Admixed Ancestry and Stratification of Quebec Regional Populations

Claude Bherer,[1,2] Damian Labuda,[1,3] Marie-Hélène Roy-Gagnon,[1,4] Louis Houde,[2,5] Marc Tremblay,[2,6] and Hélène Vézina[2,6]*

[1]*Centre de Recherche du CHU Sainte-Justine, Université de Montréal, Montréal, Québec, Canada*
[2]*Groupe de recherche interdisciplinaire en démographie et épidémiologie génétique, Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada*
[3]*Département de pédiatrie, Université de Montréal, Montréal, Québec, Canada*
[4]*Département de médecine sociale et préventive, Université de Montréal, Montréal, Québec, Canada*
[5]*Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada*
[6]*Département des sciences humaines, Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada*

KEY WORDS French Canadians; population structure; genetic contribution; genealogies; pedigree

ABSTRACT Population stratification results from unequal, nonrandom genetic contribution of ancestors and should be reflected in the underlying genealogies. In Quebec, the distribution of Mendelian diseases points to local founder effects suggesting stratification of the contemporary French Canadian gene pool. Here we characterize the population structure through the analysis of the genetic contribution of 7,798 immigrant founders identified in the genealogies of 2,221 subjects partitioned in eight regions. In all but one region, about 90% of gene pools were contributed by early French founders. In the eastern region where this contribution was 76%, we observed higher contributions of Acadians, British and American Loyalists. To detect population stratification from genealogical data, we propose an approach based on principal component analysis (PCA) of immigrant founders' genetic contributions. This analysis was compared with a multidimensional scaling of pairwise kinship coefficients. Both methods showed evidence of a distinct identity of the northeastern and eastern regions and stratification of the regional populations correlated with geographical location along the St-Lawrence River. In addition, we observed a West-East decreasing gradient of diversity. Analysis of PC-correlated founders illustrates the differential impact of early versus latter founders consistent with specific regional genetic patterns. These results highlight the importance of considering the geographic origin of samples in the design of genetic epidemiology studies conducted in Quebec. Moreover, our results demonstrate that the study of deep ascending genealogies can accurately reveal population structure. Am J Phys Anthropol 144:432–441, 2011. ©2010 Wiley-Liss, Inc.

A population is stratified when relatedness is not uniform across subgroups of this population as a result of unequal, nonrandom genetic contribution of distinct ancestors. Differential contributions occur through demographic processes such as migration, founder effect, isolation, and endogamy that lead to preferential mating and impact on the genetic structure of the population. Ancestors' genetic contributions can be traced in genealogical records in populations where data allowing for genealogical reconstructions are available. Otherwise, the ancestral connections of a population are typically inferred from its genetic diversity patterns. Population stratification is well recognized as a confounding factor in genetic association studies as it can lead to spurious associations (Cardon and Palmer, 2003; Marchini et al., 2004). Although methods are now available to detect and correct for population stratification from genome-wide data, it is preferable for researchers to be aware of potential stratification even before designing their studies. Here, we propose a new approach to analyze population structure from extensive genealogical data, which relies on the differential genetic contribution of the founders and does not require genotype data.

European colonization of the province of Quebec began four centuries ago with the foundation of Quebec City in 1608. Over the span of 150 years of French rule, approximately 8,500 settlers, mostly of French origin, established themselves in "Nouvelle-France" (Charbonneau et al., 1993, 2000). At the time of the British Conquest in 1760, the population, who numbered 70,000, inhabited mainly the shores of the Saint-Lawrence River and its principal tributary rivers. Following the Conquest, between 2,000 and 4,000 Acadians, descendants of French pioneers from Acadia (located in sectors of present-day Nova Scotia, New

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: The Canadian Institutes of Health Research, The Réseau de Médecine Génétique Appliquée of the Fonds de la recherche en Santé du Québec, Fondation de l'Hôpital Sainte-Justine et la Fondation des Étoiles, Fonds de la recherche en Santé du Québec.

*Correspondence to: Hélène Vézina, Ph.D., Groupe de recherche interdisciplinaire en démographie et épidémiologie génétique (GRIG), Université du Québec à Chicoutimi, 555 boul. de l'Université, Chicoutimi, Québec, G7H 2B1. E-mail: hvezina@uqac.ca

Received 18 June 2010; accepted 15 September 2010

DOI 10.1002/ajpa.21424
Published online 10 November 2010 in Wiley Online Library (wileyonlinelibrary.com).

©2010 WILEY-LISS, INC.

Brunswick and Prince-Edward Island), settled in Quebec after the British deportation campaign (Dickinson, 1994; Bergeron et al., 2008). A group of American Loyalists also came to Quebec after the war of Independence of the United States. In the last part of the 18th century until the end of the 19th century, the French Canadian population expanded rapidly, sustained essentially by a high fertility rate. Territories peripheral to initial settlement were colonized. During that period, immigrants came mainly from the British Isles (McInnis, 2000). In the 20th century, the immigrants to Quebec came from much more diversified locations (Piché, 2003). Today, the Quebec population numbers 7.8 million residents, of which 80% are French speaking, 8% are English speaking and 12% are allophone (www.stat.gouv.qc.ca). Eighty-two percent of the English speakers and allophones of Quebec reside in the metropolitan region of Montreal (www.statcan.gc.ca). The majority of French speakers can trace back their ancestry to the 8,500 pioneers of Nouvelle-France. Here, we focus on this portion of the population and refer to them as French Canadians.

The small number of founders—relative, for instance, to the 360,000 immigrants that left the British Isles to people the English colonies (Brais et al., 2007)—most likely contributed to the belief that the French Canadians from Quebec form a homogeneous population. However, in the past 20 years, genetic studies conducted in Quebec have demonstrated that the overall diversity of the French Canadian gene pool is not reduced compared with that of their parental European populations. For instance, mitochondrial and Y-chromosome gene diversity is nearly equal in samples from Quebec and France (Moreau et al., 2007). Genome-wide association studies for common disease did not notice substantial differences in the genetic heterogeneity of French Canadians compared to European populations. On the other hand, the patchy distribution of mutations underlying Mendelian diseases points to local founder effects (Scriver, 2001; Laberge et al., 2005), which suggests that the contemporary French Canadian population, rather than being a single randomly interbreeding entity, is stratified into genetically distinct subpopulations. In addition, genealogical studies of kinship and consanguinity in the contemporary population (Vézina et al., 2004) and of founders' genetic contribution to a cohort of couples married between 1780 and 1800 (Gagnon and Heyer, 2001) indicate some level of stratification of the Quebec gene pool.

In this study, we characterize the population structure and provide insights into the genetic diversity of the contemporary French Canadian population of Quebec using extensive genealogical data. We analyze the genetic contribution of 7,798 immigrant founders identified in the ascending genealogies of a sample of 2,221 subjects selected in all Quebec populations. We investigate the immigrant founders' characteristics and differential contribution to assess the level of diversity of the population and to address how demographic history has shaped this structure. Our study rests on the hypothesis that regional settlement histories that followed the initial founder effect in the 17th century have led to some level of genetic differentiation across regional populations and that this phenomenon can be examined through the analysis of genealogical features of the population.

MATERIAL AND METHODS

Data

Sample. A sample of 2,221 subjects married in Quebec between 1945 and 1965 was drawn randomly from the BALSAC-RETRO genealogical file, which comprises 240,000 marriages of mostly Catholic confession (Bouchard and Vézina, 2009). The sampling process was stratified according to the regional distribution of the population reported in the 1956 Canadian census. At the time, the Catholic population represented 88% of the 4,628,378 inhabitants of Quebec (Henripin and Péron, 1972). The period of 1945–65 was chosen to maximize the sample size (based on data availability) while remaining as close as possible to the present. Married individuals are expected to have contributed to the contemporary population. Out of the 2,237 initially selected, the 2,221 individuals retained have genealogies extending back at least two generations and are unrelated at the 1st and 2nd degree (kinship coefficient <0.125). They were partitioned into eight regional groupings, referred to as regions, using current geographical limits of administrative regions. The grouping was based on demographic and historical criteria in order to capture the progression of settlement (see Fig. 1).

Genealogical reconstruction. Ascending genealogies of the 2,221 subjects were reconstructed using the BALSAC population database (www.uqac.ca/balsac/) and the Early Quebec Population register (www.genealogie.umontreal.ca). Complementary sources such as marriage repositories and family dictionaries were also consulted. Genealogies were reconstructed as far as the sources would allow, essentially up to the first European settlers entering the colony (mean generation depth = 9.3). Less than 0.1% of distinct ancestral links end prematurely because of adoption or lack of information (mean generation depth = 7.9). The genealogies contain over 5 millions ancestral links connecting 153,447 distinct ancestors (see Supporting Information Table 1 for descriptive statistics of the genealogies).

Analysis

Identification of immigrant founders. Genealogical founders are defined as individuals with no parental information available. In our genealogies, they can be the actual immigrant founders, their parents, or other native individuals with no parental information such as adoptees. The BALSAC-RETRO database records information on immigrant status, thus allowing the identification of the individuals who were first to settle in Quebec among the ascending lineages. These individuals, defined as immigrant founders, were identified in 99.8% of the lineages. We included Amerindians among the immigrant founders as, from our standpoint, they introduced genetic diversity in the French Canadian gene pool. The remaining 0.2% of lineages was not considered in the present study. The geographic origin of immigrant founders was obtained either directly from their marriage records, from census data or indirectly by the place of origin/marriage of their parents. Place of origin was determined for 97% of all immigrant founders. We used the date of their first marriage in Quebec as a proxy of the time of arrival/settlement of immigrant founders. When this time could not be determined (in 647 instances, most likely because immigrants married before establishing in Quebec), we approximated the time of marriage by subtracting 30 years from the mean year of marriage of their children. Thirty years correspond to the average parent-children generation interval in the French Canadian population (Tremblay and Vézina, 2000).

Definition of ancestors' layers. We defined an ancestors' layer as a group of ancestors present in our genealogies who married within a given period of 30 years.

Genealogies of each regional sample were sliced into layers of ancestors married ±15 years around the following pivotal years 1660, 1700, 1760, 1800, 1850, and 1900. When both parents and children were found in a layer, only the parents were retained.

Founders' genetic contribution

Genetic contribution to subjects. As parents transmit half of their autosomal genome to each child, the probability that any subject received an allele from any given founder can be calculated by summing transmission probabilities over all genealogical paths connecting a founder to a subject (Roberts, 1968). We computed the genetic contribution of each founder to each subject ($GC_{f,s}$) as:

$$GC_{f,s} = \sum_{i=1}^{p} \left(\frac{1}{2}\right)^{g_i} \quad (1)$$

where f is one of the n_f founders, s is one of the n_s subjects, p is the number of genealogical paths between f and s and g_i is the number of generations separating f from s through a genealogical path i. Genetic contribution calculations were performed with the S-Plus® 8 function library GenLib (www.uqac.ca/grig/).

Relative genetic contribution to groups of descendants. We calculated the relative genetic contribution of each founder to each regional sample (rGC) as:

$$rGC_f = \frac{\sum_{j=1}^{n_r} GC_{f,j}}{n_r} \quad (2)$$

where $GC_{f,j}$ is, as described in Eq. (1), the genetic contribution of a founder f to the jth subject and n_r is the number of subjects in a regional sample. This measure describes the probability that a randomly chosen allele in a group of descendant comes from a given founder. It can be interpreted as the proportion of a group's gene pool expected to derive from a given founder. We also calculated rGC of each founder to each ancestor's layer of a given regional sample using Eq. (2), by replacing n_r by n_a, i.e., the number of ancestors in a given layer.

Homogeneity index. We calculated the "homogeneity index" (HI) for each regional sample and each ancestors' layer following Gagnon and Heyer (2001):

$$HI = \sum_{k=1}^{n_f} (rGC_k)^2 \quad (3)$$

where n_f is the number of founders contributing to a given group of descendants and rGC_k is the relative genetic contribution of the kth founder to that group (see Eq. 2). It represents the probability that two randomly chosen alleles in one group's gene pool come from the same immigrant founder. This statistic is directly related to the variance in founders' genetic contribution which determines the genetic diversity of the population.

Founders' uniform contribution number. The Founders' uniform contribution number (FUN) was calculated for each region as the reciprocal of the homogeneity index. This statistic corresponds to the number of equally contributing founders expected to produce the same genetic diversity as the actual founders in the population under study (Lacy, 1989; Gagnon and Heyer, 2001). If all founders contributed equally to a descendants' gene pool, the ratio of the FUN to the actual number of founders (n_f) equals to one. Unequal genetic contribution leads to smaller FUN/n_f ratio.

Population structure

To explore contemporary population structure observed through genealogical links, we used three different graphical methods. First, we applied a new approach that relies on a principal component analysis (PCA) of the genetic contribution of founders to subjects (GC-PCA) using Matlab. Specifically, we considered the matrix of genetic contribution of founders to subjects [see Eq. (1)] that has n_s rows and n_f columns. We retained the first two principal components (PCs) that explain the most variance as determined by the scree test (Supporting Information Fig. S1). We tested for significant differentiation of the regional samples on the first two PCs through ANOVA using the R statistical package.

Second, we performed a PCA of founders' incidence as proposed by Calboli et al. (2008) (Calboli-PCA). Specifically, we used the founder to subject incidence matrix that has 1 in row i and column j if i is descendant of founder j and 0 otherwise. Using this matrix, we also applied a K-means clustering to identify $K = 2$ clusters of subjects having a maximum number of common founders to replicate the analysis of Calboli et al. (2008).

Third, multidimensional scaling analysis (MDS) of the pairwise kinship coefficient matrix was computed, using one minus kinship coefficients as a measure of distance. Kinship coefficients were calculated using Karigl recursive algorithm (1981) as implemented in the GenLib function library (www.uqac.ca/grig/).

RESULTS

Time of arrival and origins of the founders

In the genealogies of the 2,221 subjects sampled, we identified a total of 7,798 immigrant founders (Table 1 and Fig. 2A). Among these founders, there were 2.6 times as many males as females (Fig. 2B), consistent with the skewed male-to-female ratio among the first settlers of Nouvelle-France (Charbonneau et al., 1993). Seventy-two percent of the immigrant founders settled during the French Regime (1608–1760) and 24% came in two major waves of immigration, the first between 1663 and 1673 and the second a hundred years later between 1755 and 1765 (Fig. 2B, Supporting Information Fig. S2 and Table S2). The first wave corresponds to the arrival of French women, the so-called "Filles du Roy" who were sent from France to encourage stable family-based settlement in Nouvelle-France (Charbonneau et al., 1993). The second wave coincides with the British Conquest. It included Acadians escaping deportation by the British from their original settlements in Acadia as well as French soldiers who stayed in Quebec once the war ended (Charbonneau et al., 2000).

Sixty-eight percent of the immigrant founders came from France. French founders represent the vast majority of Europeans who settled before the British Conquest (Fig. 2C - Supporting Information Table S3). Only 3% of the founders married before 1700 did not originate

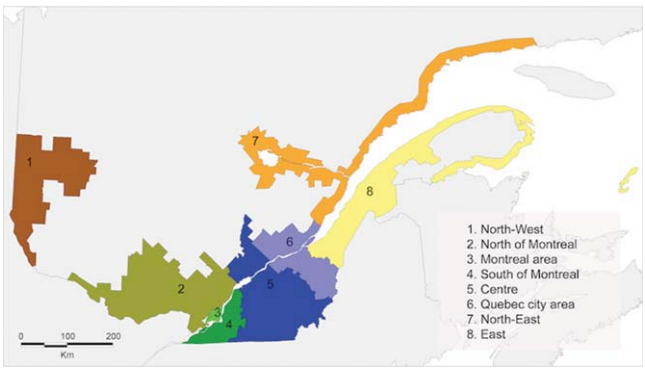

Fig. 1. Quebec regional population samples. The Quebec territory was partitioned in eight regional groupings based on geography and settlement history. The distribution of the 2,221 individuals sampled in these groupings is representative of the population repartition in 1956.

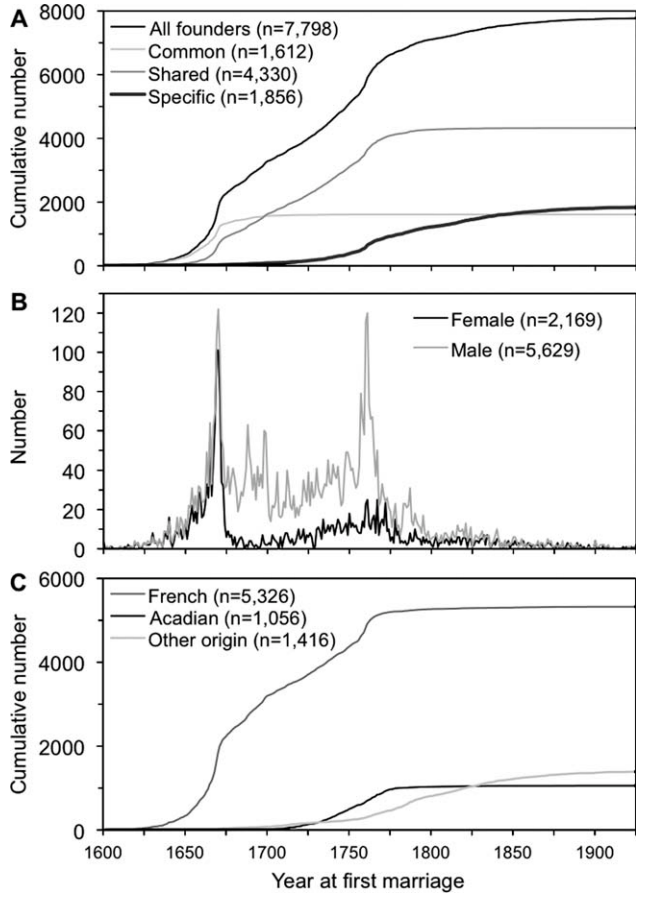

Fig. 2. Time of arrival of the 7,798 immigrant founders. We used the year at first marriage as an estimation of the time of arrival. **A**: Cumulative distribution of the immigrant founders according to year at first marriage and regional representation. Specific founders contributed to one region (n = 1,856), shared founders contributed to 2–7 regions (n = 4,330), while common founders contributed to all eight regions (n = 1,612). **B**: Distribution of immigrant founders according to year at first marriage. **C**: Cumulative distribution of the immigrant founders according to year at first marriage and origin.

directly from France. The proportion of such founders increased to 35% in the period from 1700 to 1760. Acadians represent 14% of all the immigrant founders

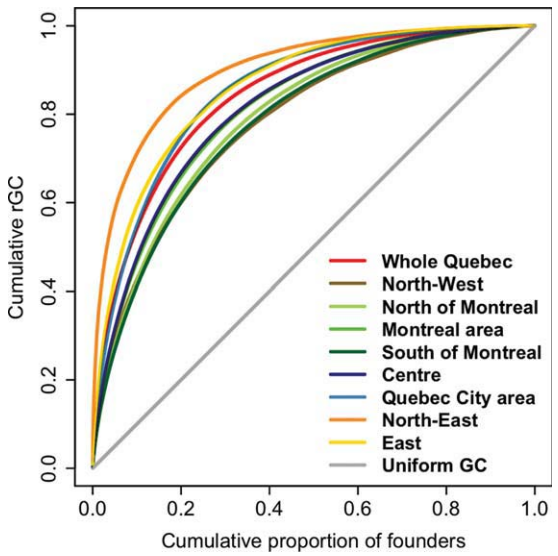

Fig. 3. Cumulative distribution of the founders' relative genetic contribution (rGC). Founders are plotted in decreasing order of genetic contribution. Under the hypothesis of uniform genetic contribution (Uniform GC), all founders would have contributed equally to a given gene pool, a linear dependence would be observed between the proportion of founders and the proportion of total genetic contribution explained by these founders.

TABLE 1. Distribution of subjects and founders per region

	Region	Number of subjects	Number of founders
1	North-West	87	3,724
2	North of Montreal	242	4,343
3	Montreal area	722	6,317
4	South of Montreal	178	4,384
5	Centre	348	4,790
6	Quebec City area	272	3,676
7	North-East	157	2,628
8	East	215	3,188
	Whole Quebec	2,221	7,798

(Fig. 2C - Supporting Information Table S3). The remaining founders of known origin came from Great-Britain (4%), Germany (2%), Ireland (3%), other European countries (1%) and other American locations (besides Acadia) (4%) (Supporting Information Table S3). Amerindian origin was documented for one per cent of founders. Overall, the period of arrival and origins of the immigrant founders appearing in the genealogical ascendance of our contemporary sample well reflected those of the pioneer immigrants that settled in Quebec prior to the British Conquest (Charbonneau et al., 2000).

Partitioning of founders among regions

Not every founder contributed descendants to all regions of Quebec. One fifth of the founders (n = 1,612; 20.7%) were common, that is, they contributed to all eight regions of Quebec (Fig. 2A). More than half of the founders (n = 4,330; 55.5%) contributed to 2–7 regions and more than one fifth (n = 1,856; 23.8%) were specific to only one region (Fig. 2A). All founders common to all eight regions married during the French rule and 97.2% of them before 1700. By contrast, 70% of the specific founders arrived after the installment of the British rule in 1760. We observed a negative correlation between

TABLE 2. Genetic contribution (%) of the founders according to their origin

	Region	Origin							
		French	British	German	Irish	Other European	Acadian	Amerindian	Other American
1	North-West	90.4	1.2	0.2	0.1	0.8	5.2	0.2	1.0
2	North of Montreal	90.0	1.4	0.6	1.4	1.0	2.6	0.2	2.2
3	Montreal area	87.5	2.0	0.4	0.8	1.5	4.1	0.3	2.2
4	South of Montreal	89.9	1.7	0.6	1.0	0.6	3.5	0.1	1.8
5	Centre	89.2	1.0	0.3	0.9	0.8	6.0	0.1	0.9
6	Quebec City area	93.8	1.2	0.2	0.4	1.0	2.3	0.2	0.6
7	North-East	90.3	2.3	0.4	0.4	1.0	3.4	0.3	1.3
8	East	76.4	3.4	0.5	1.8	0.9	11.1	0.3	3.5
	Whole Quebec	89.1	1.8	0.4	0.9	1.2	3.8	0.2	1.6
% of total number of founders		68.3	4.1	1.8	2.7	1.2	13.5	1.2	3.8

Immigrant founders of unknown origin ($n = 252$) explained a minor proportion of the total (3.2%) and of the regional gene pools (0.2–2.3%).

TABLE 3. Proportion of the genealogies (%) in which appears at least one founder of a given origin

	Region	Origin							
		French	British	German	Irish	Other European	Acadian	Amerindian	Other American
1	North-West	98.9	98.9	12.6	16.1	96.6	87.4	58.6	50.6
2	North of Montreal	99.6	91.7	31.8	41.3	96.3	63.2	48.8	75.6
3	Montreal area	98.8	90.4	18.8	28.4	92.7	73.8	44.9	69.1
4	South of Montreal	99.4	91.0	23.6	29.2	91.6	80.9	48.9	72.5
5	Centre	99.4	90.2	13.8	8.6	96.6	87.6	51.1	52.3
6	Quebec City area	100.0	97.1	12.5	7.0	98.5	75.7	61.0	29.8
7	North-East	99.4	99.4	15.3	7.0	97.5	79.0	35.7	62.4
8	East	98.1	93.0	11.6	19.5	91.6	94.4	31.2	64.2
	Whole Quebec	99.1	92.6	17.9	21.3	94.7	78.5	47.1	61.0
Number of founders		5,326	317	143	214	97	1,056	95	298

founders' arrival time and the number of regions where they have descendants (Pearson's $r = -0.6$, P-value $< 2.2 \times 10^{-16}$).

Mosaic origins of Quebec regional populations

Nearly 90% of the regional gene pools were contributed by French founders (Table 2) and 2–6% by Acadians, who are second in numerical importance. Other groups of immigrant founders each contributed 2% or less. These proportions were similar across regions except for the East where the genetic contribution of French founders was reduced (76%) to the advantage of Acadian (11%), British (3.4%), and American founders (3.5%). Despite the elevated contribution of French founders, the subjects from our sample were nearly all admixed: on average, each genealogies contained immigrant founders from 6.75 distinct origins. While virtually all genealogies (99.1%) had at least one founder originating from France (Table 3), founders of other origins also appeared in a large proportion of the genealogies: British in 93% of genealogies, Acadian in 79% and notably, Amerindian in 47%.

East-West gradient of diversity

Within each regional sample, founders did not contribute equally. Uneven contribution of the founders is shown in Figure 3 where the cumulative proportion of the genetic contribution of founders is plotted against the cumulative proportion of contributing founders. If all founders contributed equally to a gene pool, a linear dependence would be expected, as shown by a straight line in Figure 3. In contrast, the observed dependence was not linear: a small fraction of founders explains a large proportion of the gene pool while a greater fraction contributes less. For instance, 11% of the Montreal region founders explained 50% of its gene pool, while the remaining 89% founders explain the other half. The uneven contribution of founders was more pronounced in the eastern regions of Quebec (Quebec City area, North-East and East). In the North-East, half of the gene pool was explained by 3% of the founders ($n = 86$).

The lowest homogeneity index was observed in the Montreal area region and the highest in the North-East (Table 4). In the Montreal region, the FUN calculation indicates that 1,787 equally contributing founders would provide the same level of diversity as the actual 6,317 founders. In contrast, in the North-East, only 160 equally contributing founders are required to provide the same level of genetic diversity as the actual 2,628 founders. The normalized FUN/n_f ratio for the North-East of 6% shows that six equally contributing founders are equivalent to 100 actual founders. We found a higher FUN/n_f index—about 30%—in all other regions, except for the East (18.2%) and Quebec City area (21.3%), indicating a greater homogeneity of the eastern regions of Quebec. Overall, these results show a West-East decreasing gradient of genetic diversity. The FUN/n_f ratio was measured at different time points in the genealogies of each regional sample to evaluate the progression over time of the concentration of founders' genetic contribution. We observed that the FUN/n_f ratio diminishes in the first two centuries of Quebec history and stabilizes to its contemporary level for all regions in 1800 except for the North-East which does so in 1850 (Supporting Information Fig. S3).

Differential contribution of early and late founders

Figures 4A,B illustrate the progression of the founders' genetic contribution to the ancestor layers according to their period of arrival, for the Montreal area and North-East region (Supporting Information Fig. S4 shows all regions). A greater part of the regional gene pools was contributed by the early compared to the late founders as expected in a simple model of population in expansion where every individual has the same number of descendants and a constant number of new migrants arrive every discrete generation (for a graphic illustration of the model see Supporting Information Fig. S5). All regional populations of Quebec (except for the North of Montreal) had a minimum of a third of their gene pools descending from the earliest founders, defined as those founders who married before the first major immigration wave of 1660 (Supporting Information Table S4). These founders were mostly French, represented 9.3% of all founders (Supporting Information Table S2) and 93% of them were found in at least six out of eight regions (75% common to all regions). For the North-East and Quebec City ancestors, the contribution of the earliest founders was even higher, increasing over time up to 50% and 44%, respectively. This increase is not compatible with a simple model of population expansion (Supporting Information Fig. S5). Even if we assume that the rate of migration tends to zero, the contribution of the earliest founders is expected to stabilize and not to increase. For the North-East and Quebec City area, it suggests that the earliest founders had on average a higher number of descendants than expected under the assumption of uniform reproductive success of all founders. This points to a higher reproductive success of the earliest founders and their descendants throughout the period.

A substantial fraction of regional gene pools was also explained by founders married between 1660 and 1700 (Figs. 4A,B and Supporting Information Fig. S4). These founders represented 32% of all founders (Supporting Information Table S2), were mostly French (97%) and had a lower regional representation than the earliest founders with 74% appearing in at least six regions (40% common to all eight regions). However, the fraction of the gene pool explained by the 1660–1700 founders differed across regions as it decreased to the profit either of the earliest founders (in the Quebec City area and North-East) and/or of latecomers, arrived after 1700 (in all regions). Notably, the East region displayed the highest contribution of late founders (Supporting Information Fig. S4) who are more specific and have more diversified origins (Table 2). A priori, this may suggest a higher genetic diversity in the East region, but the low FUN/n_f ratio observed in this region points to a lower genetic diversity (Table 4). The FUN/n_f ratio is calculated within regions and does not take into account the origins and specificity of the founders; therefore, we cannot exclude the possibility that the late founders brought new genetic variation in the East.

TABLE 4. Founders' uniform contribution number (FUN) and its ratio to the actual number of founders in each sample (FUN/n_f)

	Region	HI ($\times 10^4$)	FUN (=1/HI)	Number of founders (n_f)	FUN/ n_f (%)
1	North-West	8.5	1,177.1	3,724	31.6
2	North of Montreal	6.9	1,452.9	4,343	33.5
3	Montreal area	5.6	1,786.6	6,317	28.3
4	South of Montreal	6.2	1,604.5	4,384	36.6
5	Centre	7.3	1,375.5	4,790	28.7
6	Quebec City area	12.8	783.1	3,676	21.3
7	North-East	62.3	160.4	2,628	6.1
8	East	17.2	580.2	3,188	18.2
	Whole Quebec	6.8	1,461.0	7,798	18.7

Stratification of Quebec regional populations

To detect population structure using ascending genealogies, we proposed a new approach based on PCA of immigrant founders' genetic contributions to the 2,221 subjects (GC-PCA). This analysis was compared to a MDS of pairwise kinship coefficients. Both methods showed graphical evidence for stratification of the contemporary French Canadian population (see Fig. 5). North-East subjects significantly clustered together on the first axis (ANOVA P-values $< 1 \times 10^{-30}$ for all pairwise comparisons of regions on PC1) and East subjects did so on the second axis albeit to a lesser extent (ANOVA P-value $< 1 \times 10^{-11}$ for all pairwise comparisons of regions on PC2) (Fig. 5 and Supporting Information Fig. S6). Thus, patterns of

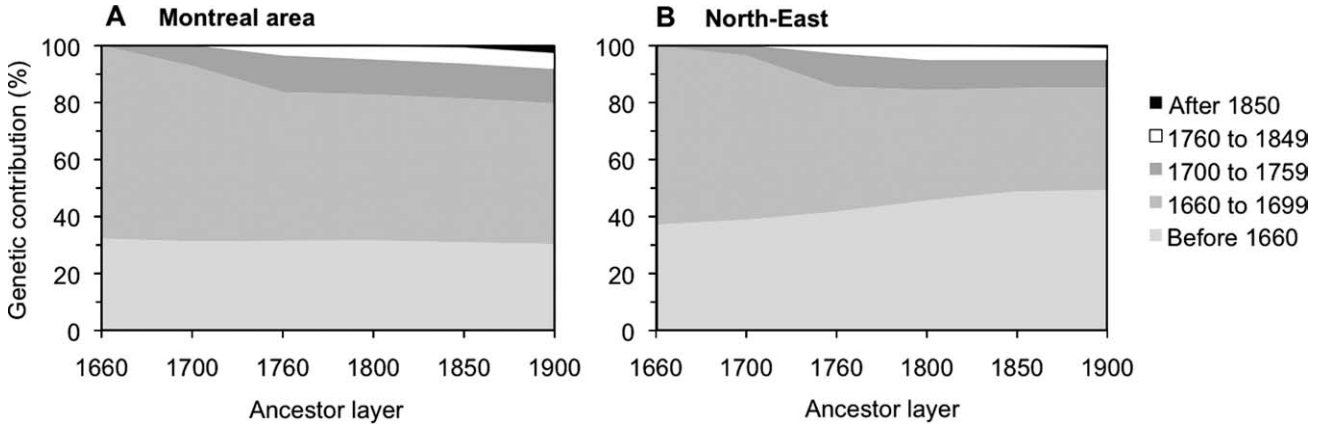

Fig. 4. Progression over time of the genetic contribution of immigrant founders. The founders' relative genetic contribution (%) to the ancestor layers is plotted according to their period of arrival for Montreal area (A) and North-East (B) regions. The ancestors in each layer were married ±15 years around the selected years.

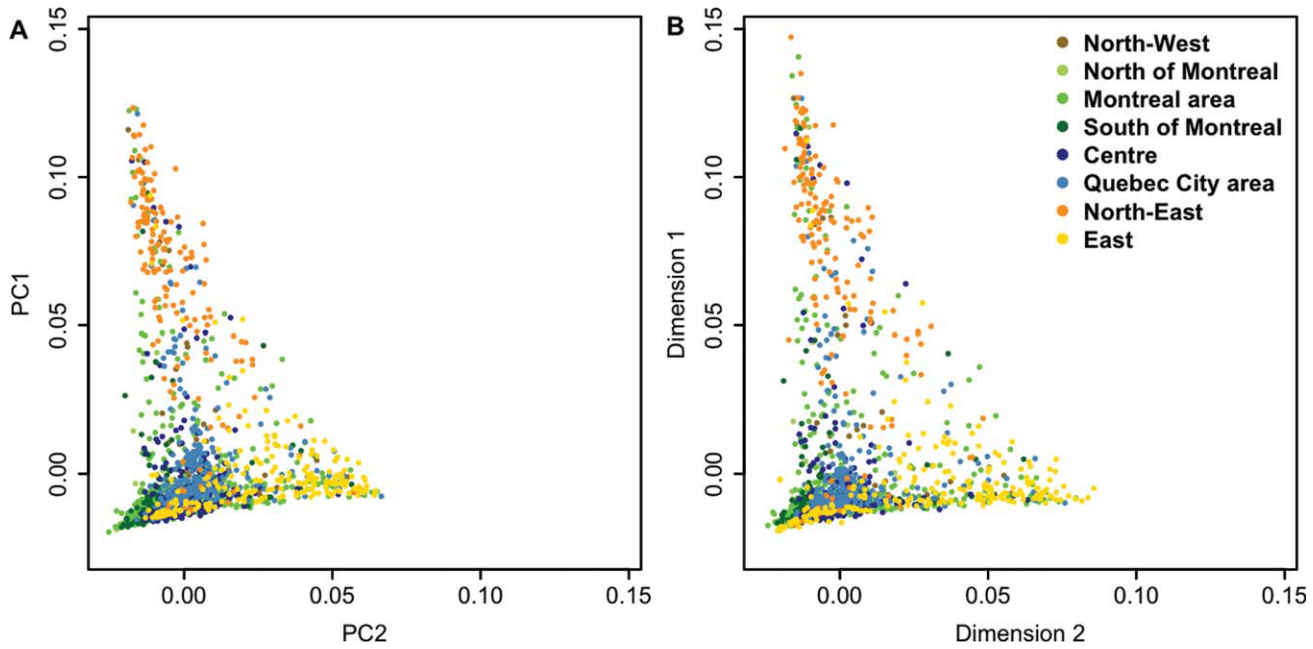

Fig. 5. Structure of Quebec regional populations. Subjects are plotted according to (**A**) the two first components of the GC-PCA (axes were rotated with the varimax method) and (**B**) the two first dimensions of the MDS based on pairwise kinship coefficients (i.e., 1-kinship). PC1 explains 6.1% of the variation and PC2 2.4%. Four outliers were excluded from the analyses (n = 2,217).

founders' genetic contribution and kinship in the North-East and East regional populations appear to be distinct from the rest of Quebec. Except for the North-West and North-East regions, subjects tended to be distributed on a West-East gradient along the second PC according to their region of marriage, reflecting their geographical location along the St-Lawrence River (Fig. 5 and Supporting Information Fig. S6). This is supported by a R^2 of 0.15 (P-value $< 2.2 \times 10^{-16}$) for the linear regression of PC2 on the subjects' region of marriage (recoded 1 to 8 from west to east). The Calboli-PCA and the K-means clustering of founder/subject incidence matrix also positioned subjects along the West-East axis but provided less information on population structure than the GC-PCA (Supporting Information Fig. S7).

Comparison between GC-PCA and MDS of kinship coefficients showed that both methods are highly correlated (PC1 and MDS1: Pearson's r = 0.99; PC2 and MDS2: Pearson's r = 0.98 - Supporting Information Fig. S8). This was expected since for two given subjects i and j, the probability of sharing an allele identical by descent from a common founder equals the sum of products of founders' genetic contribution to subject i and to j divided by two, over all founders n_f. This probability equals the kinship coefficient when the common ancestors of two subjects are not inbred and when the founders appear at the same level of generation so that they have a uniform genetic contribution to the subjects. Hence, this explains why the GC-PCA and the MDS of kinship coefficients gave very similar results. The differences between the two methods thus reflect inbreeding among common ancestors and inequality in founders' genetic contribution.

Figure 6 shows the correlation between each founder genetic contribution and the top two principal components of GC-PCA. Founders displaying a correlation coefficient with PC1 greater than 0.2 arrived for the most part before 1700 and were common to all eight regional

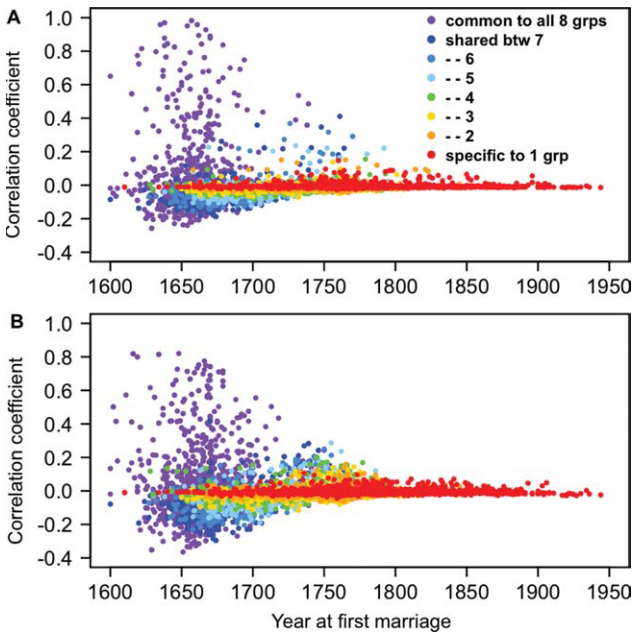

Fig. 6. Correlation between the founders' genetic contribution and (**A**) PC1 and (**B**) PC2 according to year at first marriage and regional representation. Each point corresponds to a founder and is colored according to the number of regions to which they contribute.

groupings (Fig. 6A). Moreover, founders with a high correlation to PC1 were also those who had the highest genetic contribution the North-East sample (Supporting Information Fig. S9). These results suggest that the distinction between the North-East and the other regions observed on PC1 is explained by the higher contribution

of early founders found in many regions rather than to the contribution of specific founders to the North East. Figure 6B indicates that the PC2 was also mostly influenced by early founders but on the whole the correlation was not as strong. Founders who had a greater genetic contribution to the East region tended to have a greater coefficient of correlation (Supporting Information Fig. S10). Both PC1 and PC2 were negatively correlated to founders with a higher contribution to regions located West of the Quebec City area (Supporting Information Fig. S10). Altogether, this analysis of PC-correlated founders recapitulated the results of our descriptive analysis of the immigrant founders' genetic contribution and clearly illustrated the differential impact of early versus late founders in shaping the contemporary structure of the Quebec population.

DISCUSSION

We analyzed the population structure of French Canadian from Quebec using extensive genealogical ascendance of 2,221 subjects sampled over all the territory and partitioned in eight regional groupings. We provided evidence for stratification of the French Canadian populations at the regional level and showed that this structure was correlated with their geographical location along the St-Lawrence River. In addition, we found a West-East decreasing gradient of diversity among regional populations, consistent with a previous genealogical study of kinship and consanguinity (Vézina et al., 2004). Regional populations shared a common pool of diversity contributed by the earliest founders, mostly French, but received a differential and more specific input of the latecomers, of more diverse origins. In particular, our results contrasted the regions located West of the Quebec City area from the North-East and East regions, which both show patterns of ancestry supporting their distinct genetic identity. Taken together, our results demonstrate that regional gene pools of Quebec cannot be considered homogeneous and underline the specificity of each region that can be understood in light of its settlement and subsequent demographic history.

The North-East displayed the highest homogeneity and appeared as a distinct cluster from the rest of Quebec. The southernmost part of that region, Charlevoix, was colonized at the end of the 17th century by a small number of descendants of the pioneers coming mainly from the Quebec City area (Jetté et al., 1991). In the middle of the 19th century, settlement started in the North of the region (Côte-Nord and Saguenay-Lac-St-Jean). The Saguenay-Lac-St-Jean population was founded by pioneers coming mostly (but not exclusively) from the Charlevoix region and grew rapidly due to a particularly high fertility rate in a context of relative isolation (Roy et al., 1988; Lavoie et al., 2005). Many mutations underlying Mendelian diseases, elsewhere very rare, have reached an elevated frequency in that region, thus pointing to a strong regional founder effect (Labuda et al., 1996; Scriver, 2001; Laberge et al., 2005; Yotova et al., 2005). In the ascending genealogies of patients affected by five of these diseases, it was previously shown that the 17th century founders had a high contribution, explaining nearly 80% of the cohorts' gene pool (Heyer and Tremblay, 1995). The high contributors are also the most likely to have introduced the diseases' mutations. Since three of these five diseases are specific, although not exclusive, to the Charlevoix and Saguenay-Lac-St-Jean regions, we can hypothesize that the 17th century founders with high contributions were the ones that also contributed to the genetic differentiation of the region. Here, we demonstrated that the genetic differentiation of the North-East is not explained by the input of specific founders, but by the higher contribution of a subset of earliest founders. These founders were, for the most part, common to all regions but had a higher reproductive success in the North-East. This is likely to be the result of the founder effect per se whereby the successive settlements of the three regions comprised in the North-East were biased toward descendants of the earliest settlers. This could be simply due to random sampling, but cultural explanations have also been proposed such as kin-structured migrations (Jetté et al., 1991) and social transmission of reproductive behavior (Austerlitz and Heyer, 1998).

The East region was found to be the second most homogeneous sample and was characterized by a higher contribution of later founders from Acadia, Great Britain and other parts of America. French Canadian settlement of the East, starting in Côte-du-Sud nearby Quebec City, progressively reached the Gaspe Peninsula in the 19th century, which was already occupied by descendants of the deported Acadians and of British immigrants, including Loyalists to the British Crown who took refuge on this territory after the War of Independence of the United States at the end of the 18th century (Desjardins et al., 1999). Significant genetic differentiation among self-declared descendants of the major groups of founders of the Gaspe region was found based on analysis of genome-wide diversity (Roy-Gagnon et al., unpublished results) and parental lineages (Moreau et al., 2009). The diverse ancestry of the East region, ranging from deep-rooted families in Côte-du-Sud to diverse ethno-cultural groups in the Gaspe Peninsula, was reflected in our population structure analysis, as subjects from that region are relatively dispersed even if they do form a significant cluster.

In our genealogical sample, we identified 5,623 immigrant founders who married before the British Conquest. They represent two thirds of the 8,570 settlers previously reported to have at least one married children in Nouvelle-France (Brais et al., 2007). The time of arrival and origins of these founders are found to be consistent with the composition of the pioneer immigration under the French rule (Charbonneau et al., 2000). For reasons of data availability, our sample comprised subjects married between 1945 and 1965. In the following decades, major migration movements linked to urbanization processes have taken place. In a recent study on the genealogical structure of the Lanaudiere region located in the periphery of Montreal, we showed that these movements were linked to a reduction in genetic differentiation and diversification of ancestry (Bherer et al., 2008). We therefore expect that the population structure found in this study underwent some changes. However, this effect should be more pronounced in the Montreal and Quebec City areas which are the two major poles of internal migrations.

Our results confirm the genetic importance of the French immigrants in the contemporary French Canadian gene pool: they are the most numerous founders and have the highest contribution to all regional samples. However, this study also puts in the balance arguments in favor of the heterogeneity of the pool of founders. Following the British Conquest, immigration diversified and had a variable impact on the regional populations (Bergeron et al., 2008; Tremblay et al.,

2009). Moreover, French immigrants came from all regions of France (Vézina et al., 2005) and principally landed as single member of their family (Charbonneau et al., 1993; Guillemette and Légaré, 1989). Contemporary regions of France have been shown to be genetically heterogeneous (Dubut et al., 2004; Richard et al., 2007). Assuming that this was also the case during the 17th and 18th centuries, the amalgamation of founders from different French regions together with immigrants from other European and American countries is expected to have inflated the genetic diversity introduced in Quebec.

In our sample, almost all subjects had mixed origins, including French and non-French. This indicates that admixture events have shaped the genome of nearly all French Canadians, like other post-Colombian populations of the New World. Notably, half of the subjects had at least one reported Amerindian founder in their ancestry. However, while the mestizo populations of Latin America, for example, have a mean estimated proportion of Amerindian ancestry of more than 20% (Wang et al., 2008), we found that 0.2% of the French Canadian gene pool was of Amerindian origin. Although this estimate is a lower bound because historical sources do not always allow identification of Amerindian ancestors, to our knowledge this is the first estimation of the Amerindian genetic contribution to the contemporary French Canadian gene pool based on a large genealogical sample. Overall, our study also puts forward that the French Canadians descend from a pool of founders relatively large and of diverse origins. This might be sufficient to explain why the French Canadian population of Quebec has maintained comparable levels of genetic diversity as European populations (e.g., DeBraekeleer, 1990; Moreau et al., 2007).

In this article, we showed that analysis of ascending genealogies allows population structure to be assessed without genotyping. A new approach was proposed to detect and visualize population stratification by PCA of founders' genetic contribution and was validated by its high correlation with a commonly used approach, namely the MDS of pairwise kinship coefficients. In a parallel work, we showed that the latter approach accurately mirrors the genetic structure in French Canadian pedigrees inferred from genomewide SNP data (Roy-Gagnon et al., unpublished results). By induction, GC-PCA can therefore be used to reflect genetic structure. Such PCA may not be of practical use if the only intent is to describe genetic structure of a population since MDS of pairwise kinship coefficients is more efficient and can readily be used. However, it has the advantage of pointing out which founders have the strongest weight in creating the structure. We illustrated that the PC-correlated founders, by analogy with PC-correlated markers, can be used to inform how the demographic history of a population shapes its genetic structure. Our approach could be used in historical studies to identify and further describe the highly PCA-correlated founders. Extensive genealogies are found in a number of human populations, such as the Icelanders (Helgason et al., 2005) and the Utah residents (Cannon-Albright, 2008). Deep pedigrees are also found in many other species with important economic incidence such as dogs (Calboli et al., 2008) and cattle. Our approach could also be of interest for the development of breeding strategies in captive populations aiming to maintain genetic diversity.

Regarding the Quebec population, we demonstrated that genetic epidemiology studies must take into account the characteristics of regional populations to optimize their study design. For instance, in order to minimize genetic heterogeneity, research aiming to identify rare variants implicated in complex traits might be better suited in the North-East or the East of Quebec. Stratification of Quebec regional populations highlights the need to detect and correct for genetic structure in genetic association studies, especially when sampling without regard to the geographic origin of individuals. Moreover, since population structure in Quebec is linked to geography, optimal design of studies should include the place of birth of parents and grandparents to guide the selection of individuals to sample and to genotype.

In the past decade, much effort has been devoted to analyze population structure from genotype data. In this study, we show that the analysis of deep ascending genealogies can effectively reveal population structure and therefore be a useful tool to explain the consequences of historical demographic processes in structuring of genetic variation and to develop more powerful research design in population association studies.

ACKNOWLEDGMENTS

The authors are grateful to Ève-Marie Lavoie from the Interdisciplinary Research Group on Demography and Genetic Epidemiology (Chicoutimi) and Michèle Jomphe from the BALSAC project (Chicoutimi) for technical assistance and Laurent Richard from the Historical Geography Laboratory at Laval University (Québec) for cartography work. They also thank Julie Hussin and two anonymous reviewers for their comments on an earlier version of the manuscript.

LITERATURE CITED

Austerlitz F, Heyer E. 1998. Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. Proc Natl Acad Sci USA 95:15140–15144.

Bergeron J, Vézina H, Houde L, Tremblay M. 2008. La contribution des Acadiens au peuplement des régions du Québec. Cah Quebec Demogr 37:181–204.

Bherer C, Brais B, Vézina H. 2008. Impact des récentes transformations démographiques liées à l'urbanisation sur le bassin génétique de la region de Lanaudière. Cah Quebec Demogr 37:211–235.

Bouchard G, Vézina H. 2009. Projet BALSAC - Rapport annuel 2008–2009. Chicoutimi.

Brais B, Desjardins B, Labuda D, St-Hilaire M, Tremblay M, Vézina H. 2007. The genetics of French Canadians. In: Cavalli-Sforza L, Feldman M, editors. Human population genetics: evolution and variation. London: The Biomedical & Life Sciences Collection, Henry Stewart Talks Ltd (online at available at:http://www.hstalks.com/?t=BL0251614-Desjardins).

Calboli FCF, Sampson J, Fretwell N, Balding DJ. 2008. Population structure and inbreeding from pedigree analysis of purebred dogs. Genetics 179:593–601.

Cannon-Albright LA. 2007. Utah family-based analysis: past, present and future. Hum Hered 65:209–220.

Cardon LR, Palmer LJ. 2003. Population stratification and spurious allelic association. Lancet 361:598–604.

Charbonneau H, Desjardins B, Guillemette A, Landry Y, Légaré J, Nault F. 1993. The first French Canadians. Pioneers in the St. Lawrence Valley. Newark, London and Toronto: University of Delaware Press and Associated University Presses.

Charbonneau H, Desjardins B, Légaré J, Denis H. 2000. The population of the St-Lawrence Valley, 1608–1760. In: Haines MR, Steckel RH, editors. A population history of North America. Cambridge: Cambridge University Press. p 99–142.

DeBraekeleer M. 1990. Homogénéité génétique des Canadiens français: mythe ou réalité? Cah Quebec Demogr 19:29–48.
Desjardins M, Frenette Y, Bélanger J, Hétu B. 1999. Histoire de la Gaspésie (2e édition). Sainte-Foy: Institut québécois de recherche sur la culture et Presses de l'Université Laval.
Dickinson JA. 1994. Les réfugiés acadiens au Canada, 1755–1775. Études canadiennes/Canadian Studies 37:51–61.
Dubut V, Chollet L, Murail P, Cartault F, Beraud-Colomb E, Serre M, Mogentale-Profizi N. 2004. mtDNA polymorphisms in five French groups: importance of regional sampling. Eur J Hum Genet 12:293–300.
Gagnon A, Heyer E. 2001. Fragmentation of the Quebec population genetic pool (Canada): evidence from the genetic contribution of founders per region in the 17th and 18th centuries. Am J Phys Anthropol 114:30–41.
Guillemette A, Légaré J. 1989. The influence of kinship on seventeenth-century immigration to Canada. Continuity Change 4:79–102.
Helgason A, Yngvadóttir B, Hrafnkelsson B, Gulcher J, Stefánsson K. 2005. An Icelandic example of the impact of population structure on association studies. Nat Genet 37:90–95.
Henripin J, Péron Y. 1972. The demographic transition of the Province of Quebec. In: Glass DV, Revelle R, editors. Population and social change. London: Edward Arnold. p 213–231.
Heyer E, Tremblay M. 1995. Variability of the genetic contribution of Quebec population founders associated to some deleterious genes. Am J Hum Genet 56:970–978.
Jetté R, Gauvreau D, Guérin M. 1991. Aux origines d'une région: le peuplement fondateur de Charlevoix avant 1850. In: Bouchard G, De Braekeleer M, editors. Histoire d'un génome. Québec: Presses de l'Université du Québec. p 75–106.
Karigl G. 1981. A recursive algorithm for the calculation of identity coefficients. Ann Hum Genet 45:299–305.
Laberge AM, Michaud J, Richter A, Lemyre E, Lambert M, Brais B, Mitchell GA. 2005. Population history and its impact on medical genetics in Quebec. Clin Genet 68:287–301.
Labuda M, Labuda D, Korab-Laskowska M, Cole DE, Zietkiewicz E, Weissenbach J, Popowska E, Pronicka E, Root AW, Glorieux FH. 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.
Lacy RC. 1989. Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. Zoo Biol 8:111–123.
Lavoie E-M, Tremblay M, Houde L, Vézina H. 2005. Demogenetic study of three populations within a region with strong founder effects. Community Genet 8:152–160.
Marchini J, Cardon LR, Phillips MS, Donnelly P. 2004. The effects of human population structure on large genetic association studies. Nat Genet 36:512–517.
McInnis M. 2000. The population of Canada in the nineteenth century. In: Haines MR, Steckel RH, editors. A population history of North America. Cambridge: Cambridge University Press. p 371–432.
Moreau C, Vezina H, Labuda D. 2007. Founder effects and genetic variability in Quebec. Med Sci (Paris) 23:1008–1013.
Moreau C, Vezina H, Yotova V, Hamon R, de Knijff P, Sinnett D, Labuda D. 2009. Genetic heterogeneity in regional populations of Quebec-Parental lineages in the Gaspe Peninsula. Am J Phys Anthropol 139:512–522.
Piché V. 2003. Un siècle d'immigration au Québec : de la peur à l'ouverture. In: Piché V, Le Bourdais C, editors. La démographie québécoise. Enjeux du XXIe siècle. Les Presses de l'Université de Montréal. p 225–263.
Richard C, Pennarun E, Kivisild T, Tambets K, Tolk HV, Metspalu E, Reidla M, Chevalier S, Giraudet S, Lauc LB, Pericic M, Rudan P, Claustres M, Journel H, Dorval I, Müller C, Villems R, Chaventré A, Moisan JP. 2007. An mtDNA perspective of French genetic variation. Ann Hum Biol 34:68–79.
Roberts DF. 1968. Genetic effects of population size reduction. Nature 220:1084–1088.
Roy R, Bouchard G, Declos M. 1988. La première génération de Saguenayens: provenance, apparentement, enracinement. Cah Quebec Demogr 17:113–134.
Scriver CR. 2001. Human genetics: lessons from Quebec populations. Annu Rev Genomics Hum Genet 2:69–101.
Tremblay M, Letendre M, Houde L, Vézina H. 2009. The contribution of Irish immigrants to the Quebec (Canada) gene pool: an estimation using data from deep-rooted genealogies. Eur J Popul/Revue européenne de démographie 25:215–233.
Tremblay M, Vézina H. 2000. New estimates of intergenerational time intervals for the calculation of age and origins of mutations. Am J Hum Genet 66:651–658.
Vézina H, Tremblay M, Desjardins B, Houde L. 2005. Origines et contributions génétiques des fondatrices et des fondateurs de la population québécoise. Cah Quebec Demogr 34:235–258.
Vézina H, Tremblay M, Houde L. 2004. Mesures de l'apparentement biologique au Saguenay-Lac-St-Jean (Québec. Canada) à partir de reconstitutions généalogiques. Annales de démographie historique 2:67–84.
Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM, Camrena B, Nicolini H, Klitz W, Barrantes R, Molina JA, Freimer NB, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Dipierri JE, Alfaro EL, Bailliet G, Bianchi NO, Llop E, Rothhammer F, Excoffier L, and Ruiz-Linares A. 2008. Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet 4:e1000037.
Yotova V, Labuda D, Zietkiewicz E, Gehl D, Lovell A, Lefebvre JF, Bourgeois S, Lemieux-Blanchard E, Labuda M, Vézina H, Houde L, Tremblay M, Toupance B, Heyer E, Hudson TJ, Laberge C. 2005. Anatomy of a founder effect: myotonic dystrophy in northeastern Quebec. Hum Genet 117:177–187.