

# Native American Admixture in the Quebec Founder Population

Claudia Moreau<sup>1</sup>, Jean-François Lefebvre<sup>1</sup>, Michèle Jomphe<sup>2</sup>, Claude Bhérier<sup>1</sup>, Andres Ruiz-Linares<sup>3</sup>, Hélène Vézina<sup>2</sup>, Marie-Hélène Roy-Gagnon<sup>1,4\*</sup>, Damian Labuda<sup>1,5\*</sup>

**1** Centre de Recherche, CHU Sainte-Justine, Université de Montréal, Montréal, Québec, Canada, **2** Projet BALSAC, Université du Québec à Chicoutimi, Chicoutimi, Québec, Canada, **3** Department of Genetics, Evolution and Environment, University College London, London, United Kingdom, **4** Department of Epidemiology & Community Medicine, University of Ottawa, Ottawa, Ontario, Canada, **5** Département de Pédiatrie, Université de Montréal, Montréal, Québec, Canada

## Abstract

For years, studies of founder populations and genetic isolates represented the mainstream of genetic mapping in the effort to target genetic defects causing Mendelian disorders. The genetic homogeneity of such populations as well as relatively homogeneous environmental exposures were also seen as primary advantages in studies of genetic susceptibility loci that underlie complex diseases. European colonization of the St-Lawrence Valley by a small number of settlers, mainly from France, resulted in a founder effect reflected by the appearance of a number of population-specific disease-causing mutations in Quebec. The purported genetic homogeneity of this population was recently challenged by genealogical and genetic analyses. We studied one of the contributing factors to genetic heterogeneity, early Native American admixture that was never investigated in this population before. Consistent admixture estimates, in the order of one per cent, were obtained from genome-wide autosomal data using the ADMIXTURE and HAPMIX software, as well as with the fastIBD software evaluating the degree of the identity-by-descent between Quebec individuals and Native American populations. These genomic results correlated well with the genealogical estimates. Correlations are imperfect most likely because of incomplete records of Native founders' origin in genealogical data. Although the overall degree of admixture is modest, it contributed to the enrichment of the population diversity and to its demographic stratification. Because admixture greatly varies among regions of Quebec and among individuals, it could have significantly affected the homogeneity of the population, which is of importance in mapping studies, especially when rare genetic susceptibility variants are in play.

**Citation:** Moreau C, Lefebvre J-F, Jomphe M, Bhérier C, Ruiz-Linares A, et al. (2013) Native American Admixture in the Quebec Founder Population. PLoS ONE 8(6): e65507. doi:10.1371/journal.pone.0065507

**Editor:** Dennis O'Rourke, University of Utah, United States of America

**Received:** February 25, 2013; **Accepted:** April 25, 2013; **Published:** June 12, 2013

**Copyright:** © 2013 Moreau et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported by Réseau de Médecine Génétique Appliquée of Fonds de Recherche en Santé du Québec (FRSQ) (HV and DL) and by FRSQ grant (MHRG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [damian.labuda@umontreal.ca](mailto:damian.labuda@umontreal.ca) (DL); [mroygagn@uottawa.ca](mailto:mroygagn@uottawa.ca) (MHRG)

## Introduction

A major goal of medical and population genetics is to understand phenotypic consequences of genetic variation [1]. For years, studies of founder populations and genetic isolates represented the mainstream of the genetic mapping effort in targeting rare single gene defects held to cause Mendelian disorders [2]. Searching for genetic determinants of complex disorders changed the focus from rare deleterious mutations to susceptibility variants of common frequencies [3] and shifted attention towards association studies requiring very large and diversified cohorts. However, the importance of rare variants in genetic susceptibility to common diseases has been vindicated [4]. This paradigm change, from the causal common to causal rare susceptibility variants, has renewed interest in founder populations [5]. In populations arising from a founder event or in a genetic isolate, an initially "rare" mutation may gain in frequency to become more "mappable" [6–9].

Numerous founder events accompanied European colonization of the Americas, creating populations that remained isolated due to geographic barriers and/or distinctive cultural/religious/ethnic identities [10–13]. Their putative cultural and genetic homogeneity (enhanced by demographic bottlenecks typical of New World

settlements) are considered as important advantages in association studies [14]. But is this really so? The population descending from settlers of *Nouvelle-France*, of European and mainly French origins, forming the majority of today's province of Quebec, Canada, is known for a number of recessive diseases that are endemic, of increased frequency and/or due to a population specific mutation(s) [15,16]. Because of its relatively limited number of European founders, the genetic homogeneity of French Canadians has been implicitly assumed, due to a founder effect reinforced by a demographic spurt in the 19<sup>th</sup> century [17,18]. However, non-disease oriented genealogical and genetic studies have shown that the population of Quebec is more genetically diversified than previously anticipated [19,20]. This can be ascribed to highly variable genetic contribution of distinct founders and to uneven geographic expansion of their descendants within the new colony. While designing association studies, one must be aware of genetic stratification within the patient and the corresponding control group, reflecting demographic history of the sampled population. Therefore, there is a need to understand the relative effects of demographic and genetic forces on the apportionment of genomic diversity among individuals and populations, and to be able to

distinguish ancient ancestral relations from more recent admixture [21].

In addition to a diverse contribution of European founders [22], including a minute African origin [23], the resulting population of today's Quebec was also genetically enriched by Native American admixture [19]. Indeed, the presence of Native Americans among the founders of *Nouvelle-France* is documented in historical records [24–26]. Their contribution was revealed by genetic studies of uniparentally transmitted markers showing the existence of Native American mitochondrial DNA lineages in the contemporary Quebec population [27,28]. However, the extent of Native American contribution to the genetic makeup of the contemporary Quebec population is largely unknown. Genetic information limited to uniparentally transmitted Native lineages, and especially maternal lineages, is insufficient to quantify the extent of nuclear DNA admixture [28,29]. Native ancestry is underreported in genealogical records although the extent of missing information remains an unsettled issue [26]. On the other hand, for a given individual, full Native ancestry is assumed once it is recorded, so that if this is not truly the case, it may skew our genealogical estimates of the Native genetic contribution. In genetic epidemiological surveys, the genetic homogeneity of the Quebec founder population was often taken for granted. However, assuming no admixture when in fact there is one, may lead to erroneous results in genetic mapping and association studies. In light of our results, the Quebec population is not different from other New World populations of European descent in being enriched in alleles of Native American origin, although the extent of admixture is much lower than in Central and South American populations [21,30–32]. Similarly as well, the admixture was mainly through marriages with Native women [27,28].

The aim of the present study was to obtain a more reliable estimate of the Native American ancestry in the Quebec genome using autosomal DNA diversity as well as genealogical data. This study builds on single nucleotide variations (SNVs) from 205 individuals representing different regional groups of the contemporary Quebec population (Figure 1), some of which were previously analysed in a different context [20,28]. For all but ten of these individuals ( $n = 195$ ; Table 1), the ascending genealogies were reconstructed up to the Quebec founders. The genomic data of our reference populations were from Reich et al. [33].

## Materials and Methods

### Population Sample and Ethics Statements

We analyzed 205 unrelated (up to the 3<sup>rd</sup> generation) individuals from 10 groups from Quebec (see map Fig. 1 and Table S1). Five regions were described in [20]: North Shore, Saguenay, and the areas of Quebec City and Montreal, as well as 3 Gaspesian ethnocultural groups (French Canadians, Loyalists and Acadians). In this study, we added 3 and 9 samples from the Montreal and Quebec City areas, respectively, in addition to samples from 3 new regions: Abitibi (18 samples), Outaouais (15 samples) and the Gaspesian-Channel Islander subpopulation (20 samples). All participants provided informed written consent, and the study was approved by the CHU Sainte-Justine Ethics Committee. Regional/ethnic affiliation was self-described by the participants. DNA was obtained for all participants as previously described [20,34] and sent to the McGill University and Genome Quebec Innovation Center to be genotyped on Illumina HumanHap650Y and 610Quad arrays according to the recommended protocols. Quality control filters were applied at the individual and Single Nucleotide Variation (SNV) levels using the PLINK software v1.07 [35,36] following the same criteria than in

Roy-Gagnon et al. [20]. Information was collected to reconstruct the genealogy of each participant. Genealogies were reconstructed as far back as possible in a total of 195 individuals (genealogical data were missing for 5 individuals from the Montreal area, 3 from the Quebec City area, and 1 from both Outaouais and Saguenay) using the BALSAC population register [37] and the Early Quebec Population Register [38].

Unless stated otherwise, the Native American reference sample ( $n = 52$ ) includes individuals from Northern America (Aleutians, Algonquin speakers, Chipewyans, Cree, Ojibwa as well as West and East Greenlanders), whereas the European reference sample (CEU+French) consisted of HapMap CEU ( $n = 108$ ) and French ( $n = 28$ ) from the Human Genome Diversity Panel (HGDP). Extended reference sample of Europeans contained in addition, HGDP Italians ( $n = 12$ ) and Tuscans ( $n = 8$ ), as well as HapMap Tuscans TSI ( $n = 88$ ) (Table S2).

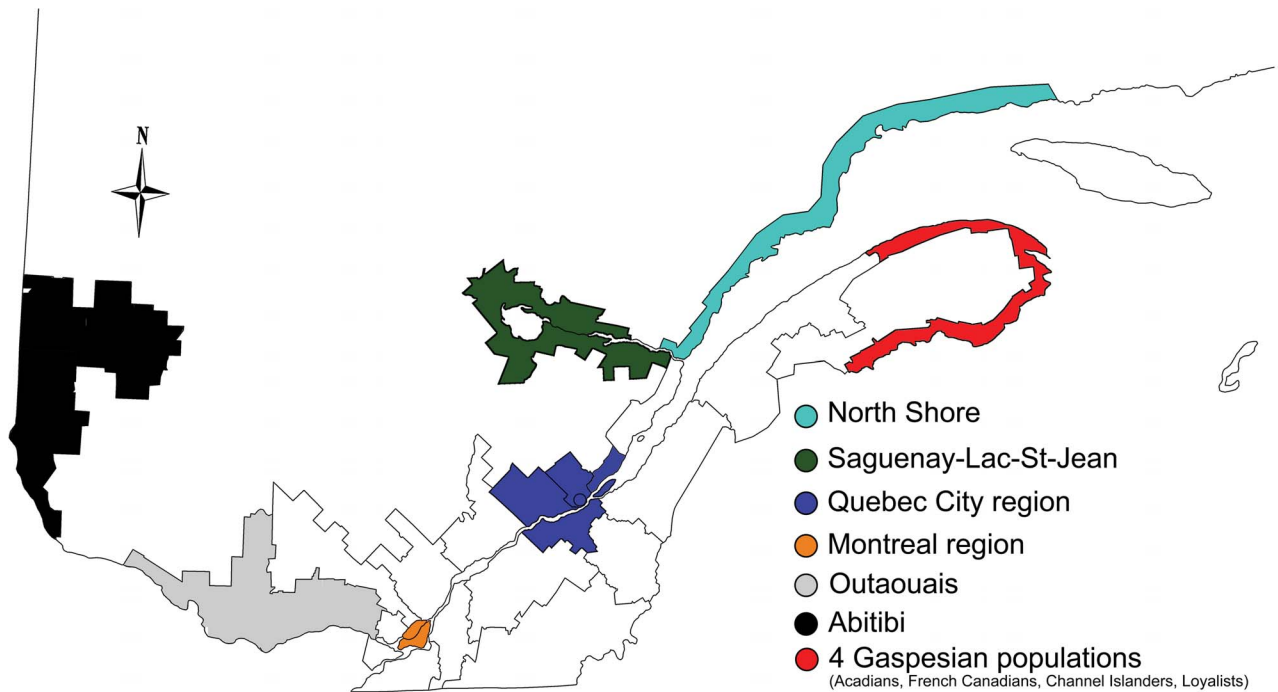
### Statistical Analysis

To infer local ancestry under a haplotype-based model, we used the HAPMIX software version 1.2 [39,40] that estimates, at each locus, the probability of having 0, 1 or 2 alleles transmitted by one of the 2 source populations. We used an approach similar to Reich [33] for masking the European and African segments in the Native Americans masked dataset. We retained all loci with  $\geq 0.95$  probability that 1 or both alleles originated from the Native American source population. For this analysis we selected subsets of 50 individuals to match the sample size of the Native North American populations. Native American and European populations were phased together using the BEAGLE software version 3.3.2 [41,42].

We estimated the global ancestry using the model-based approach implemented in the ADMIXTURE software version 1.22 [43,44] with  $K = 3$  ( $K$  being the number of ancestral populations assumed in the model) to distinguish between Native North American and European and Siberian ancestry in the Quebec sample.  $K = 3$  allowed us to distinguish between old and new Asian-Native American ancestry. We used the PLINK software to select single nucleotide variations (SNVs) in approximate linkage equilibrium (pairwise  $r^2 < 0.1$  in sliding windows of size 50 shifting every 10 SNVs), yielding a subset of 46,344 SNVs.

We used fastIBD from the BEAGLE software version 3.3.2 [41] to find shared Identity-by-Descent (IBD) segments between individuals of Quebec and the Native Americans to investigate shared ancestry [45–47]. The fastIBD method is based on estimating frequencies of shared haplotypes allowing for phase uncertainty. Results from this method were shown to be well correlated with genealogical kinship coefficients in Quebec individuals. Following the authors' recommendation, we performed 10 runs of fastIBD (*ibscale* = 4) that we merged using the scripts provided by the authors. We performed the analysis with the unmasked complete data and for each Native American individual separately, we discarded afterwards shared IBD segments located in masked genomic regions. We retained the remaining shared IBD segments of at least 1 cM. All IBD segments shared with Native Americans were finally pooled for each Quebec individual.

We calculated the  $f_3$  statistic implemented in the ADMIXtools software [40,48], which is based on patterns of allele frequency correlations across populations, to estimate the Native American ancestry proportion lower and upper bounds in the Quebec sample. The proportion of Native American ancestry in the Quebec sample was also evaluated using the ALDER software version 1.0 [49,50]. Both ALDER and ADMIXtools (*rolloff* analysis) were used to estimate the linkage disequilibrium (LD)



**Figure 1. Map of Quebec subpopulations.** In colors are the 10 regions/subpopulations included in the analyses.  
doi:10.1371/journal.pone.0065507.g001

decay due to admixture. ALDER decay curves obtained using only one reference population were fitted starting at 0.8 and 0.7 cM (determined by the software) for the Quebec and French+CEU sample respectively using the masked Native North Americans as reference. The *rolloff* decay curve for the Quebec sample was estimated with two reference populations, Native North Americans (unmasked) and Europeans, and was fitted starting at 0.5 cM.

The R statistical environment version 2.15.0 Patched [51] was used for additional statistical testing and graphing.

### Genealogical Analysis

Of the 8424 founders (individuals whose parents could not be traced in Quebec parish and civil records) identified in the 195 genealogies of the Quebec individuals, 39 were of documented Native American origin. The genetic contribution (GC), based on genealogical data, is the expected proportion of the genome transmitted by an ancestor to a given individual. The GC was calculated using the GENLIB package, a genealogical analysis tool developed at BALSAC for the S-PLUS environment and transferred to the R environment for internal use. The mean GC was obtained by summing the GC of all Native American founders to all Quebec individuals and dividing by the number of individuals.

## Results

### Admixture Proportions Inferred using Genomic Data

Using HAPMIX [39], the overall degree of Native American admixture in the Quebec population was estimated at  $1.3 \pm 1.3\%$ , with 3-fold differences (0.84 to 2.50%) between population groups (Table 1) and much greater variation among individual genomes (Table S1). The results obtained by ADMIXTURE [43] were less straightforward. While this analysis captured differences in the extent of Native admixture among Quebec individuals (see

Figures 2 and 3, and Table S1), it also detected some “background admixture” reflecting ancient common ancestry of Europeans and Native Americans, prior to European colonisation of the New World and presumably due to an earlier gene-flow between Eurasian populations [52] (Figure 2). In a simple  $K = 2$  analysis (i.e., assuming two ancestral populations), we obtained an estimate of about 8% of Native American admixture both in Europe and in Quebec (Figure 2a). Adding Siberian populations in the  $K = 3$  analysis significantly reduced this background effect, but the estimated degree of Native American admixture in Quebec, of  $2.1 \pm 1.4\%$ , still remained above that evaluated by HAPMIX and did not exceed the overall estimate in our European reference sample ( $2.3 \pm 1.2\%$ ). Nevertheless, in the ADMIXTURE graph it is easy to remark higher level of Amerindian admixture in some Quebec individuals, which appears as a signature of recent admixture unseen in European samples.

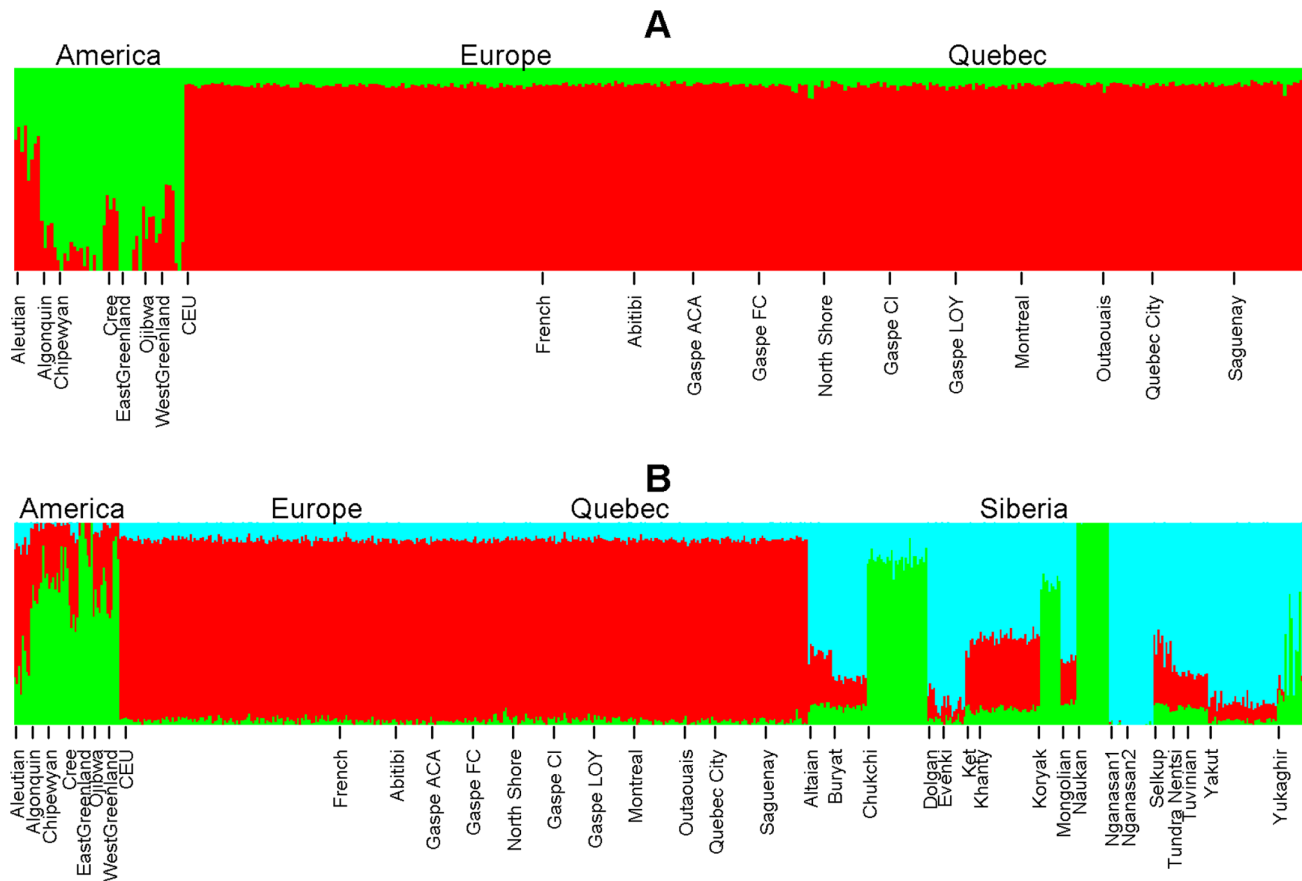
In this paper we also explore the use of Identity-by-Descent (IBD) estimation by fastIBD [41] to evaluate the extent of sharing of genomic segments between the Native genomes and Quebec individuals of primarily European descent. Of note, in this new application of fastIBD, rather than comparing individual Quebec and Native genomes one by one in a pairwise fashion, we analysed individual Quebec genomes against the entire set of Native North American samples. Based on this approach, the mean haplotype sharing in the whole Quebec sample was evaluated at  $26 \pm 37$  cM, roughly corresponding to about 0.8% of the genome, varying from 12 to 54 cM among different Quebec population groups (Table 1).

Importantly, estimates of Native American ancestry by the three methods that use genomic data are in very good agreement (Figure 3a, b and c). Discrepancies appear to be mainly due to individuals with a very low level of admixture where the signal-to-noise ratio is obviously very low. In the case of the estimates by ADMIXTURE, the effect of stochastic noise is additionally exacerbated by the genetic background mentioned above, which is much smaller but still present in the  $K = 3$  analysis including

**Table 1.** Native American ancestry proportions in the Quebec regions.

Regional/ethnocultural population group (n genotypes; n genealogies)	Expected genetic contribution of Native Americans (genealogical data)		Native American ancestry (ADMIXTURE)		Native American SNVs (HAPMIX)		Total length of haplotypes shared with Native Americans (fastIBD)	
	Mean (%)	SD	Mean (%)	SD	Mean (%)	SD	Mean (cM)	SD
Abitibi (18;18)	0.18	0.42	1.33	1.11	0.84	0.47	12	9
Outaouais (15;14)	0.24	0.43	1.94	1.43	1.19	0.97	24	32
Montreal (25;20)	0.12	0.13	1.76	0.97	1.09	0.57	16	8
Quebec City (25;22)	0.16	0.22	1.78	0.99	0.96	0.63	14	7
Saguenay (22;21)	0.17	0.18	1.95	1.69	1.41	1.73	27	62
North Shore (20;20)	0.42	0.57	2.42	1.51	2.05	1.64	47	47
Gaspesian Channel Islanders (20;20)	0.82	0.98	2.77	1.38	1.38	1.09	33	27
Gaspesian Acadians (20;20)	0.10	0.15	1.77	0.88	1.06	0.82	19	19
Gaspesian Loyalists (20;20)	0.14	0.31	2.70	1.16	0.87	0.48	16	15
Gaspesian French Canadians (20;20)	1.14	1.56	2.81	2.22	2.50	2.57	54	64
Quebec total sample (205;195)	0.35	0.73	2.12	1.43	1.33	1.34	26	37

Here Quebec regional/ethnocultural groups are presented with (number of individuals with genotype data; number of individuals with genealogical data). For each population, expected Native American genetic contribution was estimated using genealogical data. The Native American ancestry proportions in the Quebec subpopulations was estimated with 1) the ADMIXTURE software [43] 2) the HAPMIX software [39], and 3) IBD sharing analysis using the fastIBD software [41]. For each measure, we tested for differences among subpopulations using a Kruskal-Wallis test (Table S3).  
doi:10.1371/journal.pone.0065507.t001



**Figure 2. Barplots of ancestry proportions.** The global ancestry was estimated using the model-based approach implemented in the ADMIXTURE software [43] : in A with  $K=2$  to distinguish between Native American (green) and European (red) ancestry and in B with  $K=3$  in the presence of Siberian populations (cyan).  
doi:10.1371/journal.pone.0065507.g002

Siberian samples (Figure 2b). This is presumably the reason why the best correlation is obtained between the fastIBD and HAPMIX estimates (Figure 3c), techniques that appear to be less sensitive to ancient admixture events and/or ancient common ancestry.

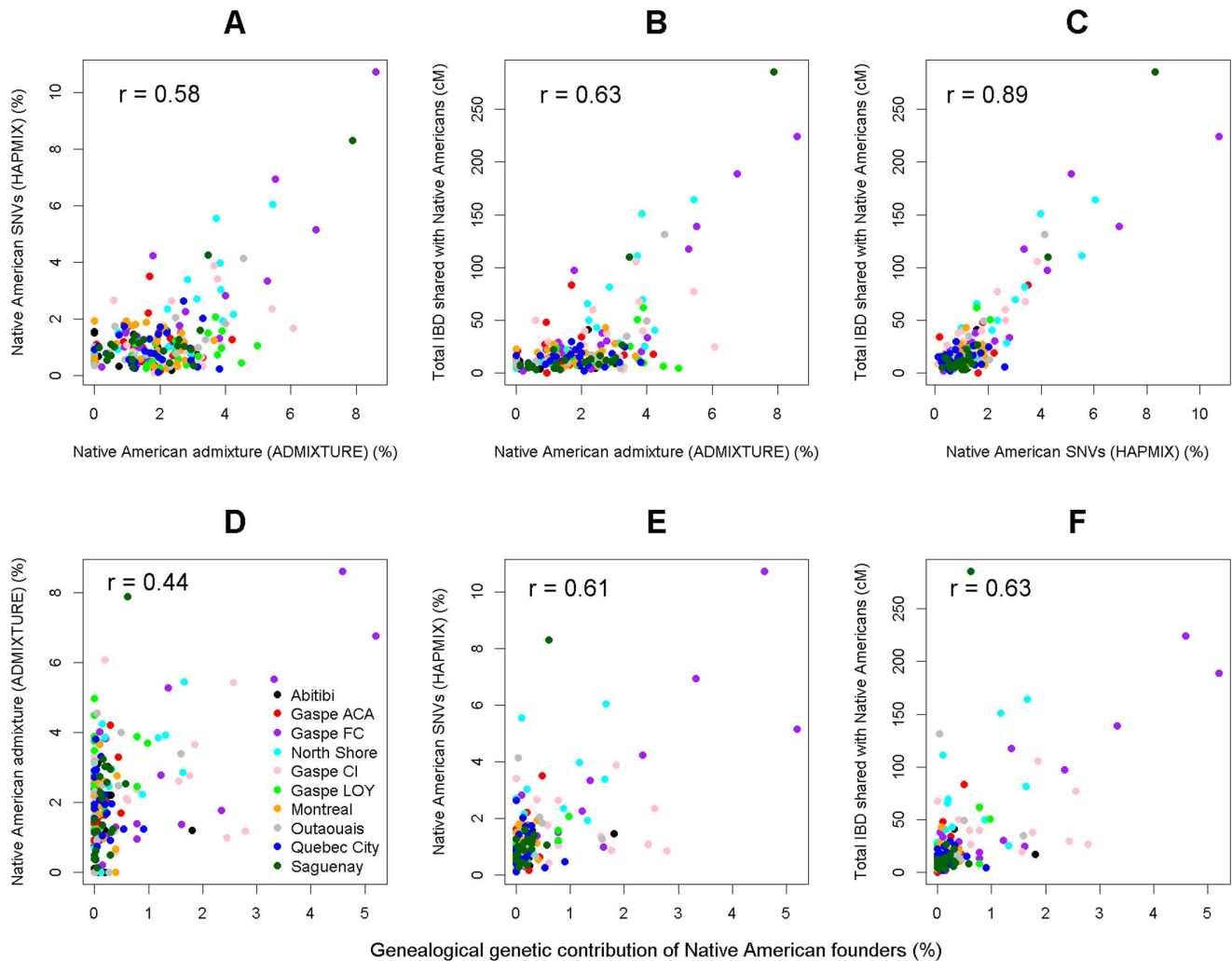
#### Admixture Proportions Inferred Using Genealogical Data

When compared to genealogical estimates of the expected genomic contribution of Native American founders (Table S1), the best correlations were again obtained with HAPMIX and fastIBD results (Figure 3e and f; Pearson correlation coefficient  $r=0.61$  and  $0.63$ , respectively). At the same time, however, we observed a fraction of individuals who display relatively low admixture in genealogical analysis but whose relatively high level of admixture was consistently detected in their genomic data (Figure 3). This is consistent with the hypothesis that genealogical information about the Native American origin of ancestors is likely incomplete, because it was either not recorded by the priest in charge of the parish registers or unknown for various reasons such as adoption or illegitimate birth. The average Native American contribution, estimated at  $0.35\%$  from the genealogical records, is thus expected to be an underestimate. Therefore, we also anticipate that having more complete genealogical information about Native American contributors would substantially improve the correlation between genomic and genealogical estimates of admixture (shown in Figure 3e and f).

#### Time Since Admixture

Using the genealogical record, we can evaluate the time period of admixture. It extends from the beginning of the colony in the early 1600s to the first half of 19<sup>th</sup> century (Figure S4), or roughly 7 to 13 generations ago. Note that Quebec individuals participating in this study and representing regional and/or ethnocultural groups of the Quebec population were unaware of and/or did not declare their admixed Native American ancestry. However, descendants of more recent mixed marriages considering themselves as Métis were not studied here. The time of admixture can be also estimated from the genomic data, from the decay of linkage disequilibrium around the reference, mutated or admixed, segment [9]. Such analysis to estimate time as well as the extent of admixture is implemented in the ADMIXtools software [48] and its recent avatar ALDER [49]. The  $f_3$  statistic estimates of the proportion of admixture were between  $0.5$  and  $2.9\%$ , and the admixture time inferred by *rolloff* was about  $13.6 \pm 2.5$  generations ago (Table 2), i.e. about  $435 \pm 80$  years ago, assuming 32 years per generation [53]. Using only Native North American as a reference, ALDER estimated the lower bound of the proportion of admixture in Quebec at  $0.8 \pm 0.2\%$  and the admixture generation at  $11.2 \pm 3.2$ , or  $358 \pm 102$  years ago (Figure 4a). The same estimates for the European sample were  $1.0 \pm 0.3\%$  and  $1890 \pm 1600$  years ago, respectively (Figure 4b). This is consistent with much older, yet not well characterized admixture events in Eurasia, such as, for example, the prehistoric gene-flow from





**Figure 3. Scatter plots of correlations.** Scatter plots showing the correlations between different Native American genetic ancestry estimates in the Quebec subpopulations (upper) and between genetic ancestry estimates and genealogical genetic contribution of Native American founders to the Quebec individuals (lower). The Pearson correlation coefficient ( $r$ ) is shown on each plot. doi:10.1371/journal.pone.0065507.g003

Siberia to Northern-Europe [52]. Importantly, the above estimates of average time of admixture in Quebec, pointing to the turn of 17<sup>th</sup> century, are in excellent agreement with historical data and the genealogical records (Figure S4), especially given the possibility of a slight upward bias in *rolloff* time estimates when the degree of admixture is low (<5%) [54].

Moreover, it is plausible that part of the “missing” genealogical information involves individuals who were admitted prior to their first registration in Quebec civil records. For instance, in a previous study [27,28] we identified Acadian settlers, whose Native ancestry was only discovered once their descendants mitochondrial DNA was investigated.

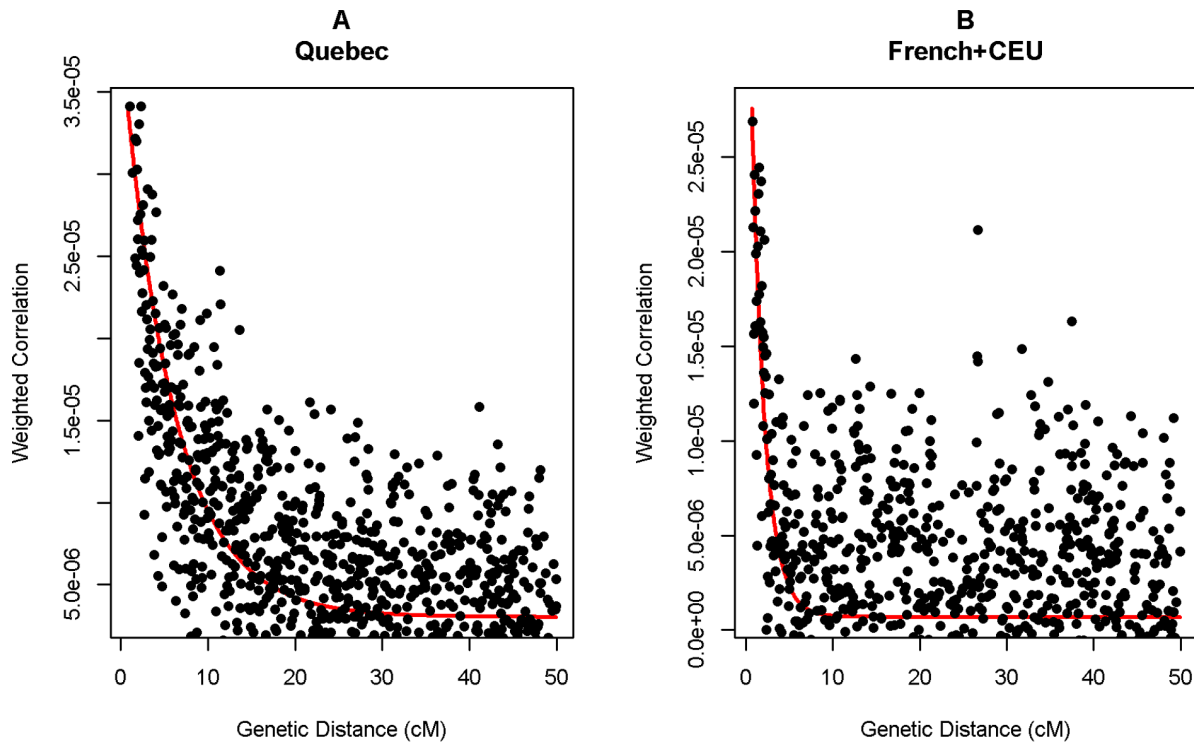
## Discussion

The advantages of using genetically isolated, founder populations in gene mapping have often been discussed in the context of studies on the genetic bases of Mendelian disorders, as well as on genetic susceptibility genes of complex diseases [14,55,56]. Some of the anticipated advantages can be compromised by a hidden population structure due to local founder effects [6,19,20,57] and/or to unrecognized admixture. We found a low level of admixture

overall. However, the variance of admixture estimates among individuals is very large (Table S1 and Figures S1, S2 and S3). Admixture estimates also vary significantly between the studied subpopulations (Kruskal-Wallis  $p < 0.001$  for genealogical and IBD estimates and  $p < 0.05$  for HAPMIX and ADMIXTURE estimates, Table S3). The four groups that consistently show the highest Native ancestry estimates are Gaspésians (ethnocultural groups of French Canadian or Channel Islander origin) as well as French Canadians in the North Shore and Saguenay regions.

These results were obtained using the sample of 52 Native North American genomes described above. In order to test how the choice of the Native American or European reference individuals affected our results, we conducted the same analyses using different reference populations. We observed a small difference in the extent of admixture using different sets of reference genomes. These differences likely reflect geographic proximity and time depth of shared ancestry of the populations we could analyze (Table 2 and Figures S1, S2 and S3).

Our study also shows that the analysis of IBD sharing performed very well in the analysis of recent admixture, comparable or perhaps even better than HAPMIX. Therefore,



**Figure 4. Plots of LD decay.** LD decay of admixed fragments was assessed using the ALDER software [49] to test for admixture in the Quebec (A) and European (B) sample using the Native North Americans as the reference population.  
doi:10.1371/journal.pone.0065507.g004

IBD sharing can be used to assist population structure analysis in association studies. Others have also shown that the analysis of IBD sharing is a promising tool for reconstructing populations' demographic history [45–47,58]. Interestingly, principal compo-

nents analysis (PCA), often used in admixture studies, was not useful here because the observed level of Native American admixture was insufficient to be revealed in PCA plots (Figure S5). In contrast, approaches based on LD and haplotype

**Table 2.** Native American ancestry proportions and age of admixture in the Quebec population sample by different methods and using different reference populations (see Table S2).

		Reference populations <sup>a</sup>		
		Native Americans, Europeans	Native North Americans, French+CEU	Unadmixed Native Americans, French+CEU
Native American ancestry (ADMIXTURE) <sup>a</sup>	Mean (%)	3.24	2.12	4.32
	SD	1.13	1.43	1.05
Native American SNVs (HAPMIX) <sup>b</sup>	Mean (%)	0.90	1.33	0.99
	SD	0.75	1.34	0.61
Length of IBD sharing with Native Americans	Mean (cM)	11	26	9
	SD	19	37	13
Admixture proportions ( $f_3$ statistic, ADMIXtools)	Lower bound (%)	1.2	0.5	0.6
	Upper bound (%)	-1.8	2.9	3.1
Admixture proportions (ALDER)	Lower bound (%)	0.8	0.8	0.7
	SD	0.2	0.2	0.2
LD decay (rolloff, ADMIXtools)	Mean (gen)	16.4	13.6	15.6
	SD	3.1	2.5	2.9
LD decay (ALDER)	Mean (gen)	13.6	11.2	13.6
	SD	4.3	3.2	4.2

<sup>a</sup>For ADMIXTURE, Siberians were used as a third reference population, whereas IBD sharing and ALDER, used only Native Americans as a single reference population.

<sup>b</sup>For HAPMIX, 50 samples were randomly selected from each of the reference populations to match the Native North American sample size.

doi:10.1371/journal.pone.0065507.t002

information used here appear sensitive enough to capture the subtle recent Native American ancestry latent in the Quebec population. Otherwise it would be difficult to discern between the genetic sharing due to common ancient population history and recent admixture (Table 2, Figure 4).

In conclusion, using dense genotypic data and deep-rooted genealogical data, we estimated the Native American ancestry in a population sample from Quebec, Canada. The Native American genetic contribution calculated using genealogical data was low. Unlike most studied admixed populations that have greater admixture proportions [30,59–61], we estimated the part of genetic ancestry coming from Native Americans to Quebec regional samples at about 1%. An individual separated by  $m$  meioses from an ancestor is expected to share  $2^{-m}$  of this ancestor's genome, i.e. on average about 0.1%, or 3.3 cM, after 10 generations. However, the length (exponentially distributed) of a shared segment has a mean of  $100\text{cM}/m$ , or about 10 cM after 10 generations, suggesting that some individuals carry fairly long shared fragments and others not at all, explaining the variance among individual genomes. One percent Native ancestry can be understood as if everybody shared a Native American ancestor 6–7 generations ago. Indeed, a recent study based on four Quebec regional populations indicates that between 53 and 78% of Quebecers have at least one Native American ancestor in their genealogy [24]. Because of the small size of the early Quebec population, the same ancestor often contributed through more than one line to the same contemporary genome, thus suggesting its average occurrence at more distant generations than 6 or 7. Obviously, in historical reality, this genetic contribution varied both in time and space, impacting on stratification of the population and uneven distribution of the rare variants. Hence, Native American ancestry likely played a role in a reduction in the homogeneity of the Quebec founder population. This should be taken into account in the context of mapping and association studies, especially when rare genetic variants are involved.

## Supporting Information

**Figure S1 HAPMIX results in the Quebec population groups using different reference populations.** Boxplots of the percentage of Native American ancestry from 3 runs of the ADMIXTURE software performed with different reference populations listed in the titles of the individual plots. See Table S2 for the description of the reference populations. (TIFF)

**Figure S2 ADMIXTURE results in the Quebec population groups using different reference populations.** Boxplots of total length of IBD sharing between the Quebec

individuals and different Native American reference populations from 3 runs of fastIBD. See Table S2 for the description of the Native American populations.

(TIFF)

**Figure S3 IBD sharing between the Quebec population groups and different Native American populations.**

Boxplots of total length of IBD sharing between the Quebec individuals and different Native American reference populations from 3 runs of fastIBD. See Table S2 for the description of the Native American populations.

(TIFF)

**Figure S4 Plot of the number (bars) and genealogical genetic contribution (dots) of the Native American (red) and non Native American (blue) founders by marriage year in the genealogies of the Quebec individuals.**

(TIFF)

**Figure S5 First 4 eigenvectors of the Principal Component Analysis of the genomic data.**

Note - Global ancestry was estimated by PCA on the genotypic data using the EIGENSOFT software version 3.0 [40,62]. To remove the effect of LD on the PCA, we used the subset of pruned SNPs described above.

(TIFF)

**Table S1 Genetic contribution and admixture indices among Quebec individuals.**

(DOCX)

**Table S2 Reference populations used in the analyses.**

(DOCX)

**Table S3 Genetic contribution and admixture indices in the Quebec subpopulations.**

(DOCX)

## Acknowledgments

We are grateful to Brad Loewen and Stephen Oppenheimer for their comments, to Guylaine Cloutier for her help in recruiting individuals from Abitibi, and to all study participants who consented to provide their genealogical information and/or genomic sample.

## Author Contributions

Conceived and designed the experiments: CM DL. Performed the experiments: CM JFL. Analyzed the data: CM DL. Contributed reagents/materials/analysis tools: CM JFL MJ CB MHRG. Wrote the paper: CM DL. Participant recruitment: CM JFL CB MHRG. Data collection: MJ HV. Manuscript revision: JFL CB MJ MHRG HV ARL.

## References

1. Torkamani A, Pham P, Libiger O, Bansal V, Zhang G, et al. (2012) Clinical implications of human population differences in genome-wide rates of functional genotypes. *Front Genet* 3: 211.
2. McKusick VA (2007) Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet* 80: 588–604.
3. Reich DE, Lander ES (2001) On the allelic spectrum of human disease. *Trends Genet* 17: 502–510.
4. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI (2008) Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 82: 100–112.
5. Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40: 695–701.
6. Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 8: 1913–1923.
7. Zeegers MP, van Poppel F, Vlietinck R, Spruijt L, Ostrer H (2004) Founder mutations among the Dutch. *Eur J Hum Genet* 12: 591–600.
8. Ostrer H (2001) A genetic profile of contemporary Jewish populations. *Nat Rev Genet* 2: 891–898.
9. Labuda D, Zietkiewicz E, Labuda M (1997) The genetic clock and the age of the founder effect in growing populations: a lesson from French Canadians and Ashkenazim. *Am J Hum Genet* 61: 768–771.
10. Morera B, Barrantes R (2004) Is the Central Valley of Costa Rica a genetic isolate? *Rev Biol Trop* 52: 629–644.
11. Rahman P, Jones A, Curtis J, Bartlett S, Peddle L, et al. (2003) The Newfoundland population: a unique resource for genetic investigation of complex diseases. *Hum Mol Genet* 12 Spec No2: R167–172.
12. Chong JX, Ouwenga R, Anderson RL, Waggoner DJ, Ober C (2012) A population-based study of autosomal-recessive disease-causing mutations in a founder population. *Am J Hum Genet* 91: 608–620.
13. Scriver CR (2001) Human genetics: lessons from Quebec populations. *Annu Rev Genomics Hum Genet* 2: 69–101.
14. Peltonen L, Palotie A, Lange K (2000) Use of population isolates for mapping complex traits. *Nat Rev Genet* 1: 182–190.



15. Laberge AM, Michaud J, Richter A, Lemyre E, Lambert M, et al. (2005) Population history and its impact on medical genetics in Quebec. *Clin Genet* 68: 287–301.
16. Yotova V, Labuda D, Zietkiewicz E, Gehl D, Lovell A, et al. (2005) Anatomy of a founder effect: myotonic dystrophy in Northeastern Quebec. *Hum Genet* 117: 177–187.
17. Bouchard G, De Braekeleer M (1990) Homogénéité ou diversité ? L'histoire de la population du Québec revue à travers ses gènes. *Histoire Sociale* 23: 325–361.
18. Charbonneau H (1988) Les francophones du Québec de 1608 à 1960. *L'Action Nationale* 78: 220–232.
19. Bherer C, Labuda D, Roy-Gagnon MH, Houde L, Tremblay M, et al. (2011) Admixed ancestry and stratification of Quebec regional populations. *Am J Phys Anthropol* 144: 432–441.
20. Roy-Gagnon MH, Moreau C, Bherer C, St-Onge P, Sinnett D, et al. (2011) Genomic and genealogical investigation of the French Canadian founder population structure. *Hum Genet* 129: 521–531.
21. Kidd JM, Gravel S, Byrnes J, Moreno-Estrada A, Musharoff S, et al. (2012) Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. *Am J Hum Genet* 91: 660–671.
22. Charbonneau H, Desjardins B, Légaré J, Denis H (2000) The Population of the St. Lawrence Valley, 1608–1760. *A Population History of North America*: Cambridge University Press. 99–142.
23. Moreau C, Vézina H, Labuda D (2007) [Founder effects and genetic variability in Quebec]. *Med Sci (Paris)* 23: 1008–1013.
24. Vézina H, Jomphe M, Lavoie E, Moreau C, Labuda D (2012) L'apport des données génétiques à la mesure généalogique des origines amérindiennes des Canadiens français. *Cahiers Québécois de Démographie* 41: 87–105.
25. Desjardins B (1990) [Ethnic homogeneity of the Quebec population during the French regime]. *Cah Que Demogr* 19: 63–76.
26. Beauregard Y (1993) Mythe ou réalité. Les origines amérindiennes des Québécois : entrevue avec Hubert Charbonneau Cap-aux-diamants 34: 38–42.
27. Moreau C, Vézina H, Jomphe M, Lavoie EM, Roy-Gagnon MH, et al. (2011) When genetics and genealogies tell different stories-maternal lineages in Gaspésie. *Ann Hum Genet* 75: 247–254.
28. Moreau C, Vézina H, Yotova V, Hamon R, de Knijff P, et al. (2009) Genetic heterogeneity in regional populations of Quebec-Parental lineages in the Gaspé Peninsula. *Am J Phys Anthropol* 139: 512–522.
29. Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, et al. (2006) Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci U S A* 103: 7234–7239.
30. Seldin MF, Tian C, Shigeta R, Scherbarth HR, Silva G, et al. (2007) Argentine population genetic structure: large variance in Amerindian contribution. *Am J Phys Anthropol* 132: 455–462.
31. Wang S, Ray N, Rojas W, Parra MV, Bedoya G, et al. (2008) Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet* 4: e1000037.
32. Campbell DD, Parra MV, Duque C, Gallego N, Franco L, et al. (2012) Amerind ancestry, socioeconomic status and the genetics of type 2 diabetes in a Colombian population. *PLoS One* 7: e33570.
33. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, et al. (2012) Reconstructing Native American population history. *Nature* 488: 370–374.
34. Moreau C, Vézina H, Yotova V, Hamon R, de Knijff P, et al. (2009) Genetic heterogeneity in regional populations of Quebec-parental lineages in the Gaspé Peninsula. *Am J Phys Anthropol* 139: 512–522.
35. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.
36. PLINK software website. Available: <http://pngu.mgh.harvard.edu/~purcell/plink/>. Accessed 2013 May 6.
37. BALSAC population register website. Available: <http://balsac.uqac.ca/>. Accessed 2013 May 6.
38. PRDH population register website. Available: <http://www.prdh.umontreal.ca/census/en/main.aspx>. Accessed 2013 May 6.
39. Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, et al. (2009) Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet* 5: e1000519.
40. Reich Laboratory website. Available: [http://genetics.med.harvard.edu/reich/Reich\\_Lab/Software.html](http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html). Accessed 2013 May 6.
41. Browning BL, Browning SR (2011) A fast, powerful method for detecting identity by descent. *Am J Hum Genet* 88: 173–182.
42. BEAGLE Genetic Analysis Software Package website. Available: <http://faculty.washington.edu/browning/beagle/beagle.html>. Accessed 2013 May 6.
43. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19: 1655–1664.
44. ADMIXTURE software website. Available: <http://www.genetics.ucla.edu/software/admixture/index.html>. Accessed 2013 May 6.
45. Atzmon G, Hao L, Pe'er I, Velez C, Pearlman A, et al. (2010) Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am J Hum Genet* 86: 850–859.
46. Velez C, Palamara PF, Guevara-Aguirre J, Hao L, Karafet T, et al. (2011) The impact of Converso Jews on the genomes of modern Latin Americans. *Hum Genet* 131: 251–263.
47. Bray SM, Muller JG, Dodd AF, Pulver AE, Wooding S, et al. (2010) Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. *Proc Natl Acad Sci U S A* 107: 16222–16227.
48. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, et al. (2012) Ancient admixture in human history. *Genetics* 192: 1065–1093.
49. Loh PR, Lipson M, Patterson N, Moorjani P, Pickrell JK, et al. (2013) Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* 193: 1233–1254.
50. ALDER software website. Available: <http://groups.csail.mit.edu/cb/alder/>. Accessed 2013 May 6.
51. Team RC (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna.
52. Der Sarkissian C, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, et al. (2013) Ancient DNA reveals prehistoric gene-flow from siberia in the complex human population history of north East europe. *PLoS Genet* 9: e1003296.
53. 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.
54. Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, et al. (2011) The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet* 7: e1001373.
55. Heutink P, Oostra BA (2002) Gene finding in genetically isolated populations. *Hum Mol Genet* 11: 2507–2515.
56. Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K (2005) An Icelandic example of the impact of population structure on association studies. *Nat Genet* 37: 90–95.
57. 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.
58. Palamara PF, Lencz T, Darvasi A, Pe'er I (2012) Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History. *Am J Hum Genet*.
59. Shah AM, Tamang R, Moorjani P, Rani DS, Govindaraj P, et al. (2011) Indian Siddhis: African descendants with Indian admixture. *Am J Hum Genet* 89: 154–161.
60. Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, et al. (2004) A high-density admixture map for disease gene discovery in african americans. *Am J Hum Genet* 74: 1001–1013.
61. Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, et al. (2007) A genomewide admixture map for Latino populations. *Am J Hum Genet* 80: 1024–1036.
62. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2: e190.