Intergenerational and Genealogical Approaches for the Study of Longevity in the Saguenay-Lac-St-Jean Population

Louis Houde · Marc Tremblay · Hélène Vézina

Abstract The mechanisms of longevity have been the subject of investigations for a number of years. Although the role of genetic factors is generally acknowledged, important questions persist regarding the relative impact of environmental exposures, lifestyle characteristics, and genes. The BALSAC population register offers a unique opportunity to study longevity from an intergenerational and genealogical point of view. Individuals from the Saguenay-Lac-St-Jean population who died at age 90 or older between 1950 and 1974 were selected from this database (*n*=576), along with a control group of individuals born in the same period who died between 50 and 75 years of age. For these subjects and controls, spouses' ages at death and parental ages at death and at their birth were investigated using regression analysis. Genealogical reconstructions were carried out for each individual, and various analyses were performed on both groups. Both fathers' and mothers' mean ages at death were significantly higher among the longer-lived cases than among controls whereas spouses' ages at death and parental ages at birth had no effect. Regression analysis confirmed the positive effect of both fathers' and mothers' age at death. Mean kinship coefficients for the parents' generations displayed significant differences, indicating that kinship was higher among subjects than controls (this effect was stronger among the oldest 10% of the subjects). Frequencies and genetic contributions of ancestors were very similar for the two groups, and none of these ancestors appeared more likely to have introduced genetic variants involved in longevity patterns in this French Canadian population.

Keywords Longevity · Intergenerational analysis · Genealogies · Quebec population · Kinship · Familial reconstructions · Inheritance

L. Houde · M. Tremblay · H. Vézina (✉)
Interdisciplinary Research Group on Demography and Genetic Epidemiology (GRIG),
Université du Québec à Chicoutimi, Québec, Canada
e-mail: hvezina@uqac.ca

The mechanisms underlying the variability in longevity among human beings have been the subject of study for a number of years. Although the role of genetic factors is generally recognized by researchers, important questions persist as to the respective impact of environmental exposures, lifestyle characteristics, and genes. Ongoing research is directed toward identifying and mapping the specific variants associated with aging processes and longevity phenotypes and understanding the function and modality of expression of the implicated genes (Johnson and Shook 1997; Vijg and Suh 2005).

The Familial Component of Longevity

Studies of the familial aggregation and intergenerational transmission of lifespans have been conducted for decades (see for a review Cournil and Kirkwood 2001; Gavrilov et al. 2002). The choice of the studied populations has been motivated by various factors, including exceptional longevity, characteristics of founder effects, or environmental and social homogeneity allowing control for confounding effects, and the availability of population registers (Blackburn et al. 2004; Cournil et al. 2000; Gavrilov and Gavrilova 2001; Gudmundsson et al. 2000; Ikeda et al. 2006; Kemkes-Grottenthaler 2004; Kerber et al. 2001; Mitchell et al. 2001; Pettay et al. 2005; Poulain and Naito 2004; Westendorp and Kirkwood 2001). Most of these studies have found a positive association between lifespan and parental or sibling age at death, although it remains difficult to disentangle genetic and environmental components of observed correlations.

Some authors have also examined the extent to which longevity might be transmitted differentially between parents and offspring of either sex, but the reported patterns are inconsistent. Among those who have detected sex-linked inheritance, mothers have a stronger effect in some studies (Blackburn et al. 2004; Kemkes-Grottenthaler 2004; Westendorp and Kirkwood 2001) and fathers are more important in others (Gavrilov and Gavrilova 2001; Landgren et al. 2005). Moreover, some studies find a more important heritability effect on daughters than on sons (Blackburn et al. 2004; Cournil et al. 2000; Mitchell et al. 2001; Westendorp and Kirkwood 2001) and others do not (Gavrilov and Gavrilova 2001; Ohta et al. 2004). Parental age at conception or birth of offspring was also suggested as a potential factor influencing offspring longevity. A negative correlation has been observed for father's age at birth of daughters (Gavrilov and Gavrilova 2001; Kemkes-Grottenthaler 2004), but other studies did not detect any association (Robine et al. 2003; Westendorp and Kirkwood 2001).

According to a review of twin and family studies carried out by Cournil and Kirkwood (2001), estimates for the heritability coefficient of longevity are uniformly low, rarely exceeding 30%. However, Gavrilov and Gavrilova (2001) argue that heritability could have been underestimated in numerous studies because of the existence of an age threshold above which familial transmission of longevity would have a more important effect. This threshold or acceleration effect is

supported by other studies (Cournil et al. 2000; Westendorp and Kirkwood 2001; Blackburn et al. 2004). Other approaches have tested specific models to evaluate the relationship between the lifespan of subjects and various demographic characteristics of their close relatives and spouses (Gudmundsson et al. 2000; Kerber et al. 2001). In their study on the Icelandic population, Gudmundsson and colleagues (2000) found that subjects who died at an advanced age (i.e., above the 95th percentile) were more closely related to each other than controls were and that their first-degree relatives were almost twice as likely to reach an old age compared with controls' first-degree relatives.

The BALSAC Population Database and Genealogical Studies in the Quebec Population

The BALSAC population database, developed in the past thirty years, has been used in many projects on the genetics and epidemiology of various monogenic and complex traits. These genealogical studies aim at understanding and explaining the role of demographic dynamics and population history in the introduction and spread of mutant alleles. Until the early 1990s, studies were mostly concerned with monogenic disorders. The founder effect along with genetic drift and remote consanguinity have been shown to be key factors in explaining the observed frequencies of several inherited disorders in the Quebec population (for a review see Vézina 1996). More recently, genetic demography analyses have also been applied to the study of complex traits such as Alzheimer disease (Vézina et al. 1999), bipolar disorders (Morissette et al. 1999), and hypertension (Hamet et al. 2005; Pausova et al. 2002).

The French Canadian population conveys many advantages for molecular and epidemiologic genetic studies, not only because of the availability of demohistorical data but also because the settlement history is characterized by a strong founder effect. Approximately six million French Canadians live in the province of Quebec. They are descendants of roughly ten thousand immigrants who came mostly from France and settled in "Nouvelle-France" between 1608 and 1760 (Charbonneau et al. 2000). After the British conquest, the French Canadian population expanded in a context of relative isolation and rapid natural growth, which amplified the genetic impact of the founder effect. Some consequences of this founder effect are observed in the presence of otherwise rare disorders found at an elevated frequency in the Quebec population (Laberge et al. 2005; Scriver 2001).

In this study, we propose to investigate the presence of intergenerational effects and shared genealogical characteristics among individuals in the Saguenay-Lac-St-Jean (SLSJ) region of Quebec (Fig. 1), using data from the BALSAC population register. We hypothesize that because of the genealogical structure of the French Canadian population of Quebec, individuals sharing genetic factors contributing to their longevity should also share genealogical characteristics, and that parental variables influencing age at death may also be detectable in the genealogical data.
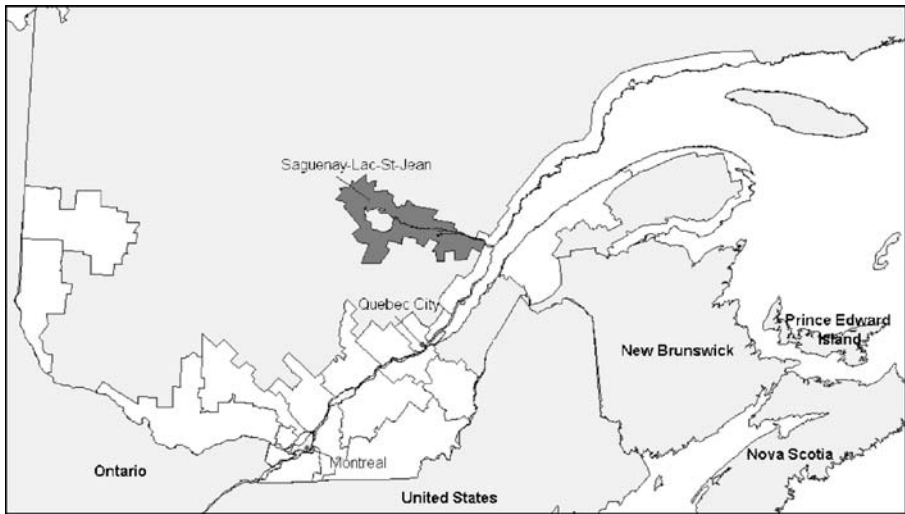
Fig. 1 Geographical location of the Saguenay-Lac-St-Jean region

Data and Methods

Selection of Cases and Familial Data

Data used in this study were retrieved from the BALSAC population register. The BALSAC register contains information on 2.1 million marriages, births, and deaths during the past two centuries in the province of Quebec (Bouchard 2005). The SLSJ region has been completely covered from the beginning of settlement in the 1830s to 1971 (up to 1986 for some death certificates). It is the only Quebec region for which all civil records have been computerized and linked. For the other regions of Quebec, only marriage records are included in the database.

Information was available for 127,294 individuals who died between 1850 and 1986. Among these, 1,439 (1.1%) died at 90 years of age or older. For the purpose of this study, we selected the 576 individuals in the database who died between 1950 and 1974 at age 90 or older. This minimal age at death was chosen as the oldest age at death in the population under study (99th percentile) for which a sufficient number of cases would be obtained to produce significant results. Since our goal was to determine whether the genealogical characteristics of these subjects were related to their longevity or merely depicted the underlying population structure, we selected 576 controls who were born in the same period as the subjects (1850–1884) but who died at between 50 and 75 years of age. This age-at-death interval for controls was chosen to minimize the effects of heterogeneity and to clearly distinguish old ages at death (90+) from intermediate ages at death (50–75). Moreover, individuals born in the same period grew up and lived in the same context in terms of broad natural, occupational, and social environments; therefore, this selection process allows for control over risk factors associated with that context (epidemics, type of health care system and access to that system, available therapy and medicine).

Dates of birth and death as well as age at death of subjects and controls, of their parents, and of their spouses (when they married) were retrieved from the database. Information on these vital statistics was available only for those events that were recorded in the SLSJ region. Age at death was calculated from dates of birth and death whenever possible. When date of birth was missing, the age at death declared on the death record was used.

Statistical Analyses of Intergenerational Effects

Mean age at death of mothers (MAD), fathers (FAD), and spouses (SAD) as well as mean age of parents at the birth of subjects and controls (FAB and MAB) were calculated, and statistical significance of differences between subjects and controls was tested using a Student's t test. To test the simultaneous effect of these variables, a logistic regression model was performed on the variable GRP (subject=0 and control=1). The relation between age at death (AD) for subjects and controls and the different parental variables was also investigated using a regression model. Two related models were considered, a general one to obtain the transformation needed for each variable and a polynomial model to estimate the effect of each parental variable on the age at death for each group. The first model is a generalized additive model (Hastie and Tibshirani 1990):

$$E(Y|\mathbf{x}) = \alpha + f_1(x_1) + \cdots f_k(x_k) + e \tag{1}$$

where Y is the variable AD, $\mathbf{x}$ the vector of parental variables, and f_i are general functions, each corresponding to a parental variable. Estimation for the nonparametric functions in (1) was obtained with a spline estimator using S-Plus (Insightful Corporation 2006). At this step, only those individuals for which data for all the variables were available were considered for the estimation of the functional relation (315 subjects and 311 controls). This approach is appropriate for an exploratory analysis because very few assumptions are made regarding the "general form" of the relation. The drawback, however, is the lack of power owing to the relatively small sample and the inability to use all the information available for each individual. After deriving the transformation f_i for each variable, the groups were integrated in the parametric model

$$E\left(Y|\mathbf{x}\colon\text{GRP}\right) = \begin{cases} \alpha_0 + f_{01}(x_1) + \cdots f_{0k}(x_k) + e & \text{if GRP} = 0 \\ \alpha_1 + f_{11}(x_1) + \cdots f_{1k}(x_k) + e & \text{if GRP} = 1 \end{cases} \tag{2}$$

where $f_{ji}(x_i)$ are polynomial functions with degree determined by the generalized additive model (1). In order to estimate parameters and to test for a group effect, equations (2) were integrated into a single equation with GRP as a dummy variable:

$$E\left(Y|\mathbf{x}\colon\text{GRP}\right) = \beta_0 + \beta_1 * \text{GRP} + \beta_2 x_1 + \beta_2 x_1 * \text{GRP} + \beta_3 x_1^2 + \beta_2 x_1^2 * \text{GRP} + \cdots$$

In this expression, each coefficient β_i associated with the dummy variable GRP tests for a specific group effect. A stepwise linear regression for all the variables in the model was carried out using S-Plus. Because FAD and MAD contain few missing values compared with FAB and MAB, the stepwise procedure (forward) was

modified to take into account all cases without missing values for all the variables in a specific model at each step in the procedure.

Genealogical Reconstruction

Reconstruction of ascending genealogies was attempted for all selected subjects and controls. Parents of 9 subjects and 10 controls could not be identified in the available sources, so these individuals were excluded from the study. For the remaining 567 subjects and 566 controls, genealogical reconstruction relied on the BALSAC population register and on the BALSAC-RETRO database which was developed in the course of various research projects making use of genealogical reconstructions (Jomphe and Casgrain 1997) and which currently contains linked genealogical information on nearly 360,000 individuals married in Quebec between the beginning of the seventeenth century and the present day (Bouchard 2005). Other sources, such as the Population Register of Early Quebec (PRDH 2006), genealogical dictionaries, marriage repertories, or microfilms of parish registers, were also consulted. Reconstruction was performed as far back as sources allowed, and in most instances, lineages were traced back to the first immigrants to Canada. Not only do these lineages include all branches in each individual tree, all genealogical links between the individuals (subjects and controls) were identified. In other words, when a given ancestor appears in more than one genealogy, the genealogical paths linking each of his or her descendants are known. After completion, the genealogies were subjected to various validation and consistency checks to minimize the occurrence of false links owing, for example, to illegitimacy or adoptions (Bouchard 1986).

Genealogical Descriptive Parameters

Genealogical analyses were performed using procedures developed by the BALSAC Project team and by the Interdisciplinary Research Group on demography and genetic epidemiology (GRIG). Descriptive parameters include the completeness index, which identifies, for each generation, the proportion of ancestors that were successfully identified among all expected ancestors (theoretically, within a genealogy, at a given generation x, 2^x ancestors can be found) (Jetté 1991). This index is also used to calculate the mean generation level at which the lineages in the genealogies come to an end (average genealogical depth). The genealogical depth can be interpreted as the mean value of the generation of the founders in a genealogical corpus.

Kinship and Inbreeding Measurements

Two individuals are biologically related if they have at least one common ancestor. In genetic terms, this means that these individuals will have a greater than zero probability of sharing identical copies of a gene coming from that ancestor. Hence, genealogical reconstructions allow for an estimate of the intensity of biological kinship among a group of individuals which depends on the number of common ancestors identified in their genealogies and on the genealogical distances (i.e., number of generations) between these ancestors and the individuals. The kinship coefficient between two individuals can be defined as the probability that one allele

(chosen at random) from the first individual is identical by descent to another allele, at the same locus, from the second individual (Thompson 1986). The mean kinship coefficient for a group of individuals is calculated by dividing the sum of all coefficients by the total number of coefficients—that is, the total number of possible pairs $[n(n-1)/2]$.

The inbreeding coefficient is the probability that an individual has inherited, at a given locus, two identical alleles originating from a common ancestor, one from his father and the other from his mother. The inbreeding coefficient of an individual corresponds to the kinship coefficient of his parents. Since each individual has his or her own inbreeding coefficient, the mean coefficient of the group is obtained by adding all coefficients and dividing by the total number of individuals in the group.

Kinship and inbreeding values strongly depend on the number of generations that is taken into account for the calculations. As the generation level increases, the probability of finding common ancestors rises if all ancestors can be identified. Mean kinship and inbreeding coefficients were thus calculated at each generation level, from the first generation (parents) to the twelfth generation (after which the coefficients stop increasing since very few new ancestors can be traced). Statistical comparisons of mean kinship coefficients were performed using the statistic proposed by Hauck and Martin (1984). *P*-values were obtained using the bootstrap method (Efron and Tibshirani 1993) with 5,000 repetitions. Because each subject has its own inbreeding coefficient, dependence is not an issue and the classic *t*-test was used to compare mean inbreeding coefficients.

Distribution and Genetic Contribution of Ancestors

For each of the 28,893 ancestors found in the genealogies, the number of subjects and controls to which they were related, the total number of times they appeared in the genealogies of each group (many ancestors appear more than once in a single genealogy), and their genetic contribution to the subjects and control groups were calculated. The genetic contribution (GC) of an ancestor to a group of subjects is calculated as follows:

$$GC = \sum_{S} \sum_{P} (1/2)^g$$

where S is the set of all subjects genealogically linked with the ancestor, P is the set of all genealogical paths between the ancestor and the subject, and g is the number of generations, in each path, between the ancestor and the subject. Ancestors with the highest genetic contributions to a group of subjects have the highest probabilities of having transmitted their genes to these subjects.

In line with our hypothesis that shared genetic factors among subjects should correspond to shared ancestors, we attempted to identify those ancestors with a high frequency among subjects in terms of number of appearances or genetic contribution. Based on our knowledge of the population structure, these ancestors should be more specific to the subjects but not exclusive to them. Hence, all ancestors who appeared in more than 350 genealogies among the subjects or controls (roughly 60% of the genealogies of either group) were selected, and their number of appearances and genetic contributions were compared.

Results

Mean ages at death and sex distributions of subjects and controls are shown in Table 1. Sex distributions show a higher sex ratio among controls (1.19) than among subjects (1.02). This can be explained by various factors, such as age differentials in migration and mortality between the two sexes (higher male mortality between the ages of 50 and 75). The subjects' highest age at death is 105 years, but most subjects died before 95 (80% of males and 76% of females). The average ages at death are 93.4 and 93.0, respectively, for the female and male subjects. Unsurprisingly, female and male controls have the same average age at death (64.9 years), which is approximately 28 years lower than the subjects'.

Age at death of mothers and fathers of subjects and controls are also compared in Table 1. For both parents and for both sexes, the average age at death is about four years higher among subjects than among controls. Nearly 41% of the subjects' mothers died at 80 years or older, compared with 30% for the controls. Among fathers, these proportions are even higher: 45% for the subjects and 34% for the controls. As shown in Table 1, these differences between subjects and controls are significant for both parents and for both male and female individuals.

Differences between subjects and controls are much smaller in the case of spouse's age at death, and they are not significant. The main difference here is between males and females: for both subjects and controls, mean age at death of spouses is about five to six years higher for females than for males. This result can be explained in part by the fact that age at marriage is higher for males than females; thus the male spouses have an advantage of a few years from the starting point of observation. Another factor to consider is the female mortality risk after giving birth, which was still relatively high at the end of the nineteenth century and the beginning of the twentieth century, when most spouses of male subjects and controls were in their highest fertility period (Gauvreau and Bourque 1990).

Parental age at birth of subjects and controls was also compared (Table 2). In both cases (maternal and paternal age), the results are very similar (with differences of less than 0.4 years in each case). These differences are very small and do not reach statistical significance.

Lastly, logistic regression confirmed these results: FAD and MAD explain the group variable ($p<0.001$) while SAD, FAB, and MAB are not in the model ($p>0.50$).

Table 1 Mean age at death of subjects and controls, and of their mothers, fathers, and spouses

	Males		Females		Both Sexes	
	Subjects (n=286)	Controls (n=308)	Subjects (n=281)	Controls (n=258)	Subjects (n=567)	Controls (n=566)
Individuals	93.0±2.3	64.9±6.7	93.4±2.7	64.9±7.1	93.2±2.5	64.9±6.9
Fathers	77.4±13.2**	72.6±14.1	75.4±13.8*	72.5±13.7	76.4±13.5**	72.6±13.9
Mothers	72.6±15.6*	68.5±16.7	74.1±16.2*	69.9±17.1	73.4±15.9**	69.1±16.8
Spouses	65.9±18.9	64.5±19.3	70.4±15.4	70.6±14.4	68.1±17.4	67.3±17.5

Student *t*-test between subjects and controls: *$p<0.05$, **$p<0.001$

Table 2 Mean maternal and paternal ages at birth of subjects and controls

	Males		Females		Both Sexes	
	Subjects (n=286)	Controls (n=308)	Subjects (n=281)	Controls (n=258)	Subjects (n=567)	Controls (n=566)
Fathers	34.7±8.2	34.5±8.4	34.6±7.6	34.2±8.3	34.6±7.9	34.3±8.3
Mothers	30.4±7.0	30.0±7.6	29.4±7.1	29.0±7.0	29.9±7.1	29.6±7.3

Student t-test between subjects and controls: $p>0.05$ for all comparisons

Intergenerational Effects

A first look at the relation between MAD, FAD, and SAD indicates that the functions for low values (ancestors' ages at death below 50) are not consistent with other values of the variables. Therefore, all values lower than 50 for the three age-of-death variables were not taken into account in the regression. The results of the generalized additive model in terms of marginal effect are presented in Fig. 2. The marginal effect for the variable i is the nonparametric estimation of the function f_i in Eq. (1). For example, the y values for f_1 represent the part of AD that can be explained by FAD considering that all other variables are in the equation. Confidence intervals around each function help interpret the general form of the functions. The relation between FAD and SAD and AD clearly seems linear, whereas the other relations may be of a cubic order.
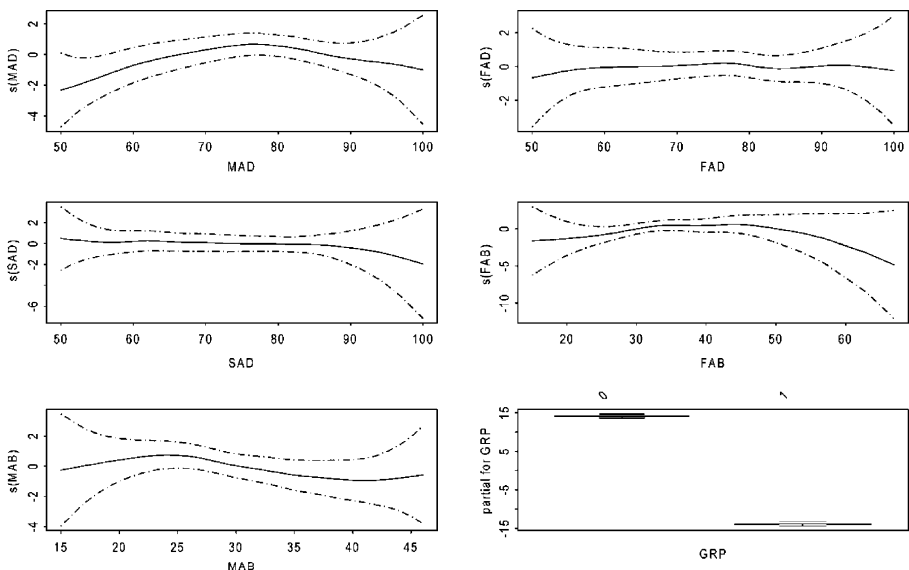

Fig. 2 Marginal relations given by the generalized linear model between AD (age at death) and MAD (mothers' age at death), FAD (fathers' age at death), SAD (spouses' age at death), FAB (fathers' age at birth) and MAB (mothers' age at birth). The bottom right graph gives the partial plot for the GRP effect (subject=0 and control=1)

Table 3 Results of the regression analysis on familial characteristics of subjects and controls

Variables	Order	Coefficient	p value
Group		−73.3028	0.00
Mother's age at death	First	−0.0462	0.86
	Second	0.0003	0.02
Group * Mother's age at death	First	1.1288	0.68
	Second	−0.0077	0.02
Mother's age at birth	First	0.35	0.49
	Second	−0.006	0.11
Father's age at death	First	0.0066	0.06
Group * Father's age at death	First	0.0605	0.13

The final model with the stepwise regression includes GRP, MAD, MAB, and FAD (in this order), with second-order terms only for MAD and MAB. The resulting regression equation is given by

$$AD = \beta_0 + \beta_1 GRP + \beta_2 MAD + \beta_3 MAD^2 + \beta_4 MAD * GRP + \beta_5 MAD^2 * GRP$$

$$+ \beta_6 MAB + \beta_7 MAB^2 + \beta_8 FAD + \beta_9 FAD * GRP$$

Coefficients and tests are summarized in Table 3. These results indicate that MAD, MAB, and FAD are predictors of AD, with MAD being the only factor reaching statistical significance in the model ($p=0.02$). They also point to the fact that subjects and controls behave differently in terms of the relation between AD and MAD ($p=0.02$). FAD has an impact on AD ($p=0.06$) but the difference between controls and subjects is not significant ($p=0.13$). The influence of MAB is not clear ($p=0.11$), and this could be explained by the loss of power following a nearly 50% cut in the sample size when this variable is included in the model.

The effects of MAD and MAB in the equation are not easy to interpret because there is a quadratic effect and because the dummy variable GRP interacts with MAD. The GRP value for the subjects is 0, so the resulting relation for subjects is:

$$AD = 79.354 - 0.0462 MAD + 0.0003 MAD^2 + 0.35 MAB - 0.006 MAB^2$$

$$+ 0.0066 FAD$$

and the resulting relation for the controls (GRP=1) is:

$$AD = 6.0512 + 1.0826 MAD - 0.0074 MAD^2 + 0.35 MAB - 0.006 MAB^2$$

$$+ 0.0671 FAD$$

The quadratic functions for MAD (controls and subjects) and MAB are presented in Fig. 3. For the subjects, the marginal effect of MAD is very small. Among controls there is a positive relation for values of MAD lower than 75, but the relation becomes negative for values greater than 80. This is explained by the fact that since controls' AD values are, by definition, below 80, the resulting function becomes asymptotic and high values of MAD can only be associated with reduced marginal effects. For MAB among subjects and controls, marginal effect increases from 15 to 30 years and decreases afterwards.
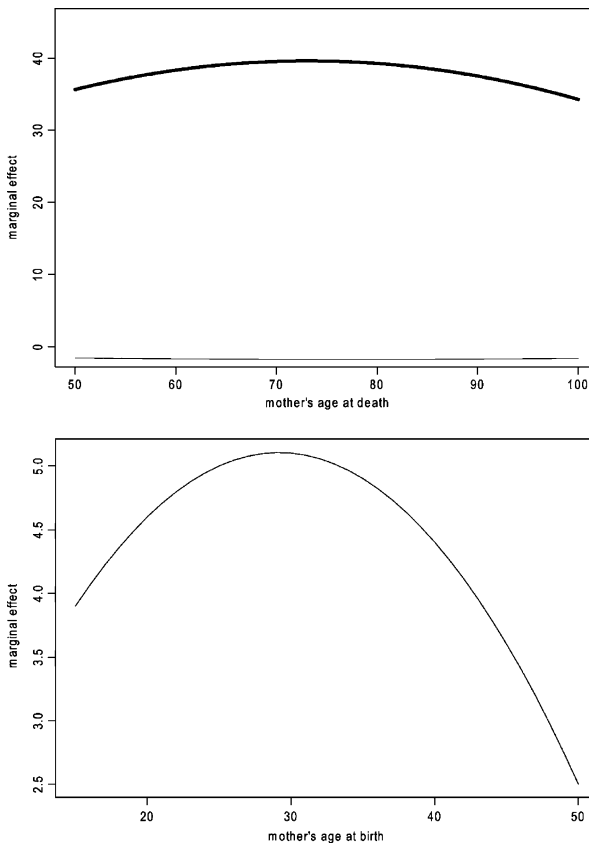
Fig. 3 Marginal effect for MAD and MAB. On the top graph, MAD effect among controls is represented by a thick line and MAD effect among subjects is represented by a solid line. The bottom graph presents MAB effect among controls and subjects

Genealogical Descriptive Parameters

Table 4 describes the main characteristics of the subjects' and controls' genealogies. A subset of the subject group, defined as the 10% "oldest" subjects and corresponding in our sample to individuals who died at age 97 or older, was analyzed separately. Genealogical reconstructions reached as far back as 13 generations, back to the beginning of the seventeenth century. On average, genealogical paths have a depth of 7.9 (subjects), 7.5 (10% oldest subjects), and 8.0 (controls) generations. Up to the sixth generation, more than 90% of ancestors are known. The completeness of the genealogies starts to drop more rapidly after the eighth generation, and after the eleventh generation, most of the genealogical branches have come to an end.

   Although the total numbers of ancestors identified in the genealogies are quite high, many of these ancestors appear more than once; hence the numbers of distinct ancestors are much lower. On average, ancestors appear 20.8 times in the subjects' genealogies, 5.3 times in the 10% oldest subjects' genealogies, and 20.9 times in the

Table 4 Characteristics of the genealogies

	Subjects	Controls	Oldest subjects
Number of genealogies	567	566	57
Mean genealogical depth	7.9	8.0	7.5
Maximal genealogical depth	13	13	12
Total number of ancestors	439,612	450,266	35,115
Number of distinct ancestors	21,125	21,581	6,636
Mean number of appearances	20.8	20.9	5.3

controls' genealogies. The average number of appearances is strongly linked to the total number of genealogies, but it is also influenced by the stratification of the SLSJ population over time. This stratification can be better described and explained with the use of kinship and inbreeding coefficients, as well as the measurement of the genetic contribution of ancestral founders.

Kinship and Inbreeding Measurements

Table 5 shows the values of the mean kinship and inbreeding coefficients at each generation level. Within both subject and control groups, the mean kinship coefficient is relatively low until the fourth generation. Much of the increase occurs between the sixth and ninth generations. These generations coincide with the arrival of many of the early French pioneers in Quebec. At the eleventh generation, kinship coefficients reach their maximum values of approximately 0.0068, 0.0055, and 0.0070, respectively, for the subjects, the 10% oldest subjects, and the controls. The leveling of the kinship curves is from the lack of genealogical information at these generational depths (marriages outside the Quebec population).

Table 5 Mean kinship and inbreeding coefficients by generation for subjects and controls

Generation*	Mean inbreeding coefficient (× 10,000)			Mean kinship coefficient (× 10,000)		
	Subjects	Controls	10% Oldest Subjects	Subjects	Controls	10% Oldest Subjects
1	0.00	0.00	0.00	0.73	0.45	3.13
2	0.00	0.00	0.00	1.93	1.45	5.09
3	2.21	0.00	0.00	3.91	3.39	7.20
4	11.71	11.87	2.74	7.59	7.28	10.71
5	23.18	23.02	13.71	13.96	14.08	16.00
6	46.49	44.93	40.95	27.55	27.97	28.35
7	75.07	72.24	65.98	48.71	49.28	44.20
8	93.92	90.87	78.13	64.43	65.44	52.94
9	98.08	95.10	80.13	67.96	69.38	54.47
10	98.52	95.55	80.36	68.38	69.86	54.63
11	98.54	95.58	80.38	68.40	69.89	54.64
12	98.54	95.58	80.38	68.40	69.89	54.64

*Generation 1=parents of subjects and controls

At the lower generation levels (<5), subjects' coefficients are higher than controls', indicating that long-lived individuals in our sample are more closely related among themselves than are controls, and that this elevated kinship is explained by shared ancestors found at the level of parents, grandparents, and great-grandparents. These differences are significant for generations 1, 2, and 3 ($p=0.021$, 0.004, and 0.04) for the subjects, and for generations 1–4 for the oldest 10% of the subjects ($p=0.016$, 0.006, 0.013, 0.044). Subjects' mean kinship coefficients are not significantly higher than controls' after the fifth generation ($p>0.23$).

Mean inbreeding coefficients for subjects and controls are very similar. The evolution of coefficients is comparable to that of kinship with, in this case, a strong increase of inbreeding values between the fifth and eighth generations. At the maximum level, inbreeding coefficients reach 0.0099 among subjects, 0.0080 among the 10% oldest subjects, and 0.0096 among controls. Differences between inbreeding coefficients of each group are not significant, even for the lowest generations ($p>0.26$).

Distribution and Genetic Contribution of Ancestors

Table 6 displays the distribution of ancestors in terms of their specificity (found in one of the two groups or common to both groups) and the number of subjects or controls to which they are related—that is, the number of genealogies in which they appear. For both subjects and controls, nearly half of all ancestors appear in only one genealogy. This proportion is, as expected, much higher for specific ancestors (80%) than for common ancestors (27%). In fact, no specific ancestor appears in more than 49 genealogies (less than 10% of the genealogies). More common ancestors (64%) appear in 2 to 49 genealogies, but a few of them are found in 450 or more genealogies (80% or more) of both groups, and up to 529 control genealogies for 6 ancestors. Again, the distributions of subjects' and controls' ancestors according to their specificity and to the number of genealogies in which they appear are very similar.

All ancestors who appeared in more than 350 genealogies in at least one of the two groups were selected and their genetic contribution to subjects and controls was compared (results not shown). One hundred and fourteen ancestors were thus

Table 6 Distribution of ancestors according to their specificity and to the number of subjects' and controls' genealogies where they appear

Number of Genealogies	Subjects				Controls			
	Specific Ancestors		Common Ancestors		Specific Ancestors		Common Ancestors	
	n	%	n	%	n	%	n	%
1	6,818	80.73	3,470	27.37	7,260	81.56	3,367	26.55
2–49	1,627	19.27	8,087	63.78	1,641	18.44	8,194	64.62
50–249	—	—	950	7.49	—	—	939	7.41
250–449	—	—	122	0.96	—	—	121	0.95
450–529	—	—	51	0.40	—	—	59	0.47
Total	8,445	100.00	12,680	100.00	8,901	100.00	12,680	100.00

identified in the subjects' group and 113 in the controls'; 112 of these ancestors were in fact the same in both groups, and no significant difference in their genetic contribution was found. No ancestors were found to be more specific to the 10% oldest subjects, either.

Discussion

The comparison of familial characteristics yielded interesting results. Average ages at death of mothers and fathers of subjects who died at 90+ years were significantly higher than those of the control group (individuals who died between ages 50 and 75). These results are consistent with previous studies on the effects of parental characteristics on the longevity of their children (Blackburn et al. 2004; Cournil et al. 2000; Gavrilov and Gavrilova 2001; Gudmundsson et al. 2000; Kemkes-Grottenthaler 2004; Mitchell et al. 2001; Westendorp and Kirkwood 2001).

In the long-lived group, parental age at death was elevated among children of both sexes and therefore no sex-linked inheritance pattern could be detected. There was no difference in our data between subjects' age at death and their spouses'. This can be interpreted as a weaker role for shared environmental factors and life habits pertaining to the adult stage of the life course among familial factors contributing to longevity (Gudmundsson et al. 2000). Lastly, in line with some other studies, no effect of parental age at birth on lifespan could be identified (Westendorp and Kirkwood 2001; Robine et al. 2003).

Logistic and stepwise linear regression analysis showed that, among all variables, mothers' mean age at death and fathers' mean age at death had the strongest effect. These two variables were found to significantly discriminate between subjects and controls. The effect of mothers' mean age at death displayed a different pattern in the two groups. Fathers' mean age at death also had a different effect, but the observed relation was not significant. Differential effects of mothers and fathers' mean ages at death remain a controversial issue since conflicting results were obtained in previous studies (Gavrilov and Gavrilova 2001; Westendorp and Kirkwood 2001; Blackburn et al. 2004; Kemkes-Grottenthaler 2004; Landgren et al. 2005).

Genealogical analysis revealed that subjects were significantly more closely related than controls, and this was explained by the sharing of ancestors at more recent generation levels (<5). This effect was more striking when a separate analysis was performed on the 10% oldest subjects. On the other hand, when ancestors found at remote generation levels were taken into account in calculations, kinship and inbreeding coefficients were similar for subjects and controls. Moreover, among ancestors with a high genetic contribution, none had a contribution that could be considered distinct enough to suggest a preferential transmission of genes to subjects. These observations point to the fact that the founder effect process and its consequences do not seem to have impacted the distribution of genetic variants involved in longevity in the SLSJ population.

These results suggest that a potential genetic component in longevity could be attributable to a combination of variants with minor or moderate effects which were most likely introduced by many founders. In this context, it is understandable that the patterns of introduction and diffusion of these variants did not leave specific

traces in the genealogies of subjects and that no differences with controls' genealogies could be detected. On the other hand, when only close common ancestors were taken into account, the level of kinship found among subjects was slightly elevated relative to that of controls, suggesting that in more recent generations (<5) transmission paths of this combination of genetic variants by specific ancestors could be perceived in the genealogical structure. However, with close kinship links, the role of shared environment and life habits must also be taken into consideration.

From an evolutionary standpoint, it is difficult to integrate natural selection effects in models attempting to explain the familial component of longevity (Allen et al. 2005). The grandmother hypothesis (Hawkes 2003) offers an interesting framework to account for postreproductive survival but is of limited interest to explain the observed variability in longevity after the reproductive period, especially when one is concerned with exceptional values. Positive selection on mutations offering an advantage earlier in life as well as favoring a longer life has also been proposed: it could, for instance, be the case for the elevated frequency of the mutation responsible for glucose 6-phosphate deidrogenase deficiency (G6PD) among Sardinian centenarians (Caselli et al. 2006).

In conclusion, results from this study provided evidence of demographic and genealogical characteristics associated with longevity, although as in previous studies on other populations, the impact of these factors seems relatively small. The understanding of the genetic basis of complex traits such as longevity and of the biodemographic factors influencing duration of life is far from comple. Further investigations based on familial and genealogical data could help to disentangle these factors and formulate hypotheses on the role of genetic factors in lifespan and mortality patterns.

References

Allen, J. S., Bruss, J., & Damasio, H. (2005). The aging brain: The cognitive reserve hypothesis and hominid evolution. *American Journal of Human Biology*, *17*, 673–689.

Blackburn, M. E., Bourbeau, R., & Desjardins, B. (2004). Hérédité et longévité au Québec ancien. *Cahiers québécois de démographie*, *33*, 9–28.

Bouchard, G. (1986). The processing of ambiguous links in computerized family reconstruction. *Historical Methods*, *19*, 9–19.

Bouchard, G. (2005). *Projet BALSAC – Rapport annuel 2004–2005*. Université du Québec à Chicoutimi (available at www.balsac.uqac.ca).

Caselli, G., Pozzi, L., Vaupel, J. W., Deiana, L., Pes, G., Carry, C., Franceschi, C., & Baggio, G. (2006). Family clustering in Sardinian longevity: A genealogical approach. *Experimental Gerontology*, *41*, 727–738.

Charbonneau, H., Desjardins, B., Légaré, J., & Denis, H. (2000). The population of the St. Lawrence valley, 1608–1760. In M. R. Haines, & R. H. Steckel (Eds.) *A population History of North America* pp. 99–142. New York: Cambridge University Press.

Cournil, A., & Kirkwood, T. B. (2001). If you would live long, choose your parents well. *Trends in Genetics*, 17, 233–235.

Cournil, A., Legay, J. M., & Schächter, F. (2000). Evidence of sex-linked effects on the inheritance of human longevity: A population-based study in the Valserine valley (French Jura), 18–20th centuries. *Proceedings of the Royal Society of London B*, 267, 1021–1025.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. London: Chapman & Hall.

Gauvreau, D., & Bourque, M. (1990). «Jusqu'à ce que la mort nous sépare»: Le destin des femmes et des hommes mariés au Saguenay avant 1930. *Canadian Historical Review*, LXXI, 441–461.

Gavrilov, L. A., & Gavrilova, N. S. (2001). Biodemographic study of familial determinants of human longevity. *Population: An English Selection*, 13, 197–222.

Gavrilov, L. A., Gavrilova, N. S., Olshansky, S. J., & Carnes, B. A. (2002). Genealogical data and the biodemography of human longevity. *Social Biology*, 49(3–4), 160–173.

Gudmundsson, H., Gudbjartsson, D. F., Frigge, M., Gulcher, J. R., & Stefansson, K. (2000). Inheritance of human longevity in Iceland. *European Journal of Human Genetics*, 8, 743–749.

Hamet, P., Merlo, E., Seda, O., Broeckel, U., Tremblay, J., Kaldunski, M., et al. (2005). Quantitative founder-effect analysis of French Canadian families identifies specific loci contributing to metabolic phenotypes of hypertension. *American Journal of Human Genetics*, 76, 815–832.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. London: Chapman and Hall.

Hauck, W. W., & Martin, A. O. (1984). A statistical test for detection of ancestral genetic contributions to disease occurrence in finite populations. *Genetic Epidemiology*, 1, 383–400.

Hawkes, K. (2003). Grandmothers and the evolution of human longevity. *American Journal of Human Biology*, 15, 380–400.

Ikeda, A., Iso, H., Toyoshima, H., Kondo, T., Mizoue, T., Koizumi, A., JACC Study Group, et al. (2006). Parental longevity and mortality amongst Japanese men and women: The JACC Study. *Journal of Internal Medicine*, 259, 285–295.

Insightful Corporation (2006). S-PLUS (Available at http://www.insightful.com/).

Jetté, R. (1991). *Traité de généalogie*. Montréal: Les Presses de l'Université de Montréal.

Johnson, T. E., & Shook, D. R. (1997). Identification and mapping of genes determining longevity. In K. W. Wachter, & C. E. Finch (Eds.) *Between Zeus and the Salmon: The biodemography of longevity* pp. 108–126. Washington, DC: National Academy Press.

Jomphe, M., & Casgrain, B. (1997). *Base de données généalogiques RETRO: structure des données*. IREP, Programme de recherches en génétique des populations, Document III-C-97. Chicoutimi, Quebec, Canada.

Kemkes-Grottenthaler, A. (2004). Parental effects on offspring longevity—evidence from 17th to 19th century reproductive histories. *Annals of Human Biology*, 31, 139–158.

Kerber, R. A., O'Brien, E., Smith, K. R., & Cawthon, R. M. (2001). Familial excess longevity in Utah genealogies. *Journal of Gerontology A*, 56(3), B130–B139.

Laberge, A. M., Michaud, J., Richter, A., Lemyre, E., Lambert, M., Brais, B., & Mitchell, G. A. (2005). Population history and its impact on medical genetics in Quebec. *Clinical Genetics*, 68, 287–301.

Landgren, O., Bjorkholm, M., Granath, F., & Ekbom, A. (2005). Familial longevity: The older you are, the older your father may have been. *Journal of the American Geriatrics Society*, 53, 357–358.

Mitchell, B. D., Hsueh, W. C., King, T. M., Pollin, T. I., Sorkin, J., Agarwala, R., et al. (2001). Heritability of life span in the Old Order Amish. *American Journal of Medical Genetics*, 102, 346–352.

Morissette, J., Villeneuve, A., Bordeleau, L., Rochette, D., Laberge, C., Gagné, B., et al. (1999). Genome-wide search for linkage of bipolar affective disorders in a very large pedigree derived from a homogeneous population in Quebec points to a locus of major effect on chromosome 12q23-q24. *American Journal of Medical Genetics*, 88, 567–587.

Ohta, T., Ogihara, R., & Nakamura, O. (2004). No sex-linked inheritance of longevity between Japanese centenarians and their relatives. *Journal of the American Geriatrics Society*, 52, 849–850.

Pausova, Z., Jomphe, M., Houde, L., Vézina, H., Orlov, S. N., Gossard, F., et al. (2002). A genealogical study of essential hypertension with and without obesity in French Canadians. *Obesity Research*, 10, 463–470.

Pettay, J. E., Kruuk, L. E., Jokela, J., & Lummaa, V. (2005). Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 2838–2843.

Poulain, M., & Naito, K. (2004). L'évolution de la longévité à Okinawa, 1921–2000. *Cahiers québécois de démographie*, 33, 29–49.

PRDH (Programme de Recherche en Démographie Historique) (2006). Programme de recherche en démographie historique (www.genealogie.umontreal.ca/en/).

Robine, J. M., Cournil, A., Henon, N., & Allard, M. (2003). Have centenarians had younger parents than the others. *Experimental Gerontology*, *38*, 361–365.

Scriver, C. R. (2001). Human genetics: Lessons from Quebec populations. *Annual Review of Genomics and Human Genetics*, *2*, 69–101.

Thompson, E. A. (1986). *Pedigree analysis in human genetics*. Baltimore: Johns Hopkins University Press.

Vézina, H. (1996). Démographie génétique et maladies héréditaires au Québec: L'état des recherches. *Cahiers Québécois de Démographie*, *25*, 293–322.

Vézina, H., Heyer, E., Fortier, I., Ouellette, G., Robitaille, Y., & Gauvreau, D. (1999). A genealogical study of Alzheimer disease in the Saguenay region of Quebec. *Genetic Epidemiology*, *16*, 412–425.

Vijg, J., & Suh, Y. (2005). Genetics of longevity and aging. *Annual Review of Medicine*, *56*, 193–212.

Westendorp, R. G., & Kirkwood, T. B. L. (2001). Maternal and paternal lines of familial longevity. *Population: An English Selection*, *13*, 223–236.

Louis Houde is a professor in the Department of Mathematic and Computer Science at the Université du Québec à Trois-Rivières. He holds a Ph.D. in mathematics from the Université de Montréal. His research focuses on the development and implementation of statistical methods for genealogical data or other data with a similar structure.

Marc Tremblay is a professor in the Department of Human Sciences at the Université du Québec à Chicoutimi. He holds a Ph.D. in demography from the Université de Montréal. His research focuses on the genetic structure of Quebec regional populations and the underlying demographic mechanisms.

Hélène Vézina is a professor in the Department of Human Sciences at the Université du Québec à Chicoutimi. She holds a Ph.D. in demography from the Université de Montréal and received her undergraduate training in anthropology at the University of British Columbia. Her research focuses on the genetic structure of Quebec regional populations and the underlying demographic mechanisms.