elles provenaient de l'extérieur, ces filles étaient moins apparentées à la population d'accueil que les hommes qui les y ont pris pour femme.

³² *Répertoire des actes de baptême, mariages, sépulture et des recensements du Québec ancien* (Charbonneau et Légaré, 1986).

Ce petit historique nous invite à émettre une première hypothèse pour expliquer l'écart entre les mesures généalogiques et isonymiques: 1) *Parmi les fondateurs de la région, il y aurait eu plus de frères que de sœurs et en général plus de liens de parenté en ligne paternelle que maternelle.* Cela aurait eu pour conséquence de réduire la diversité patronymique et d'entraîner une surestimation de la mesure de l'apparentement.

Pour vérifier en partie cette hypothèse, on peut considérer les premiers mariages célébrés dans la région et comptabiliser les différents liens de parentés entre tous les conjoints impliqués. Or, cette liste de premiers mariages n'épuise pas l'ensemble des fondateurs de la région, car plusieurs d'entre eux se sont mariés à l'extérieur et s'y sont ensuite établis en famille. Il s'agit donc ici d'une vérification partielle de l'hypothèse évoquée. Pour l'instant, nous constatons tout de même que, parmi les individus qui figurent dans la liste des 20 premiers mariages célébrés dans la région (entre 1695 et 1710), il y a plus de paires « frères-frères » que de paires « sœurs - sœurs ». Deux frères « Houde », deux frères « Boucher » et deux frères « Lambert » se sont mariés durant cette période. Par contre, chez les femmes, seulement deux sœurs « Bergeron » ont convolé à la même époque. On trouve aussi d'autres types de liens de parenté parmi ces conjoints : un frère et une sœur « Houde », neveux des précédents, un frère et une sœur « Dubois », et ainsi de suite. Dans l'ensemble, seulement 7 hommes et 9 femmes ne sont apparentés à aucun autre des 40 conjoints de la liste. Ainsi, l'énumération des patronymes (et des numéros de lignée) est un peu plus courte et plus asymétrique chez les hommes que chez les femmes. Autrement dit, les hommes partagent un stock patronymique moins diversifié que ne l'est celui de leurs femmes. Tout ceci expliquerait en partie la surestimation que nous avons constatée. Toutefois, cette différence de longueur et d'asymétrie entre les listes masculines et féminines (symptôme de ce que les liens de parentés mettent surtout en relation des hommes et passent principalement par eux) ne suffit pas. Il est d'autres circonstances, celles-là peut-être plus déterminantes encore, qu'il faut considérer.

Dans les sociétés européennes traditionnelles, comme dans celles qui en sont issues, les conjoints exogames habitent la plupart du temps dans la paroisse de l'époux. *À la longue, cette « règle » de résidence patrilocale conduit à une concentration locale des patronymes et, dans de telles conditions, la mesure isonymique ne peut que surestimer l'apparentement.* Nous supposons qu'il en fut ainsi à Lotbinière et c'est là notre deuxième hypothèse. Cette fois, nous pouvons la vérifier avec plus d'assurance. Pour ce faire, nous avons calculé les proportions attendues de conjoints isonymes lorsque les mariages ont lieu au hasard: - d'abord pour l'ensemble de la région; - et ensuite pour chacune des cinq paroisses qui la composent.

Une telle proportion s'exprime comme suit: $P_a = \sum p_i q_i$, où p_i est la fréquence du patronyme i chez les hommes et q_i la fréquence de ce patronyme chez les femmes. En observant ces proportions attendues et celles qu'on observe (P_o), on peut avoir une idée assez nette de la concentration locale des patronymes. Le tableau 2.1 présente ces proportions dans l'ensemble de la région et dans chacune des 5 paroisses qui la composent, pour la période 1780-1799.

Tableau 2.1 - Proportions attendues (PA) et observées (PO) d'unions isonymes à Lotbinière, 1780-1799 (%)

	Lotbinière (Ensemble)	St-Nicolas	St-Louis-de- Lotbinière	Ste-Croix	St-Antoine- de-Tilly	Deschaillons
P_a	1.85	3.09	4.33	4.60	3.33	5.63
P_o	3.70	1.68	6.73	6.17	2.72	0.00

On remarque que dans l'ensemble de la région, la proportion *observée* d'unions isonymes ($P_o = 3,70\%$) est deux fois supérieure à celle qui est attendue ($P_a = 1,85\%$). S'agit-il d'une « homogamie patronymique » positive? En fait, puisque la proportion attendue d'unions isonymes est plus élevée dans chacune des paroisses que dans l'ensemble de la région, nous sommes plutôt, comme nous l'avons supposé, en présence d'une stratification spatiale de la

distribution des patronymes. Cette stratification a dû apparaître dès le départ. Toutefois, cela est difficilement démontrable parce que les cinq paroisses de la région ont été fondées à des époques différentes. Le premier registre de Lotbinière fut ouvert entre 1683 et 1685 et desservait l'ensemble de la région. Saint-Antoine-de-Tilly est devenue indépendante et a ouvert ses propres registres en 1702, Sainte-Croix en 1727 et Deschaillons en 1741 (Inventaire des registres paroissiaux catholiques du Québec (1621-1876); Bélanger et Landry, 1990). D'autre part, le tableau montre également qu'à l'échelle paroissiale, il semble y avoir parfois homogamie patronymique positive (Saint-Louis, Sainte-Croix) et parfois homogamie patronymique négative (Saint-Nicolas, Saint-Antoine-de-Tilly et Deschaillons). Étant donné que ces deux groupes ont un poids démographique sensiblement égal, l'homogamie patronymique ne peut expliquer la surestimation de la mesure isonymique constatée ici pour l'ensemble de la région.

Nous allons enfin faire appel à une troisième hypothèse. Elle a trait au *remariage des hommes, dont la période reproductive est plus longue que celle des femmes*. Ce phénomène introduit le même biais que la polygamie, et il faut remonter aux postulats de la méthode pour en apprécier la teneur. Cela est rarement explicité, mais l'estimation isonymique suppose que les conjoints isonymes ont *nécessairement deux ancêtres communs* : l'homme qui leur a transmis son nom, ainsi que sa femme. Pour cette raison, on peut prétendre que le rapport de la proportion de cousins germains isonymes (1/4) au coefficient de parenté de ces cousins (1/16) est de 1/4. Or, deux cousins en ligne paternelle, l'un issu du premier mariage d'un homme, l'autre du second, ont le même patronyme, mais leur coefficient de parenté est de 1/32 et non de 1/16. L'utilisation de la méthode isonymique à l'endroit d'une population où les remariages des veufs sont fréquents donnera donc une surestimation.

Dans le cas présent, parmi les 1327 mariages célébrés dans Lotbinière avant 1800, on en compte 156 qui unissent un *veuf* et une *célibataire*. Or, seulement 89 scellent une alliance

entre une *veuve* et un *célibataire*. Avant 1720, sur les 53 mariages célébrés, ces chiffres sont respectivement de 6 et de 3. Bien entendu, chez les hommes, les premières et les secondes noces ne sont pas toutes également fécondes (une telle répartition équivalente conduirait à un biais maximal). Mais plusieurs le sont. Nous avons donc là une autre source de biais qui, avec les deux autres, entraîne la mesure vers une valeur deux fois trop élevée.

CONCLUSION

Au terme de cette étude, deux constats s'imposent. Le premier frôle l'évidence: la méthode isonymique est relativement fiable si elle est appliquée à une population de grande taille, mais dès qu'on se penche sur de petites populations, elle peut devenir délicate. En passant du général au particulier, on voit apparaître ces écarts si souvent constatés dans les études portant sur de petites populations (entre autres : Crow et Mange, 1965; Vernay et al., 1998). À cet égard, certains auteurs nous proposent des analyses avec une assurance qui paraît démesurée en regard de l'échantillon utilisé. Avec environ 500 unions pour la région de Lotbinière entre 1780 et 1800, nous obtenons une valeur nettement trop élevée. Que dire de ces listes qui ne dépassent pas 100 unions!

Grâce à l'utilisation des numéros de lignée, nous avons pu constater que dans Lotbinière, l'écart ne provenait pas de l'origine polyphylétique des patronymes, mais de circonstances socio-démographiques particulières. Il aura suffi que trois sources de biais y additionnent leurs effets : (1) prévalence de liens de parentés en ligne masculine parmi les fondateurs, (2) concentration locale des patronymes et (3) remariage plus fréquent des veufs que des veuves. Encore fallait-il, et c'est là notre deuxième constat, que les biais en question soient introduits très tôt dans le peuplement pour qu'ils portent à conséquence. C'est un principe

sur lequel il nous faut insister une fois de plus : une population conserve la trace des conditions qui ont entouré ses origines.

L'étude que nous avons menée ici n'est évidemment pas reproductible dans la plupart des populations. Il faut en effet pouvoir disposer d'information sur les conditions initiales pour s'y atteler. Toutefois, elle peut profiter à d'autres en ce qu'elle attire l'attention sur certaines sources d'erreur dont on n'a pas souvent idée. Ici, nous nous en sommes tenus à une analyse partielle qui avait pour but de paver la voie à d'autres travaux. Aussi, dans un avenir proche, nous tenterons de développer des outils qui permettront de corriger quelques sources de biais (facteurs correcteurs pour le déséquilibre des sexes parmi les fondateurs, pour les remariages, etc.). Ainsi, l'expérience acquise dans ce contexte très particulier pourra-t-elle prétendre à un peu plus de généralité.

CHAPITRE 3 - FRAGMENTATION OF THE QUÉBEC POPULATION GENETIC POOL (CANADA): EVIDENCE FROM GENETIC CONTRIBUTION OF FOUNDERS PER REGION IN THE 17TH AND 18TH CENTURIES³³

ABSTRACT

The 6 million French-Canadians of Québec derive from a relatively small number of founders. Consequently, some hereditary diseases, which may or may not present a world-wide distribution, have been detected in high frequency in this population. Several studies, however, indicate a non-uniform distribution of these diseases through the population, suggesting that the French Canadian founder effect has been geographically stratified. Here we explore this stratification by using a demographic database, The Population register of Early Québec, that contains almost all birth, marriage and death certificates (>712,000) recorded in parish registers between 1608 and 1800. In this database, every genealogical link has been traced back to the founders of the population, so that we can compute the genetic contribution of founder per region, and then account for the early events that have shaped the distribution of diseases. Ten regions, comprising varying numbers of parishes, have been selected. We first describe each region in terms of homogeneity and concentration of its gene pool. For this purpose, a new concept is introduced, the Founders' Uniform contribution Number (FUN), i.e. the number of founders a population would have if all its founders had an equal contribution. Second, we estimate genetic similarity between regions on the basis of differential genetic contribution. To classify the regions, we use Principal Component and Cluster Analysis. Our results show a

³³ Par A. Gagnon et E. Heyer. Accepté avec correction mineure par l' *American Journal of Physical Anthropology*.

tripartite clustering of the population, and invite us to reconsider the results obtained from biomolecular and clinical studies, which show a bipartite clustering.

INTRODUCTION

Between 1608 and 1800 about 14,000 pioneers settled in the Saint-Lawrence Valley, once known as "Nouvelle-France". Among them, many had few or no descendants up to present, while very few had many descendants. For instance, Charbonneau et al. (1993) estimated that 2,600 pioneers, who settled before 1680, account for two-third of the modern French-Canadians' gene pool. Moreover, 575 of these pioneers contribute approximately to one-third of this genetic endowment (*ibid.*).

Not surprisingly, many have emphasized the genetic homogeneity of the French-Canadians, who represent approximately 80% of the of the modern Québec population. Biomolecular studies, however, have shown that French-Canadians are in fact remarkably well diversified. According to a study of De Braekeleer (1990), for example, the frequency of the 5 most common HLA-A alleles in the French-Canadians is approximately 72%, while it is 69% in France. Among the Hutterites, Dariusleut and Kel Kummer (highly isolated populations), the percentages are, respectively, 88% and 79% (*ibid.*). Other analyses based on haplotype frequencies and on several VNTR loci also show the relative heterogeneity of the French-Canadian gene pool (*ibid.*; Ohayon and Cambon-Thompsen, 1986; Monson et al., 1995). This can be explained, to some extent, by the fact that even if most immigrants were from France, they came from different regions of that country (Charbonneau et al. 1993). Also, the time elapsed since the foundation of the population is relatively small in comparison to the number of its founders; hence genetic drift may not have significantly reduced the genetic diversity (De Braekeleer, 1990).

There is nonetheless evidence of a founder effect in the Eastern regions of Québec, especially in Charlevoix and Saguenay-Lac-Saint-Jean, where some inherited disorders have a fairly high carrier frequency (up to 1/20). At least two of those disorders have only been found in this part of the world (Spastic ataxia and Sensorimotor Polyneuropathy), and others are very rare elsewhere (Tyrosinemia type 1, and Pseudovitamin D Deficient Rickets, etc) (Bouchard and De Braekeleer, 1991a; De Braekeleer, 1991; De Braekeleer and Dao, 1994a; Bouchard et al., 1995). On the other hand, there is no evidence of high frequency of any genetic disorders in the western part of Québec. Rather, here we encounter diseases common in populations of Europe or of European ancestry, such as cystic fibrosis, Friedreich ataxia or phenylketonuria (De Braekeleer, 1991; Bouchard et al. 1995). Hence, according to these studies, the province seems to be separated into two distinct geographical entities in terms of gene frequency distributions, with a frontier located upstream from Québec City.

These features suggest a geographical stratification of the “French Canadian founder effect” (Bouchard et al., 1995). With the use of an exhaustive demographic and genealogical database, The population register of Early Québec (1608-1800), we can account for early historic and demographic events that have shaped the distribution of diseases over the population. In a recent paper, we have explored the evolution of kinship between mates per region in the 17th and 18th centuries (Gagnon et al, 1998). The coefficients of kinship between mates were shown to be much higher in the Eastern part than in the Western one, which explains to some extent the cleavage between Eastern and Western Québec’s distribution of diseases. A study of a regionally-based sample of genealogies of 20th century individuals (listed in the RETRO database of the Interuniversity Institute for Population Research) has also shown this East-West gradient of kinship coefficients (Tremblay et al., 1998). Although in agreement with the general thesis of these studies, we hope to bring new insights by offering this study

based on genetic contribution. More precisely, the focus of our study lies on genetic contribution of founders, per region, to a cohort made of spouses married between 1780 and 1799.

First, we describe the form of the distribution of this contribution for the entire colony and for each region. For this purpose, we introduce a new concept, the Founders' Uniform Contribution Number (FUN), to account for the homogeneity of a genetic pool. Simply stated, it is the associated number of founders of equal contribution required to yield the same homogeneity to a population genetic pool as the actual number. By comparing this number to the actual number of founders, we have an indication of the concentration of the founders' genetic contributions. Second, we classify the regions on the basis of differential genetic contribution of founders per region. This is done by principal component and cluster analyses. The aim is to see if there is a spatial fragmentation in terms of ancestral composition, and to see if there is any significant clustering. Finally, we explore the relationship between the classification of the regions and the homogeneity of founders' contribution to each of those regions.

3.1 DATA AND METHOD

3.1.1 Historical background and geographical location

The French colonization of Canada began in 1608 with the foundation of Québec City and was dramatically interrupted by the English Conquest in 1759. Of the >30,000 colonists (mostly) from France (Boleda, 1984), only an estimated 8,000-10,000 established themselves permanently in the colony (Charbonneau et al. 1993). The strongest wave of immigration spans between 1663 and 1680, when Louis XIV himself, worrying about the low number of entries into the colony, took its destiny in hand. After 1680, the population had to rely largely on its own natural increase, which was supported by a high fertility rate (Charbonneau, 1987; Nonaka et al. 1990). By the end of the 18th century, the population had reached approximately 200,000 individuals.

As shown in Figure 3.1 (voir Carte 1; Introduction), in the course of the 17th and 18th centuries, Canadian settlements were still restricted to the banks of the Saint-Lawrence's and to its principal tributary shores. Indeed, to make the map easier to read, we have intentionally extended the surface of those settlements. The peopling of the Saguenay-Lac-Saint-Jean, a well-known region from an epidemio-genetic point of view, began as late as 1840 (Bouchard et De Braekeleer, 1991a), which precludes its inclusion in our study. However, the regions where most of its immigrants came from, the Charlevoix and (to a lesser extent) Beaupré counties, are included. On the map, they are grouped into one region, labeled number '9'. More specifically, we have divided the colony into 10 regions, made up of different numbers of parishes and individuals, according to such criteria as geographical proximity, sharing a common settlement history, etc. Those 10 regions are located in the three colonial authorities' "governments" (Montreal, Trois-Rivières, and Québec).

3.1.2 Database

The Population Register of Early Québec contains almost every birth, marriage and death certificate (>700,000) recorded in parish registers between 1608 and 1800. Originally designed for automatic family reconstitution by the Research Program in Historical Demography (PRDH, University of Montreal), this database encompasses the vital events, as well as the kinship network of every individual of European ancestry born in the Saint-Lawrence Valley between 1608 and 1766. As such, it reconstitutes exhaustively the birth of a population. But the database also includes a wealth of information on genealogy for a longer period. The ancestors of every individual who married in the colony before 1800 have been traced back to the founders (for more information on the database, see : www.genealogy.umontreal.ca).

More specifically, we focus here on individuals included in a marriage cohort between 1780 to 1799, which is the last period for which data are available for the entire colony. The

reason we chose to look at individuals who married (instead of considering the whole population at a given date) is twofold: first, since we want to account for the contemporary distribution of inherited disorders, it is more appropriate to rely on people who married (because they have an overall better chance of transmitting their genes). Second, the total reconstitution of the population (i.e. total linkage of all birth, death and marriage records for each individual and family) does not extend beyond 1768. Thus, using the 1780-99 marriage cohort allows us to analyze the population at a more recent time. To avoid double counts, we considered only the first marriage of individuals who married twice or more. In total, 44,510 men and women, linked to 9,659 founders, are included in our analysis. The exact place of wedding is known for approximately 95% of the couples. When this information was not available, we replaced it by the parish of birth of the first child. So in total, we ignore only 1.5% of the wedding places (see Table 3.1 for the resulting numbers of individuals and founders involved for each region and for the whole colony). We believe that such an observation on the birth of a new population, including its early geographical settlement, is unparalleled elsewhere. This gives us the opportunity to compute the genetic contribution of founder per region with a high level of confidence.

Table 3.1 - Number of spouses and founders involved for each region and for the whole colony (1780-99 marriage cohort).

Region	Number of spouses	Number of founders
1) North of Montreal	7654	4161
2) Montreal Island	5976	4471
3) South of Montreal	10474	4809
4) Gvt of Trois-Rivières	4348	3104
5) Portneuf	1332	1407
6) Lotbinière	870	1073
7) Quebec city area	2976	3004
8) Orléans Island	776	958
9) Beaupré and Charlevoix	1914	1303
10) Bas-St-Laurent and Beauce	8190	2782
Total (whole colony)	44510	9659 ³⁴

3.1.3 The measure of founders' Genetic Contribution (GC)

If one chooses a gene at random out of the two genes carried by an individual at a locus, the probability that it came from his father is $\frac{1}{2}$, or from his mother is $\frac{1}{2}$; the probability that it came from a given grandparent is $\frac{1}{4}$, provided that his parents are not first degree relatives, and so on. By tracing back the genealogy of an individual, one finally reaches his most distant known ancestors, i.e. the individuals whose parents are unknown. Those "last known ancestors" are defined as the founders of the population. They may genuinely be the first immigrants to settle in the population, or they may have arrived later on. Adding all the probabilities for each of the genealogical paths between a founder f and each individual of a descendant cohort d gives the expected number of copies of a particular gene carried by f in the descendant cohort d , that is the genetic contribution (GC) of f to d .

³⁴ Note that the sum of this column exceeds the total number of founders of the colony ($N_f = 9659$) because many founders have had descendants in many regions

In other terms:

$$GC = \sum_{i=1}^p \sum_{j=1}^c \left(\frac{1}{2}\right)^{g_c}$$

where p is the number of individuals in a given group genealogically related to the founder; c the number of genealogical paths between the founder and the individual; and g_c the number of generations separating the founder from the individual, for each path. Adding the genetic contribution of a specific founder to a given group of individuals and dividing the result by the number of individuals in the group, we obtain the expected proportion of the gene pool attributable to this specific founder. Then, a population can be described by its 'origin vector', whose GC_i elements represent either the probability that, at a given locus, a gene chosen at random in the population comes from the founder i , or the expected proportion of the entire genome that come from that i^{th} founder. Note that because of random genetic sampling, there is a variance around this expectancy (see Roberts, 1968; Jacquard, 1974, Cazes, 1986; O'brien and al., 1988; Heyer, 1995; Heyer et al. 1999).

3.1.4 The Founders' Uniform Contribution Number (*FUN*)

The effective number of alleles (n_e) maintained in a population has been defined by Crow and Kimura (1964; 1970: 323) as the reciprocal of the sum of the squares of allele frequencies:

$$n_e = \frac{1}{\sum p_i^2} = \frac{1}{F_r}$$

where, p_i is the frequency of the i^{th} allele. Indeed, $\sum p_i^2$ is the probability of homozygosity under the assumption of random union of gametes (F_r). The idea is to find the number of alleles of *equal* frequency that would result in the same amount of heterozygosity as the actual number does. In fact, many alleles will be represented only once or twice in the population and contribute very little to the average heterozygosity or genetic variance. For example, with three alleles ($n = 3$) whose frequencies are 2/3, 1/6 and 1/6, the mean allele frequency is, of course, $1/n = 1/3$.

But, $\sum p_i^2 = 1/2$. Thus, the actual number of alleles is three, but the amount of homozygosity is the same as would be found in a population with two alleles of *equal* frequency (with two alleles whose frequencies are 1/2 and 1/2, we have $\sum p_i^2 = 1/2$, as expected).

In a similar manner, we can define the *Founders' Uniform contribution Number* (FUN) as the number of founders of *equal* contribution that would yield the same probability that two genes, drawn at random in the cohort, come from the same founder (note that we don't refer to probability of identity by descent at a given specific locus). The rationale behind this is that the ratio of genetic contribution of a founder to the total number of individuals in the cohort is the probability that a randomly chosen gene comes from this specific founder. As the drawing of a second gene is an independent event of the first draw, the square of the probability that one gene comes from one specific founder is the probability that two genes come from this founder. Hence, for a given region X, $FUN_X = 1/\sum x_i^2$, where x_i is the GC of the i^{th} founder. FUN alone is an homogeneity index, and accounts for the '*genetic homogeneity*' of a population by assessing the number of founders a population would have if all the founders had contributed equally. The ratio of this number to the actual number of founders (FUN/N_f), gives the relative '*concentration*' or '*asymmetry*' of founders' contribution. The lower this ratio is, the higher is the concentration, and the stronger is the influence of the socio-demographic differentials (as fertility and migration) in increasing the homogeneity of the population. For example, a highly variable family size from one generation to the next will result in a small FUN/N_f ratio.

3.1.5 The measure of genetic similarities between regions and their classification

Over the years, population geneticists have suggested a number of genetic distance measures that are specifically designed to be proportional to the time since divergence between separate populations, whether they are a function of gene frequencies or are based on codon substitutions per locus (see Nei, 1987, chap. 9). As Relethford (1996: 30) pointed out, in many

studies, “the similarity between regions is assumed to be a reflection of branching history; the more genetically similar two regions are, the more recently they have had a common ancestor”. Therefore, those measures are not appropriate for genealogical data of a population with many exchanges of migrants between its subdivisions, such as the population we are concerned with here. The measures that we use are of another kind, because they attempt to classify regions in terms of the differential contributions of founders to those regions.

The results of a classification depend on the measure of similarity and on the method used. Some give more weight to founders whose overall contribution is low; others give the advantage to high contributors. In the first instance, we have decided to filter out, as much as possible, the influence of low contributors. Today's distribution of hereditary diseases is very unlikely a function of the low contributors to the 1780-1799 marriage cohort. Deleterious mutations that have been introduced by those founders may exist. But their frequency must be very low and therefore they are probably undetectable. Hence, each one of the presently known mutations were introduced by one or two high contributors (Austerlitz and Heyer, 1998; Heyer et al., 1997).

Principal Component Analysis (PCA) is particularly well designed to filter out the effect of low contributors on similarity measures. This method allows a projection on a low dimensional space (for example two or three dimensions) of the information originally contained in a high dimension space. Here, we wish to reduce a 9,659 lines X 10 columns matrix \mathbf{G} (9659 founders X 10 region), in which each element g_{ij} is the contribution of founder i to region j . For this aim, each element g_{ij} is first standardized by the following transformation: $g_{ij} \Rightarrow z_{ij} = (g_{ij} - \mu_j) / N_f \sigma_j$, where μ_j and σ_j are the mean and the standard deviation of founders' contribution to region j . Note that since the contributions of founders to each region add up to 1, $\mu_j = 1/N_f$, $\forall j$. After the transformation, the regions are located on a 9,659 dimensions (= N_f) hyper-sphere of unity

radius, and the cosine of the angle between any two regions is their correlation. Then, the coordinates of the regions on the projection axes are given by the product of \mathbf{Z}' by the eigenvectors associated to the highest eigenvalues of \mathbf{Z} . Since the projection is made by maximizing the variance of the dispersion of points (here the regions, set as variables) on the firsts axes of PCA, the position of regions on those axes will not reflect the influence of low contributors (because of their overall weak GC, the variance in their GC per region is also weak). This method enables us to see if there are subgroups of regions which lie close to one another according to the most variable founders' GC.

To see if the position of regions on the projection planes is not spurious, we use cluster analysis. Unlike PCA, this method allows a classification that is not based on the implicit filtering out of low contributors. This is simply done by including all founders in the analysis ($N_f = 9,659$). But it can also be used to classify the regions according to the most important founders (or the highest contributors). For this purpose, we also classify the regions according to the founders whose GC to the entire colony rank from 1 to 300 (their GCs add up to about 28% of the total genetic contribution). This way, the filtering of contributors is directly made by the level of total GC, rather than by its variance (in the PCA, the filtering is done according to the variance). Thus, we have two classifications (one using all the contributors, the other using only the 300 highest contributors) that can be compared. The question is: When we consider only the highest contributors, do we obtain a clearer separation between the eventual sub-groups than when we use all the founders? Or does classification according to the highest contributors display a less clear separation between the clusters?

The choice of the 300 "firsts founders" is, to some extent, an arbitrary choice. We could use the first 200, or the first 400, etc. We need a number of this order of magnitude because in the largest study so far of the geographical distribution of hereditary disorders in the Québec

population (including 28 different diseases), 249 founders have been identified as ancestors of all cases (De Braekeleer and Dao, 1994a). All of these presumed “disease introducers” settled in the Saint-Lawrence Valley in the course of the 17th century, so they had a high overall contribution to Québec’s population. By using a number of highest contributors that is close to 249, we believe that we can accurately account for the present distribution of diseases throughout the population, thus realizing one of our primary aims.

Different measures of similarity can be used in the cluster analysis. Here, we first use the *covariation around each founder’s total’ GC*, i.e.:

$$C_{XY} = \sum(x_i - t_i)(y_i - t_i)$$

where x_i and y_i are the GC (in percentage) of the i^{th} founder to regions X and Y, and t_i is the GC (in percentage) to the whole colony of this i^{th} founder (or his average weighted GC over all regions). The more covariation between two regions with regard to each founder’s total GC, the more similar those two regions are, and vice-versa. This crude measure of covariation does not have the simple predefined range of variation of a correlation coefficient, which varies from -1 to 1 . It is not a measure of similarity *per se*. In order to have such a range of variation, we could divide C_{XY} by the square root of $\sum(x_i - t_i)^2 \sum(y_i - t_i)^2$, which would standardize the measure. However, we believe that it would not be appropriate. Assume for example that C_{XY} and $\sum(x_i - t_i)^2 \sum(y_i - t_i)^2$ are both very high. In other words, the two populations being compared present a *high covariation* (high numerator) and a *high variation* (high denominator) about each founder’s total GC. This results in a relatively low standardized measure of similarity. But the fact that both regions deviate from the whole population (high denominator) is indeed a sign of their high affinity, if, of course, their covariation is positive. Hence, the use of a standardizing factor could make them appear more distant from each other than two other regions that have a small covariation but that are close to the overall population origin vector.

The second measure of similarity used in the cluster analysis is based on the FUN concept developed above. Its purpose, as it will be shown below, is to classify the regions on the basis of the likelihood that they share, in a high frequency, a common inherited disorder derived from a common founder. We define it as the probability that two genes, one chosen in population X, the other in population Y, come from the same founder. This measure is somewhat related to the probability of identity of genes by descent in two populations, except that it is not based on allele frequencies but on founders' GCs. Following Nei's notation (1972, 1987), it can be written as

$$J_{XY} = \sum x_i y_i$$

where, again, x_i and y_i are the GC of founder i to region X and Y. Considering the definition of the FUN, it may also be written as $J_{XY} = 1 / \text{FUN}_{XY}$. Hence, FUN_{XY} is the associated number of equally contributing founders that would give the same probability that two randomly chosen genes, one in each region, come from the same founder. The lower the figure FUN_{XY} , the higher is that probability. Suppose that we have two regions X and Y whose "origin vectors" are, respectively, $X = (1/6, 1/6, 2/3)$ and $Y = (2/3, 1/6, 1/6)$. Thus, $J_{XY} = (1/6)(2/3) + (1/6)(1/6) + (2/3)(1/6) = 1/4$ and $\text{FUN}_{XY} = 4$. In other words, in terms of the probability of drawing two genes that come from the same founder, four equally contributing founders in each region would have been equivalent to the three observed founders.

Again, J is not a measure of similarity *per se*, since $J_{XX} \neq 1$, unless $N_f = 1$. Indeed, an appropriate and desirable coefficient of similarity should always be equal to 1 when a population is compared with itself. In order to have such a coefficient, it would be necessary to divide J_{XY} by $(\sum x_i^2 \sum y_i^2)^{1/2}$, which would give the cosine of the angle between the origin vectors of regions X and Y (a measure similar to Nei's "normalized identity of genes" or I_{XY} ; *ibid.*). But here, recall that our goal is to classify the regions according to the probability that they share, in a high frequency, a common inherited disorder. A simple example should help to summarize our

intention. Suppose that *five* founders have contributed equally to regions X and Y. In this case, $J_{XY} = 5(1/5)(1/5) = 1/5$. Now, suppose that *ten* founders have contributed equally to populations U and V. In this other case, $J_{UV} = 10(1/10)(1/10) = 1/10$. In both cases, the two populations are exactly the same: X has exactly the same origin vector as Y, and U has the same origin vector as V. But J_{XY} is higher than J_{UV} . Thus, a dendrogram using the J measure would display shorter branch lengths between X and Y than between U and V. This is precisely what we want to display. Since they derive from a small number of equally contributing founders, it is likely that X and Y share, in a high proportion, some common inherited disorders. U and V may also share inherited disorders, but each of those disorders is likely to be found in a lower frequency in both regions U and V.

To combine clusters, the UPGMA (“unweighted pair-group method using arithmetic averages”) and the “average linkage within group method” have been used. The first defines the similarity between two clusters as the average of the similarities between all pairs of cases and combines clusters by maximizing the distance *between* clusters. The second, which is a variant of the first, combines clusters so that the similarity between all cases in each resulting cluster is as high as possible.

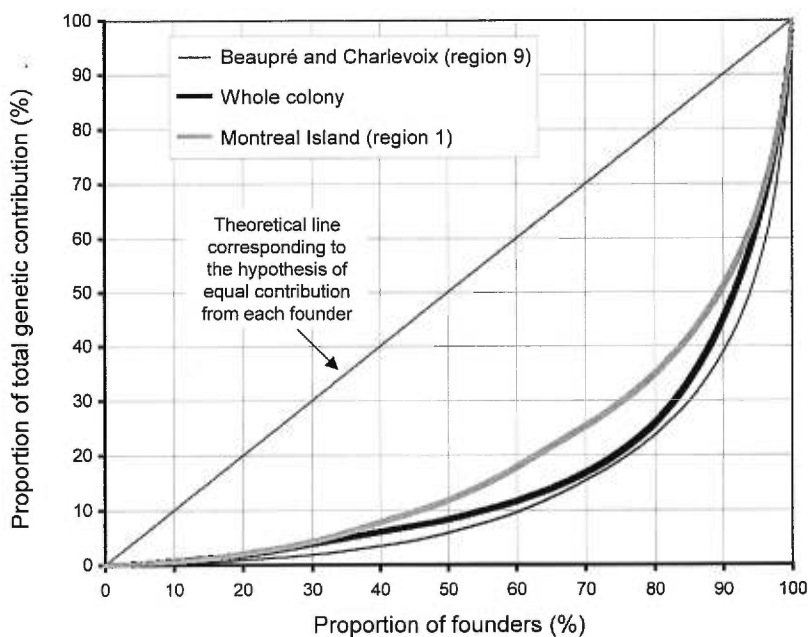
3.2 RESULTS

3.2.1 Distribution of founders’ contribution and the FUN

Figure 3.2 clearly shows that a large proportion of the total genetic contribution to the whole colony comes from a small proportion of the 9659 founders. For instance, when looking at the whole colony curve, we see that approximately 92% of the founders only account for 50% of the total genetic contribution, meaning that the remaining 8% of the founders account for the remaining 50% of the genetic contribution. For some regions, the inequality of the distribution is even higher. This is the case of Region 9 (Beaupré and Charlevoix), where about 95% of

founders explain 50% of the gene pool (the remaining 5% account for the remaining 50%). For other regions, like region 1, the inequality is lower. As a first interpretation, one may presume that the more a curve deviates from the theoretical straight line of equal contribution from each founder, the more the region in concern is “homogeneous”.

Figure 3.2 - Distribution of founders’ contribution to the 1780-99 marriage cohort.



But assume, for instance, that only 2 individuals, a founding couple, founded a population many generations ago. By definition, their contribution would be perfectly equal, i.e. both of them would contribute half of the gene pool. Hence, the resulting curve would be a straight line. However, as one can easily figure, this population would not be heterogeneous, but, to the contrary, quite homogeneous. In fact, a graph such as the one depicted in Figure 2 does not represent the *homogeneity* of a population’s gene pool. It only gives us an indication of the relative *concentration* of founders’ genetic contribution. Here, one may conclude that the three gene pools being compared do not differ significantly in this respect, since the three curves do not

deviate very much from one another. We could also calculate, for any region, the area between the “theoretical straight line” and the observed curve in order to obtain a numerical evaluation of the concentration. But the genetic meaning of such an evaluation would not be very clear. Hence the advantage of the FUN concept.

Table 3.2 shows, for each region and for the whole colony, sums of squares of founders' contribution ($\sum p_i^2$), FUNs' corresponding values ($1/\sum p_i^2$), relative index of homogeneity, actual number of founders (N_f) and FUN/ N_f ratios. This table is, in our opinion, a better way to emphasize the relative homogeneity of the gene pool and the concentration of its founders' contribution. For example, the actual number of contributors to Region 1 is 4161. But about 1256 equally contributing founders would have given the same “genetic homogeneity” (in terms of the probability of drawing two genes that come from the same founder) in the descendant marriage cohort. In Region 9, which has been set here as a standard for comparison, the homogeneity is much higher since its FUN is approximately equal to 185. In fact, Region 9 is 6.79 times more homogeneous than Region 1 (= 1256/185). In general, the eastern regions are more homogeneous than the western ones, as might be expected from the results of the latest bio-molecular studies conducted in the province.

Table 3.2 - Founders' uniform contribution number (FUN) and its ratio to actual number of founders per region (FUN / N_r) in the 1780-99 marriage cohort.

Region	$\sum p_i^2$	FUN (=1/ $\sum p_i^2$)	Relative (to region 9) heterogeneity	Actual number of Founders(N _r)	FUN / N _r
1) North of Montreal	7,96 ^E -04	1256,4	6,79	4161	30,2%
2) Montreal Island	8,97 ^E -04	1115,0	6,03	4471	24,9%
3) South of Montreal	7,11 ^E -04	1406,3	7,60	4809	29,2%
4) Government of Trois-Rivières	1,20 ^E -03	831,8	4,50	3104	26,8%
5) Portneuf	3,04 ^E -03	329,2	1,78	1407	23,4%
6) Lotbinière	4,27 ^E -03	234,2	1,27	1073	21,8%
7) Quebec city area	1,10 ^E -03	912,9	4,94	3004	30,4%
8) Orléans Island	3,60 ^E -03	277,4	1,50	958	29,0%
9) Beaupré and Charlevoix	5,41 ^E -03	184,9	1,00	1303	14,2%
10) Bas-St-Laurent and Beauce	1,78 ^E -03	560,8	3,03	2782	20,2%

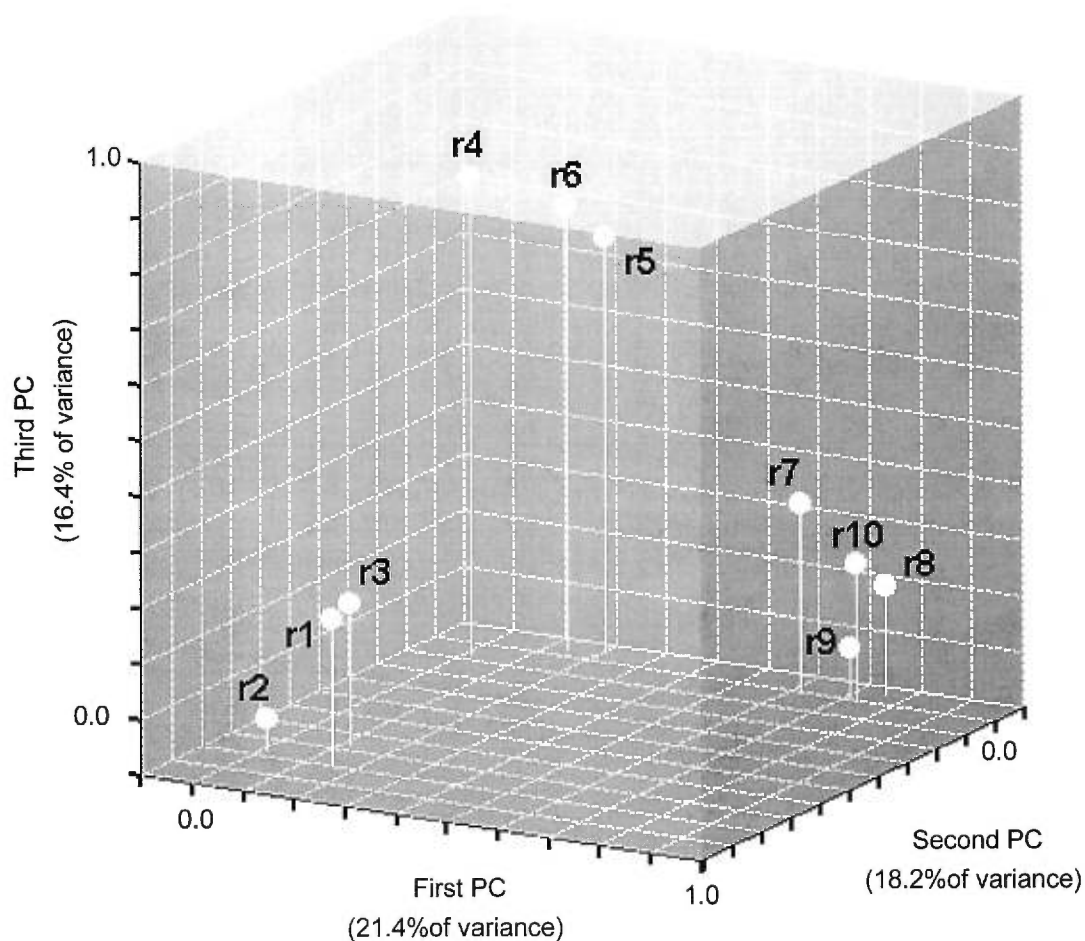
We may consider FUN/N_r ratios as indexes of concentration. They give indications of the 'potential of concentration' in each region, independently of the number of founders. As an example, only about 14 equally contributing founders in Beaupré and Charlevoix (Region 9) are equivalent to 100 actual founders. In other regions demographic processes had a much lower effect. Although Orléans Island (8) is quite homogeneous in terms of the probability that two randomly chosen genes come from the same founder (FUN \cong 277, the second lowest), its FUN/N_r ratio is quite high (= 29%, among the highest). If this region had been characterized by the same concentration index as region 9 (i.e. a ratio of 14%), it would have been much more homogeneous. Its FUN would have been 136 instead of 277 (see discussion).

3.2.2 The classification of regions

The PCA performed on the 1780-99 marriage cohort shows interesting results (Figure 3.3). We see that none of the regions is negatively correlated with any other. In other words, in

comparison with mean values, high values of contribution for some regions are not systematically paralleled with small values for other regions.

Figure 3.3 - Plot of the first three principal components (PC) of founders' contribution per region (1780-99 marriage cohort). The three PC account for 56.0% of the variation. Axes have been rotated by varimax method.

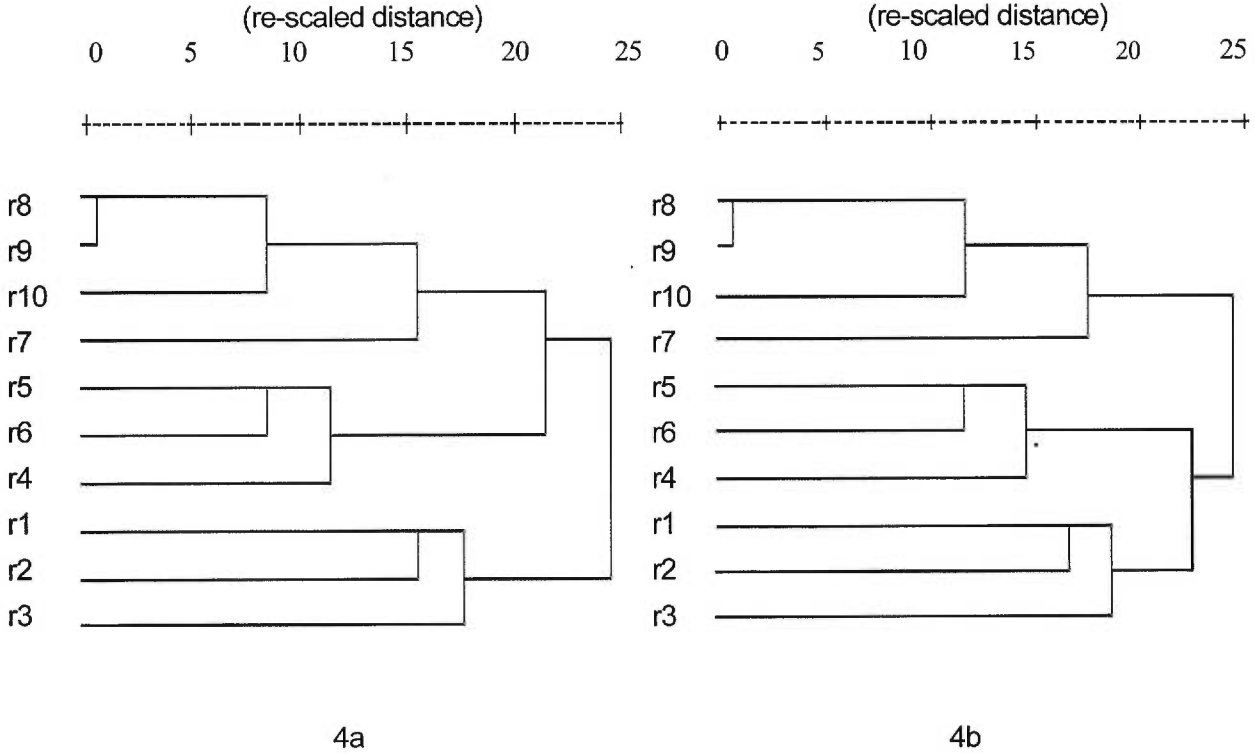


Nevertheless, there are three subgroups of regions (see Figure 3.1 for geographic location). Each of those three clusters is located on high positive values of its specific principal component so that the three subsets are mutually de-correlated (they form right angles). First, the 'Western

cluster', contains the three regions of the Government of Montreal; they are numbered 1,2 and 3. Second, regions 4, 5 and 6 group together and form the 'Central cluster'. Finally, the third subgroup is made of regions 7,8,9 and 10, the 'Eastern cluster'. Since the three subgroups are not negatively correlated, it can be said, however, that the whole population is not extensively stratified into a three-fold structure. As shown below, this can be explained by the fact that the highest contributors have had descendants all over the colony.

In order to take into account the impact of all founders (low contributors included), we use a cluster analysis. Figure 3.4a shows that the hierarchical classification of regions according to the covariation around the total GC of each of the 9659 founders (using 'UPGMA') displays the same three clusters.

Figure 3.4 - Cluster analysis of regions according to GC of founders to the 1780-99 marriage cohort. Measure of similarity: standardized covariation about total GC. 3a) $N_f = 9659$ 3b) $N_f = 300$. Grouping made by UPGMA.



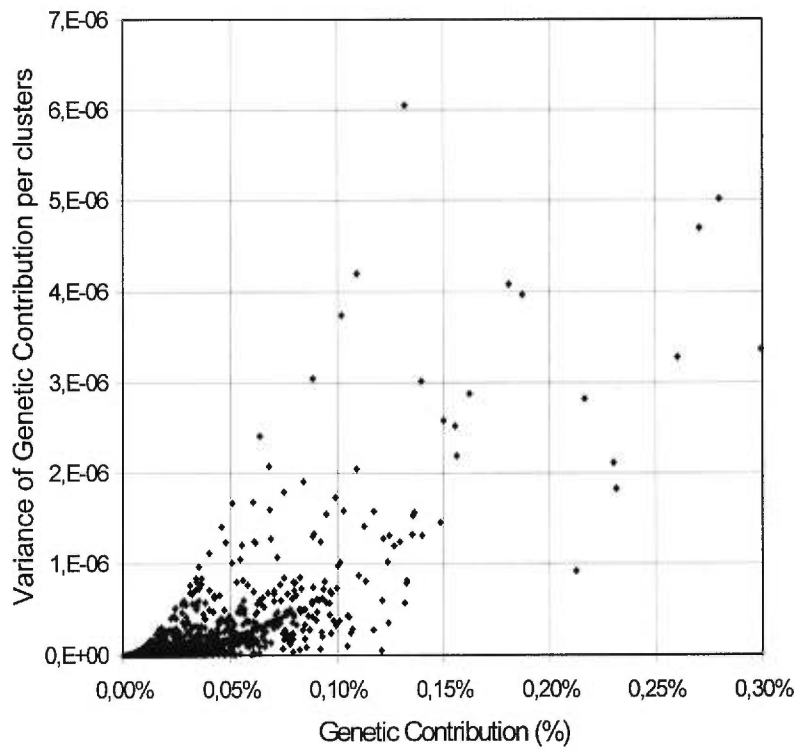
Note that the link between Québec city (region 7) and the three other regions of the Eastern cluster is somewhat tenuous; in fact, this city is also “attracted” by regions 5 and 6 (Portneuf and Lotbinière), which are geographically close to it (see Fig. 3.1). The PCA also shows this feature (see the distance between region 7 and the Central cluster in Fig. 3.3). This could explain why there is a second level of clustering that groups together the Central and the Eastern areas. But this second order link is not very strong since it has not always been displayed in other analyses, using other methods of clustering and similarity measures (see Fig. 3.4b for example).

In figure 3.4b, regions are clustered according to the founders whose total GC ranks from 1 to 300 (all the others have been excluded from this analysis), again using the C_{XY} measure. The aim is to account for the structuring impact of the highest contributors. Looking only at these 300 ‘first founders’, the Center aggregates, at a second level, with the West, rather than with the East. But the most important feature we observe here is the overall slight decrease of the “inside integrity” of each cluster. The three clusters seem to be more ‘attracted’ to each other than they were in the previous outcome (Fig. 3.4a), so that now most branch lengths are longer (the “within group method” of clustering displays approximately the same results; not shown here). This tendency (lengthening of branches) is even more marked in analyses based on the ‘first 200’ founders (not shown here). This implies that, compared to all founders, the highest contributors have spread their genes in a less structured way (one that induces a three-cluster arrangement). This seems to contradict the results given by PCA, which shows a clear distinction between the clusters. However PCA is based on correlation, not on covariation. Moreover, PCA projects variables according to highest variances, not highest values.

In Figure 3.5, we plot the variance of founders’ contribution *to the three clusters* as a function of their overall genetic contribution. In general, the higher the GC of a founder, the

higher is its variance of contribution per region. But the link between those variables is not perfect ($r=75\%$).

Figure 3.5 - Variance and overall genetic contribution to the three clusters of the 9659 founders.

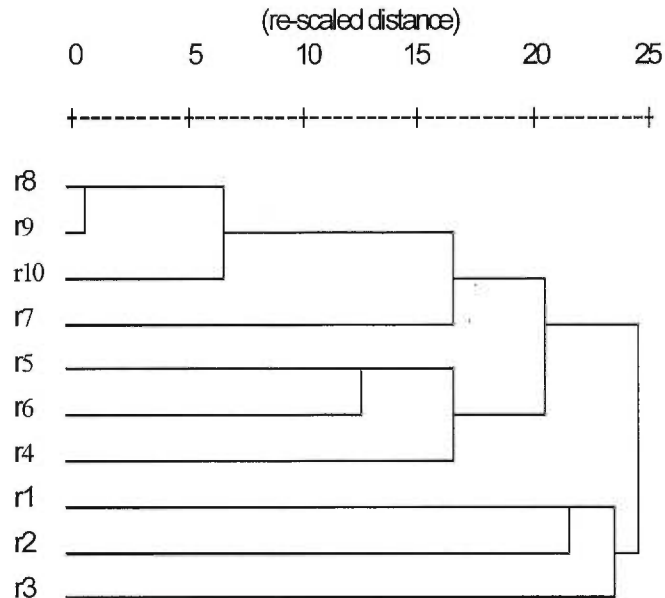


There are many founders who may be labeled “middle range” founders who have a higher variance of contribution than many of the highest contributors. Be as it may, the three same clusters are still displayed even if we consider only the “truly” highest contributors. For instance, an analysis of the first 100, or even the first 50 contributors display approximately the same results as an analysis based on the first 300 (not shown). On the other hand, analyses that include the first 500, 1000 or 5000 first founders display dendrograms that are in close agreement with the one depicted in figure 3.4a.

3.2.3 The classification of regions as a function of their homogeneity

Figure 3.6 shows the classification of regions according to the J_{XY} measure. In this respect, it can be inferred that the most (and maybe the only) significant clustering is the one that involves the Eastern regions, especially regions 8, 9 and 10, which are located to the east of Québec city (region 7), downstream along the Saint-Lawrence river. The Western cluster has much longer branch lengths. Hence, in comparison to the Eastern feature, the probability of choosing two genes that come from the same founder in any two regions of the Western cluster is much smaller.

Figure 3.6 - Cluster analysis of Regions according to GC of founders to the 1780-99 marriage cohort. Measure of similarity: probability of choosing two genes that come from the same founder. $N_f = 9659$. Grouping made by UPGMA.



As stated above, J_{XY} depends on the number of founders that simultaneously account for the regions being compared. More precisely, it depends solely on the FUN_{XY} of the two regions ($J_{XY} = 1/FUN_{XY}$). For instance, since the regions of the Western cluster have the highest numbers in this respect, it is expected that they also show the lowest probability of having genes from the *same* founder. To the contrary, this probability is the highest in the eastern cluster, where the numbers of equally contributing founders are the lowest.

Table 3.3 supplements this view. It clearly shows that going from the West to the East, the gene pools become more homogeneous (2nd column) and also more concentrated (4th column). Note that this feature also parallels the results of Table 3.2. There are, however, some exceptions, particularly in the Central cluster, in which region 6 (Lotbinière) is among the most homogeneous and concentrated regions.

Table 3.3 - Founders' uniform contribution number (FUN) and its ratio to actual number of founders per clusters (FUN/ N_f) in the 1780-99 marriage cohort.

Cluster	FUN	Actual number of founders (N_f)	FUN/ N_f ratio
West (regions 1, 2 and 3)	1 924.0	7 154	26.9%
Center (regions 4, 5 and 6)	914.8	3 446	26.5%
East (regions 7, 8, 9 and 10)	770.9	4 345	17.7%

DISCUSSION

FUN and underlying socio-demographic processes

Initial conditions of the settlement have played a major role in shaping today's distribution of inherited disorders in Québec. At the end of the 18th century, the most homogeneous genetic pools (in term of FUN) are seen in the eastern part of the province. Also, it is worth mentioning that the FUNs calculated here for every region are almost perfectly correlated ($r = 0.95$) to coefficients of kinship between mates calculated in a previous paper (Gagnon et al., 1998). Both of these measures, indeed, account for the homogeneity of a population's gene pool.

The ratio of Founders' uniform contribution number (FUN) to the actual number of founders (N_f) has been used to describe the concentration of the distribution of founders' contribution. But it also helps to summarize the impact of underlying historic and demographic processes. For instance, we have stated that if region 8 (Orléans Island) had had the same ratio as region 9 (Beaupré and Charlevoix), it would have been much more homogeneous (see table 3.2). The reduction of heterogeneity could have been achieved in two ways that are not mutually exclusive and may indeed be conceptually related. First, the years of entry of the 958 inventoried founders could have been much more delayed. As a matter of fact, history tells us that the Island was populated in one large move, in the early years of the Nouvelle-France colony (Landry,

1992). Few immigrants came later on as it was already fully occupied (Charbonneau, *personal communication*). So, if they had not come in a restricted period of time as they did, their genetic contribution would have been more asymmetric, and the first immigrants would have left a higher number of copies of their genes to future generations. In that case FUN/N_f ratio of this region would have been lower, and its genetic homogeneity higher. Second, a higher intergenerational fertility differential could have given the same result. This could have taken the form of a more variable effective family size (EFS) and/or a higher intergenerational correlation of EFS (see Austerlitz and Heyer, 1998). Of course, both of the scenarios hypothesized here may be simultaneously considered. Thus, FUN/N_f can be seen as a summary of the genetic impact of demographic differentials, merging all their constituents into one synthetic index. Speaking of these demographic differentials, we may note that, at a general level, they have played a greater role in the Eastern area (see Table 3.3).

A three cluster grouping

In the second part of this paper, our results show a tripartite classification of Québec's population in terms of the differential of founders' contribution per region. This is of interest because until now, biomolecular studies of Québec's population have stated that the province was divided into two distinct genetic entities, instead of three: The West and the East (see introduction). Here, we clearly see that a third cluster (the central one) slips between them. This can be easily understood. Like Montreal and Québec City, Trois-Rivières (the chief-town of the central area) was also a port of entry to Nouvelle-France. Moreover, the three-cluster arrangement does not exactly parallel the administrative divisions of the colony. Portneuf and Lotbinière (regions 5 and 6) are located in the Québec Government. But those two regions are more akin to Trois-Rivières' Government.

It must be said that to some extent, classification depends on the choice of the method and the similarity measure used. For instance, the correlation, which has been used in the PCA, is a measure of the 'association' between two genomes. It is high if knowledge of one of the values taken by one variable (here the GC of a given founder to a given region) helps to predict with a high level of confidence the value taken by the other variable (the GC of the same founder to the other region). But we can figure populations that share approximately the same proportion of GC from numerous founders without being highly correlated (and vice versa). In a correlation, the covariation of regions about their mean is divided by each region's standard deviation. Hence, when those standard deviations are low, the regions will appear highly correlated if their covariation is positive, and highly negatively correlated if their covariation is negative. To confirm our results, we have used another method of classification, and other measures of similarity. For instance, the covariation about each founder GC (C_{XY}) has also displayed a three-cluster grouping, in good agreement with the PCA.

By considering only the highest contributions, however, we have seen a general lengthening of branches, i.e. a less clear separation of the three clusters. This can be explained, to some extent, by the history of the Nouvelle-France settlement. The three governments were not simultaneously created at the beginning of the colonization. Québec City was founded first, in 1608, several decades before the advent of Trois-Rivières (1634) and Montréal (1642). But many descendants of Québec City's first settlers, along with new pioneers from France, went up the St-Lawrence river to settle in these areas, as soon as they were open. As a result, the most ancient founders, who had on average the highest contributions, had descendants in the first three settlements, and consequently, everywhere in the colony. This may explain why, according to their contribution, the three-fold structuring is less clear. Now, as the colonization process has gone further on, the two new settlements naturally became administrative centers like Québec city. New pioneers then started to settle in their respective areas, providing more "integrity" to

these areas. Since they arrived later on, they had an overall lesser contribution than the previous ones, so that they may be labeled as 'middle range contributors', but their impact on the separation of the three clusters is high.

Finally, an analysis based on the probability that two randomly chosen genes come from the same founder (J_{XY}) has also resulted in a three-clusters arrangement. However, according to this measure, the Eastern cluster shows much smaller branch lengths than the two others. Thus, the probability that the eastern regions (especially regions 8, 9 and 10) share, in a high frequency, a common inherited disorder is higher. This probability is lower in the Central area, and lowest the Western area. These features are paralleled by the results given in Table 3.3 and clearly account for the known distribution of diseases over the population.

Past and present classification

Why is there a difference between our classification and the one revealed by studies of present distributions of hereditary disorders and gene frequencies? There are several possible explanations for this discrepancy. To begin with, clinical and biomolecular studies were probably insufficiently accurate for the sake of a systematic classification. As a matter of fact, they were not specifically designed toward this end. Most of them were developed for the clinical study of particular diseases. Since many rare diseases have been located downstream from Québec City (East), and very few elsewhere in the Province, it first appeared to researchers that Québec's genome was made-up of two contrasting features (see De Braekeleer and Dao, 1994a). In a certain sense, they were right, and our results parallel their findings since, according to the J_{XY} measure, there is a close clustering in the East, and no evidence for 'high integrity clusters' elsewhere. But as far as the total ancestral composition of the population is concerned, there are effectively three clusters. Two given populations may well have a low probability of having two genes that come from the same founder. But their origin vectors may be quite similar, and this is

a sign of their affinity in genetic terms, since they may share many mutations, even if the frequencies of those mutations are low.

Secondly, the three-cluster structure may have been partly erased (or blurred) in the last two centuries of Québec history (from 1800 until today). The key factor here is migration. For example, exchanges between the Center and the two other areas may have been intense enough to erase its genetic 'integrity'. We also know that many (about 700,000) French Canadians emigrated to the United-States in the 19th century, a mass movement that may have had some effects, if migrants were not an independent sample relative to their ancestors' geographical origins. Moreover, the process of colonization of new territories did not stop at the end of the 18th century. We mentioned earlier that the Saguenay-Lac-Saint-Jean was founded in the middle of the 19th century. But other territories have been open to colonization up until 1950. As a result, the comparison of our data with gene frequency data may be jeopardized by the non-correspondence of geographic divisions.

Another difficulty is that a classification based on genealogy is of a different nature than a classification based on gene frequency. One cannot, for instance, use the differential of GC to compare two populations that have no known founders in common. This would be senseless because the two sets would be seen as populations with no genetic affinities, even if there is a genetic similarity between them that can be revealed by marker frequencies. The different nature of the two approaches can also be illustrated by the following fact: according to ancestral composition (or origin vector), the genetic distance between a brother and a sister is zero. But two sibs are not genetically identical. Their high affinity is only revealed when comparing their genetic materials to many other individuals at many loci. This also applies to comparison of populations or regions. Finally, the difference between the two approaches that may have the biggest impact is that the genealogical approach implicitly assumes as many alleles as there are

founders. Here, we assume a set of 9659 different alleles (the total number of founders). Of course no marker is so polymorphic as to be occupied by this amount of alleles. Yet, on the other hand, the genes introduced into the population were certainly not all identical at every locus. To what extent were those alleles, introduced by each founder, different? This is a puzzling question that goes beyond present knowledge.

Genetic contribution as a proxy for the distribution of inherited disorders

Given these limitations, we believe, however, that a study of numerous highly polymorphic markers would display three clusters, and that our study may help to predict, although crudely, the distribution of hereditary diseases in today's population. Even if we do not use actual gene frequencies, our work is based on the whole population, from its foundation through the next two centuries. Since we have found three clusters, we expect three different profiles in terms of epidemiological genetics. The eastern part of the province has already been described by many clinical and molecular studies, and our results are in agreement with those studies. Its main feature is the presence, in high frequencies, of some inherited disorders that are rare elsewhere. This is paralleled here by small FUNs and close relationships between regions in terms of the probability that two genes come from the same founder (small FUN_{XY}). To the contrary, the West shows a diametrically inverted feature, and again, this corroborates our results. Unfortunately, the Central profile is more complicated and prior to this investigation no study has established its existence. In light of our results, however, it may be seen as intermediate with respect to the two others (see its middle range FUN and FUN/N_f ratio in Table 3.3). Nevertheless, there is a notable exception: the Lotbinière region. Its FUN is almost as low as the one calculated for Beaupré and Charlevoix (region 9). For some reason, geneticists have never extensively surveyed this county. Hence, we don't know if its inhabitants carry any rare or specific deleterious mutations in high frequency. But if so, we expect from our classification that

those mutations are different, in a general manner, from those detected in Charlevoix and Saguenay-Lac-Saint-Jean. These regions, indeed, do not fall into the same cluster.

ACKNOWLEDGEMENTS

We deeply thank Professor Hubert Charbonneau and the two anonymous reviewers for their helpful comments on an earlier draft of this paper. We also thank Bertrand Desjardins for database access, Stephan Gilbert for his technical assistance, Edwin John Sullivan and Owen Dyer for their help with English usage. This research was supported by the *Fonds pour la formation et l'aide à la recherche* (FCAR) of Québec and by the *Social science and humanities research council of Canada* (SSHRC).

CHAPITRE 4 - INTERGENERATIONAL CORRELATION OF EFFECTIVE FAMILY SIZE IN EARLY QUÉBEC (CANADA)³⁵

ABSTRACT

The use of a comprehensive demographic database of the early French Canadian population (1608-1800) reveals an almost null impact of parents' fertility on children's fertility, which contradicts the commonly held view that family size has a tendency to run in families. However, in this population, there is a clear transmission from one generation to the next of the *effective family size* inside a given geographical area (EFS, here defined as the number of children that settle per reproducing individual). Three types of correlations between parents and children EFS are presented, in order to account for the impact of migrations, taking place in one generation or the other. Results show that individuals who belong to a large sibship settled in a given subdivision tend to encourage the settlement of a high number of their own children in the same subdivision. An additional correlation is introduced to see if geographically-based differentials of EFS can account for the differential of founders' regional genetic contribution. The analysis shows that correlation of EFS has a definite impact on the concentration of a population's gene pool, and accounts, at least to a certain extent, for the differences between subdivisions in this regard.

³⁵ Par A. Gagnon et E. Heyer. Article soumis au *American Journal of Human Biology*.

INTRODUCTION

In many studies, positive correlations have been found between the number of children born to a family and the number of children within the husband's and/or wife's family. In other words, it would appear that family size has a tendency to run in families. This phenomenon was first interpreted as evidence of the inheritance of *fecundity* (i.e. biological capacity to reproduce) and attempts have been made to assess the contribution of inheritance to the total variation in individual reproduction. The earliest study goes back to the turn of the century when Pearson et al. (1899) used records of the British peerage to investigate fertility correlations. Working with several thousand pairs of mothers and daughters, they found a positive correlation between daughters and mothers in fertility and took this to indicate an inherited capacity to reproduce. Some 30 years later, Fisher (1930) used these data to estimate the genetic variance in fertility, and concluded that about 40% of the observed variance in fertility was due to heredity.

However, from a re-examination of Pearson's data, Williams and Williams (1974) have shown that secular change, i.e. emergence of birth control practices, was contributing a significant proportion of the correlation between sibship size and progeny number. They concluded that attempts to measure the heritable component in fertility are neither valid nor reliable because of the overwhelming contribution of the social environment to fertility in human populations. Later on, these statements were reinforced by a study by Langford and Wilson (1985), who showed, with 16th to 19th century family reconstitutions of some seven British parishes, that there was no evidence of *fecundity* transmission.

Concomitant to this failure to identify an underlying biological determinant of intergenerational fertility, there has been a shift toward a concern with socio-economic factors that affect the transmission of fertility behaviour (Ben-Porath, 1975; Johnson and Stokes, 1976;

Thornton, 1980; Thornton et al. 1986). Among cohorts of marriages where the prevalence of birth control could be assumed, a certain traditional continuity in family-building habits was also offered as an explanation of fertility differentials (Anderton et al., 1987).

In sum, it appears that when parental family size exerts a definite, even small, differentiating effect on *fertility* (i.e. the demographic measure of the total number of children born from an individual), a large portion of the correlation comes from socio-cultural rather than biological factors (apart from the above-mentioned references, see also: Huestis and Maxwell, 1932; Berent, 1953; Duncan et al., 1965; Imazumi et al., 1970; Neel and Schull, 1973; Wise and Condie, 1975).

As in other populations, no biological source of intergenerational demographic differentials have been found in the early French Canadian population of Quebec. Entirely reconstituted from its first permanent settlement in 1608 to the end of the 18th century by the *Programme de recherche en démographie historique* (PRDH), this semi-closed and homogeneous population provides a mean to measure fertility transmission at a time when birth control was unknown. Studies based on the PRDH computer file displayed a relatively low (if not null) impact of biological factors on some demographic measurements of interest, such as the heritability of longevity (Desjardins and Charbonneau, 1990), the transmission of the capacity to reproduce, i.e. the transmission of fecundity (Desjardins et al., 1991) and the hypothesis of a trade-off between early fecundity and longevity (Le Bourg et al, 1993). In fact, more recent studies exhibit a somewhat inverse relationship, i.e. an impact of environmental factors (such as climate) on biologically related reproductive variables (e.g. sex ratio, Nonaka et al., 1999).

A specific part of the province, namely the Saguenay-Lac-St-Jean (SLSJ) region, has been intensively studied from an epidemio-genetic, as well as from a socio-demographic point of

view by the Inter-University Institute for Population Research (IREP) (see Bouchard and De Braekeleer, 1991; Heyer, 1995; Labuda et al. 1997; Tremblay and Vezina, 2000). Many studies have emphasised the impact of fertility differentials on population renewal and on the evolution of its genetic pool (Tremblay and Heyer, 1993, Heyer, 1995; Heyer and Tremblay, 1995). Also, one contribution has revealed a positive correlation between parents' number of *married* children and the mean number of children of the married children (Tremblay, 1997). In opposition to the above-cited papers, these contributions did not focus on the determinants of fertility differentials, but took them for granted, focusing on the demographic and genetic implications of those differentials.

More recent investigations emphasised the long term genetic impact of the correlation of *effective family size* (EFS, i.e. the number of children that reproduce in a population per reproducing individual; Heyer and Cazes, 1999). Taking over the early seventies studies of MacCluer, Neel and Chagnon (MacCluer et al., 1971; Neel, 1971), showing that socio-cultural factors increase the expected amount of inbreeding in some Indian tribes of Central America, Austerlitz and Heyer (1998) pointed out that the variance of EFS, despite its significant effect, was not sufficient to explain the high frequency of recessive inherited disorders in the SLSJ population. To account for carrier frequencies around 1/21 (e.g. spastic ataxia disorder), it is necessary to include a correlation parameter into the simulations. Adding a relatively low level of correlation ($r = 0.27$) for successive generations (from the early 17th century up to present) gave a good fit to the data of the carrier frequencies. Moreover, in a paper yet to be published, Austerlitz and Heyer show that the correlation affects the recombination rate and increases the expected proportion of individuals carrying the ancestral haplotype in present generations (Austerlitz and Heyer, submitted).

The authors pointed out that with respect to the evolution of a specific regional gene pool, the important feature is not family size *per se*, but EFS, i.e. the number of children that marry and reproduce in the region. Moreover, since the correlation between one generation and the next for family size, i.e. number of children born, is very low (0.07), they also stressed that the correlation in EFS reflects mainly a socio-demographic phenomenon. This involves vertical transmission of such factors as mortality, nuptiality (e.g. age at marriage) and fertility. But it also includes a migration factor, exemplified by the clear exogamous tendency of the less deeply rooted families in the population. In other words, the population may be seen as a core surrounded by a fringe of immigrants and their descendants; core individuals have children that also belong to the core as they produce more “effective” children than do the fringe individuals (see also Heyer, 1993). In total, then, the socio-demographic intergenerational processes involved likely increase the correlation.

In their research, Auzterlitz and Heyer also found that the correlation of EFS was higher during the 17th century ($r = 0.34$) than in the 19th and 20th centuries ($r = 0.17$), and suggested that, since EFS accounts for the number of children who reproduced in the population, the discrepancy between the two values could be explained by differential migration in the 18th century. What is intended here is precisely to explore these reproduction and migration differentials on the whole territory of early settlement of the Québec population (1608-1800), along with aspects related to what has been termed ‘multiplying effects’, i.e. socio-cultural factors that amplify the consequences of founder effect in increasing population homogeneity (Bouchard and Roy, 1990; Bouchard, 1991). Using the PRDH database, we present a geographically-based study of the variance and correlation of EFS for two successive generations.

Before focusing on geographical variations of EFS, we first computed family size (i.e. when fertility alone is considered) and EFS correlations for the whole Saint-Lawrence Valley. Secondly, we compared the three major sub-territories (or 'areas') of the Saint-Lawrence Valley settlement, identified in a previous study by a classification based on the differentials of founders' regional genetic contribution (Gagnon et Heyer, submitted): the Western, the Central and the Eastern areas. For instance, the founders' genetic contribution was shown to be the most *concentrated* in the Eastern area (in the East, few founders had a very high contribution, while many have a small contribution). This area was also the most *homogeneous* as it derives from a relatively small number of founders. The aim of the present analysis is to account for the impact of EFS on the geographical variations of concentration and homogeneity. Hence, thirdly, the same analysis was done for regions, instead of areas. This helped to account for the extent to which the phenomenon is related to the size of the populations under investigation.

4.1 DATA AND METHODS

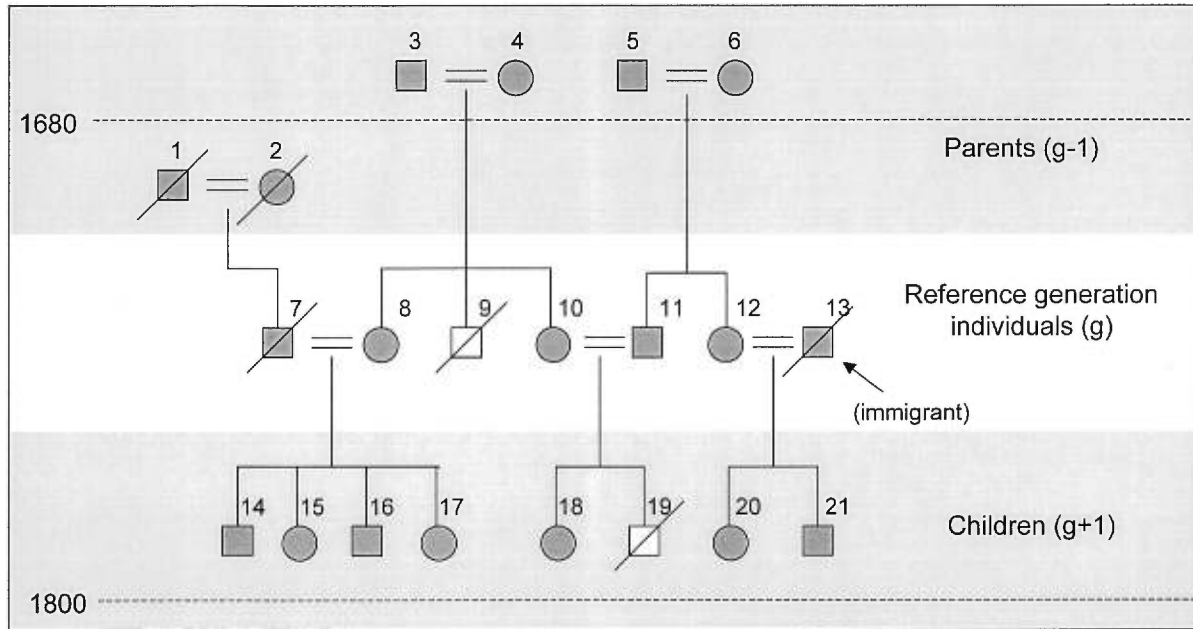
4.1.1 Database and selection of cases

Taking advantage of a set of exceptional conditions the Research Program in Historical Demography (PRDH) has set up a computer file that reconstitutes the birth of a population (Charbonneau et al., 1993). Based on the exhaustive collection of baptism, marriage and death certificates (>712,000) recorded in parish registers, the Population Register of Early Quebec encompasses the vital events, as well as the kinship network of every individual of European ancestry who lived in the Saint-Lawrence Valley between 1608 and 1765. For genealogical reconstitution purposes, the period of observation covered by the Register is a little longer, since the ancestors of every individuals who married within the colony before 1800 have been traced back to the founders of the population. It is worth to mention that the importance of family and religion in this population helps to prevent any falsification of reports (for more information on the database, see : <http://www.genealogy.umontreal.ca>).

With the help of these family and genealogical reconstitutions, we focus here on the Effective Family Size (EFS) of two successive generations. The EFS is defined as the number of children that settle per reproducing individual. The aim is to verify if the number of settled siblings of an individual is correlated with the number of his/her own settled children. For this purpose, data on the nuptiality of three fully linked generations are needed: the parental generation ($g-1$), the reference generation (g), and the children generation ($g+1$). This implies a careful scrutiny of the generations' time-span. For instance, the children's generation ($g+1$) must have completed its nuptiality before 1800, the last year for which comprehensive observation of nuptiality is available in the PRDH computer file. Therefore, we decided to include, in the parents' generation ($g-1$), every reproducing individuals who married before 1680. It can be noted that 1680 corresponds to the period when fertility began to have a higher impact than immigration on the population increase (Charbonneau et al. 1993), and may be seen as a significant turning point of its development.

Figure 4.1 depicts an extract of some families' pedigrees and clarify our criteria of inclusion. It can be seen that individual 7 is not included in the analysis because his parents (1 and 2) married after 1680 (of course, the latter are also excluded for the same reason). However, 14, 16, and 17 are all counted as effective children: they were all settled, and, at least, one set of their grandparents married before 1680, here the maternal grandparents, labelled 3 and 4. Similarly, the fact that 13 is an immigrant does not preclude the inclusion of his married children, 20 and 21, because 12, the mother of the latter (i.e. 13's wife) satisfies all the conditions. Since individual 9 did not settle, he is not added in the EFS of 3 and 4. To the contrary, his sister (8) and his brother (10) are added (both also had at least one effective child).

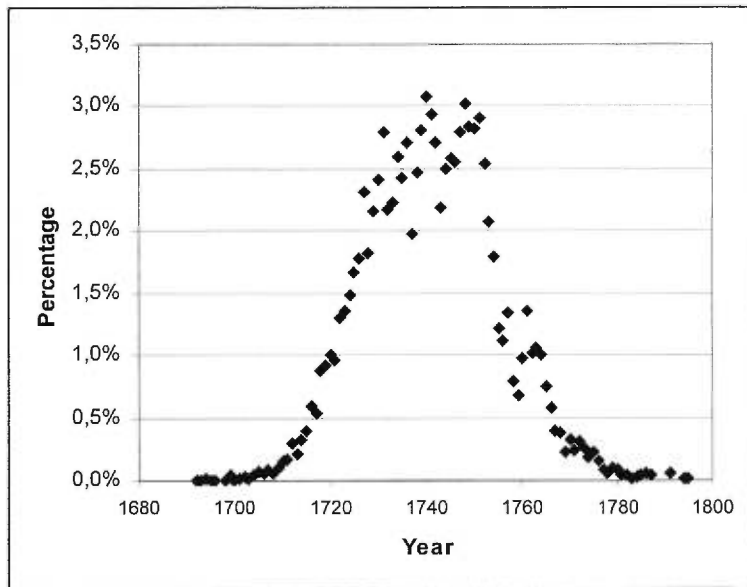
Figure 4.1 - Schematic representation of some families' pedigrees showing which individuals are included in the analysis and to which generation they belong.



- or ■ = married male or female individuals
- or □ = unmarried male or female individuals

In order to confirm that the chosen limit of 1680 would ensure complete observation of three full generations, we plotted (fig. 4.2) the distribution of the marriage years of the grand sons (g+1) whose paternal grand father (g-1) was married between 1650 and 1680. Note that because men do marry and have children at a later age than women, the paternal lines cover the longest possible time periods among the four possible types. A precise study of intergenerational time intervals conducted in the SLSJ population supports this view. For example, the age of fathers at the birth of their last child reaches higher limits than the mothers' (average: 44 years) and shows more variability (Tremblay and Vézina, 2000). Here, since there are very few marriages in the last years of the 18th century, we can consider that no third generation marriage took place after 1800, and, consequently, that the 1680 limit for the first generation marriage year is correct.

Figure 4.2 - Distribution of the year of marriage of the grand sons whose paternal grandfather was married between 1650 and 1680. (Number of grandfathers: 3110; number of grandsons: 16309).



The assigned place of settling of married individuals is the place where his or her first child was born. If this information was not available (about 10% of all the selected unions), we selected his/her place of wedding. The parish of birth of the first child is a better indication of where the couple settled than the parish of marriage because the latter were generally celebrated in the woman's parish of residence, even if the new spouses were to raise their family in the groom's parish or elsewhere (Mayer, 1981; Gradie et al, 1991; Tremblay et al, 2000).

4.1.2 The correlations

To compute correlations, we selected every married individual (males and females of generation g) whose parent's marriage took place before 1680, and then looked for the number of married siblings and the number of married children each of them had. Because of the unbalanced sex-ratio of both immigration and remarriage, we considered separately male and female parents of generation g-1. In other words, we tabulated correlations for both father-

offspring and mother-offspring pairs. Most of the time, nevertheless, the results were not significantly different for both sexes.

All the geographically-based correlations (based on areas or regions) of parent-offspring pairs were done as follow.

- (1) First, for each generation g individual settled in the same subdivision as his parent, we computed the correlation between the number of his/her siblings settled in the same subdivision, and the number of his own children ($g+1$) settled wherever in the territory of the colony. This is called 'Type 1 correlation'.
- (2) Second, we accounted for the impact of the children generation exogamous marital practices by using the same procedure as in (1), except that we subtracted the $g+1$ generation children settled outside their parents' subdivision of settling. Then, in order to be included, the members of all three generations must have settled in the same subdivision. This "intra-territorial" or *in situ* correlation is the 'Type 2 correlation', and corresponds to the procedure used for the SLSJ population in the above-cited Austerlitz and Heyer study (1998).
- (3) Third, for a given subdivision, we selected each individual whose parents (of generation $g-1$) settled in another subdivision or outside the colony, and we computed the correlation between the number of his/her siblings settled in the subdivision, and the number of his own children (generation $g+1$) also settled in the same subdivision. This 'Type 3 correlation' account for the process by which a territory is invested by new incoming groups of siblings. It allowed us to answer the following question: does a large new incoming sibship tend to occupy a territory as to ease the settling of its descendants in this specific territory? In other words: will individuals who have many brothers and sisters settled into a new territory have many of their own children also settled in this territory?

- (4) Finally, for a given subdivision, we selected only individuals whose one or both parents were founders, the latter being defined as generation g-1 immigrants, married before 1680, wherever the place of marriage or settling. The correlation is then taken between the EFS of founders and the EFS of their children. This “Type 4” correlation is used to see if the correlation of EFS had an impact on the concentration of the founders’ genetic contribution. The assessment of this impact is described in the next section.

For the Type 1 and 2 correlations, the tabulations are based on 10776 parent-offspring pairs, i.e. 5388 father-offspring and 5388 mother-offspring pairs (the number are equal because both parents of all the selected individuals are known). At the geographical level, the numbers are 1380 parent-offspring pairs in the Western area, 1262 in the Centre, and 6434 in the Eastern. For the Type 3 correlation tabulation, the number of available pairs is a little higher (11406; respectively 3932, 2290 and 5184) mainly because it allows the inclusion of parents who married outside the colony, as well as parents who married inside. Finally, Type 4 correlations is based on 11019 cases, i.e. 6606 male and 4413 female founder-offspring pairs (2/3 of the founders of the French Canadian population were male). It must be noted that this procedure imposes no constraint on the place of settling of the founders (generation g-1): their wedlock may have took place in the same subdivision as their offspring or elsewhere; what counts is solely their immigrant status.

In the four cases, the sample is slanted towards large sibships, because the tabulations are based on generation g individuals (for example, a parent who had 8 settled children will be represented 8 times). Another solution would have been to use the mean EFS of g generation individuals. This way, we would have had one parent-offspring pair for each g-1 generation individual. But having as many cases as there are parent-offspring pairs represents a better solution. First, families with more children account for a larger proportion of the gene pool; so it

is justified that their weight on correlations is higher. Moreover, using all parent-offspring pairs allowed us to keep track of the intra-group variation of EFS among brother and sisters (while taking the children's mean EFS would not, and would consequently result in a higher correlation). As a matter of fact, a high variation of this sort is a sign of low transmission of EFS, and so it must be taken into account. Finally, the measure used here is more suited for population simulation based on branching process (see Austerlitz and Heyer, 1998).

4.1.3 The effect of variance and correlation of EFS

The relationship between population effective size and variance of family size is a well known matter since Wright's pioneering work (Wright, 1931; Crow and Kimura, 1970; Jacquard, 1974b). But the rigorous mathematical treatment of the effect of the correlation of EFS on the effective size of a population (N_e) remains to be done. Although one can easily figure that the former reduces the latter, the exact relationship is difficult to trace algebraically. Some attempts have been made to shed some light on the phenomenon in terms of branching processes (e.g. Berlin et al, 1992), but these advances remains difficult to apply to real populations.

Nevertheless, the impact of the correlation can be accounted for by simply measuring the increase that it exerts on the probability that two randomly chosen genes in the grand children's genome come from the same grandparent. In a recent paper (Gagnon and Heyer, *submitted*), we presented the calculation of this probability. Starting from genetic contribution, which gives the probability x_i that a randomly chosen gene in a descending cohort comes from the i^{th} founder, we defined the probability that two of his genes would be chosen at random as $P_x = \sum x_i^2$, which is called the 'homogeneity index'. The inverse of this sum was then defined as the Founder's Uniform contribution Number (FUN), i.e. the number of founders a population would have if all its founders had the same contribution (allowing for the same homogeneity in terms of $\sum x_i^2$)

Here, we will focus on the probability that two genes chosen at random in the children (generation $g+1$) gene pool come from the same grandparent ($g-1$). This probability is again defined as $P_x = \sum x_i^2$, where x_i is, for the i^{th} grandparent, the number of his grand children multiplied by $1/4$. However, since the listed grandparents do not necessarily account for the total of the third generation's ($g+1$) gene pool, $\sum x_i^2$ measures in fact the probability that two randomly chosen genes *in the fraction of the genome contributed by the listed grandparents* come from the same of these grandparents (hereafter the “accounted homogeneity index”; AHI). As a matter of fact, one of the parents of the children generation individual may be an immigrant and two of his grandparents may have been married after 1680, so that two grandparents of the individual may be absent of the list.

To allow for the impact of correlation of EFS on the above mentioned probability (i.e. $\sum x_i^2$), we have to account first for the effect of the variance of EFS. This was done by simulation as follows: we first reassigned an EFS for each g generation individual by randomly drawing a number, without replacement, in the g generation's observed EFS distribution. This way, the new g generation's ‘fictional’ distribution of EFS is not linked to the $g-1$ generation's EFS distribution, but the mean and the variance are kept constants. The procedure was repeated 1000 times, which gave 1000 sets of re-distributed EFSs, each of them being independent of the first generation EFSs. Finally, we calculated the $\sum x_i^2$ for each of those 1000 fictional sets, and then averaged for the whole to obtain an estimation of the level of the “accounted homogeneity index” in the absence of correlation ($AHI_{\text{calculated}}$), but with the same mean and variance of the g generation's EFS distribution. The amount of the increase associated to the correlation is then given by: $(AHI_{\text{observed}} - AHI_{\text{calculated}})/AHI_{\text{calculated}}$.

4.2 RESULTS

4.2.1 Transmission of fertility and of EFS (whole colony)

The number of children born to daughters by the number of children their mother had (for all the mothers married before 1680 in the entire colony) are shown in Table 4.1. As it is readily seen, the fertility of French Canadians was quite high; many daughters had more than 15 children, some even reaching 20 or so. But more importantly for what is of concern here, the correlation between mother's and daughter's fertility calculated from this tabulation is almost null ($R=0.05$). Analysis based on men's family size (not shown here) show similar results: the correlation is also practically null ($R=0.01$).

Table 4.1 - Number of children born to daughters by the number of children their mothers had (mothers married before 1680).

Daughter's children	Mother's children																				Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
0	1	2	3	2	3	10	25	18	23	13	23	20	16	8	10	4	2	4			187
1		4	2	2	12	14	15	18	19	24	22	24	17	11	3	1	1	1			190
2	2	2	2	4	3	9	11	18	18	21	13	17	20	6	9	1	1	1			158
3		2	1	3	4	9	14	13	24	18	18	16	15	5	5	2	2		1		152
4	2	1	2	2	5	18	14	16	16	21	24	21	21	10	5	3	1	2	1		185
5	3	5	3	2	11	17	22	14	20	24	23	26	15	9	2	1	4	4			205
6		2	4	2	9	13	27	16	30	35	33	29	9	13	10	1	4	2			239
7	1	2	4	6	7	15	24	25	27	36	35	36	25	6	17	2	1	1		1	271
8		6	1	5	13	18	19	22	35	35	36	36	35	11	19	4	3	2			300
9	3	4	6	9	11	24	33	32	37	46	48	47	26	9	13	2	8	4			362
10	2	5	8	8	14	13	30	37	44	58	42	52	32	19	13	7	11	2		2	399
11	1	2	3	5	10	11	24	28	51	46	45	50	30	16	19	3	9				353
12	3	3	5	7	10	19	20	26	37	38	40	51	22	14	11	3	5	4			318
13		1	4	3	6	17	17	21	20	29	29	29	20	9	12	5		2			224
14	1		2	4	3	4	8	18	16	18	21	19	14	12	7	4	2	3		1	157
15	3		2	1		7	7	6	12	5	6	12	17	3	4	1	1	2			89
16			1	2	1	1	1	4	6	14	5	13	4		2	1	4				59
17				1	1	3	3		2	4	1	4	2	3	1	1		2			28
18		1			2	1		1	3		3	3		2			1				17
19							1					2		1	1						5
20								2		1			2								5
21													1								1
23															1						1
Total	22	42	53	68	125	223	315	335	440	486	467	507	343	167	164	46	60	36	2	4	3905

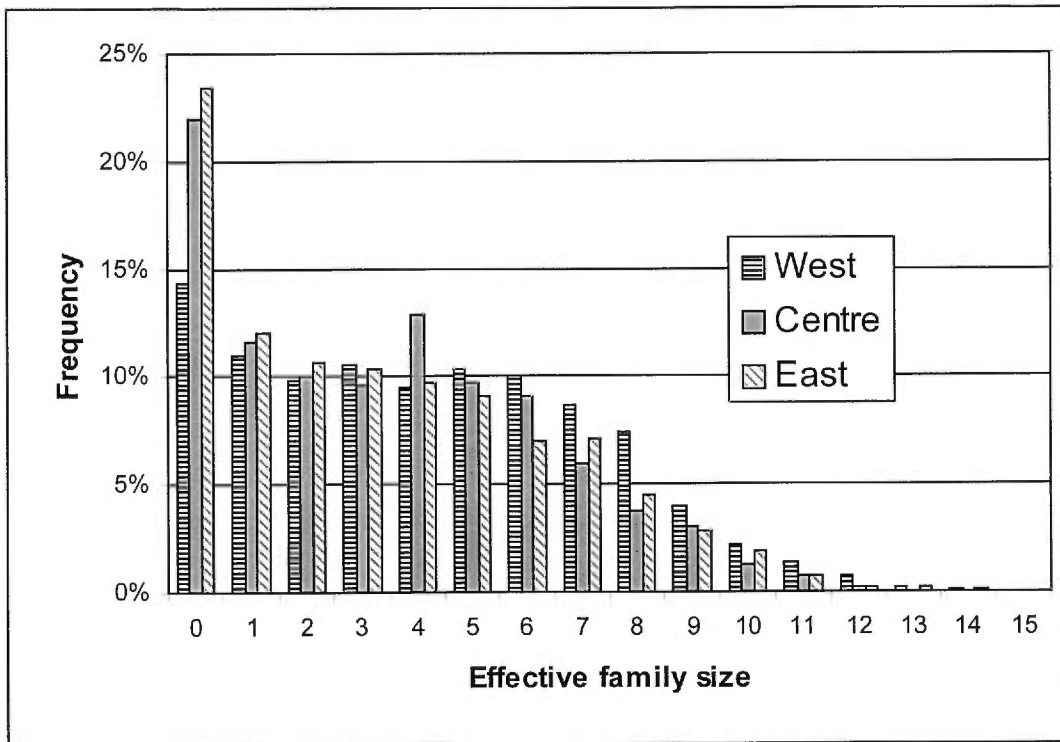
$R = 0.05$

If we focus on EFS, instead of fertility alone, we get higher coefficients of correlation of 0.11 for father-offspring and 0.12 for mother-offspring pairs (table not shown). Thus, having a parent who had many children is not linked to having many children; what matters is the number of settled children, i.e. EFS. Let us turn now to the geographical variation of EFS.

4.2.2 The distribution of EFS

The frequency distributions of EFS for women married before 1730 for the three areas are presented in Figure 4.3. The three distributions are quite similar, except for the zero class frequency, which shows a lower value in the Western area (because there was relatively few out migrations from this area). Overall, the variances and the means parallel each other, and thus the ratio of the two values are similar (the men's distributions, not included here, show approximately the same results). Thus, the variation of EFS alone cannot explain the differentials of homogeneity and concentration indexes between the three clusters (see introduction).

Figure 3. Distribution of EFS of women married before 1730, per area



4.2.3 Area-based analysis of EFS

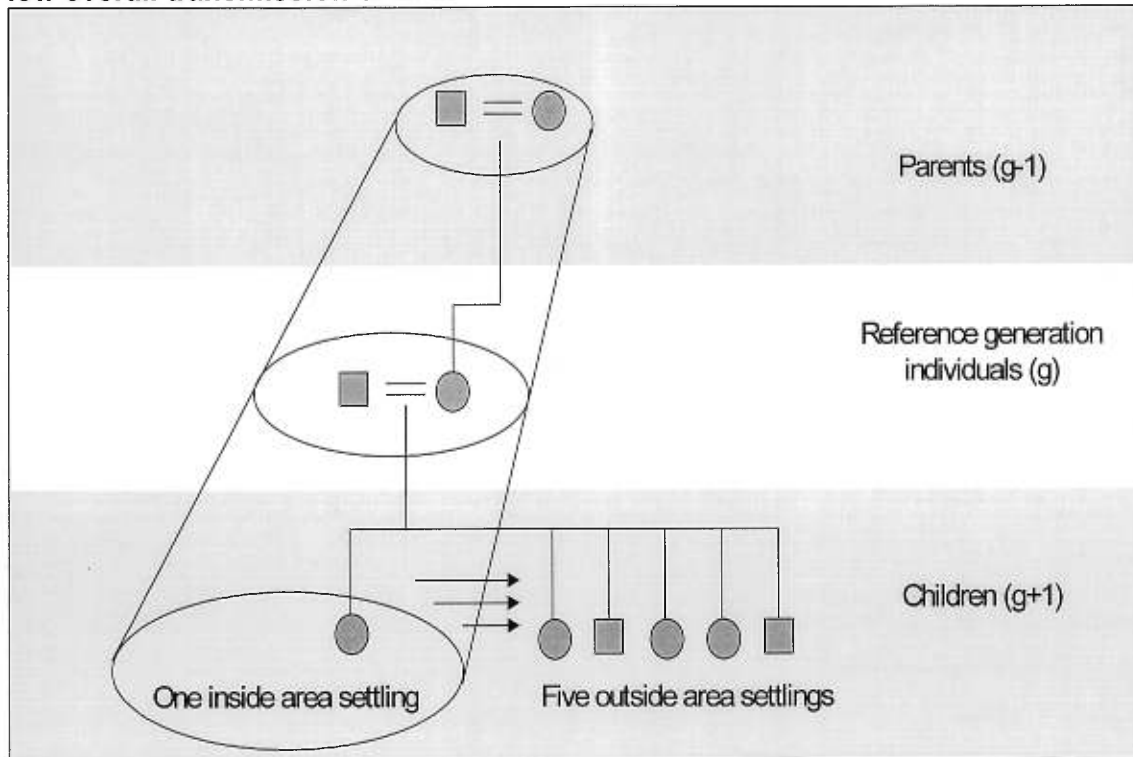
Table 4.2 presents the three types of correlation, applied to the three subdivisions of the colony. Although the Type 1 values are relatively low, it can be inferred that generation g individuals whose parents ($g-1$) had many settled children in their own area had a better chance to have many settled children also, *wherever in the territory of the colony*. This tendency is clearest in the Eastern area, where Type 1 correlation is equal to 0.17 for father-offspring pairs, and 0.18 for mother-offspring pairs. The results are, respectively, 0.13 and 0.09 in the Central area, and 0.09 and 0.11 in the Western area.

Table 4.2 - Three types of EFS correlations according to three geographic subdivisions (area) in early Quebec

Area	Father-offspring pairs			Mother-offspring pairs		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
West	0.09	0.10	0.22	0.10	0.11	0.22
Centre	0.13	0.22	0.17	0.09	0.15	0.18
East	0.17	0.20	0.28	0.18	0.21	0.28

By subtracting the generation $g+1$ individuals who settled outside their parents' (generation g) area, i.e. by calculating the *in situ* correlation (Type 2), we get an indication on the generation $g+1$ propensity to migrate according to the g generation's settled sibship size. It is seen that sibship size seems to be more easily transmitted *into a given subdivision* from one generation to the next. For the three areas, the Type 2 correlations are higher than the Type 1, although the increase is small in the Western area. This means that individuals (g) who belonged to small settled sibship had children ($g+1$) who were more likely to migrate outside their area and settle elsewhere. Otherwise, Type 1 and Type 2 correlations would be approximately equal. A simple example will make it clear. Figure 4.4 shows an extract of a pedigree of a family that does not seem to "transmit" its EFS. The parent had only one settled child, but the latter had 5 settled children. If only the third generation children settled in the same area of their parents are taken into account, the transmission of intra-territorial EFS is quite high: the parent have one settled child, and so has this child (the four other children settled outside the area). Thus, this family decreased the Type 1 correlation, and increased the Type 2.

Figure 4.4 - Extract of a pedigree representing a high in situ transmission of EFS, but a low overall transmission of EFS.



The highest values of Table 4.2 are found for Type 3 correlations, which underlines the process of sibship establishment into a new subdivision. What strikes the most here is the increase from Type 1 and 2 to Type 3 in the Western area (with a Type 3 correlation of 0.22 for both father-offspring and mother-offspring pairs). However, the Eastern area shows the highest values (respectively, 0.28 and 0.27 for father-offspring and mother-offspring pairs). For those two areas, it may be said that what matters the most is not the exclusive *in situ* transmission of EFS into the area, but the fact that new incoming sibships are numerous or not. To the contrary, in the Central area, the *in situ* process of transmission of EFS is more important (in this area, the Type 2 correlation is stronger than the Type 3).

The Type 4 correlations, which assess the link between the EFS of founders and the EFS of children (Type 4 correlations) are presented in Table 4.3, along with the 'accounted index of

concentration' associated with the observed correlation of EFS, the percentage of increase and the concentration indexes taken from a previous study (Gagnon and Heyer, *submitted*). Those concentration indexes refer to the ratio of Founders' Uniform contribution Number (FUN) of the 1780-1799 marriage cohort to the actual number of founders.

Table 4.3 - Percentage of homogeneity increase associated to the Type 4 correlation per area and FUN/N_f ratios calculated from the genetic contribution of founders to the 1780-1799 marriage cohort.

Area	$\sum x_i^2$ without correlation (AHI _{calculated})	Type 4 correlation	$\sum x_i^2$ with correlation (AHI _{observed})	% of increase	FUN/N _f ratio
West	1.47x10 ⁻³	0.18	1,80x10 ⁻³	23%	26.9%
Centre	2.57x10 ⁻³	0.17	3.30x10 ⁻³	21%	26.5%
East	1.08x10 ⁻³	0.22	1.37x10 ⁻³	27%	17.7%

It is clear that even if the observed correlation is relatively low, it exerts a certain effect in increasing the homogeneity of a population. In the Eastern area, a Type 4 correlation of 0.22 increased the 'accounted index of concentrations and homogeneity' by 27% (with a zero correlation, but with the same first and second generations means and variances, the index would have been 27% lower). The table also shows that the increase accounts, to a certain extent, for the concentration of the founders' genetic contribution to the 1780-1800 marriage cohort. The two measures are inversely related, as expected, since we suppose that a high increase produced by a correlation would decrease the ratio of the FUN to the observed number of founders (N_f). For instance, the most concentrated area (the East) is also the one that shows the highest increase due to correlation. Thus, the correlation seems to be clearly associated to the concentration of the gene pool. However, the regionally-based analysis that follow shows that this is not always the case.

4.2.4 Region-based analysis of EFS

Table 4.4 is the same as Table 4.2, except that the analysis focuses on regions, instead of areas. Type 1 and Type 2 correlations could not be computed for every regions. Regions 1 (North of Montreal), 5 (Portneuf), 6 (Lotbinière) and 10 (Bas-Saint-Laurent and Beauce) were founded too late to yield an observation of three full generations. But for regions for which data are available, we see, again, that the correlations rise from Type 1 to Type 3. Thus, the processes that act at a higher level of geographic subdivision also act at this lower level. It cannot be said that the intergenerational links in the transmission of EFS are stronger or weaker in the case of regions (the correlations are neither higher nor lower). But it may be suspected that some variations were masked at the area level of observation. As an example, the value of Type 3 correlation is about 0.26 for Region 1, while it is 0.22 for the whole Western area. Moreover, the values of the eastern regions (around 0.20) are notably lower than the whole area values (0.28), except for region 9 (Beaupré and Charlevoix), which shows the highest value of all the tabulations (0.30). Hence, being along with many settled brothers and sisters in region 7 (Quebec City area), 8 (Orléans Island) and 10 (Bas-St-Laurent and Beauce) does probably ease the settlement of many children in the same region. But this “sibship” effect is stronger in regards to the whole Eastern area. Note also that in region 8, the Type 2 correlation is much higher than Type 3.

Table 4.4 - Correlations of EFS per region in early Quebec.

Area	Region	Father-offspring pairs			Mother-offspring pairs		
		Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
West	1) North of Montreal	--	--	0.26	--	--	0.26
	2) Montreal Island	0.09	0.11	0.21	0.12	0.17	0.20
	3) South of Montreal	0.03	0.08	0.19	0.03	0.03	0.17
Centre	4) Gvt of Trois-Rivières	0.10	0.16	0.12	0.06	0.07	0.11
	5) Portneuf	--	--	0.18	--	--	0.20
	6) Lotbinière	--	--	0.16	--	--	0.14
East	7) Quebec city area	0.11	0.20	0.21	0.13	0.22	0.18
	8) Orléans Island	0.16	0.29	0.16	0.13	0.27	0.20
	9) Beauré and Charlevoix	0.05	0.16	0.30	0.07	0.15	0.30
	10) Bas-St-Laurent and Beauce	--	--	0.18	--	--	0.19

As mentioned earlier, the link between the homogeneity index and the correlation is weaker in the case of regions (Table 4.5). In region 5 for instance a correlation of 0.22 increased the indexes by 36%, a higher value than the increase produced by the same correlation in region 1 (24%). This is because more than one value of 'accounted homogeneity indexes' can be associated to the same value of correlation. The results must then be interpreted with care (see discussion). However, for every region, whatever the level of correlation, there is always an increase of the indexes. Moreover, the maximum increase (45%) is found for region 9, which also displays the highest correlation (0.28).

Table 5. Percentage of homogeneity increase associated to the Type 4 correlation per region and FUN/N_f ratios calculated from the genetic contribution of founders to the 1780-1799 marriage cohort.

Area	Region	$\sum x_i^2$ without correlation (AHI _{calculate})	Type 4 correlation	$\sum x_i^2$ with correlation (AHI _{observed})	% of increase	FUN/N _f ratio
West	1) North of Montreal	10.7x10 ⁻³	0.22	13.3 x10 ⁻³	24%	30,2%
	2) Montreal Island	2.29 x10 ⁻³	0.17	2.90 x10 ⁻³	27%	24,9%
	3) South of Montreal	4.03 x10 ⁻³	0.14	4.80 x10 ⁻³	19%	29,2%
Centre	4) Trois-Rivières	4.39 x10 ⁻³	0.13	5.18 x10 ⁻³	18%	26,8%
	5) Portneuf	7.04 x10 ⁻³	0.22	10.6 x10 ⁻³	36%	23,4%
	6) Lotbinière	33.2 x10 ⁻³	0.21	4.15 x10 ⁻³	25%	21,8%
East	7) Quebec city area	1.92 x10 ⁻³	0.19	2.60 x10 ⁻³	37%	30,4%
	8) Orléans Island	4.84 x10 ⁻³	0.22	6.46 x10 ⁻³	33%	29,0%
	9) Beaupré/Charlevoix	5.17 x10 ⁻³	0.28	7.90 x10 ⁻³	45%	14,2%
	10) Bas-St-Laurent/ Beauce	5.33 x10 ⁻³	0.20	6.64 x10 ⁻³	24%	20,2%

DISCUSSION

In the early French Canadian population, family size does not run in families since there is an almost null impact of parents' fertility on children's fertility (R = 0.05 for mother-daughter pairs and 0.01 for father-offspring pairs). Similar low correlations were also found in the SLSJ population (R = 0.07; Austerlitz and Heyer, 1998) and in another "natural fertility" population: from published genealogies of Amish families, Neel and Schull (1973) gave a correlation of 0.09 between mother's sibship size and her completed reproduction (the comparable figure for fathers is 0.07). In fact, what runs in families is the *effective family size*, which depends mostly on socio-demographic factors, as some studies have pointed out (see introduction) Here, what must be emphasised is the impact of 'in and out migration' between the subdivisions of a semi-closed population.

First, we have shown that when the subdivision of settling of the children ($g+1$) generation is taken into account, there is an increase of the correlation of EFS (in our data, each case showed an increase from Type 1 to Type 2 correlation, whatever the level of subdivision of the territory). As said earlier, this means that, in a given subdivision, third generation individuals who's parents belong to small settled sibships were more likely to settle outside the subdivision. It also means that emigrants constituted a non random sample with respect to the number of collaterals (i.e. uncles, aunts and cousins) they are related to within their natal subdivision. In fact, more individuals belonging to large families emigrate, but this is simply because they outnumber the individuals belonging to small families: in terms of "probability" of departure, the value is higher for the latter than for the former. Overall, the process involved is likely to increase the average consanguinity among the remaining individuals, in comparison with what would be expected under random migration. This *in situ* correlation does not explicitly refers to the so-called "kin-structured-migrations" (Fix, 1978; 1981; 1997). But it may be said that it reveals a kind of "number-of-kin-migrations".

In his potential mate analysis study of St. Barthelemy (West Indies), Leslie (1980) explored the relationship between internal migration and genetic differentiation. He pointed out that individuals who married out of their natal community tended to be more closely related to members of that community than were those who remained and married endogamously. In Leslie's study, the picture is somewhat inverted, since the probability of leaving is higher for those who have less collaterals. But in all, the models that predict random exchanges of individuals between subdivisions of a population must be revisited, not only in light of the "kin-structured-migrations", but also in light of "number-of-kin-migrations".

Second, when numerous siblings settled into a subdivision from which their parents were absent, they tended to settle many of their own children in this same subdivision (Type 3

correlations). This phenomenon, which can also be termed as a “number-of-kin-migration”, must have had a high impact on the stratification of the early French Canadian gene pool. Although, to our knowledge, it has never been shown in other studies of other populations, it can be hypothesised here that it generally took place in the early times of the colonisation of a new territory. However, we may suspect that as the time went, as the territory became wholly occupied, Type 3 correlation erased itself, while Type 2, which although had a lesser impact, took over the relay.

In fact, this interplay between Type 2 and Type 3 correlations can be involved to explain why, in the Austerlitz and Heyer study, the SLSJ correlation was higher for the 17th and 18th century than for later periods (see introduction). In the Beaupré and Charlevoix region (labelled 9), which gave birth to the SLSJ region in the 19th century, the *in situ* (Type 2) correlation is only 0.15, while the “sibling” correlation (Type 3) is 0.30. This means that the correlation was stronger in the early times of colonisation than afterward. As a matter of fact, the mean year of marriage of the reference generation individuals (g) in Type 2 correlations is 1703, while it is 1685 for Type 3 correlation, almost 20 years earlier. Thus, the high correlation in old SLSJ genealogies, which include all ancestors (wherever they came from) of the sample of individuals under study, reflects the Type 3 correlation.

It is well known that the very beginning of the peopling of the Beaupré and Charlevoix region was organised by a landlord who was successful in attracting several immigrants from the small region of Perche (France), immigrants who were thus presumably interrelated to a certain extent. This truly kin-structured-migration from France to Nouvelle-France, along with the founder effect, has been held responsible for the present high frequency of some rare inherited disorders (Bouchard and De Braekeleer, 1991; De Braekeleer, 1994). But, as G. Bouchard (1990) also pointed out, some ‘multiplying effect’ took over and extended the scope of the

founder effect. For example, in the SLSJ population, compared to other descendants, those related to Charlevoix settlers more often remained in the Saguenay, they had more children who stayed in the region up to their marriage and beyond, and they married preferentially within the bonds of kinship. "They also benefited in many ways from the assistance provided by the family group, which helped them to improve their homestead and develop roots" (p.181). The correlations involved here are clearly related to this phenomenon.

How much do the correlations explain the geographic differentials of the concentration of the founders' contribution ?

Although the correlations had an effect in increasing the homogeneity and concentration indexes, it cannot explain alone all the geographic differentials of the founders genetic contribution to the 1780-1799 marriage cohort. Other factors had an influence. First, the time of the beginning of the peopling. Quebec City (the main city of the Eastern area) was founded first in 1608, although the first family settled in 1617. But it took 2 or 3 decades before the advent of Trois-Rivières (1634) and Montreal (1642), respectively the main cities of the Central and the Western areas. Then, if the Eastern area is more homogeneous and concentrated at the end of the 18th century, it is also because its peopling began first. Therefore, the fertility differentials (including the effects of both variance and correlation of EFS) had approximately one more generation to play.

Second, a high correlation implies a high generation time length. The transmission of high EFS requires a longer period because having many children takes time. In other words, the intergenerational interval is much higher for those families who contribute to increase the correlation by transmitting their high capacity to settle children. Hence, many individual descending from large families settled much later than those who's parent and grandparents' EFS where small. Consequently, the effect of correlation is weaker if generations overlap than if

they do not. And in fact, it is likely that if a time-based correlation was measured, i.e. if the correlation of the number of children married within a given limit of time for each parent was calculated, lower results would be obtained.

The third reason has already been invoked; even if the means and the variances of EFS are kept constant, to a given value of correlation, many homogeneity indexes (i.e. $\sum x_i^2$) can be associated. A clear demonstration of this fact is difficult, but a roundabout way can be used. Let's suppose, for instance, that we compute the correlation between the EFS of parents and the *mean* EFS of their children. We want to assess its effect on the probability that two randomly chosen genes in the third generation came from the same grandparent (or founder). If w_i is the number of grandchildren of the i^{th} founder out of n founders, this probability is:

$$\begin{aligned}\sum x_i^2 &= \sum (1/4 w_i / \sum 1/4 w_i)^2 \\ &= \frac{1/n \sum w_i^2 / n}{(\sum w_i / n)^2}\end{aligned}$$

But $w_i = e_i \times m_i$, where e_i is the EFS of i , and m_i is the mean EFS of i 's children, so that:

$$\begin{aligned}\sum x_i^2 &= \frac{1/n \sum (e_i m_i)^2 / n}{(\sum e_i m_i / n)^2} \\ &= \frac{\text{Cov}(e^2, m^2) + M(e^2) \cdot M(m^2)}{n (\text{Cov}(e, m) + M(e) \cdot M(m))}\end{aligned}$$

where Cov is the covariance, and M is the average. The last relation can also be written as:

$$\sum x_i^2 = \frac{(\text{Var}(e^2) \cdot \text{Var}(m^2))^{1/2} \cdot R(e^2, m^2) + M(e^2) \cdot M(m^2)}{n ((\text{Var}(e) \cdot \text{Var}(m))^{1/2} \cdot R(e, m) + M(e) \cdot M(m))}$$

where R stand for 'correlation'. Hence, it is seen that, in this case, the index of homogeneity is proportional to the correlation of the *square* EFSs, but inversely related to the

correlation of EFSs. Most of the time, a high $R(e, m)$ is associated to a high $R(e^2, m^2)$, so that a high observed correlation will result in a high $\sum x_i^2$. But still, two populations with the same mean, variance and correlation of EFS can have different homogeneity indexes, and this may explain some apparently odd results. Even a null correlation of EFS can be associated with different correlations of square EFS, and this is why we chose to take the average of many simulated resulting homogeneity indexes to assess the base value for a null correlation. Of course, this interpretation is correct only if we can transpose the relationship between the correlation and the homogeneity to the case of parent-offspring EFS correlations (instead of parent-mean offspring EFS correlation as done here). But intuitively, it may be deduced that a sort of covariance of squared values would also appear in this complicated equation.

Nevertheless, the correlation has an impact. It is worth mentioning that at the regional level, it is higher in Beaupré and Charlevoix (region 9), and also higher in the East at the area level, where the concentrations are higher at both level. But it also acted in every territorial sub-units, which supports the generality of the “core-fringe model” (Heyer, 1993; Austerlitz and Heyer, 1998). In this regard, the frontiers of a population need not be hermetic, and, in fact, we know that the whole territory of the colony has always witnessed numerous inter-subdivisions migrations flows. But, since the probability of departure was higher for individuals who belonged to small families (who then had less collaterals), the phenomenon was likely to favour the maintenance of divergence between subdivisions. Consequently, the population may have kept the memory of the stratification that occurred at the outset of settlement. At a more general level, it may be hypothesised that similar features are always involved in the peopling of a new territory.

ACKNOWLEDGEMENTS

The authors wish to thank Pr. Hubert Charbonneau for helpful comments. We also thank Bertrand Desjardins and Edwin John Sullivan for their help with English usage. This research was supported by the *Fonds pour la formation des Chercheurs et l'Aide à la Recherche* (Québec) and by the *Social Sciences and Humanities Research Council of Canada*.

CONCLUSION GÉNÉRALE

Pour conclure cette étude, nous avons jugé utile de passer brièvement en revue ses principaux résultats tout en suggérant de nouvelles avenues de recherche qui pourraient répondre aux questions qu'elle laisse en suspens. Pour ce faire, nous allons procéder par ordre chronologique en indiquant en quoi chaque article, bien que constituant une unité en soi, est lié de façon structurelle à chacun des autres. En effet, le second article constitue un complément méthodologique au premier. Le troisième propose une nouvelle manière d'aborder l'homogénéité d'une population, un thème qui sous-tend les deux premiers articles. Il aborde par ailleurs plus en profondeur la stratification spatiale du génome québécois. Quant au quatrième, il tente d'explicitier les résultats obtenus dans les articles précédents au moyen d'un mécanisme de transmission intergénérationnelle.

Apparentement et choix du conjoint (Chapitre 1)

Le premier article nous a permis de montrer, si besoin en était, que l'apparentement des conjoints est beaucoup plus élevé dans l'Est de la province que dans l'ouest. Dans l'ensemble toutefois, les mariages entre très proches parents (oncle-nièce ou cousins germains) constituent l'exception plutôt que la règle: ils ne comptent en effet que pour 0,2% du total. La hiérarchie des régions suivant le coefficient moyen de parenté n'est donc pas tributaire de l'importance différentielle de ces mariages. Aussi ne relève-t-elle pas d'une propension plus ou moins grande à choisir son conjoint dans la parenté immédiate. En fait, le facteur clé est celui de la *disponibilité* des conjoints non apparentés. Cette disponibilité s'avère tout simplement plus réduite dans les régions les plus isolées, comme celle de Charlevoix. Les individus ne partageant aucun lien de parenté (proche ou éloignée) s'y font globalement plus rares qu'ailleurs au Québec. C'est ainsi qu'ils sont en quelque sorte « forcés » d'épouser un parent, même si le

parent est de degré éloigné. Cela corrobore d'ailleurs des études semblables, menées à l'endroit du Saguenay-Lac-Saint-Jean (Bouchard et De Braekeleer, 1991a).

Certains résultats obtenus dans cet article rejailliront peut-être dans le domaine de l'épidémiologie génétique. Par exemple, aucune enquête n'a jusqu'à maintenant ciblé la région de Lotbinière en vue d'y repérer des génopathies. Or, nous avons pu montrer que l'apparement des conjoints y était à peu près aussi important que dans les régions de Beauré et de Charlevoix (qui ont donné naissance au Saguenay-Lac-Saint-Jean, où les maladies héréditaires ont une incidence certaine). Ainsi, les généticiens épidémiologistes, qui prendront acte de ce résultat, seront peut-être tentés d'y regarder de plus près.

Par ailleurs, étant donné la richesse des données dont nous disposons, nous aurions pu entreprendre des études beaucoup plus poussées sur la nuptialité et les déterminants socio-démographiques du choix du conjoint. C'est par la juxtaposition d'un ensemble de choix matrimoniaux que se constitue en partie le pool de gamètes qui forme les générations successives. C'est une phase essentielle du processus de renouvellement des gènes qui est mise en branle au moment où les conjoints se choisissent l'un l'autre, et il est aisé de comprendre pourquoi les généticiens des populations s'y sont toujours intéressés. Ces choix ne sont ni complètement déterminés par des contraintes normatives, que mettraient en œuvre les stratégies familiales, ni totalement le fruit de décisions individuelles. Ils obéissent toutefois à une certaine logique qui peut être décrite au moyen de certains modèles. Nous pensons, entre autres, au modèle du « voisinage », qui permet d'évaluer l'aire géographique au sein de laquelle se contractent les unions et d'expliquer en partie la création d'une stratification régionale au sein d'un bassin génétique (neighborhood knowledge; Boyce, 1967, Swedlund, 1972). L'intérêt du RPQA à cet égard est bien évidemment son caractère exhaustif. À notre avis, il permettrait de rendre compte de la représentation des limites d'une aire matrimoniale avec une précision

inégalable. Un aspect, qui aurait certes mérité plus d'attention de notre part, serait sans doute mis en évidence dans une telle étude : la linéarité du peuplement le long de la vallée laurentienne, qui a sans doute largement orienté le cours de la stratification du bassin génétique québécois. D'une manière générale également, nos efforts auraient pu porter sur l'étude des conjoints potentiels (« potential mates analysis »). Encore une fois, ces efforts auraient bénéficié de l'observation exhaustive de la population. Mais il nous fallait opérer des choix, et la possibilité de tester la méthode isonymique nous est apparu comme prioritaire, étant donné sa grande popularité.

Apparemment et isonymie (Chapitre 2)

C'est ainsi que nous avons pu montrer, au deuxième chapitre, que cette méthode semble bien fonctionner à l'endroit de l'ensemble de la population, mais qu'il en va tout autrement au niveau régional. À Lotbinière, par exemple, les coefficients mesurés au moyen de la fréquence des mariages entre conjoints portant le même patronyme sont deux fois plus élevés que ceux mesurés à partir de la méthode généalogique. Grâce aux numéros de lignée, nous avons démontré que la source de ce biais n'était pas l'origine polyphylétique des patronymes. Elle est plutôt à mettre sur le compte de différentiels socio-démographiques ayant trait au rapport de masculinité. D'abord, les liens de parenté des premiers habitants de Lotbinière passaient surtout par les hommes. Ensuite, tout au long de la période d'observation, les hommes s'y remariaient plus souvent que les femmes. Enfin, les hommes étaient moins mobiles que les femmes, ce qui a amené une concentration locale des patronymes (la stabilité des hommes a déjà été mise en évidence par F. Mayer dans la population de La Serpentine, voir introduction).

Ces trois sources de biais sont rarement invoquées dans les études d'isonymie. Nous croyons qu'il valait la peine de le relever. Toutefois, nous aurions pu faire beaucoup plus. Il aurait été possible de développer des moyens pour corriger les estimations, en fonction de

certaines informations concernant le peuplement. Par exemple, nous savons que le rapport de masculinité des fondateurs de la population québécoise avantage nettement les hommes (Charbonneau et al. 1987; voir aussi chapitre 2), ce qui laisse supposer un plus grand nombre de liens en ligne maternelle. Ainsi, à l'endroit de la méthode isonymique, on pourrait calculer un effectif efficace de fondateurs qui tiendrait compte de ce déséquilibre. Néanmoins, cela ne serait pas sans difficulté. Nous venons d'y faire allusion, à l'échelle locale, les gènes empruntent plus volontiers les lignées masculines, parce que les hommes émigrent moins que les femmes. D'un côté, à l'échelle de l'ensemble de la population, on peut s'attendre à ce que les boucles de consanguinité éloignées se nouent préférentiellement autour des fondatrices. De l'autre, à l'échelle locale, ce sont plutôt par les hommes que les conjoints se trouvent des ancêtres communs. Ce fait à lui seul justifierait une thèse! Sans doute y apprendrait-on, par exemple, une foule de choses sur les probabilités de survie de l'ADN mitochondrial (voir à ce sujet : Heyer, 1995).

En somme, même si les études de consanguinité et d'apparentement ont déjà fait couler beaucoup d'encre, les perspectives à peine esquissées ici montrent que le sujet est loin d'être épuisé. Or, bien qu'elles aient jeté un éclairage nouveau sur l'évolution de l'apparentement au Québec, nos contributions en ce sens n'ont apporté, somme toute, qu'une vision bien partielle de la stratification du bassin génétique québécois à ses origines. Les deux premiers chapitres ne nous ont pas dit comment les différentes régions se distinguent les unes des autres en regard de leur « composition ancestrale », celle-ci ayant une importance certaine sur le plan épidémiogénétique. Certes la consanguinité accroît le potentiel d'expression d'une génopathie particulière au sein d'une population, mais encore faut-il que le gène mutant associé à cette génopathie y ait été introduit. Or, les fondateurs de la population canadienne-française n'ont pas tous contribué de la même manière aux différentes régions du Québec. Certains ont eu beaucoup de descendants à Montréal, d'autres en Beauce, d'autres encore à Trois-Rivières. Pour des raisons

évidentes toutefois, les régions limitrophes auront reçu des apports relativement semblables, auxquels pouvaient être associées certaines mutations particulières. Il y avait donc lieu, pour compléter les profils esquissés aux deux premiers articles, de procéder à une analyse de la répartition spatiale de la contribution génétique des fondateurs.

Contribution génétique et classification des régions (Chapitre 3)

Mais d'abord, dans une première étape, le troisième article a surtout cherché à développer une nouvelle manière d'appréhender l'homogénéité et la concentration d'un bassin génétique, ce qui, en cela, rejoignait les objectifs des deux premiers articles. Nous avons été ainsi amenés à nous poser la question suivante : Pour un niveau d'homogénéité donné, combien de fondateurs aurait eu une population si tous les fondateurs avaient eu une contribution égale ? Nous avons pu ensuite fournir une notion claire, d'un point de vue génétique, de la concentration de la contribution des fondateurs en rapportant le *Nombre utile de fondateurs à contribution uniforme* (NUF) au nombre observé de fondateurs (N_f). Du même coup, cela permettait de mettre en évidence le jeu des facteurs socio-démographiques qui modulent cette concentration.

Ce faisant, nous avons tout de même préservé le mode de représentation traditionnel de la concentration d'une distribution, pour comparer nos résultats avec ceux des études antérieures. Ainsi, en observant la « courbe de Lorentz » du troisième article, on constate qu'à la toute fin du XVIIIe siècle, la contribution s'avère beaucoup plus inégale que celle calculée pour l'année 1730 dans *Naissance d'une population* (Charbonneau et al. 1987): Dix-sept pour cent des fondateurs fournissent environ 70% des gènes des individus mariés entre 1780 et 1800. En fait, Charbonneau et ses collaborateurs comparaient la contribution génétique des seuls pionniers arrivés avant 1680, tandis qu'ici, nous avons considéré l'ensemble des fondateurs, quelle que soit leur période d'arrivée (pas seulement les pionniers). Ainsi, l'asymétrie de la

contribution peut-elle apparaître plus accentuée par le fait que les fondateurs qui arrivent plus tard ont généralement une contribution plus faible que ceux arrivés plus tôt. Il serait utile de refaire nos analyse de concentration en n'y incluant que les fondateurs arrivés avant 1680. Tout ceci pourrait par ailleurs être accompagné d'une étude de l'évolution du NUF. En suivant l'évolution du NUF, par exemple, on apprend que celui-ci n'a de cesse de diminuer tout au long du XVIIIe siècle (résultats non fournis dans cette thèse). Il peut paraître curieux que le nombre de fondateurs diminue même s'il y a encore des entrées dans la population. En fait, comme le NUF est fonction de l'homogénéité de la population (il évolue en raison inverse de celle-ci), il est attendu qu'il diminue au fil du temps.

Dans la seconde partie du troisième article, nous avons été amenés à formuler certaines hypothèses, relativement audacieuses, qui touchent à la stratification actuelle du bassin génétique québécois. Compte tenu de ce que l'on sait de l'importance des conditions initiales en génétique des populations, il est permis d'envisager que le profil des distances génétiques entre régions du Québec actuel reflète en grande partie celui qui caractérise le Québec ancien. Ainsi, puisque l'analyse de la contribution génétique des fondateurs par région a révélé l'existence de trois bassins de population différents, on peut s'attendre à ce qu'il en aille de même pour le Québec actuel. Une telle découverte mettrait en cause une idée communément admise (à l'effet que le Québec soit divisé en deux bassins; Bouchard et DeBraekeleer, 1991a; Bouchard et al. 1995). En fait, l'idée d'une division bipartite relève d'une observation relativement sommaire de la distribution des génopathies sur le territoire. On trouve en effet dans les régions de l'est des maladies dont la fréquence est très élevée, alors qu'ailleurs, les distributions sont plutôt caractéristiques de celles que l'on trouve en général dans les populations caucasiennes. Mais, si les régions du Centre et de l'Ouest sont similaires en regard des données actuelles de la biologie moléculaire, elles le sont peut-être par défaut: on ne croit pas utile de les distinguer parce qu'elles ne présentent rien de particulier pour l'instant. Une meilleure connaissance de la

structure génétique du Québec permettrait d'y voir plus clair. Car cette connaissance est pour l'instant très fragmentaire. Il n'existe en effet aucun répertoire exhaustif des maladies répertoriées dans la population. Celui-ci permettrait d'ajouter du relief à la carte de l'épidémiologie génétique du Québec, tout en précisant ses contours pour les études à venir.

Mais, sans avoir recours à de telles données, l'utilisation du RPQA nous permettrait d'en apprendre encore beaucoup sur l'évolution de la structure génétique du Québec, et de formuler encore d'autres hypothèses. Voici quelques exemples de ce qu'on pourrait envisager :

- On pourrait observer l'évolution des distances génétiques entre régions : quand apparaissent les trois clusters; demeurent-ils stables par la suite ?
- On pourrait également procéder à une analyse par sexe de la contribution génétique, à l'exemple de ce qui a déjà été fait pour la population du SLSJ (Heyer, 1995b). Du même coup, on pourrait classifier les régions à partir de la contribution des fondateurs de sexe masculin, puis de sexe féminin. Les résultats obtenus pourraient alors être comparés avec ce que l'analyse de l'apparentement des conjoints selon le sexe des ancêtres communs aurait révélé (voir plus haut).
- On pourrait encore classifier les régions en utilisant le lieu de provenance des fondateurs. Le concept de NUF, développé dans le troisième article, pourrait alors être aménagé pour représenter la probabilité que deux gènes, choisis au hasard, viennent de la même région en France, de la même paroisse, etc.

Ainsi, avons nous à peine profité des nombreuses possibilités qu'offrent le RPQA pour le calcul et le développement théorique de la mesure de la contribution génétique. Toutefois, malgré son intérêt, la contribution génétique à elle seule ne permet pas d'identifier clairement l'ensemble des processus socio-démographiques qui expliquent l'évolution de la structure

génétique d'une population. Il faut donc faire appel à d'autres méthodes qui soient à même de cerner d'encore plus près ces facteurs.

Corrélation intergénérationnelle de la fécondité utile (Chapitre 4)

Au chapitre 4, nous avons tenté de mettre en évidence (à l'exemple de Tremblay, 1997; Austerlitz et Heyer, 1998) l'existence d'un lien intergénérationnel de la fécondité utile et d'en tirer les conséquences sur les processus d'homogénéisation et de concentration d'un bassin génétique. Sur un « terroir neuf », les stratégies familiales visent essentiellement à étendre les assises foncières de la famille en vue d'y établir un maximum de descendants (Bouchard, 1991; 1996). Évidemment, tous n'y parviennent pas avec un même succès (Tremblay 1997; Austerlitz et Heyer, 1998), sans quoi il n'y aurait aucune *Corrélation intergénérationnelle de la fécondité utile* (CIFU). Il a été montré qu'au SLSJ (Bouchard, *ibid.*), ceux qui implantent le plus grand nombre de leurs enfants se rattachent par de nombreux liens généalogiques aux plus anciens fondateurs de la région, venus pour la plupart de Charlevoix. En marge de ce noyau figure une population plus ou moins flottante, celle des migrants ou étrangers installés depuis peu, qui ne parviennent pas à reproduire les leurs dans une aussi grande proportion que ceux de la « vieille souche ». Ce modèle d'implantation, et ses conséquences sur le plan génétique, est-il valable pour l'ensemble de la Vallée laurentienne à l'époque de la Nouvelle-France ?

Nous pouvons dire qu'il l'est en général. Certes notre démonstration est pour l'heure indirecte et partielle. Mais nous la trouvons convaincante, bien que d'autres travaux devront être engagés afin d'en examiner avec plus de détails les ressorts et les particularismes régionaux. La validité du modèle tient au fait que la corrélation de Type 3 (qui circonscrit le processus d'implantation de nouvelles fratries) est, sauf exception (région 8; Île d'Orléans), la plupart du temps supérieure à la corrélation de Type 2 (qui circonscrit le processus de la reproduction familiale *in situ*). Cela veut dire que tout se joue au début. C'est en quelque sorte ce que G.

Bouchard (1996) appelle le « phénomène des pionniers accapareurs », par lequel les premiers arrivants prennent littéralement possession du territoire et parviennent à y implanter un plus grand nombre de leurs enfants que les autres. En un certain sens, l'exception confirme la règle. Si la corrélation mettant en cause le processus d'implantation de nouvelles fratries est plus faible à l'Île d'Orléans, c'est simplement, comme nous l'a suggéré H. Charbonneau, parce que cette île a été fondée d'un seul jet, par des fondateurs récemment débarqués dans la colonie (voir chapitre 1). Ainsi la maxime « premier arrivé premier servi » ne pouvait-elle y être mise en application. Toutefois, là comme ailleurs, « ... la famille nombreuse était le facteur qui à la fois provoquait ces stratégies (de reproduction) et en assurait la réussite. » (Bouchard, 1996 : 480), et c'est pourquoi on y observe la plus grande corrélation de Type 2 (qui circonscrit le processus *in situ*) de tout le territoire de la Nouvelle-France.

Dans ce chapitre, nous avons également essayé de voir dans quelle mesure les processus de reproduction différentielle des fondateurs et de leur descendance (tels que mesurés par la CIFU) pouvaient rendre compte de l'homogénéisation de la population. Nous avons vu qu'en général, plus la corrélation est élevée, plus l'homogénéisation est grande. Toutefois, la correspondance n'est pas parfaite. D'autres développements devront donc être apportés afin de circonscrire plus précisément les facteurs qui en expliquent la variation. Dans une première phase, cela pourrait se faire, à l'exemple de Austerlitz et Heyer (1998), aux moyens de simulations. La question posée pourrait être du type : Avec une distribution donnée de la fécondité utile et une corrélation intergénérationnelle de cette fécondité, quelle distribution des gènes devrait-on avoir au bout de tant de générations ? Après plusieurs « essais » ou simulations, on pourrait ainsi avoir une meilleure connaissance des facteurs sociaux et démographiques qui ont modulé l'évolution, sur le plan génétique, de la population québécoise depuis ses origines. Mais il faudra tôt ou tard proposer des développements analytiques qui donneraient une « prise » plus solide sur le phénomène. Des travaux sont d'ailleurs

actuellement en cours à cet égard au Laboratoire d'anthropologie biologique du Musée de l'Homme.

Il faut insister sur le caractère largement « en friche » du *Registre de population du Québec ancien* à l'endroit des études de démographie génétique. Ici, nous avons bien modestement tenté d'en tirer le maximum, compte tenu des limites qui nous étaient imparties. Mais beaucoup reste à faire. Au moment de conclure, il nous vient d'ailleurs à l'esprit qu'il eût été intéressant de comparer l'intervalle intergénérationnel du Québec ancien à celui calculé par M. Tremblay et H. Vézina (2000) pour la population du SLSJ. À partir de l'armature généalogique de cette population, ces derniers calculent des intervalles qui tournent autour de 30 ans (avec quelques variations entre les hommes et les femmes), ce qui invalide les estimations la plupart du temps utilisées pour calculer l'âge des mutations dans les études d'anthropologie génétique (soient 20 ou 25 ans). Sans doute y aurait-on trouvé une estimation semblable. Or, il pourrait être intéressant de vérifier si cet intervalle varie selon le découpage territorial adopté. On pourrait par exemple calculer l'intervalle en ne prenant en compte que la descendance utile *in situ* (un peu comme nous l'avons fait ici fait dans le calcul de la CIFU) et voir s'il y a des variations inter-régionales. On pourrait ensuite comparer les résultats par région avec ceux obtenus pour l'ensemble de la population. Cette étude, par ailleurs, nous entraînerait du même coup vers des problématiques qui touchent de près à la question de l'endogamie intergénérationnelle, bien étudiée à l'IREP (entre autres : Tremblay et Heyer, 1993; Tremblay et al. 2000).

Nous espérons avoir montré la pertinence de l'emploi du RPQA et l'intérêt des nouvelles méthodes que nous avons développées. À un niveau plus général, on peut dire que l'intérêt de la démographie génétique tient à ce qu'elle nous permet de comprendre la diffusion des maladies héréditaires et d'établir un politique épidémiogénétique pour l'ensemble de la

population. Le vieillissement a déjà commencé à mettre à jour un certain nombre de maladies dont la composante est en partie héréditaire (voir par exemple la thèse de H. Vézina sur la maladie d'Alzheimer, 1998). Avec la diminution importante de la fécondité et la réduction de l'effectif efficace qu'elle entraîne, on peut prévoir que ces maladies augmenteront en incidence et que d'autres apparaîtront, tout au moins en région. Il est remarquable que ces deux variables inter-reliées que sont la diminution de la fécondité et le vieillissement de la population soient spécifiquement du ressort de la démographie, ce qui en justifie encore la pertinence pour l'épidémiologie génétique.

BIBLIOGRAPHIE

- Adams, J. (1990) « Introduction : genetics and demography and historical information ». Dans *Convergent Issues in Genetics and Demography*, sous la direction de Adams, J., Lam, D.A., Hermalin, A.I. et Smouse, P.E., New York et Oxford, Oxford University Press : 3-13.
- Allen, G. (1965) « Random and nonrandom inbreeding ». *Eugenics Quarterly* 12 (4): 181-198.
- Anderton, D.L., Tsuya, N. Bean, L.L. et Minau, G.P. (1987) « Intergenerational transmission of relative fertility and life course patterns ». *Demography* 24 (4) : 467-480.
- Auger, F. (1966) « La corrélation comme statistique descriptive ». *Biométrie Humaine* : 147-162.
- Austerlitz, F. et Heyer, E. (1998) « Social transmission of reproductive behaviour increases frequency of inherited disorders in a young-expanding population ». *Proc. Natl. Acad. Sci* 95 : 15140-15144.
- Austerlitz, F. et Heyer, E. (2000) « Linkage disequilibrium depends on socio-demographic factors ». Soumis.
- Barbeau, A. (1966) « The syndrome of hereditary late onset ptosis and dysphagia in French Canada » dans *Symposium uber progressive Muskeldystrophie*, sous la direction de Kuhn, E, Berlin, Springer Verlag : 102-109.
- Barbeau, A., Coiteaux, J., Trudeau, G. et Fulleim, G. (1964) « La chorée de Huntington chez les Canadiens français : étude préliminaire ». *L'Union médicale du Canada* 93 :1178-1183.
- Beauchamp, P., Charbonneau, H., Desjardins, B. et Légaré, J. (1977) « La reconstitution automatique des familles: un fait acquis ». La mesure des phénomènes démographiques. Hommage à Louis Henry, *Population* 32, no spécial : 375-399.
- Bélanger, P. et Landry, Y. (1990) *Inventaire des registres paroissiaux catholiques du Québec (1621-1876)*. Montréal, Presse de l'Université de Montréal.
- Benoist, J. et Auger, F. (1965) « Note sur l'anthropométrie des Canadiens-Français ». *Anthropologica* 7 : 5-11.
- Ben-Porath, Y. (1975) « First generation effect on second-generation fertility ». *Demography* 12 : 397-405.
- Berent, J. (1953) « Relationship between family sizes of two successive generations ». *The Milkbank Memorial Fund Quarterly* 31: 39-50.
- Berlin, Y.A., Drobnitsky, D.O., Goldanskii, V.I., et Kuz'min, V.V. (1992) « On inherited fertility in biological systems: a model of correlated fluctuations in the stochastic branching process ». *Biosystems* 26: 185-192.
- Blanchard, R. (1935) *L'est du Canada français, province de Québec*, Montréal, Beauchemin, vol 1.

Boisvert, M., et Mayer, F. (1994) « Mortalité infantile et consanguinité dans une population endogame du Québec ». *Population* 49 (3): 685-724.

Boleda, M. (1984) « Les migrations au Canada sous le régime français (1608-1760) ». *Cahiers québécois de démographie* 13: 23-40.

Bouchard, G. (1991) « Reproduction familiale et effets multiplicateurs ». Dans *Histoire d'un Génome : Population et génétique dans l'est du Québec*, sous la direction de Bouchard, G. et DeBraekeleer, M., Québec, Presse de l'Université du Québec : 213-254.

Bouchard, G. (1996) *Quelques arpents d'Amérique : population, économie et famille au Saguenay (1838-1871)*, Boréal, Montréal.

Bouchard, G. et Roy, R. (1990) « Effet fondateur et effet multiplicateur dans la population de Saguenay (Québec) ». Dans *Approche pluridisciplinaire des isolats humains / Pluridisciplinary Approach of Human Isolates*, Congress and Colloquia no 3, sous la direction de Chaventré, A. et Roberts, D.F, INED / University of New Castle: 163-182.

Bouchard, G. et De Brakeleer, M. (1991a) *Histoire d'un Génome*, Québec, Presse de l'Université du Québec.

Bouchard, G. et De Brakeleer, M. (1991b) « Mouvements migratoires, effets fondateurs et homogénéisation génétique ». Dans *Histoire d'un Génome*, sous la direction de Bouchard., G et De Braekeleer M., Québec, Presse de l'Université du Québec : 281-321.

Bouchard, G., Charbonneau, H., Desjardins, B., Heyer, E. et Tremblay, M. (1995) « Mobilité géographique et stratification du pool génique canadien-français sous le Régime français ». Dans *Les chemins de la migration en Belgique et au Québec du XVIIe au XXe siècle*, sous la direction de Landry, Y., Dickinson, J.A., Pasleau S. et Desama C., Louvain-la-Neuve, Éditions Académia : 51-59.

Boyce, A.J., Kuchemann, C.F. et Harrison, G.A. (1967) « Neighbourhood knowledge and the distribution of marriage distances ». *Ann Hum Genet*, 30 (4):335-8.

Brais, B., Bouchard, J.P., Jomphe, M., Desjardins, B., Dubé, M.-P., Gosselin, F., Xie, Y.-G., Tomé, F.M.S., Fardeau, M., Brunet, G., Samson, F., Mathieu, J., Prévost, C., Duranceau, A., Codère, F., Pericak-Vance, M., Bulman, D.E., Tapscott, S., Feasby, T., Munsat, T.L., Pilarski, R., Morgan, K. et Rouleau, G.A. (1997) « When genetics and history converge : the fine-Mapping, North American introduction and diffusion of the French Canadian oculopharyngeal muscular dystrophy mutation ». Thèse de doctorat de Bernard Brais, Université McGill, Chapitre 3 : 57-86.

Bry, X. (1995) *Analyses factorielles simples*, Paris, Éditions Economica.

Caron, I. (1916) *La colonisation du Canada sous la domination française*, Québec.

Cazes, M.H. (1986) « Genetic origins of the Dogon population in the arrondissement of Boni (Mali) ». *Am J of Hum Gen* 39 : 96-111.

Chapman, A. et Jacquard, A. (1971) « Un isolat d'Amérique centrale: Les Indiens Jicaques du Honduras ». Dans *Génétique et populations; Hommage à Jean Sutter*, INED, Travaux et Documents no 60, Paris, PUF.

- Charbonneau, H. (1970) *Tourouvre au Perche aux XVIIe et XVIIIe siècles : Étude de démographie historique*, INED, Travaux et Documents no 55, Paris, PUF.
- Charbonneau, H. (1984) « Essai sur l'évolution démographique du Québec de 1534 à 2034 ». *Cahiers québécois de Démographie* 13 (1) : 5-21.
- Charbonneau, H. (1990) « Le caractère français des pionniers de la vallée laurentienne ». *Cahiers québécois de démographie* 19 (1) : 49-62.
- Charbonneau, H. et Légaré, J. (1986) *Répertoire des actes de Baptême, Mariages, Sépulture et des Recensements au Québec ancien*. Montréal, Presses de l'Université de Montréal.
- Charbonneau, H. et Harris, R.C. (1987) « Resettling the St Lawrence valley ». Dans *Historical Atlas of Canada*, sous la direction de Harris, R.C. et Matthews, G.J., Toronto, University of Toronto Press, V 1, pl.45.
- Charbonneau, H., Desjardins, B., Guillemette, A., Landry, Y., Légaré, J., et Nault, F. (1987) *Naissance d'une population: Les Français établis au Canada au XVIIe siècle*. INED, Travaux et Documents no 118, Paris et Montréal, PUF et PUM.
- Charbonneau, H., et Desjardins, B. (1990) « Vivre cent ans dans la vallée du St-Laurent avant 1800 ». *Annales de démographie historique* : 217-226.
- Charbonneau, H., Desjardins, B., Guillemette, A., Landry, Y., Légaré, J., et Nault, F. (1993) *The First French Canadians. Pioneers in the St. Lawrence Valley*. Newark, London et Toronto, University of Delaware Press and Associated University Presses, 236 p. (version anglaise de Naissance d'une population).
- Charbonneau, H., et Guillemette, A. (1994) « Provinces et habitats d'origine des pionniers de la vallée laurentienne ». Dans *Langue, espace, société : Les variétés du français en Amérique du Nord*, sous la direction de Poirier, C., Boivin, A., Trépanier, C. et Verreault, C., Sainte-Foy, Presses de l'Université Laval.
- Crow, J.F. (1980) « The estimation of inbreeding from isonymy ». *Hum Biol* 52 (1): 1-14.
- Crow, J.F. (1983) « Discussion ». *Hum Biol* 55 (2): 383-397.
- Crow, J.F., et Mange, A.P. (1965) « The measurement of inbreeding from the frequency of marriages between persons of the same surname ». *Eugenics Quarterly* 12 (4): 199-203.
- Crow, J.F. et Kimura, M. (1970) *An Introduction to Population Genetics Theory*, New York, Evanston, London, Harper and Row Publishers.
- Darlu, P., et Cazes, M.H. (1988) Bias in the estimation of inbreeding coefficient and probability of origin of genes with errors in genealogy. *Hum Biol* 60 (6) : 901-908.
- De Braekeleer, M. (1990) « Homogénéité génétique des Canadiens français du Québec: mythe ou réalité ? ». *Cahiers Québécois de Démographie* 19 (1) : 29-48.

- De Braekeleer, M. (1994a) « Hereditary disorders in the french canadian population of Quebec. I. In search of founders ». *Hum Biol* 66: 205-223.
- De Braekeleer, M. (1994b) « Hereditary disorders in the French Canadian population of Quebec. II. Contribution of Perche ». *Hum Biol* 66: 225-249.
- Deniel, R. et Henry, L. (1965) « La population d'un village du Nord de la France, Sainghin-Mélantois, de 1665 à 1851 ». *Population* 20 (4) : 563-602.
- Desjardins, B. (1990) « Homogénéité ethnique de la population québécoise sous le Régime français ». *Cahiers québécois de démographie* 19 (1) : 63-76.
- Desjardins, B. (1991) *Le Registre de la population du Québec ancien : Genèse, fonctionnement et applications*, Thèse de doctorat, Université Lumière Lyon 2.
- Desjardins, B., Beauchamp, P. et Légaré, J. (1977) « Automatic family reconstitution: The French Canadian seventeenth-Century experience ». *Journal of Family History* 2 (1) : 56-76.
- Desjardins, B., Bideau, A., Heyer, E. et Brunet, G. (1991) « Intervals between marriage and first birth in mothers and daughters. *Journal of Biosocial Sciences* 23: 49-54.
- Desjardins, B., Bideau, A. et Brunet, G. (1994) « Age of mother at last birth in two historical populations ». *Journal of Biosocial Science* 26 (4) : 509-516.
- Dugergil, D.F. (1998) « Population démographique, population biologique: quelques concepts ». Dans *Le choix du conjoint*, sous la direction de G. Brunet, A. Fauve-Chamoux et M. Oris, Recherches en sciences humaines, Programme Rhône-Alpes: 13-26.
- Duncan, O.D., Freedman, R., Coble, J.M. et Slesinger, D.P. (1965) « Marital fertility and size of family of orientation ». *Demography* 2 : 508-515.
- Dyke, B. (1983) « Estimation of random isonymy ». *Annals of Human Biology* 10 (3): 295-298.
- Ellis, W.S. et Frield, J. (1976) « Inbreeding as measured by isonymy and by pedigrees in Kippel, Switzerland ». *Social Biology* 23 (2) : 158-161.
- Ellis, W.S. et Startmer, W.T. (1978) « Inbreeding as measured by isonymy, pedigrees, and population Size in Törbel, Switzerland ». *Am J Hum Genet* 30 : 366-376.
- Fisher, R.A. (1930) *The Genetical Theory of Natural Selection*. 1958 Dover Pub. New York.
- Fix, A.G. (1978) « The role of kin-structured migration in genetic microdifferentiation », *Annals of Human Genetics* 41: 329-339.
- Fix, A.G. (1981) « Endogamy in settlement populations of Semai Senoi: Potential mate pool analysis and simulation ». *Social Biology* 28: 62-64.
- Fix, A.G. (1997) « Gene frequency clines produced by kin-structured founder effects ». *Hum Biol* 69 (5): 663-673.

- Gagnon, A. (1998) *L'incidence des facteurs socio-démographiques sur la structure génétique d'une population*. Examen général de synthèse. Département de démographie. Université de Montréal.
- Gagnon, A., Charbonneau, H. et Desjardins, B. (1998) « Apparement des conjoints par région au Québec ancien ». Dans *Le choix du conjoint*, sous la direction de Brunet, G., Fauve-Chamoux, A. et Oris, M., Les chemins de la recherche, no 43, Lyon, BOSC: 133-150.
- Gauvreau, D. et Jetté, R. (1992) « Histoire démographique et génétique humaine dans une région du Québec avant 1850 ». *Annales de Démographie Historique* : 245-267.
- Godbout, A. (1946) « Nos hérédités provinciales françaises », *Les Archives du folklore*, t. 1, Montréal, Fides : 26-40.
- Gomila, J. et Guyon, L. (1969) « Étude comparative de petites communautés rurales : méthode et premiers résultats à propos de la consanguinité à Bois-Vert (Québec) ». *Population* 24 (6) : 1127-1153.
- Gradie, M.I., Jorde, L.B., et Bouchard, G. (1988) « Genetic structure of the Saguenay, 1852-1911: evidence from migration and isonymy matrices ». *Am J of Phys Anthropol* 77: 321-333.
- Gradie, M.I., Jorde, L.B., et Bouchard, G. (1991) « La structure génétique de la population du Saguenay » Dans *Histoire d'un Génome*, sous la direction de Bouchard, G et De Braekeleer M., Québec, Presse de l'Université du Québec : 253-277.
- Henripin, J. (1954) *La population canadienne au début du XVIIIe siècle*. INED, Travaux et Documents no 22, Paris, PUF.
- Henry, L. (1956) *Anciennes familles genevoises : Études démographiques, XVIe – XXe siècle*. INED, Travaux et Documents no 26, Paris, PUF.
- Heyer, E. (1993) « Population structure and immigration: a study of the Valserine valley (French Jura) from the 17th century until present ». *Annals of Human Biology* 206: 565-573.
- Heyer, E. (1995a) « Genetic consequences of differential demographic behaviour in the Saguenay region, Québec ». *Am J Phys Anthropol* 98: 1-11.
- Heyer, E. (1995b) « Mitochondrial and nuclear genetic contribution of female founders to a contemporary population in Northeast Quebec ». *Am J Hum Genet* 56 : 1450-1455.
- Heyer, E. et Tremblay M. (1995) « Variability of the genetic contribution of Quebec population founders associated to some deleterious genes ». *Am J Hum Genet* 56 (4):970-978.
- Heyer, E., Tremblay, M., et Desjardins, B. (1997) « Seventeenth-century European origins of hereditary diseases in the Saguenay Population (Quebec, Canada) ». *Hum Biol* 71, 91-101.
- Heyer, E., et Cazes, MH. (1999) « Les "enfants utiles": Une mesure démographique pour la génétique des populations ». *Populations* 54 (4-5): 677-692.
- Huestis, R.R. et Maxwell, A. (1932) « Does family sizes run in families? ». *Journal of Heredity* 23:77-79.

- Imaizumi, Y., Nei, M. et Furusho, T. (1970) « Variability and heritability of human fertility ». *Ann Hum Genet* 33 : 251-259.
- Jacquard, A., et Nadot, R. (1968) « Mariages consanguins et fertilité différentielle ». *Population* 23 (2) : 325-332.
- Jacquard, A. (1974a) *Généétique des populations humaines*. Paris, PUF.
- Jacquard, A. (1974b) *The Genetic Structure of Populations*. Biomathematics Volume 5, New York, Heidelberg et Berlin, Springer-Verlag.
- James, A.V., (1984) « Isonymy and mate choice on St.Bart, French West Indies : computer simulations of random and total isonymy ». *Hum Biol* 55 (2) : 297-303.
- Jetté, R., Gauvreau, D. et Guérin, M. (1991) « Aux origines d'une région: le peuplement fondateur de Charlevoix avant 1850 ». Dans *Histoire d'un Génome : Population et génétique dans l'est du Québec*, sous la direction de Bouchard, et DeBraekeleer, M., Québec, Presse de l'Université du Québec : 75-106.
- Johnson, N.E. et Stokes, C.S. (1976) « Family size in successive generations: the effects of birth order, intergenerational change in lifestyle, and familial satisfaction. *Demography* 13 : 175-187.
- Johnston, F.E. et Albers A.E. (1973) « Computer simulation of demographic processes ». Dans *Methods and Theories in Anthropological Genetics*, sous la direction de Crawford, M.H. et Workman, P.L., University of New Mexico Press, Albuquerque : 201-217.
- Klinger, K., Horn G.T., Stanislovitis, P., Schwartz, R.H., Fujiwara, T.M. et Morgan K. (1990) « Cystic fibrosis mutations in the Hutterite Brethren ». *Am J of Hum Genet* 46 : 983-987.
- Laberge, C. (1967) « La consanguinité des Canadiens français ». *Population* 22 (5): 861-896.
- Laberge, C. (1969) « Hereditary tyrosinemia in a French Canadian isolate ». *Am J of Hum Genet* 21 (1) : 36-45.
- Laberge, C. (1976) « Population genetics and health care delivery : the Quebec experience ». Dans *Advances in human genetics*, Volume 6, sous la direction de Harrisson, H. et Hirschorn, K., New York, Plenum Press : 323-375.
- Labuda, M., Labuda, D., Korab-Laskowska, M., Cole, D.E.C, Sietkewicz, E., Weisenback, J., Popowska, E., Pronicka, E., Root, A.W. et Glorieux, F.H. (1996) « Linkage disequilibrium analysis in young populations : pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians ». *Am J Hum Genet* 59 : 633-643.
- Landry, Y. (1992) « Population et reproduction sociale à l'île d'Orléans aux XVIIe et XVIIIe siècles ». *Revue d'histoire de l'Amérique française* 45 (3): 403-413.
- Langford, C.M. et Wilson, C. (1987) « Is there a connection between a woman's fecundity and that of her mother? ». *Journal of Biosocial Sciences* 17: 437-443.

Larochelle, J., Privé, L. et Mortezaï, A. (1965) « Study of 29 cases of cirrhosis in the infant of which 14 were familial ». Présenté à la 42e réunion annuelle de la Société de Pédiatrie Canadienne, Ottawa, 28-30 juin 1965.

La Rose, A. (1976) *Les registres paroissiaux au Québec aux XVIIe et XVIIIe siècles : Introduction à l'étude d'une institution religieuse et civile*. Mémoire de maîtrise. Université de Montréal.

Lasker, G.W. (1985) *Surnames and Genetic Structure*. Cambridge, Cambridge University Press.

Lasker, G.W. (1988) « Repeated surnames in those marrying into British one-surname lineages: an approach to the evaluation of population structure through analysis of the surnames in marriages ». *Hum Biol* 60 (1) : 1-9.

Le Bourg, E., Thon, B., Légaré, J., Desjardins, B., et Charbonneau, H. (1993) « Reproductive life of French-Canadians in the 17-18th centuries: A search for a trade-off between early fecundity and longevity ». *Experimental Gerontology* 28 : 217-232.

Légaré, J., Larose, A., et Roy, R. (1973) « Reconstitution de la population canadienne au XVIIe siècle: méthodes et bilan d'une recherche ». *Recherches sociographiques* 14 (3) : 383-400.

Légaré, J. (1981) « Le Programme de recherche en démographie historique de l'Université de Montréal: fondements, méthodes, moyens et résultats ». *Études canadiennes/Canadian Studies* 10 : 149-182.

Leslie, P.W. (1980) « Internal migration and differentiation in St Barthelemy, French West Indies » Dans *Genealogical Demography*, sous la direction de Dyke B. et Morrill W., New York, Academic Press : 167-178.

Leslie, P.W. (1990) « Demographic behavior, mating patterns, and the distribution of inbreeding ». Dans *Convergent Issues in Genetics and Demography*, sous la direction de Adams, J., Lam, D.A., Hermalin, A.I. et Smouse, P.E., New York et Oxford, Oxford University Press : 63-79.

MacCluer, J., Neel, J.V., et Chagnon, N.A. (1971) « Demographic structure of a primitive population: a simulation ». *Am J Phys Anthropol* 35: 193-207.

Magnan, E. et Benoist, J. (1969) « Les groupes sanguins chez les Canadiens français ». Dans *Anthropologie biologique au Québec*, sous la direction de J. Benoist : 33-47.

Mayer, FM. (1981) « Histoire démographique et destin biologique de deux villages québécois ». *Anthropologie et sociétés* 5 (2) : 17-67.

Mayr, E. (1963) *Animal Species and Evolution*. Cambridge, Harvard University Press.

Monson, K., Moisan, J.P., Olivier, P., McSween, M., Aubert, D., Guisti, A., Budowle, B. et Lavergne, L. (1995) « Description and analysis of allele distribution for four VNTR markers in French and French Canadian populations ». *Human Heredity* 45: 135-143.

- Morissette, J. (1991) « La consanguinité dans la population de Charlevoix (1680-1852) ». Dans *Histoire d'un Génome : Population et génétique dans l'est du Québec*, sous la direction de Bouchard, G. et DeBraekeleer, M., Québec, Presse de l'Université du Québec : 107-120.
- Naud, J.F. (1997) *Lieux d'établissement des immigrants en Nouvelle-France*. Mémoire de Maîtrise, Département de Démographie, Université de Montréal.
- Naud, J.F., Desjardins, B. et Charbonneau, H. (1998) « Les composantes de la parenté biologique en régime de fécondité naturelle. L'exemple du Québec ancien ». *Population* 53 (1-2) : 291-310.
- Neel, J.V. (1970) « Lessons from a "primitive" people ». *Science* 170 (3960) : 815-822.
- Neel, J.V. et Schull, W.J. (1972) « Differential fertility and human evolution ». *Evolutionary Biology* 6 : 363-379.
- Nonaka, K., Desjardins, B., Légaré, J., Charbonneau, H. et Miura, T. (1990) « Effects of maternal birth season on birth seasonality in the Canadian population during the seventeenth and eighteenth centuries ». *Hum Biol* 62 (5): 701-717.
- Nonaka, K., Desjardins, B., Légaré, J., Charbonneau, H. et Miura, T. (1999) « Human sex ratio at birth and mother's birth season : A multivariate analysis ». *Hum Biol* 71 (5) : 875-884.
- Nei, M. (1972) « Genetic distance between populations ». *The American Naturalist* 106, no 949 : 283-292.
- Nei, M. (1987) *Molecular Evolutionary Genetics*. New York, Colombia University Press.
- O'Brien, E., Jorde, L.B. et Ronloff, B. (1988) « Founder effect and genetic disease in Sottunga, Finland ». *Am J Phys Anthropol* 77: 335-346.
- O'Brien, et al. (1994) « Founder Effect : Assessment of variation in genetic contribution among founders ». *Hum Biol* 66 (2) : 185-204.
- O'Brien, E., Rodgers, A.R., Beesly, J. et Jorde, L.B. (1994) « Genetic structure of the Utah Mormons: comparison of results based on RFLPs, blood groups, migration matrices, isonymy and pedigrees ». *Hum Biol* 66 (5): 743-759.
- Ohayon, E. et Cambon-Thompsen, N.D. (1986) *Génétique des populations humaines*, Paris, INSERM.
- Pearson, K.A. et Bramley-Moore, L. (1899) « On the inheritance of fertility in mankind ». *Phil Trans Roy Soc Lond* 192 : 257-330.
- Philippe, P. (1972) « Inbreeding Effects in a French Canadian Isolate. II. Sex Ratio at Birth ». *Z Morph Anthropol* 64 : 54-59.
- Philippe, P. (1990) « Twinning causative origin investigated by Sartwell's biometrical method ». *American Journal of Human Biology* 2 : 107-115.

- Philippe, P. (1991a) « Box-Jenkins modeling of chronological series of twin births and single births ». *Hum Biol*, 63 (3) : 367-387.
- Philippe, P. (1991b) « Twinning in susceptible mothers : an exploratory study of international data by Payami's models suggests "reproductive maturity" as a risk factor ». *Acta Geneticae Medicae et Gemellologicae* 40 : 269-289.
- Philippe, P. et Gomila, J. (1971) « Structure de population et mariages consanguins à l'Isle-aux-Coudres (Québec) ». *Population* 26 : 707-716.
- Philippe, P. et Gomila, J. (1972) « Inbreeding effects in a French Canadian isolate. I. evolution of inbreeding ». *Z Morph Anthropol* 64 : 54-59.
- Pontier, J., Dufour, A.B. et Normand, M. (1990) *Le modèle euclidien en analyse des données*, Bruxelles, Éditions de l'Université de Bruxelles et Éditions Ellipses.
- Preston, S.H. (1976) « Family sizes of Children and Family Sizes of Women ». *Demography* 13 (1) : 105-114.
- Prost, M. (1998) « Évolution comparée de l'apparement dans les deux paroisses de la vallée de Vallouise en Briançonnais (XVIIe et XVIIIe siècles) ». Dans *Le choix du conjoint*, sous la direction de Brunet, G., Fauve-Chamoux, A. et Oris, M., Les chemins de la recherche, no 43, Lyon, BOSC : 151-166.
- Relethford. (1996) « Genetic drift can obscure population history: problem and solution ». *Hum Biol* 68 (1): 29-44.
- Roberts, D.F. (1968) « Genetic effect of population size reduction ». *Nature* 220 : 1084-1088.
- Samson, R. et Héroux, A. (1996) *Histoire de Lévis-Lotbinière*. Québec, Institut québécois de recherche sur la culture, Collection « Les régions du Québec » no 8.
- Scriver, C.R., Siverberg, M. et Clow, C.L. (1967) « Hereditary tyrosinemia and Tyrosyluria : Clinical report of four patients ». *Can Med Assoc J* 97 : 1047.
- Scriver, C.R., Laberge, C., Clow, CL. et Fraser, FC. (1978) « Genetic and medicine : An Evolving Relationship ». *Science* 200 : 946-952.
- Stefansson, K. (1999) « La banque de données anonymes du système islandais : risques et avantages ». *Centre d'études du polymorphisme humain (CEPH)*, VIIe congrès annuel de Génétique Humaine, Paris, juin 1999.
- Swedlund, A.C. (1972) « Observations on the concept of neighbourhood knowledge and the distribution of marriage distances ». *Ann Hum Genet*, 35 (3):327-30.
- Tay, J.S.H. et Yip, W.C.L. (1984) « The Estimation of inbreeding from isonymy: relationship to the average inbreeding coefficient ». *Annals of Human Genetics* 48: 185-194.
- Thornton, A. (1980) « The influence of first generation fertility and economic status on second generation fertility ». *Population and environment* 3: 51-72.

- Thornton, A., Freedman, R., Sun, T.H. et Chang, M.-C. (1986) « Intergenerational relations and reproductive behavior in Taiwan ». *Demography* 23 : 185-197.
- Tremblay, M. (1997) « Transmission intergénérationnelle de la reproduction au Saguenay à la fin du XIXe siècle ». *Cahiers québécois de démographie* 26 (1) : 129-145.
- Tremblay, M. et Heyer, E. (1993) « Reproduction régionale différentielle des femmes nées entre 1860 et 1870 au Saguenay ». *Cahiers québécois de démographie* 22 (2) : 263-283.
- Tremblay, M., Jomphe, M. et Vézina, H. (1998) « Comparaison de structure patronymiques et génétiques dans la population québécoise ». Dans *Le Patronyme: Histoire, Anthropologie et Société*. Communication présentée aux 11e Entretiens Centre Jacques Cartier. Lyon. Décembre 1998.
- Tremblay, M., et Vézina, H. (2000) « New estimates of intergenerational time intervals for the calculation of age and origins of mutations ». *Am J Hum Genet* 66 (2) : 651-8.
- Tremblay M., Heyer E. et Saint-Hilaire M. (2000) « Comparaisons intergénérationnelles de l'endogamie à partir des lieux de mariage et de résidence. L'exemple de la population du Saguenay ». *Cahiers québécois de démographie* 29 (1) : 119-146.
- Vernay, M., Bideau, A., et Brunet, G. (1998) « Flux matrimoniaux et identité génétique: choix du conjoint et patronymes dans le Haut-Bugey au XIXe siècle ». Dans *Le choix du conjoint*, sous la direction de Brunet, G., Fauve-Chamoux, A. et Oris, M., Les chemins de la recherche, no 43, Lyon, BOSC : 187-208.
- Vézina, H. (1996) « Démographie génétique et maladie héréditaire au Québec ». *Cahiers québécois de démographie* 25 (2) : 293-322.
- Vézina, H. (1997) *Aspects démographiques et généalogiques de la maladie d'Alzheimer*, Thèse de doctorat, Département de démographie, Université de Montréal.
- Weiss, K.M., Chakraborty R., Buchanan, A.V. et Schwartz, R.J. (1983) « Mutations in names : implications for assessing identity by descent from historical records ». *Hum Biol* 55 (2) : 313-322.
- Williams, L.A. et Williams, B.J. (1974) « A re-examination of the heritability of fertility in the British peerage ». *Social Biology* 21: 225-231.
- Wise, J.M. et Condie, S.J. (1975) « Intergenerational fertility throughout four generations ». *Social Biology* 22 : 144-150.
- Wright, S. (1931) « Evolution in Mendelian Populations ». *Genetics* 16: 97-159.

ANNEXE

Tableau A - Effectifs des unions utilisées dans les calculs d'apparement des conjoints par décennie

Décennie	Méthode isonymique	Méthode généalogique
1620-1629	2	4
1630-1639	21	29
1640-1649	59	79
1650-1659	232	263
1660-1669	583	733
1670-1679	614	845
1680-1689	907	1028
1690-1699	1286	1441
1700-1709	1455	1598
1710-1719	2055	2244
1720-1729	2787	2982
1730-1739	3669	3851
1740-1749	4868	5105
1750-1759	6467	6889
1760-1769	8470	8929
1770-1779	9033	9270
1780-1789	10893	11098
1790-1799	14364	14479
Total:	67765	70869

Tableau B - Effectifs des unions utilisées dans le calcul de l'apparement des conjoints par région au moyen de la méthode généalogique.

PÉRIODE	RÉGIONS											Total	
	1	2	3	4	5	6	7	8	9	10	*		
Avant 1620												2	2
1620-39				2			27					4	33
1640-59		63		46			219		1			13	342
1660-79	1	228	98	142	8		742	124	161	2		72	1578
1680-99	42	485	222	226	149	6	602	245	289	155		48	2469
1700-19	133	790	407	319	163	64	874	307	382	327		76	3842
1720-49	485	1364	902	519	320	172	1216	390	528	841		96	6833
1740-59	1264	2240	1691	979	488	330	1795	504	749	1808		146	11994
1760-79	2875	2927	3605	1603	652	402	1567	481	904	2985		198	18199
1780-99	4414	3415	5968	2453	736	489	1634	431	1084	4610		343	25577
Total:	9214	11512	12893	6289	2516	1463	8676	2482	4098	10728	998	70869	

* : Région indéterminée

Tableau C - Effectifs des unions utilisés dans le calcul de l'apparement des conjoints par région au moyen de la méthode isonymique (patronymes et numéros de lignée)

PÉRIODE	RÉGIONS											Total	
	1	2	3	4	5	6	7	8	9	10	*		
1620-39							23						23
1640-59		61		22			208						291
1660-79		217	29	56	1		665	93	133	2	1		1197
1680-99	30	471	191	198	125	4	533	228	281	132			2193
1700-19	120	757	366	298	145	49	848	287	359	281			3510
1720-49	466	1351	877	489	317	143	1189	378	511	735			6456
1740-59	1196	2185	1649	916	477	301	1727	500	726	1657	1		11335
1760-79	2816	2848	3560	1503	620	387	1509	477	880	2878	25		17503
1780-99	4424	3396	5958	2449	733	487	1606	429	1077	4604	94		25257
TOTAL:	9052	11286	12630	5931	2418	1371	8308	2392	3967	10289	121		67765

* : Région indéterminée

Sont présentées ici les différentes régions, les paroisses qu'elles englobent et la date d'ouverture des registres de ces paroisses. La première colonne se réfère à une codification adoptée dans le RPQA, la seconde à la paroisse correspondante. Le chiffre qui suit représente l'année d'ouverture des registres paroissiaux. Enfin, le cas échéant, « mi » désigne « mission amérindienne ».

Tableau D. Les régions et les paroisses.

GOVERNEMENT DE MONTRÉAL			
RÉGION 1 : Nord de Montréal			
075	St-Cuthbert, 1770	282	Lachenaie, 1697
071	Ile Dupas, 1727	591	Terrebonne, 1727
073	Berthier, 1733	285	St-Henri-de-Mascouche, 1750
074	Lanoraie, 1727	285	Ste-Anne-de-Mascouche, 1788
072	Lavaltrie, 1735	593	Ste-Thérèse-de-Blainville, 1789
283	St-Sulpice, 1706	152	St-Eustache, 1769
284	L'Assomption	151	Lac-des-Deux-Montagnes, 1721, mi
375	St-Pierre-du-Portage, 1724	153	St-Benoît, 1799
281	Repentigny, 1679	291	St.-François de Sales, 1687
286	St-Roch-de-l'Achigan, 1787	292	St-Vincent-de-Paul, 1743
361	St-Jacques-de-la-Nouvelle-Acadie, 1774	293	Ste-Rose, 1745
241	St-Paul-de-Lavaltrie, 1786	294	St-Martin (de l'Île-Jésus), 1774
RÉGION 2 : Île de Montréal			
191	Pointe-aux-Trembles, 1674	234	St-Laurent, 1720
192	Rivière-des-Prairies, 1687	231	Lachine, 1676
193	Longue-Pointe, 1724	233	Pointe-Claire, 1713
194	Fort-Lorette, 1696, mi	235	Ste-Geneviève, 1741
195	Sault-au-Récollet, 1721	514	La Présentation, 1688, mi
391	Notre-Dame-de-Montréal, 1642	232	Ste-Anne-du-bout-de-l'Île, 1686
392	Hôpital Général de Montréal, 1720	236	Mission St-Louis, 1704, mi
393	Hôtel-Dieu de Montréal,	612	Île Perrot, 1786
394	Mission de la Montagne, 1680, mi	611	Vaudreuil, 1773
201	St-Régis, 1762, mi	561	St-Joseph de Soulanges, 1752

RÉGION 3 : Sud de Montréal

131	Châteauguay, 1736	621	Contrecoeur, 1679
272	Sault-St-Louis, 1667, mi	471	Sorel, 1669
273	St-Constant, 1752	472	St-Ours, 1727
271	Laprairie, 1670	624	St-Antoine, 1741
274	St-Philippe, 1753	512	St-Denis, 1741
102	Longueuil, 1698	626	St-Marc (sur Richelieu), 1794
101	Boucherville, 1668	511	St-Charles, 1741
622	Varenne, 1693	625	Beloeil, 1772
623	Verchères, 1724	493	St-Hilaire, 1799
521	Fort-St-Jean, 1757	491	Pointe-Olivier, 1739
522	L'Acadie, 1784	103	Chambly, 1706
513	St-Hyacinthe, 1777	19d*	St-Jean-Baptiste, 1797
633	Yamaska, 1727		

GOVERNEMENT DE TROIS-RIVIÈRES

RÉGION 4 : Gouvernement de Trois-Rivières

114	Sainte-Anne-de-la-Pérade, 1670	532	Yamachiche, 1722
113	Batiscan, 1679	331	Rivière-du-Loup, 1714
115	Ste-Geneviève-de-Batiscan, 1727	332	Maskinongé, 1728
116	St-Stanislas, 1787	636	Mission des Abenakis, 1698, mi
112	Champlain, 1665	631	St-François-du-Lac, 1687
111	Cap-de-la-Madeleine, 1660	632	Baie-du-Febvre, 1715
601	Trois-Rivières, 1635	412	Nicolet, 1716
602	Hôpital des Ursulines, 1796	411	Bécancour, 1704
533	Forges du St-Maurice, 1740	414	Gentilly, 1784
531	Pointe-du-Lac, 1742	413	St-Pierre-les-Becquets, 1734

GOUVERNEMENT DE QUÉBEC

RÉGION 5 : Portneuf

444	Saint-Augustin, 1691	441	Cap-Santé, 1679
442	Neuville, 1679	445	Deschambault, 1705
446	Les Écureuils, 1742	443	Grondines, 1679

RÉGION 6 : Lotbinière

324	Deschaillons, 1741	322	Sainte-Croix, 1727
321	Lotbinière, 1697	323	St-Antoine-de-Tilly, 1702
		302	St-Nicholas, 1694

RÉGION 7 : Région de Québec

465	Charlesbourg, 1679	463	L'Ancienne Lorette, 1676
464	Jeune-Lorette, 1697	452	Hôtel-Dieu de Québec, 1723
461	Sillery, 1638, mi	451	Notre-Dame-de-Québec, 1616
466	Ste-Foy	453	Hôpital Général de Québec, 1728

RÉGION 8 : Île d'Orléans

386	St-Pierre, Île d'Orléans, 1666	388	St-Jean, Île d'Orléans, 1679
383	Ste-Famille, Île d'Orléans, 1666	385	St-Laurent, Île d'Orléans, 1679
387	St-François, Île d'Orléans, 1679	462	Beauport, 1673

RÉGION 9 : Beaupré et Charlevoix

501	Tadoussac, 1646, mi	123	Petite-Rivière-St-François, 1733
125	La Malbaie, 1774	389	St-Joachim, 1684
122	Les Éboulements, 1733	381	Ste-Anne-de-Beaupré, 1657
124	Île-aux-Coudres, 1741	382	Château-Richer, 1661
121	Baie-St-Paul, 1681	384	L'Ange-Gardien, 1670

RÉGION 10 : Beauce et Bas-Saint-Laurent

042	Ste-Marie, 1745	313	St-Jean-Port-Joli, 1767
041	St-Joseph, 1738	312	St-Roch-des-Aulnaies, 1734
043	St-François-d'Assise, 1765	252	Ste-Anne-de-la-Pocatière, 1715
303	St-Henri (de Lauzon), 1766	251	Rivière-Ouelle, 1685
301	Pointe-de-Lévy, 1679	253	Kamouraska, 1727
061	Beaumont, 1692	****	St-André, 1791
063	St-Vallier, 1713	583	L'Île-Verte, 1766
064	St-Charles, 1749	581	Trois-Pistoles, 1713
065	St-Gervais, 1780	481	Rimouski, entre 1700-1769
062	St-Michel, 1693	185	Mont-Louis, entre 1700-1759
373	Berthier, 1710	181	Île-Percée, avant 1700
374	St-François, 1733	182	Pabos, entre 1700-1759
375	St-Pierre, 1727	084	Paspébiac, entre 1760-1800
371	St-Thomas, 1679	083	Bonaventure, entre 1760-1800
372	Cap-St-Ignace, 1679	082	Carleton, entre 1760-1800
311	L'Islet, 1679	221	Havre-Aubert, entre 1760-1800
081	Ristigouche, registre éphémère	082	Tracadieche
