NOTE TO USERS

Page(s) not included in the original manuscript and are unavailable from the author or university. The manuscript was scanned as received.

missing pgs. 146-157

This reproduction is the best copy available.



Exploring Population Structure and Migration with Surnames: Quebec, 1621-1900

Kevin A. Henry Department of Geography McGill University Montréal

1

August 2004

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Doctor of Philosophy © Kevin Henry, 2004



Library and Archives Canada

Published Heritage Branch

395 Wellington Street Ottawa ON K1A 0N4 Canada Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington Ottawa ON K1A 0N4 Canada

> Your file Votre référence ISBN: 0-494-12855-0 Our file Notre référence ISBN: 0-494-12855-0

NOTICE:

The author has granted a nonexclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or noncommercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.



Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Abstract

This research uses isonymy (same-surname) methods and models to examine the population structure and migratory history of Quebec, Canada. Through a case study using 1765 and 1881 census and marriage records from 1621-1900, I explore the accuracy of sources as well as develop, test and apply different statistical methods, and experiment with mapping techniques that reveal paths and patterns of French Canadian surnames. Each investigation explores and evaluates a particular method. I noted that multivariate methods, including cluster analysis, relevance networks, and correspondence analysis, not traditionally used in surname analysis offer reliable and informative results, and insights into the hierarchical structure of populations not easily gleaned from traditional surname methods. The spatial and temporal components of Quebec surname distributions revealed that groups of names which populate and distinguish certain regions were in place by 1800, and cross-river relatedness became less significant as the population expanded upstream away from the St. Lawrence River. I also found that surnames unique to certain regions remained strongly clustered until the midnineteenth century when urbanization and the settlement of new territory led to the fusion of name pools (diversification) in and around urban areas, while at the same time causing losses of names in some rural areas. The marriage records provided evidence, through their measure of random mating, that surnames within different regions in Quebec continually diversified throughout the nineteenth century. Overall, I found surnames to be an informative variable for inferring population relatedness and migratory paths. Because surnames are readily available in a number of sources researchers involved with historical migration research should find that the methods presented in this work will provide a time-saving technique which can overcome the restrictions of spatial and temporal scale and provide a broader, more systematic and balanced picture of long and short distance migration and population structure.

Résumé

Cette thèse utilise les méthodes et modèles dits d'isonymie (dit d'un même patronyme) dans le but d'examiner la structure et l'histoire migratoire de la population du Québec. À partir des recensements des années 1765 et 1881 et des registres de mariage de la période 1621-1900, et ce après avoir vérifié l'exactitude des sources en question, j'ai développé, testé et appliqué diverses méthodes statistiques et me suis employé à révéler les trajectoires et les distributions des patronymes Canadiens-Français à l'aide de techniques cartographiques expérimentales. Chacune des enquêtes explicitées par la suite s'emploie ainsi à explorer et évaluer une méthode en particulier. J'ai noté que les méthodes d'analyse multivariée (parmi lesquelles les analyses regroupées, les réseaux de pertinence, et les analyses de correspondances), dont l'usage est traditionnellement peu répandu dans le cadre d'analyses patronymiques offrent une alternative fiable et informative aux méthodes habituellement utilisées. Elles permettent par ailleurs de mettre en exergue la structure hiérarchique de la population d'une manière non encore révélée par ces mêmes méthodes. L'étude de la distribution spatiotemporelle des patronymes au Québec démontre que des regroupements régionaux distincts de patronymes étaient en place dès 1800. Elle démontre également que les correspondances patronymiques observées de part et d'autre du fleuve Saint-Laurent s'éteignirent à mesure que la population crût dans la partie haute du fleuve. J'ai par ailleurs découvert que les patronymes uniques à certaines régions restèrent regroupés régionalement jusqu'au milieu du dix-neuvième siècle, date à laquelle l'urbanisation et l'établissement de la population dans de nouveaux territoires entraîna à la fois la fusion des groupes patronymiques (diversification) au sein et autour des centres urbains et la disparition de patronymes dans certaines régions rurales. A l'aide de mesures de mélange patronymique déduites des registres de mariage, j'ai pu finalement mettre en évidence une diversification soutenue des patronymes au Québec au cours du dix-neuvième siècle. J'ai donc pu valider, pour conclure, l'utilité des patronymes en tant que variable informative permettant de déduire les relations entre diverses populations et leurs chemins migratoires. Au jour où ces données sont rendues facilement accessibles, les chercheurs intéressés par les migrations historiques trouveront les méthodes présentées ici particulièrement efficaces et rapides à la lumière des restrictions spatiotemporelles typiquement inhérentes à ce type de recherche. Enfin, ces méthodes ont pour ambition mesurée de permettre d'envisager les phénomènes migratoires, les structures démographiques et les interactions entre populations de manière plus systématique et plus équilibrée.

Acknowledgements

I take this time to express my gratitude to all those who have assisted me in the research and writing of this dissertation. First, I would like to thank my supervisor, Sherry Olson, who has provided support, guidance and intellectual stimulation during the course of my Ph.D degree. I am especially grateful for her patience and her willingness to meet and talk whenever I had a question.

My dissertation work would not have been possible without the generous support of McGill and its alumni and faculty. First, I would like to thank Roger Warren for his Geographical Information Systems fellowship, and for the words of encouragement he provided during his annual visits to the Geography Department. I also thank the alumni of McGill for their "Friends of McGill" fellowship; Sherry Olson for supporting me through her "Montréal: l'Avenir du Passe" GEOIDE grant and the Canadian Social Science Research Council grant; and David Brown for hiring me as a GIS lab instructor in the Department of Planning.

During the course of my degree at McGill I have benefited from the knowledge and help of a number of people. I thank Pat Thornton (Concordia University) for her patience in reading the thesis and her helpful comments on migration history and demography; Marc St-Hilaire (Université Laval) for his trips to Montréal and his help with French Canadian history, Bertrand Desjardins (Université de Montréal) for his knowledge of French Canadian surnames; and his advice on collecting the samples; and finally John Relethford (State University of New York) for his rapid email responses to what now seem very basic questions.

Many thanks are also due to my colleagues. I thank Jason Gilliland for his advice about the Ph.D process, for helping me adjust to life in Montréal, and finally for all those Canadiens games. I thank Leif Burge for the many breakfast conversations and phone calls during which we discussed multivariate statistics and what eventually ended up as an article on bootstrapping. I thank April James for her math expertise and patience in helping me remember and master matrix algebra. I am also grateful to my friends and fellow graduate students and employees at McGill for their help. In particular, I thank Nate Basiliko, Rosalyn

iv

Trigger and Dao Nguyen.

٠

.

Finally, I wish to thank my family (and dog) for their support and understanding when I did not make it home for Thanksgiving (four years in a row) and for watching my dog whenever a conference came up and when I was teaching in Virginia.

Contents

Abstract		i
Ac	Acknowledgements	
Ta	Table of Contents List of Tables	
Li		
List of Figures		x
Appendix		xiv
1	Introduction	1
2	Quebec surname data sources	17
	Census of 1765	
	Canadian Census of Canada 1881	
	Quebec marriage records, 1621-1900	
	United States Census for Vermont: 1840, 1870, and 1880	
	Geographic and miscellaneous sources for Quebec	
3	Surname methods for population and migration research	34
	Measures of inbreeding, isonymy, kinship and genetic distances	
	Measures of inbreeding by same-surname marriages	
	Same-surname quotients (isonymy equations)	
	Measures of kinship	
	Measures of genetic distance from surnames	
	Micro-differentiation	
	Isolation by distance and surnames (distance decay models)	
	Ordination and multivariate methods for surnames	
	Clustering and fuzzy clustering	
	Relevance networks	
	Bayesian probability	
	Conclusion	

4 18th Century French-Canadian surname population structure 91

- Historical geography
- Eighteenth century same-surname results (isonymy)
- Between population same-surname values
- Eighteenth-century relevance networks
- Eighteenth-century cluster analysis
- Conclusion: eighteenth-century surnames

5 The 1881 surname population structure of Quebec

129

158

- Nineteenth-century historical geography
- Quebec 1881: Isonymy
- 1881 relevance networks
- 1881 cluster analysis
- Discussion/Conclusion: Quebec 1881 surnames

6 Case Studies

- 1 Urban Choices: Montréal and Quebec City and inter-provincial movements
- 2 French-Canadian Migration to Vermont, New Hampshire and Maine
- 3 Lumberjacks and the settlement of the Mauricie: a possible link to a known strain of tuberculosis
- 4 Repeated-name and same-surname marriages
- 7 Conclusion1978 Bibliography213

List of Tables

- 2.1 Quebec parishes and opening dates used for the seventeenth and eighteenth century surname analysis
- 3.1 Total surnames in Vermont from 1800-1920, and total names lost and gained by each decade
- 3.2 Total surnames and same-surname values (isonymy) for all heads of household in 1840 and 1880 Vermont
- 3.3 1840 and 1880 Vermont conditional kinship values determined from surnames.
- 3.4 Measures of micro differentiation determined from surnames for 1840 and 1880 Vermont
- 3.5 Detrended correspondence analysis of surnames for 1880 Vermont
- 3.6 Correlation coefficients of environmental variables for 1880 Vermont
- 3.7 Results of canonical correspondence analysis of surnames for 1880 Vermont
- 3.8 Vermont 1880 example: probability (p) values for CCA
- 4.1 1765 Quebec measures of micro-differentiation determined from surnames.
- 4.2 Quebec surrogate 1765 census sample
- 4.3 Between population surname similarity values for the 30 largest parishes of Quebec, 1765 and the four most related parishes (based on between population isonymy values)
- 4.4 Between population surname similarity values for the first 30 parishes opened in Quebec and the four most related parishes (based on between population isonymy values)
- 4.5 Summary of names and places in clusters.
- 4.6 Summary of the parishes in the 15 clusters
- 4.7 Participation rates: percentage of names in a cluster belonging to a set of place clusters.
- 5.1 Micro differentiation determined from surnames for 1765 and 1881 Quebec.

- 6.1 Clusters for the sample of 32 surnames
- 6.2 Rank order of percent of each of the five surname clusters in Montréal and Quebec City
- 6.3 Percent of all marriages for each of the 5 clusters of single-founder surnames by region, 1870-1900 (n=32)
- 6.4 Percent of all phone users for each of the 5 clusters of single-founder surname 32 single-founder surnames 2002
- 6.5 Spearman's rank correlation of the probability values for possible migrants from Quebec
- 6.6 Measures of same-surname marriages (inbreeding) for Quebec,1750-1790
- 6.7 Observed and randomly simulated coefficient values obtained from the repeated surname pairs method.
- 6.8 Proportion of same-surname marriages (23 name sample)
- 6.9 Observed and randomly simulated coefficient values from the repeated surname pairs method.

List of Figures

- 2.1 Number of dit-names versus the total number found in the PRDH marriage records between 1621-1799
- 2.2 Date single founder surnames arrive in Quebec
- 2.3 Quebec parishes used for the seventeenth and eighteenth century surname analysis.
- 2.4 Hexagon map used for the 1881 Quebec census surname analysis.
- 2.5 Total population 1881 Quebec Census (sample)
- 2.6 Quebec's 23 cultural regions
- 3.1 Log-log distribution of the frequency of occurrence of surnames in Vermont, 1840 and 1880
- 3.2 Ancestor paths (chains) showing the probability of homozygosity of a child between two cousins.
- 3.3 Vermont map illustrating same-surname values (local isonymy) for 1840.
- 3.4 Vermont map illustrating same-surname values (local isonymy) for 1880.
- 3.5 Vermont hexagon map illustrating 1880 same-surname values (local isonymy).
- 3.6 Strongest between population surname similarity (isonymy) connections in Vermont.
- 3.7 Second strongest between population surname similarity (isonymy) connections in Vermont.
- 3.8 Multidimensional Scaling results explaining the between population surname similarity (isonymy) values for 1880 Vermont.
- 3.9 Between population 'a priori' kinship values for Washington, Addison, Orange and Windsor counties.
- 3.10 Between population 'a priori' kinship values for small areas within Washington, Addison, Orange and Windsor counties.
- 3.11 Multidimensional scaling results explaining the 1880 Vermont 'a priori'

kinship values.

- 3.12 First and second axes (dimensions) of the detrended correspondence analysis positioned onto map
- 3.13 Ordination diagram (tri-plot) of results of the CCA, Vermont 1880
- 3.14 Vermont hexagons clustered to various degrees: 15, 10, 5, 3.
- 3.15 Relevance networks correlations set to 4.8, created 34 features, 1 network
- 3.16 Relevance networks correlations set to 6.0, created 2 networks
- 3.17 Relevance networks set to 8.5 for surnames, created 8 networks
- 4.1 Geographic distribution of marriages for five surnames, 1621-1799
- 4.2 Surname diversity by opening date of parish prior to 1780
- 4.3 Surname diversity for the 1765 Quebec surrogate census.
- 4.4 Between population surname similarity values (isonymy) between Rivière-du-Loup (Louiseville) and all other parishes.
- 4.5 Multidimensional scaling of between parish surname similarity (isonymy) values for the surrogate 1765 census
- 4.6 Parish opening date versus the number of parishes where at least one surname is shared (interconnectedness).
- 4.7 Multidimensional scaling of between parish surname similarity values (isonymy) for the 1765 surrogate census
- 4.8 Between population surname similarity values (isonymy) between selected parish (star) and all others.
- 4.9 Relationship between the correlation values and the numbers of networks and parishes
- 4.10 Relevance networks illustrating the (correlation) connections among 1765 parishes.
- 4.11 Relevance networks illustrating the (correlation) connections among 1765 parishes
- 4.12 Relevance networks illustrating the (correlation) connections among

1765 parishes.

- 4.13 Relevance networks illustrating the (correlation) connections among the 832 single-founder surnames.
- 4.14 Relevance networks illustrating the (correlation) connections among the 832 single-founder surnames
- 4.15 Geographic distribution of the relevance networks of the 832 single-founder surnames
- 4.16 Map illustrating the name clusters. The locations represent the top locations where the majority of marriages within that cluster occurred
- .4.17 Sequential maps illustrating the clusters of places at different levels of hierarchy. Clusters were created using grooms from the set of 832 single-founder names for the years 1771-1799 (30,20 and 15 clusters).
- 4.18 Sequential maps illustrating the clusters of places at different hierarchies Clusters were created using groom from the set of 832 single-founder names for the years 1771-1799 (10,5 & 3 clusters).
- 5.1 Local same-surname values (isonymy) for 1881 Quebec
- 5.2 Multidimensional-scaling results of same-surname analysis (isonymy) for the 1881 census single-founder sample
- 5.3 Same-surname (isonymy) values for La-Grande-Riviére (72) and all other locations, 1881.
- 5.4 Same-surname (isonymy) values for Baie-St-Paul (51) and all other locations, 1881
- 5.5 Same-surname (isonymy) values for Yamachiche (37) and all other locations, 1881.
- 5.6 Same-surname (isonymy) values for Buckland (42) and all other locations, 1881.
- 5.7 Same-surname (isonymy) values for Ste-Adéle (16) and all other locations, 1881.
- 5.8 Same-surname (isonymy) values for Stanstead (21) and all other locations, 1881.
- 5.9 Relevance network. Total networks versus total number of hexagons included at different correlation values.
- 5.10 Relevance network. Total networks versus total number of parishes included at various correlation values.

- 5.11 Relevance networks determined from the 1881 census single-founder surname sample
- 5.12 Relevance networks determined from the 1881 census single-founder surname sample transposed onto the results of the between population same-surname (isonymy) analysis.
- 5.13 Close-up view of hexagon 33 (Halifax) and surrounding between-population surname values (isonymy)and relevance network
- 5.14 Relevance networks derived from 1881 census subdistricts.
- 5.15 Seventeen clusters extracted from the 1881 census single-founder surname sample.
- 5.16 Seventeen name clusters mapped onto the results of the same-surname (isonymy) multidimensional-scaling.
- 5.17 Thirty clusters defined from the single-founders surname sample extracted from the Quebec census, 1881.
- 6.1 Growth rate of marriage sample of 32 names versus total population
- 6.2a Geographic distribution of marriages from the 32 single-founder sample
- 6.2b Geographic distribution of marriages from the 32 single-founder sample.
- 6.2c Geographic distribution of marriages from the 32 single-founder sample.
- 6.3 Number of parishes represented in sample by decade (31 Single Founder Surnames)
- 6.4 Percent of total marriages from the 32-name sample occurring in Montréal and Quebec City
- 6.5 Percent of total marriages from each cluster occurring in Montréal and Quebec City
- 6.6 Distribution of phone users with surname Vachon, Quebec, 2000
- 6.7 Distribution of phone users with surname Beauchamp, Quebec, 2000
- 6.8 Distribution of phone users with surname Laberge, Quebec, 2000
- 6.9 Probable origins of migrants who settled in Maine before 1880

- 6.10 Probable origins of migrants who settled in New Hampshire before 1880
- 6.11 Probable of origins of migrants who settled in Vermont before 1880
- 6.12 Geographic distribution of between population same-surname values (isonymy) for the population of single-founders (131 single founders of 832) associated with the single-founder surnames of the 1861 timber population
- 6.13 Geographic distribution of between population same-surname values (isonymy) for the population of all 1881single-founders associated with the single-founder surnames of the 1861 timber population
- 6.14 Probability of finding a surname matching the 221 single-founder surnames defined in the 1861 Mauricie lumberjack population

Appendix A

Table A1. Summary of the 832 surname single-founder marriage sample used for the cluster analysis (1771-1799)

Table A2. Summary of the 1881 census single-founder surname sample used in the same-surname (isonymy) analysis.

Chapter 1

Introduction

The subject of this research is the use of surnames as markers for examining population structure and for analysing and mapping migration. Through various examples and case studies covering Quebec between 1621 and 1900, I explore the accuracy of sources, test and critique various surname methods (models), and experiment with new statistical methods (algorithms) with the goal of quantifying change in surname distributions and mapping the migrations which can be inferred. The major contribution is methodological, providing knowledge about how lists of surnames, which are readily available, can be employed to study historical patterns of settlement and migration. Through its exposition of surname methods this study will, of course, provide some insights into the geographical characteristics of surname distributions in Quebec and help unravel the great puzzle of movements since the first landings in the St. Lawrence Valley in the seventeenth century.

Researchers studying past migration almost always encounter problems in locating, organizing and appraising the accuracy of data that might incidentally record migration. Those involved in historical research know all too well that data limitations sometimes confine our research to locations where data exists, or force our hypotheses to focus on specific inquiries at a limited spatial resolution and a short time period. Research on the historical geography of migration has advanced thanks to creativity and ambition of researchers in overcoming these problems by discovering archival materials, developing unique statistical and sampling strategies, and using cartographic display (Southall et al 1997, Tobler 1997).¹ In North America, data that fully record historical migration behaviour are scarce and as a result, we have an abundance of studies focussing on changing social-cultural processes, the origin of immigrant/migrant communities and their cultural assimilation, the migration process itself and the choices which govern spatial and temporal characteristics of migration such as its volume, length and direction. One

reason the study of immigrant assimilation processes has become so popular is that the Canadian and U.S. censuses record place of birth and nationality, the two variables used to describe and define the immigrant community in their new homes. 'Disconnect' between the newly defined immigrant (migrant) population and their original home (location) is due to the lack of specificity of the census and other sources to define more explicitly where they previously lived. Birth places are aggregated into broad categories like nation-states or administrative provinces, and this type of data permit only a crude and general interpretation of migrant paths. For example, the tens of thousands of 'British' that immigrated to North America came from different parts of their homelands, but the census groups them into a single category. Additionally, these censuses provide only a window every ten years from which to examine an individual's location.

To overcome these limitations, historical geographers and demographers have become very familiar with and quite successful at using surnames for matching individuals from successive censuses to estimate and pinpoint rates of mobility and exact migrant paths, as well as integrating other sources (i.e tax lists) to fill in the gaps between decennial censuses. Researchers also employ family reconstitution from vital records (births, deaths and marriages) to compile longitudinal family histories or genealogies from which we can infer migration and rates of fertility, mortality, and natural increase (Bouchard 1992; Stephenson 1974, 1980, Adams et al. 2002). For Quebec, this technique has dominated due to their superb records and has mainly been applied to the Saguenay population, a region north of Quebec City settled after 1840 (St-Hilaire 1996, 1996; Gauvreau et al. 1987; Haines and Steckel 2000). Both methods produce useful and reliable origin and destination data, but they are labor intensive. Family reconstitution is not possible in most of North America because of a lack of complete vital records prior to the twentieth century.

The inflexibility and sparseness of sources have resulted in many gaps in migration literature and have most probably pulled potential students of historical migration towards other research areas where better data exist. I believe one way of trying to fill the gap in historical migration research and supplement the already

proven methods is by using the very surnames that researchers have been using for decades, but to use them as markers (variables) for defining and describing populations, and for mapping paths of human migration. Surname methods invite us to think about and use the sources in new ways. The lists of names themselves should be thought of as variables that have particular statistical, demographical and genetic characteristics.

The earliest recognized use of surnames as a marker for describing a population was by Charles Darwin's son George (1875) for a study of marriages between first cousins in England. Beddoe (1883, 1898) followed by using surnames from a 'Police Gazette', listing military deserters, to categorize and locate those individuals with indigenous and exotic surnames for Irish, Scottish and English, and determine the number in each group with varieties of skin pigmentation and eye color. This work is a good example of how the names we bear can label us with a culture, religion, race, ethnicity, place or history. In 1908 Arner employed a method that would eventually become the most popular technique applied to surnames. He used marriage licenses to define cousin marriages in eighteenth-century New York and nineteenth-century Ohio, and to derive probabilities of same-surname marriages. Surname studies over the firsthalf of the twentieth century specifically treated surnames as variables to quantify race and blood type groups and were mostly published in anthropological journals. Despite publications focussing on names, studies did not take a systematic approach until after the important publication by geneticists Crow and Mange (1965) where they introduced a systematic isonymy (same-surname) based model, intended as a surrogate genetic variable. Since this publication, others have employed surname lists in new ways, and it has become a special topic of human biology, physical anthropology and human genetics.

What attracts the attention of biologists and geneticists to surnames is that they are transmitted through the paternal line in many Western societies. One can think of the family name as an allele carried on the Y chromosome. Surnames, like DNA, embody a record of who we are, where we came from and to what degree we are related. Your father's Y chromosome is the same as his father's, his

grandfather's, and any of his male descendants (Jobling 2001; Sykes and Irven 2000; Underhill et al. 2001; Barbosa et al. 1997; Hill et al. 2000).

The objective of the biologist researching surname distributions is to define population structure and interpret the influences of demographic components (births, deaths, relatedness, migration) which are responsible for the spatial and temporal variation in gene and allele frequencies which ultimately play a role in the 'micro-evolution' of populations. Population structure is generally defined as the deviation from non-random-mating (Hardy-Weinberg Equilibrium), and includes components of inbreeding, consanguinity, associative mating, and geographical subdivision. Biologists' explorations of population structure have focused primarily on historical migrations and geographic distributions because migration directs the fissions and fusions, and the isolation of distinctive surname (gene) 'pools' with important consequences on genetic drift and the rates of genetic disease. According to Lasker (1985), the main point of surname studies "is to measure the different probabilities of finding the same surnames in different times, places, groups and, especially in marital partners."

The usual strategy for using surnames to define population structure and migration is to define one or more populations with a list of names which is representative of that population at a certain time: drawn from a census, for example. From the frequencies of surnames in a population one can calculate diversity, estimate levels of relatedness within that population, and infer whether recent in or out-migration has occurred. From lists of surnames from two populations, we can estimate their relatedness and the rates of admixture through migration. Computing surname frequencies at given time intervals can yield information about demographic rates, surname constancy, and past migrations (Lasker 1985, 1998; Relethford 1988).

Like historical geographers, biologists are aware that surnames are readily available. Lists of names are available in land records, probate records, marriage, birth and death records, tax rolls, voter lists, phone books, ship passenger lists, naturalization records and military records. Without surnames as raw material, biologists would have to rely solely on expensive and time-consuming genetic

analyses and theoretical models to answer questions about population structure. For earlier generations, for whom DNA is rare, spotty, and difficult to decode accurately, geneticists are often obliged to resort to meticulous reconstruction of family genealogies. Reconstructing genealogies and comparing them to specific DNA alleles of present-day relatives provides by far the best results for determining past rates of consanguinity, and kinship, but is time-consuming, expensive, and dependent on the quality of vital records.² Surname methods provide therefore a viable and necessary preliminary tool that can be employed before investing in a more comprehensive genetic study (Mascie-Taylor and Lasker 1988, 1995). In *Consanguinity, Inbreeding, and Genetic Drift in Italy*, Cavilli-Sforza et al. (2004) describes the use of lists of surnames from telephone books, electricity bills and voter lists as a means to characterize clusters of populations sharing similar names, and then uses these populations for selective sampling of blood for genetic testing.

When explaining surname analysis to those unacquainted with the methods, I am often met with scepticism. The methods are in fact well documented and tested and published in a great number of journals of biology and physical anthropology. It is true, of course, that errors creep into surname spellings, and cases of illegitimacy would make the assumptions false. But problems also creep into census name matching, family reconstitution, theoretical population genetics and molecular genetic analysis. Researchers are aware of potential problems with surnames, and most models take these problems into consideration. For large lists of names, the law of large numbers prevails, and these problems have little impact on the results. Lasker (1985) has suggested, for example, the use of large samples to reduce standard error rates, and I will suggest throughout this work other techniques which should be considered to improve the reliability of surname methods.

I am also aware of the criticism within the field of human biology, including Roger's (1991) article "Doubts about isonymy," where he raises in particular the problem of polyphyletic surnames—that is the presence of several founders with the same surname from different origins (or different Y chromosome). This

complication tends to overestimate the relatedness of a population. I minimize this problem because for the Quebec French Canadian records we know the number of founders associated with each surname and have adequate control of the spelling variants of each name. This type of information does not exist for many nations or regions and other steps need to be taken when processing data to ensure accurate results. Despite the salutary scepticism, the number of studies has increased, and have recently begun to incorporate larger study region. Surname studies have been strengthened by findings of researchers such as Sykes and Irven (2000), who in a randomly selected sample of British males with the same surname found that about half carried the same Y chromosome haplotype (Cavalli-Sforza and Bodmer 1991; Cavalli-Sforza, 2001; Menozzi et al. 1994). Other studies on a single surname have revealed similar results, indicating that many men of identical (or variant) surnames do in fact carry the same Y chromosome haplotype even if they know of no distant genealogical links (Jobling 2001; Paabo 1995).³ A recently published review article summarizing isonymy studies concluded that the number of studies has increased covering even more regions. The authors described an increase in large and small (micro) scale studies, but pointed to the failure to combine scales and get at the underlying hierarchy of population structures. In this study, I will work at different scales and try to add new context to this deficiency (Colantonio et al. 2003).

While historical and cultural geographers have used surnames for record matching, they have shown only a fleeting interest in their use as markers for mapping migration and cultural diffusion. In the 1950s, cultural geographers were inquiring about the "discovery and application of efficient and sensitive measures of the extent to which the cultural system and its major components have varied through space and time" (Zelinsky 1970). They were looking for a surrogate variable that could identify a cultural 'system' and show how features on the humanized landscape resulted from movements and distributions of different cultures. Among the variables that cultural geographers tested were religious affiliation and practice, and the settlement landscape, which included cemeteries, fences, barns, bridges, churches, street patterns, town centres, and place names (Zelinsky 1970).

Zelinsky's (1955, 1970) articles on the historical distribution of forenames and toponyms have advanced the furthest the uses of names (by geographers) as a variable for making empirical inferences about America's past cultural and ethnic makeup. Using samples from census records, Zelinsky explored the effectiveness of using forenames to distinguish and map cultural groups in the northeastern United States. Zelinsky recognized in the notes of his 1970 article the potential of surnames for research in ethnic geography and migration history, but little has been published by geographers on the topic.

A decade later, Porteous (1982, 1987), a geographer, demonstrated the value of surnames for England for the period 1539-1979, showing both a micro and macroperspective on the movement of the surname 'Mell'. He employed the number and percentage of Mell births per county, and calculated the diffusion and location quotients of Mells throughout the country. Another work highlighting the use of surnames was West's (1986) *Atlas of Louisiana Surnames of French and Spanish Origin.* Here West uses telephone directories to record French and Spanish surnames and puts them into historical context by creating maps to show the present-day surname distributions resulting from past migration.

Quebec makes an especially useful test region for application of surname methods. First, few places possess such comprehensive primary sources. Due to the work of organizations like the Programme de recherche en démographie historique of the Université de Montréal and BALSAC of the Université du Québec à Chicoutimi, records available are nearly complete, well indexed and accompanied by documentation of errors and data collection methods (Bouchard 1992). Through the excellent documentation, we have knowledge about the number of founders who carried each surname. This allows for the creation of specific subsets (samples) restricted to single founders (surname with one 'pioneer') and provides an easier and more accurate way of describing the French Canadian population at one moment in time or tracking the movements of people (surnames) over several time periods. This disposes of the objection of Rogers (1991) where we risk overestimating relatedness of certain surnames if we cannot

distinguish polyphyletic surname (different founders with the same surname) from monophyletic (single-founders). In chapter 2, I provide information on the degree of such overestimation, and throughout this work I provide the numerical differences between using full name lists (e.g. full census) versus selections from single origins.

A further reason for focusing on Quebec is that the genealogical sources have already yielded important insights into processes of migration, settlement, and the regionalization of names, yet they render only pieces of a larger picture that has yet to be put together (Brunet et al. 2001). Many of the questions about the founding population, and about the demographic characteristics of the French Canadian population before 1800, have been fleshed out, but questions remain about nineteenth-century demography, the French Canadian exodus into New England, Ontario and the prairies, inter-provincial migration, regional surname organization (networks), and rates of urbanization (Bouchard 1991). Specialists of migration and settlement have employed detailed local records such as life histories, genealogies and aggregations from parish and nominal censuses (St Hilaire, 1988, 1996, 2002). These data are usually examined within the context of understanding the effects of life cycles, job skills and kinship and marriage networks.

Quebec scholars have developed international expertise in historical demography and have created two computerized family reconstitution databases. These two databases fueled research in demography and genetics as well as inferences about processes of migration and settlement. The first database, developed by the Programme de recherche en démographie historique (PRDH) of the Université de Montréal, contains almost all parish registers (baptisms, marriages and burials) for the period prior to 1800, which have been matched and turned into family genealogies for 95 percent of the population (Desjardins 1998; Bélanger and Landry 1990). This near-perfect tool has been used for a wide variety of studies in history, genetics, linguistics, anthropology and genealogy. Topics examined in detail about the ancien régime include longitudinal analysis of biological kinship, epidemics, fertility, infant, child and adult mortality,

orphanhood, effects on marriage behaviour, and an examination of founders and their contribution to unique genetic disorders.⁴

The second computerized family reconstitution is being completed by BALSAC at the Université du Québec à Chicoutimi under the direction of Gérard Bouchard. The database initially contained almost all records of births (baptisms), marriages and deaths from the Saguenay over the period 1838 to 1971, and like the PRDH database it has been linked to generate continuous family genealogies and contains a subset known as RETRO which contains a province-wide sample of genealogies. Already extended to most of eastern Quebec, the BALSAC database will cover the entire province, connecting to the PRDH database. Together they will ensure near-complete genealogical history of the entire provincial population of French colonial origins. Perceived as a tool to study demography and genetics, the BALSAC database has resulted in numerous publications on the population structure, many of them useful in analysis of rare genetic disorders (Bouchard 1992; Tremblay et al. 2001, Scriver 2001).

Both projects (PRDH, BALSAC) have contributed to specific works on isonymy, including the geographic organization of surnames, migration and settlement (Brunet et al. 2001). From telephone directories Bouchard (1985) collected data for the 15 most popular French Canadian surnames and applied surname measures to distinguish fourteen geographic regions of Quebec. He found differences between surname affinities in eastern and western Quebec, noting that Eastern Quebec regions (lles-de-la Madeleine, Charlevoix, Saguenay, Bas Saint-Laurent and Beauce) have considerably lower diversity (greatest homogeneity). He also noted less dispersion of surname concentrations (more mixing) in the west, and a greater concentration of surnames on the Gaspésie peninsula which masked the diversity resulting from the subsequent settlement by the Acadians and English.

Gagnon and Toupance (2002), using pre 1800 data from the PRDH, compared kinship estimates given by surnames and genealogies in paternal and maternal lineages. Results of using same-surname (isonymy) measures proved to be very close to the genealogical estimates, and thus were in favor of using surnames as

markers in the Quebec population.

Besides the genetic and isonymy studies from which migration can be inferred, considerable research addresses the origins of French settlers in Canada and the French Canadian exodus to New England (Bouchard et al. 1995; Landry 1990 and Choquette 1985, 1999; Ramirez, 1995). St-Hilaire has published evidence for migration into and out of Quebec City, and compiled considerable evidence of marriage fields (2002). Quebec's rich primary records have not, however, been fully exploited to study migration between provinces of Canada or within the province. The exodus of French Canadians to New England has been estimated, and close attention has been given to its political and ideological meaning, but little research has been done to map the specific origins and destinations of migrants, or to put these moves into larger historical context. Exceptions include the work of Ramirez (1991), Vicero (1968) and Roby (1968) and Freneette (1989) who indicate intricate, precise, local pathways between villages of origin and destinations in New England.

Some of the most promising historical migration research that employs geographic mapping for Quebec is found in the *Historical Atlas of Canada*, and the *Atlas historique du Québec: Population et Territoire*. Both series feature maps showing movements of people within Quebec and from overseas. In the Québec atlas, for example, the Beauchamp surname is shown expanding from its origin around Montréal into the surrounding region between 1720 and 1899 (Olson 1996) and the Boisvert surname found around Quebec City is shown leaving for New England (1853-1900) and concentrating in specific locations of New England (Boisvert 1996). These two maps illustrate how migration to certain regions, both within and outside Quebec, was not random, but instead was directed along definite paths, each surname of groups or surnames (a population) following a consistent current.

Despite all of the research completed on Quebec relating genetic, settlement and migration history, there are still some gaps. My research on surname methods will incidentally provide some new information to unravel movements of settlers from first landings in the St. Lawrence Valley inland, but its primary purpose is to expand the methodological foundations for historical geography and other academic fields where migration is an important consideration, but which have yet to employ surname methods. Like the unrivalled computerized genealogical databases of Quebec which have provided data and methods to a large interdisciplinary group, I believe surnames and the associated methods I present could be potentially useful in many disciplines. While new lists of names become available online and through history societies, the need for tested and standardized methods will increase. In addition, through the use of cartographic mapping (visualization) I intend to provide an illustration of Quebec's changing population structure. By simplifying statistical results maps allow us to express complex and elaborate observations in a comprehensible way. Maps to geographers, like piecharts and graphs to statisticians, are an indispensable tool for reasoning about quantitative information. They are instruments that make it easier to describe, explore and summarize massive amounts of spatial data. I will use maps throughout to summarize my findings.

Throughout this work I refer to migration in the broadest of terms, as a movement of an individual name or a population of names. I am, however, aware of the many definitions of migration and the important implications migration has on changing communities politically, culturally, socially and genetically (Lewis 1985; Fix 1999). Migration can take many spatial and temporal forms and can be interpreted in different ways. The most common methods include examining local or inter-urban movement, regional movement (intra or interstate), and global (intercontinental) movement. Because migration is a dynamic process, involving limitless number of variables, the definition is often situational and dependent on the investigator's needs and types of data available (Fix 1999; Newman and Matzke 1984; Lewis 1982; White and Woods 1982).

I intend to show that surname methods provide flexible tools that allow for examining migration at various spatial and temporal scales. This is not always an option with other sources and methods. Throughout this work I will change spatial-temporal scales by aggregating or de-aggregating data and show how easily surnames can be adapted to this approach.

As migration is one of the components of the demographic model, researchers from various disciplines have provided a number of theoretical and mathematical models for describing and generalizing the migration process. In particular historical geographers are interested in the spatial components and the effects of these characteristics on varying aspects of origin and destination communities. Historical geographers and others have framed most of their hypotheses regarding migrant behaviour around cartographer/statistician E.G. Ravenstein's "Laws of Migration" (1885), which were the result of research on nineteenth-century British and western European internal migration. His laws are based on the notion that migration is economically motivated. He concludes that most migrants move only short distances; the volume of migration increases with expansion of industry and commerce; primary migration currents produce counter-currents; migration proceeds in steps; females tend to migrate short distances; and rural residents are more likely to migrate than urban residents (Grigg 1977). Geographers have further refined these concepts to include a spatial examination of the movements. For instance, Olsson (1965) examined variation in migration distances as a function of a hierarchy of central places. The spatial structure of migration has been further integrated into gravity models which explains migration in the context of places pushing and pulling people based on certain economic, political, cultural, environmental conditions (Haggerstrand 1975). Though I do not go into detail about applying these models to surnames I hope you will see the potential of using names as a variable for the various models.

Through this work I will also provide evidence of how surname methods can be used as a surrogate variable to test or examine the many migration theories, models and hypotheses that have been developed.⁵

In chapter 1, I review the primary sources and the name samples used throughout. I include a review of how I derived the samples from sources such as marriage records, censuses and phone records, and I discuss the reliability of the sources and the samples, as well as common problems like name misspellings, and techniques to address these problems. I follow with a review of the geographic sources available for geocoding and mapping the statistical results; and finally I provide information as to how I aggregate data and available sources in order to cope with boundary changes over the long period between 1621 and 1900.

In chapter 2, I review surname methods and introduce statistical methods from other fields which provide alternative ways of assessing regional surname organization and migration. Each method is treated by an application from the population of Vermont between 1840 and 1880. The case is useful because it has a smaller and simpler structure than Quebec. The example will highlight the changes associated with a high rate of out-migration of Vermont born and in-migration of French Canadians. My intention in this section is to examine the various surname methods that have been used for other studies and provide context to different multivariate methods that are well suited for surname data, but unfortunately used little. The chapter is divided into five parts: 1) isonymy and related methods, 2) ordination and multivariate statistics, 3) cluster and fuzzy cluster analysis, 4) Relevance networks, and 5) Bayesian probabilities.

The purpose of chapter 3 is to fill some gaps in what we know about Quebec migration, settlement and population structure before 1800. This chapter applies three methods from the previous chapter, and each method is chosen to answer a distinct array of questions about the geographic structure of names. It begins with the use of same-surname (isonymy) methods to examine diversity, regional genetic differentiation, and name affinities between regions. As we shall see, same-surname methods are useful for extracting the relative similarities among all locations because the results provide a measurable value for each location. I follow with the use of relevance networks to identify the strongest paths and explore the links between surnames and parishes. I conclude with the use of cluster analysis to interpret the hierarchical organization of 'relatedness' of populations from surname similarity. In each of these cases I compare the results of the several methods and, when possible, use available literature on Quebec history and demography to appraise its accuracy and its limits.

In chapter 4, we move a century ahead using the 1881 census, and I apply the same three methods as in the previous chapter. Use of the same methods with a subsample of the same single-founder surnames allows for comparisons between

1765 and 1881. I shall pursue many of the same questions regarding the geographic structure of names, and in addition, I will argue that earlier surname structure directed later settlement, migration and overall geographic name organization.

In the remainder of the thesis, I present case studies which apply the several methods at different spatial and temporal scales. Each study is intended to answer a specific question using one or more of the methods, though not necessarily to provide a comprehensive review of any one topic. I begin by using surnames to calculate the rates at which local French Canadian populations urbanized, and their likelihood of moving to Quebec City or Montréal and eventually to other Canadian provinces. This case study uses marriage records to provide new details on the rapid urbanization of urban areas during the late nineteenth-century, focussing specifically on French Canadians rather than the thousands of trans-Atlantic emigrants. In addition, using phone records, I provide a brief glimpse into the possible migratory paths of the surnames moving from Quebec, and I evaluate whether there might be certain names or regions more inclined to move from Quebec. The evidence I can provide from surname frequencies would be hard to confirm using other methods due to a lack of sources. In a second case study I examine the nineteenth-century long-distance migrations of French Canadians to Vermont, Maine and New Hampshire, and some present day distributions of French Canadian surnames in the other Canadian provinces. In the third case study I track the migration of a population of lumberjacks to the Mauricie region in relation to the curious present day overrepresentation of a drug-resistant strain of tuberculosis believed to be more than 150 years old and possibly linked to this population. In the final case study I use marriage records from 1700-1900 to track the geographic distribution of same-surname marriages and the likelihood of intermarriage between regions. Here I use the earlier results of regional isonymy to examine how surname diversity affects the percentage of same surname marriages, and how researchers can use simple diversity measures as a way of extrapolation.

My hope throughout is that you will understand the many different ways that lists of surnames can be used to infer different population characteristics and

migratory patterns. You will be exposed to methods that have been used extensively by researchers to determine population surname structure (and genetic structure) in different regions throughout the world. I intend to describe the strengths and weaknesses of these methods, as well as provide a description of new tools and methods which I believe add a new perspective in examining population simultaneously at several spatial scales, a perspective missing in many same-surname studies.

NOTE

¹ For work on migration also see Thornton 1998; Widdis 1987; Stephenson 1980; Haines and Steckel 2000

² Information on rates of intermarriage, mating (kinship) patterns, relative isolation, demographic characteristics and migration can provide information on past and presentday geographic variation in gene frequencies. ³ For examples of one name studies see : 'The Guild of One-named Studies: http://www.one-name.org/

 ⁴ See Bideau 1997; Charbonneau et al. 1967; Charbonneau 1985; Choquette 1997; Desjardins 1997a, 1997b, Desjardins et al.2000; Landry and Légaré 1987; Nault et al. 1990
⁵ For a review of migration concepts and models see: Fix 1999; and Lewis 1982.

Chapter 2

Quebec surname data sources

In this chapter I review the primary records and geographic data used throughout the thesis, in particular the logic behind my selection of names, as well as data problems (e.g. spellings), sample sizes and related technical considerations and modifications.

Biologists and geneticists studying population structure usually attempt genealogical links painstakingly by family reconstitutions. Throughout this study (1621-1900) I will infer genealogical links using only surnames, by restricting my application to a specifically selected set of single founders. Knowing which surnames are associated with a single founder provides a reasonable basis for assessing population structure, tracking population movements and settlement patterns. By using single founders you can think of each name as a complete ascending genealogy that contains all of the branches (members) of the family tree.

Our immense knowledge of the French Canadian founder population is a result of the excellent baptismal, marriage and burial records that were diligently kept by the Catholic clergy. Based on the methods used in France, record-keeping was practised from the first days of the colony by curates and missionaries located in the parishes strung along the St Lawrence River to track the population growth and monitor marriages births and deaths. A utopian ambition, a disciplined bishop and the presence of Counter-Reformation orders of well-educated priests (notably Sulpicians and Jesuits) ensured entries according to the rules of the Council of Trent (1545-1563), and the colonial government formalized the parish registers as the *état civil*. Catholic registries were used as means of record keeping, and Cavalli-Sforza et al. (2004) explains, as a means to keep track of levels of consanguinity. Individuals had to seek approval from priests that they were not closely related before they could be married (usually relationships of third cousins were acceptable, but nothing less).

Despite the introduction of a few new regulations in registration procedures,

record-keeping was continued after the British conquest and expanded to cover Protestants and others. Using records which are so detailed, so precise, and so nearly complete, the PRDH, with the help of many genealogists, created a computerized family reconstitution database which provided users convenient access to well documented and accurate data (Jetté 1983; Kuczynski 1930; Charbonneau et al 1993, Nault et al. 1990; Choquette 1997; PRDH).¹

The Quebec 'founder population' consists of approximately 6800 men who meet the defining criteria: each of them arrived from France prior to the eighteenth century and had at least one son who married in Quebec. These 6800 immigrants carried about 3300 surnames, of which about 2300 were unique; that is the pioneer did not share his surname with any other pioneer (Charbonneau 1993).

The term founder deserves special attention as various authors employ it in different ways. The traditional use of the words 'founding population' refers to the entire number of original colonizing members including women. In this research, the term is restricted to a single-founder population of men who did not share their surname with another founder. Jean Bessette, for example, who settled in 1642, would be considered a single founder since all of the Bessettes are believed to descend from Jean. All of the male Bessettes except those adopted or illegitimate presumably carry a similar Y chromosome. In contrast is the term multiple founders, which refers to unrelated immigrant men who bore the same surname, such as the six men named Auger who arrived in the seventeenth century from different parts of France, each carrying the same surname but quite possibly a different Y chromosome. We have, in fact, little information about the earlier descendents of the Auger families in the Old World or how many unique genealogical lines exist: they may go back to the thirteenth⁻century.

The main Quebec datasets used in this study include a surrogate 1765 census, marriage records covering the years 1621 to 1800 from the PRDH database, marriage records from 1800 to 1900 from the Drouin collection, the 1881 Census of Canada for the Province of Quebec, and 2001 Bell Canada telephone directories.

Removing surnames that have more than one founder removes some of the
noise from multiple founders, but I also have to consider surnames that become extinct and the ones that have a large number of aliases or nicknames, such as 'Brien dit Desrochers'or 'Ouellet dit Roch'. These aliases or compound names, translated as "said names," were alternative last names used to distinguish extended families bearing the same name. Alternate names were in extensively used throughout the seventeenth century and there widespread use must be considered when selecting the sample. Figure 2.1 describes the distribution of ditnames and illustrates their distribution in the pre-1800 set of single founders. Examining their distribution we find a varying range with the majority (824) of names with only one dit-name while Ouellet had 72 different combinations. Desjardins et al. (2000) describe the distribution of Quebec family names and explains that around 1870 government administrators issued a decree that all families with more than one surname (dit-name) to choose one. The selection of the name was done arbitrarily and this does contribute to the eventual decline in certain names.

To deal with the problem of dit-names, I reduced my sample (for some analyses) to a smaller, 'cleaner' set of 832 single-founder surnames by introducing more restrictive criteria. As described above some names had more dit-names than others before 1800 and theire relative numbers might provide a means of selection. To make my sample more specific I performed two tasks. First, I created a database containing all 2300 single-founder surnames. I summed the number of times each of the surnames showed up in the key datasets (1765 c ensus, marriages 1621-1799, 1881 census) at different times. The final sample had to meet the following criteria: surnames which showed up at least once in the surrogate 1765 census, had at least two marriages in the interval 1780-1799, and had fewer than 10 reported aliases (dit). The final sample included 832 single-founder surnames (Figures 2.1, 2.2.). The objective was to select names which existed and had adequate representation during the study period. Bias caused by the use of dit names is certaintly a possibility, but would be most pronounced when using the 1881 census, as a result of families changing their names.

Because I can select surnames that represent only single founders, and

because surname spellings have been standardized, there is much better control over the meaning of the results. Most studies cannot control for this parameter and therefore risk over-estimating the number of same surnames and relative relatedness of populations. We shall see later that the 1881 data is less tidy, but its value is nevertheless immensely reinforced by the availability of the information about the initial founding population.

1765 Census

A census of heads of household was compiled by British authorities between 1760 and 1765 to cover the entire colony upon transfer to British jurisdiction. For the two large urban parishes of Notre Dame de Montréal and Notre-Dame-de-Québec (Québec City), however, records have never been found, and in about 20 other rural parishes there was serious undercounting. The problems with the 1765 census have been reviewed by Landry (1975). To produce a more comprehensive data set, I created a surrogate census by extracting all fathers mentioned in baptismal records between 1765-1775 (extracted mother's and father's full names and identification number for 58,093 baptisms). Using the unique identifier created by the PRDH, duplicates were removed (couples who had more than one child in the interval) and the unique fathers were extracted (head of household). The final set had 25,034 individuals (fathers), with a total of 5239 different surnames (including double names) dispersed throughout 121 parishes. For the 1765 surname analysis, I took the subset of 1347 single-founder names, removed the second of the double names, and kept only the 832 names which matched my criteria. To consider whether the numbers of baptisms were representative, they were compared with Tanguay's population totals published in the 1871 Canadian Census.² In each case the numbers of baptisms were within 10% of Tanguay's figure for each place considered. The most frequent surnames in my surrogate 1765 census were Roy (151 times), Morin (132) Gauthier (116) and Gagnon (112).

1881 Canadian Census

The 1881 Canadian census as used in this research is a digital version created by the Institute of Canadian Studies at the University of Ottawa, the Church of Jesus Christ of Latter-day Saints, and volunteers from family history societies. The

database contains 4.3 million records, covering the entire enumerated Canadian population in 1881. I received an early version of the Quebec 1881 census database from Prof. Dillon from the Université de Montréal. The Quebec file contains 1,358,651 records (people) and 46 fields.

Although Dillon (2002) and her team of graduate students had conducted a series of cleaning steps to improve the data set, there were still a number of final tasks remaining before I could use it. The group confirmed and cleaned the spellings of district and sub district names, and created an identifying number for each household, but only modest work had been done to standardize spellings of surnames, religion, birthplace, and nationality. ³ A release by the Church of Jesus Christ of Latter-day Saints (LDS) contains the same surname spelling problems (Dillon and Desjardins 2002).

The 1881 Quebec census database has many of the common spelling mistakes associated with census transcription. Surname spelling mistakes arise through a combination of original census compilation as well as the transcription process from paper (microfilm) to computer database. Among the common problems in the initial transcription were that illiterate people would not know how to spell their own names, enumerators would misspell names because of speech accents, and enumerators would misspell when copying information from one form to another. Further problems were introduced because transcribers could not read handwriting or microfilm was not clear. In the 1881 database it is the French Canadian names that are most often misspelled, and the organization of the volunteer effort was flawed by failing to find French Canadian volunteers and provide templates and software with French accents.

From the 1881 population of Quebec I selected 235,165 individuals (approximately 20%) who were described as married or widowed men older than 15. Because of problems with surname spellings, I carried out a surname standardization and spell check. To simplify the correction process, I separated the sample by 'ethnic origins' and split it between French Canadians and other ethnic origins (e.g. American, British, Irish).⁴ French Canadian names were standardized according to the list of the 3300 unique founders provided by the PRDH. To make the spelling corrections, I used a dictionary containing the names of unique founders as well as a list of recognized spelling variants reported by the PRDH.

I explored the statistical properties of the entire set of names, but the French Canadian single-founder sample discussed here and used throughout totaled 48,000 individuals (5% of total population, 24% of men 15 and older) and included 1300 of the 2300 total single-founder surnames. The most frequent surnames in the 1881 subsample were Gagnon (1491 times), Roy (1398) Cồté (1383), Tremblay (1232), and Bélanger (1091).

Quebec Marriage Records, 1621-1900

For the selected 832 single-founder surnames discussed above, I obtained from the PRDH all 17,000 marriage records for the years 1621-1799. These records contain the groom's and bride's surnames, name of the parish where the wedding was celebrated, and the marriage date. To develop a sample of nineteenth-century marriage records, I used the Drouin index (series 1 and 2) to collect marriages of men bearing one of 35 single founder surnames, totalling around 33,000 marriages. The Drouin lists almost all of the French Canadian marriages in present-day Quebec, 1760-1935. Currently, the Drouin volumes are available on micro-film and in bound volumes, but the American-French Genealogical Society (AFGS) recently began a project to transcribe the entire series of approximately 1,000,000 marriages into a computerized database. In exchange for my volunteer hours transcribing Drouin marriage records, project leader Roger Beaudry provided me with completed portions of the database.

I can find no documentation confirming exactly how accurate and complete the Drouin records are, but discussion with Quebec's research community has provided assurance that besides the normal spelling mistakes, the records are almost complete for Catholic marriages in Quebec, while there is only sporadic coverage of marriages outside the province (Ontario, New England and New Brunswick). The source provides date of marriage, parish where the marriage was celebrated, the husband's surname, wife's maiden name, the father's name and the mother's maiden name for both bride and groom (see http://drouininstitute.com/faq-en.html).

United States, Vermont Census: 1840, 1870, 1880

The 1840 Vermont census as used for this study was transcribed by Ancestry.com and extracted via the world-wide-web. Selecting heads of household age 21 and older, I obtained 53,497 persons (19% of the total 1840 population). The 1880 Vermont census was transcribed by the Church of Jesus Christ of Latterday Saints (LDS), and the entire set of males age 21 and older totalled 78,879 (24% of the total 1880 population).

To improve the accuracy of the 1840 and 1880 data, I 'corrected' the rare spellings listed 5 or fewer times using surname dictionaries which were created for this research. Correction reduced the list of unique names from 11,281 to 10,300 for the 1840 data and for 1880 from 15,002 to 12,101.⁵

I employ the 1840 and 1880 Vermont census in the methods chapter as an aid to explain surnames techniques, and in a later case study the 1870 and 1880 censuses for Maine, Vermont and New Hampshire are employed to track French Canadian migrations to New England. The census records for Vermont, Maine and New Hampshire were collected from Ancestry.com via the world-wide-web using an automated data extraction program written in Perl (computer language). Although Ancestry.com had processed the surname spellings with a soundex spell checker, problems particularly remained for French Canadian names.

The Church of Jesus Christ of Latter-day Saints (LDS) has transcribed the United States Census of 1880 and organized a computerized database containing 50,475,366 inhabitants of the 38 states and eight territories of 1880. It is the largest complete census thus far to be entered into a fully searchable database. To ensure the integrity of the original content and organization, the LDS received assistance from the Minnesota Population Center of the University of Minnesota. The database has a very flexible search engine and the information on surnames, age, nationality and occupations can be easily extracted. The database is available from the LDS as a set of 56 cd's or can be accessed via the world-wide web through the LDS website.

From the 1870 and 1880 Vermont, Maine and New Hampshire census records I drew heads of household 21 and older. The data include information on

birthplace, age, county, and township. The surname spellings were post-processed by the LDS (1880) and Ancestry.com (1870) using a soundex spellchecker, and further processing was completed with the custom spell dictionary. Since I was planning to use this census to corroborate movements of people (surnames) from Quebec, as part of the spell-check process the French Canadian names were standardized, as described earlier.

Geographic and miscellaneous sources for Quebec

A Geographic Information Systems (GIS) geodatabase was the backbone of the project, providing data organization, querying, analysis and mapping. A geodatabase is similar to a traditional database (e.g. Oracle, MS-Access), but adds a geographic component to the tabular data and preserves a coordinate system and geographic projection. What is so useful about using a geodatabase is the flexibility for moving systematically between spatial scales and modeling the spatial and tabular data simultaneously. This flexibility will complement my intention to study surname distributions at multiple scales with data that are easily mapped when analysis is complete.

To handle both local and regional scales, between 1621 and 1900, there is a need for geographic sources at a number of scales to match the numerous census and Catholic parish records. For regional analysis of Quebec, eastern Ontario and New England, digital layers at 1:1,000,000 scale are employed. To map the parishes (1621- 1900) point layers from two sources were combined. The first, provided by St. Hilaire (2002), contained parish opening dates and associated geographic coordinates for 2700 parishes. St Hilaire refers to these as population nodes, and they were digitized by hand from the 1:250,000 topographic sheets (NRCAN). The 'points' are village centers or road intersections which can be interpreted as locations of churches and cemeteries (shown explicitly on larger scale topographic maps). The second point source was compiled from paper maps and a list of parishes found in the *Inventaire des registres paroissiaux catholiques du Québec 1621-1876*. Coordinates for each parish were linked to marriage data (1621-1900), the 1765 and 1881 census. The name and location of the parishes

used throughout this work for the pre nineteenth century analysis are illustrated in Figure 2.3 (see Table 2.1).

The 1881 Quebec census is organized into 61 districts and 816 sub districts. Census districts for this project were created by combining maps from Drouilly's *Atlas des élections fédérales au Québec, 1867-1988,* and a set of GIS boundaries provided by St Hilaire for 1871 and 1891. Drouilly's 1881 paper maps were georeferenced to St-Hilaire's computerized (GIS) maps to match the boundaries.⁶ The map revealed few boundary changes between 1871 and 1881, and no changes between 1881 and 1891. Using the 1891 GIS file the 1881 file was recoded.

After examining the irregularity (uneven boundaries and size) of the 1881 boundaries through a series of statistical tests and initial surname analysis, I concluded it would be difficult to interpret unabashedly, so I formulated a set of geographic boundaries providing a more practical and simpler spatial unit. To do this ninety-one hexagons were overlaid (approximately 40 km across) on a map of the 815 subdistricts. Each hexagon absorbed the name data for the set of population 'nodes' it enclosed, so that I could calculate measures for each of the 91 hexagons. This procedure is similar to hexagon binning found in the Splus (and R), statistical libraries and is a form of bivariate histogram useful for visualizing the structure in datasets with large numbers (Ihaka and Gentleman 1996). A 40 km hexagon was used because St Hilaire (2002) had found this distance approximates median marriage distances in the mid-nineteenth century approximated this distance. Figure 2.4 illustrates the hexagon map used throughout the work. To help you find your way around the map, I have identified the largest population node belonging to each hexagon and I refer to this as an identifier throughout the work.

For all of the households reported in the manuscript census of 1881, Dillon and her team verified Canada-wide the 815 sub districts. To alleviate the problem of accounting for very small populations in the statistical analysis (often 'unorganized territories'), 135 of the 815 population nodes were aggregated with their nearest geographic neighbour (Figure 2.5).

An additional geographic source used throughout is a map of 23 cultural

regions created by St Hilaire 2000 and BALSAC. The boundaries were created from various historical sources and have been used in numerous publications . They are intended to represent an approximate a boundary distinguishing cultural regions in Quebec. The 23 regions are used only in one case study, but throughout I refer to these regions in my explorations and interpretation, so you may need to refer back to Figure 2.6 if you are not familiar with Quebec geography.

To corroborate surname movements, I employ a large array of classical historical sources. For example, characteristics of the population are well documented for the seventeenth and eighteenth-century regional patterns including settlement, parish opening dates, organization of seigniorial lands, and demographic characteristics of the population (Courville, Robert and Séguin 1995; Courville 1996). What we know about the regional demographic history is based on the work of Tanguay and Jetté and the family reconstitution database created by the PRDH. Tanguay's work, in particular, resulted in demographic information being published in the 1871 Canadian Census (Henripin 1972a, 1972b): yearly (1610-1875) county-level data on the numbers of baptisms, marriages and deaths for French Catholics up to 1871. Tanguay's figures were originally interpreted and summarized by Henripin in 1961 and revised in 1973 (Charbonneau et al.1973) The revised publication provides a provincial summary of births, deaths and marriages (1680-1950) in five-year intervals. Recent publications related to the PRDH project have resulted in more accurate interpretations of the regional demographic characteristics before 1800, but the work of Tanguay and Henripin remains valuable for the period 1800-1871.⁷

NOTE

² The 1765 census is also found in the Rapport de l'Archiviste de la Province de Québec and has been transcribed by the PRDH and entered into their master database.

³ Dillon and Desjardins (2003) reviewed progress on cleaning the 1881 census in a paper titled "The Historical Demography Research Infrastructure."

⁴I then summarized the surname and nationality fields to provide an aid for correcting the surname spellings. The first nationality query produced about 1200 different nation variations. I then standardized the nations and after another summary reduced the list to about 280. There were about 44,000 different surname variations. I proceeded first by correcting the obvious typographical errors and cut the list by about 1000. I then joined the list to the 3300 founders surnames that I received from the PRDH. Only 2000 of the 44,000 surnames matched. After realizing that many of the surnames were misspelled or did not conform to the standardized spellings, I split the list into French Canadian and Non French Canadian (i.e. American, British) using the nationality field. Using a PERL program (package wwwlib) I automatically matched the list of 14,000 non-French Canadians (30000 FC) with all of the phone users in Canada to see which surnames were matches and were potentially spelled correctly (Bell Canada 2002). This step was approached with the assumption that if the surname did not exist anywhere in Canada in 2002 then it was possibly misspelled and should be checked. There are few options for doing this without having to go back to the original census page and even then the enumerator may have spelled the name wrong. Nine thousand surnames checked out as true or showing up somewhere in the Bell Canada 411 online phone directory. The remaining 5000 names that did not match in the phone directory were manually spell checked. Using three surname spell-check dictionaries, all 44,000 thousand different surnames were corrected. The first dictionary contained the 3300 French Canadian standardized surnames; the second contained only anglonames created from a file from the U.S census bureau and Guppy (1890) H.B.Guppy, "Homes of Family Names of Great Britain" Harrison and Sons, London, 1890; the third file would combine the previous two files. Each of these dictionary files provided a semi-automatic way of correcting and standardizing the surnames. After surname corrections the database was reduced from 44,000 different surnames to 26,000.

⁵ Palgrave (2004) discusses possible linguistic aspects that led to surname deviants. Also see Bardsley (1996)

⁶ For more information on editing and correcting name spelling see Christen et al. 2002.

¹ Because record-keeping was precise in Quebec from the time of settlement, and because the *Programme de recherche en démographie historique* (PRDH) followed rigorous reconstituting methods, errors make up only a small percentage of the total number of entries in the PRDH database. Theoretically, the population register is a nearly perfect tool, especially since it is exhaustive and continuous. http://www.genealogie.uMontréal.ca/en/main.htm

⁷ Henripin's figures have been used in recent publications, such as the *Atlas du Quebec* and the *Historical Atlas of Canada*. They have provided the basis of what researchers know about natural increase and fertility for the Province of Quebec before the nineteenth century (Nault et al 1990; Courville, Robert and Séguin 1995, Courville 1996). Henripin's data provide accurate coverage for Quebec before 1800. After this date, because of increasing regional variability in natural increase and no provincial level family reconstitution databases like the PRDH we have little nominal information on the demographic characteristics of the population after 1871.



Figure 2.1 Number of dit-names versus the total number found in the PRDH marriage records between 1621-1799. For example, the marriage records of single-founder surnames (1621-1799) reveal that 824 surnames have a single dit-name that, and 1 surname (OUELLET) has a dit-name 72 different dit-names.



Figure 2.2 Date single founder surnames arrive in Quebec



Figure 2.3. Quebec parishes used for the seventeenth and eighteenth century surname analysis (see table 2.1). Note: Some parishes with low numbers were aggregated for analysis.



Figure 2.4 Hexagon map used for the 1881 Quebec census surname analysis. The place name used for each hexagon represents the largest population as defined from the single-founder sample extracted from the 1881 census



Figure 2.5. Total population 1881 Quebec Census (Log 10) (Sample total, 250,000)



Table 2.1. Quebec parishes and opening dates used for the seventeenth and eighteenth century surname analysis (see figure 2.1 for map).

Parish Name	Parish Id	Opening Date	Parish Name	Parish	Id Opening Date
St- losenh-de-Beauce	401	1738	Pointe-aux-Trembles	1901	1674
Ste-Marie-de-Beauce	402	1745	Rivière-des-Prairies	1902	1687
St Eropcois de Beauce (Beauceville)	402	1745	Longue-Pointe	1903	1724
Service (Deauceville)	403	1700	Sault-au-Récollet	1905	1736
Beaumon	001	1092	Lachine	2301	1676
La Durantaye	602	1693	Ste Anne de Bellevrue	2302	1703
St-Vallier	603	1713	Deinte Cleire	2002	1703
St-Charles (Bellechasse)	604	1749	Pointe-Claire	2303	1713
St-Mathieu-de-Beloeil	625	1772	St-Laurent	2304	1720
L'ile-Dupas	701	1704	St-Laurent	2304	1720
Lavaltrie	702	1732	Ste-Geneviève (Pierrefonds)	2305	1741
Berthier-en-Haut	703	1751	Rivière-Ouelle	2501	1695
Berthier-en-Haut	703	1751	La Pocatière	2502	1715
Lanoraie	704	1732	Kamouraska	2503	1727
Saint-Cuthbert	705	1770	St-André (Kamouraska)	2504	1791
Boucherville	1001	1668	Laprairie	2701	1670
Longueuil	1002	1701	St-Constant	2703	1752
Chambly	1002	1706	St-lean-Francois-Régis (St-P	hilinne) 2704	1753
Cap de la Madaleiro	1003	1700	Repentiony	2801	1670
Chomplain	1101	10/3	l schensie	2802	1675
Batiasan	1102	16/9	St-Sulpice	2002	1706
Dauscan La Dérada	1103	1680		2000	4704
La Perade	1104	1693	LAssomption	2804	1724
Ste-Geneviève-de-Batiscan	1105	1728		2805	1750
Bale-St-Paul	1201	1681	St-Roch-de-l'Achigan	2806	1787
Les Eboulements	1202	1733	St-François-de-Sales (lie-Jés	us) 2901	1702
Petite-Rivière-St-François	1203	1733	St-Vincent-de-Paul	2902	1743
St Francois-Xavier	1203	1733	Ste-Rose	2903	1745
L'Isle-aux-Coudres	1204	1741	St-Martin-de-Ile-Jésus	2904	1774
La Malbaie	1205	1774	Pointe-de-Lévy (Lauzon)	3001	1679
Châteauguay	1301	1738	St-Nicolas	3002	1694
Oka	1501	1727	St Henri-de-Lauzon	3003	1679
St-Eustache	1502	1769	L'Islet	3101	1679
Parish Name	Parish Id	Opening Date	Parish Name	Parish id Op	ening Date
St-Roch-des-Aulnaies	3102	1734	St-Augustin	4404	1693
Saint-Jean-Port-Joli	3103	1767	Deschambault	4405	1705
Lotbinière	3201	1697	Les Écureuils	4406	1742
Ste-Croix (Lotbinière)	3202	1727	Québec	4501	1621
St-Antoine-de-Tilly	3203	1702	Beauport	4602	1673
Deschaillons-sur-Saint-Laurent	3204	1741	L'Ancienne-Lorette	4603	1676
Rivière-du-Loup (Louiseville)	3301	1714	Charlesbourg		1679
Meskinongé	0001	11.14	01- 5	4605	
St. Incourse de l'Achigan	3303	1708	Ste-roy	4605 4606	1699
ol-vacques-ue-r Achigan	3302	1728	Ste-Foy Sorel	4605 4606 4701	1699 1675
Montmonny	3302 3601 3701	1728 0	Ste-Foy Sorel St-Ours	4605 4606 4701 4702	1699 1675 1750
Montmagny	3302 3601 3701	1728 0 1679	Ste-⊧oy Sorel St-Ours Rimouski	4605 4606 4701 4702 4801	1699 1675 1750 1701
Montmagny Cap-St-Ignace Bother en Bos	3302 3601 3701 3702	1728 0 1679 1679	Ste≻oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias)	4605 4606 4701 4702 4801 4901	1699 1675 1750 1701 1739
Montmagny Cap-St-Ignace Berthier-en-Bas St. Francis de la Pivitàm du Suid	3302 3601 3701 3702 3703 2704	1728 0 1679 1679 1710	Ste-roy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu	4605 4606 4701 4702 4801 4901 5101	1699 1675 1750 1701 1739 1740
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud Cl. Picture de la Divière du Oud	3302 3601 3701 3702 3703 3704	1728 0 1679 1679 1710 1713 1733	Ste-roy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu	4605 4606 4701 4702 4801 4901 5101 5102	1699 1675 1750 1701 1739 1740 1740
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Dierre-de-Decument	3302 3601 3701 3702 3703 3704 3705	1728 0 1679 1679 1710 1733 1727	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denie-sur-Richelieu St-Denie-sur-Richelieu St-Hyacinthe	4605 4606 4701 4702 4801 4901 5101 5102 5103	1699 1675 1750 1701 1739 1740 1740 1777
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré	3302 3601 3701 3702 3703 3704 3705 3801	1728 0 1679 1710 1733 1727 1657	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie	4605 4606 4701 4702 4801 5101 5102 5103 5202	1699 1675 1750 1701 1739 1740 1740 1740 1784
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer	3302 3601 3701 3702 3703 3704 3705 3801 3802	1728 0 1679 1679 1710 1733 1727 1657 1661	Ste-roy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301	1699 1675 1750 1701 1739 1740 1740 1740 1747 1784 1782
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O.	3302 3601 3701 3702 3703 3704 3705 3801 3802 3802 3803	1728 0 1679 1879 1710 1733 1727 1657 1661 1666	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302	1699 1675 1750 1701 1739 1740 1740 1777 1784 1784 1782
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804	1728 0 1679 1710 1733 1727 1657 1661 1666 1669	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Lee Cèdres	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601	1699 1675 1750 1701 1739 1740 1770 1740 1777 1784 1742 1742 1752
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O.	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805	1728 0 1679 1770 1710 1733 1727 1657 1661 1666 1669 1679	Ster-roy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St Lienz Boniste de l'Vie Verte	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802	1699 1675 1750 1701 1739 1740 1740 1740 1747 1784 1742 1722 1752 1766
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O.	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3804 3805 3806	1728 0 1879 1879 1710 1733 1727 1651 1651 1666 1669 1679 1679	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'ile-Verte Tarebone	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901	1699 1675 1750 1701 1739 1740 1740 1777 1784 1742 1722 1752 1756 1756 1757
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre I.O. St-Pramille 1.O. St-François 1.O.	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3806 3806 3807	1728 0 1679 1710 1733 1727 1657 1661 1666 1669 1679 1679 1679	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Danis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'Ile-Verte Terrebonne	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902	1699 1675 1750 1701 1739 1740 1740 1740 1747 1784 1742 1742 1742 1742 1766 1766 1766 1769
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Pierre I.O. St-Pierre I.O.	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3806 3805 3806 3807 3808	1728 0 1679 1770 1710 1733 1727 1657 1661 1666 1669 1679 1679 1679 1679	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denie-sur-Richelieu St-Dyachathe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'lie-Verte Terrebonne Ste-Anne-des-Plaines	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5902	1699 1675 1750 1701 1739 1740 1740 1740 1747 1784 1742 1722 1752 1766 1727 1788
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Jacent I.O. St-Pierre I.O. St-Jacobi I.O. St-Jacobi I.O. St-Jacobi II.O.	3302 3601 3701 3703 3704 3705 3801 3802 3803 3804 3805 3806 3806 3806 3806 3808 3808	1728 0 1679 1679 1710 1733 1727 1661 1661 1666 1669 1679 1679 1679 167	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'ile-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thrése	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 5904	1699 1675 1750 1701 1739 1740 1740 1740 1742 1752 1766 1752 1766 1727 1788 1789
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Ange-Gardien St-Famille I.O. St-Jauernt I.O. St-François I.O. St-François I.O. St-Joachim Montréal	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3806 3806 3807 3808 3809 3809	1728 0 1679 1770 1733 1727 1657 1661 1666 1669 1679 1679 1679 1679 167	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'lie-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thérèse Trois-Rivières	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 6001 6104	1699 1675 1750 1701 1739 1740 1740 1740 1747 1784 1742 1742 1742 1752 1766 1727 1766 1727 1788 1789 1634
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Pierre I.O. St-Joachim Montréal Bécancour	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3805 3805 3805 3806 3807 3808 3809 3809 3901	1728 0 1879 1879 1710 1733 1727 1857 1861 1869 1879 1879 1879 1879 1880 1725 1842 1716	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Denis-sur-Richelieu St-Dyachthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'Ile-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thérèse Trois-Rivières Vaudreuil	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 6001 6101	1699 1675 1750 1701 1739 1740 1740 1740 1747 1784 1742 1722 1752 1752 1766 1727 1788 1789 1634 1773
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Pierre I.O. St-Jachtim Montréal Bécancour Nicolet	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3806 3806 3806 3806 3806 3806 3809 3809 3901 4101	1728 0 1679 1679 1710 1733 1727 1657 1651 1666 1669 1679 1679 1679 1679 1680 1725 1642 1716 1716	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'lis-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thérèse Trois-Rivières Vaudreuil L'Île-Perrot	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 6001 6101 6102 802	1699 1675 1750 1701 1739 1740 1740 1740 1777 1784 1742 1752 1766 1727 1766 1727 1768 1789 1634 1773 1788
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Jean I.O.	3302 3601 3701 3702 3703 3704 3804 3802 3804 3805 3804 3805 3806 3807 3808 3809 3809 3901 4101 4102 4103	1728 0 1679 1770 1733 1727 1657 1661 1666 1669 1679 1679 1679 1679 167	Ster-oy Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'Ile-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thérèse Trois-Rivières Vaudreuil L'Île-Perrot Contrecoeur	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5301 5302 5901 5902 5903 6001 6101 6102 6201	1699 1675 1750 1701 1739 1740 1774 1740 1777 1784 1742 1742 1752 1756 1766 1727 1788 1789 1634 1773 1786 1789
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Pierre I.O. St-Joachim Montréal Bécancour Nicolet St-Pierre-les-Becquets Cap-Santé	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3805 3805 3805 3806 3807 3808 3809 3901 4101 4102 4103 4401	1728 0 1879 1879 1710 1733 1727 1857 1861 1686 1869 1879 1679 1679 1679 1680 1725 1642 1716 1716 1716 1734 1879	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denie-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'Ile-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thriftse Trois-Rivières Vaudreuil L'Île-Perrot Contrecceur	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 6001 6101 6102 6201 6202	1699 1675 1750 1701 1739 1740 1740 1740 1747 1784 1742 1722 1752 1766 1727 1788 1789 1634 1773 1786 1631 1693 1693
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Pierre I.O. St-Jean I.O.	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3805 3805 3806 3805 3806 3806 3809 3901 4101 4102 4103 4401	1728 0 1679 1679 1710 1733 1727 1657 1651 1666 1669 1679 1679 1679 1680 1725 1642 1716 1716 1734 1679 1679	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'lis-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thérèse Trois-Rivières Tois-Rivières Vaudreuil L'Île-Perrot Contrecoeur Varennes Verchères	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 6001 6101 6101 6102 6201 6202 6203	1699 1675 1750 1701 1739 1740 1740 1747 1742 1742 1742 1752 1766 1727 1766 1727 1768 1789 1634 1773 1786 1681 1693 1724
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-Beaupré Château-Richer Ste-Framille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Joachim Montréal Bécancour Nicolet St-Pierre-les-Becquets Cap-Santé Neuville Grondines	3302 3601 3701 3702 3703 3704 3804 3802 3803 3804 3805 3805 3805 3805 3805 3805 3806 3809 3901 4101 4102 4103 4401	1728 0 1679 1770 1733 1727 1657 1661 1666 1669 1679 1679 1679 1679 167	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'lie-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thérèse Trois-Rivières Vaudreuil L'Île-Perrot Contrecoeur Varennes Verchères St-Antoine-sur-Richelieu	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 6001 6101 6102 6201 6202 6203 6204	1699 1675 1750 1701 1739 1740 1774 1740 1777 1784 1742 1722 1752 1766 1727 1788 1789 1634 1773 1788 1789 1634 1773 1786 1681 1693 1724
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-Beaupré Château-Richer Ste-Arne-de-Beaupré Château-Richer St-Franille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Pierre I.O. St-Joachim Montréal Bécancour Nicolet St-Pierre-les-Becquets Cap-Santé Neuville Grondines	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3805 3806 3805 3806 3807 3808 3809 3901 4101 4102 4103 4401 4402	1728 0 1879 1879 1710 1733 1727 1857 1861 1686 1869 1879 1679 1679 1679 1680 1725 1642 1716 1716 1716 1716 1734 1679 1679 1679	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denie-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'Ile-Verte Ternebonne Ste-Anne-des-Plaines Ste-Anne-des-Plaines Ste-Thriftese Trois-Rivières Vaudreuil L'Île-Perrot Contrecoeur Varennes Verchères St-Antoine-sur-Richelieu St Marc-de-Coumoyer	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 6001 6101 6101 6102 6201 6202 6203 6204 6206	1699 1675 1750 1701 1739 1740 1740 1747 1784 1742 1722 1752 1766 1727 1788 1789 1634 1773 1786 1633 1773 1786 1693 1724 1741 1794
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud Ste-Anne-de-Beaupré Château-Richer Ste-Famille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Pierre I.O. St-Jean I.O. St-Jean I.O. St-Joachim Montréal Bécancour Nicolet St-Pierre-les-Becquets Cap-Santé Neuville Grondines St-Augustin	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3806 3805 3806 3806 3806 3806 3809 3901 4101 4102 4103 4401	1728 0 1679 1679 1710 1733 1727 1657 1661 1666 1669 1679 1679 1679 1680 1725 1642 1716 1734 1679 1679 1679 1679 1679 1679 1679	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'lie-Verte Terrebonne Ste-Anne-des-Plaines Ste-Andes-Plaines Ste-Andes-Plaines Ste-Andes-Plaines Ste-Thérèse Trois-Rivières Vaudreuil L'Île-Perrot Contrecoeur Varennes Verchères St-Antoine-sur-Richelieu St Marc-de-Coumoyer St-François-du-Lac	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5601 5802 5901 5902 5903 6001 6101 6101 6102 6201 6202 6203 6204 6206 6301	1699 1675 1750 1701 1739 1740 1740 1747 1742 1742 1742 1752 1766 1727 1766 1727 1768 1789 1634 1773 1786 1681 1693 1724 1741 1794 1687
Montmagny Cap-St-Ignace Berthier-en-Bas St-François-de-la-Rivière-du-Sud St-Pierre-de-la-Rivière-du-Sud St-Pierre-de-Beaupré Château-Richer Ste-Framille I.O. L'Ange-Gardien St-Laurent I.O. St-Pierre I.O. St-Pierre I.O. St-Jean I.O. St-Santé Neuville Grondines St-Augustin	3302 3601 3701 3702 3703 3704 3705 3801 3802 3803 3804 3805 3805 3805 3805 3806 3807 3808 3809 3901 4101 4102 4103 4401 4403	1728 0 1679 1770 1733 1727 1657 1661 1666 1669 1679 1679 1679 1679 1680 1725 1642 1716 1734 1679 1679 1679 1660 1693	Ster-by Sorel St-Ours Rimouski Pointe-Olivier (St-Mathias) St-Charles-sur-Richelieu St-Denis-sur-Richelieu St-Hyacinthe L'Acadie Pointe-du-Lac Yamachiche Les Cèdres St-Jean Baptiste-de-l'Ile-Verte Terrebonne Ste-Anne-des-Plaines Ste-Thérèse Trois-Rivières Vaudreuil L'Île-Perrot Contrecceur Varennes Verchères St-Antoine-sur-Richelieu St Marc-de-Coumoyer St-François-du-Lac La-Baie-du-Febvre	4605 4606 4701 4702 4801 5101 5102 5103 5202 5301 5302 5801 5802 5901 5902 5903 6001 6101 6102 6201 6202 6203 6204 6206 6301 6302	1699 1675 1750 1701 1739 1740 1740 1747 1784 1742 1722 1752 1766 1727 1788 1789 1634 1773 1788 1789 1634 1773 1786 1681 1693 1724 1741 1794 1687 1715

Chapter 3

Surname methods for population and migration research

Though human populations exhibit some level of organized spatial (geographic) structure because of their tendency to mate non-randomly and take part in specific migration and kinship patterns, their geographic structure are not always easy to decipher. The difficulty often results from pursuing analysis at pre-set spatial and temporal scales without exploring the hierarchical population order by aggregating at different spatial scales. Most isonymy studies thus far have failed to capitalize on the use of surnames simultaneously at different spatial scales to uncover the extent to which populations are organized in a hierarchical manner. In this chapter I systematically review routine surname methods as well as multivariate statistical methods whose application to surname data would provide avenues to pursue population hierarchies. The additional techniques have been extensively used in biology, genomics, and community ecology to organize, classify and visualize the structure of complex multivariate data including gene expression profiles, ecological communities and microhabitats. The value of these techniques in deciphering patterns in large complex data sets makes them a natural choice for exploring populations of names and their spatial distributions.

The five types of methods discussed are: 1) same-surname measures (or isonymy) kinship and genetic distance; 2) ordination methods such as correspondence analysis, canonical correspondence analysis and multidimensional scaling; 3) hierarchical and fuzzy clustering; 4) relevance networks; and 5) Bayesian probability. The same-surname methods have been developed and applied primarily by human biologists and have also provided an alternate means of estimating kinship and genetic distances. The ordination and clustering techniques have multidisciplinary foundations with applications in both the social and physical sciences. Relevance networks, a new method, has its roots in bioinformatics; and Bayesian probability, which involves making inferences under various 'prior' probability assumptions, has its roots in the fields of mathematics and statistics. These techniques resolve into rather standard statistical parameters, and I am using all of them as exploratory tools. In each case, we have to be careful about the assumptions and constraints (linearity, colinearity, model assumptions and sample size). I provide an explanation of each method, and where possible, examples from 1840 and 1880 name frequencies in Vermont. I do not in all cases provide the numerical computations or algorithms, and often I provide only a brief explanation because the methods are described elsewhere. After the examination I pick the appropriate methods and apply them to the Quebec study. The results will follow in the next three chapters. .

The Vermont case study is developed and presented here for three reasons. First, the state consists of only 14 counties and 300,000 people by 1880, a modest dataset when compared with the Quebec sample with hundreds of places and a million people. We have therefore a simpler vehicle to explain the limits and strengths of alternative methods for interpreting surname population structure and settlement and migratory patterns. Second, the US census data are available in their entirety, in a relatively 'clean' and convenient form, and they present fewer uncertainties in pre-treatment. Finally, the years 1840 and 1880 offer an interesting historical contrast. In 1840 Vermont had only a few thousand French Canadians, and tens of thousands by 1880, when it became one of the largest "French Canadian" populations in the United States. The large population movement portrayed in this case study adds to the overall perspective on French Canadian settlement history and provides evidence of the role French Canadians played in shaping the population structure of Vermont.¹ We will visit in a future chapter the movement of French Canadians to Vermont.

The earliest French Canadian settlements in Vermont were located in the counties of Grand Isle, Franklin, Chittenden and Addison, all of which border Lake Champlain to the south of Quebec. Many French Canadian settlers migrated from the Richelieu Valley, travelling by water routes into Lake Champlain or by roads along the same waterway.² Some of the early connections between Vermont and Quebec were associated with timbering taking place in the Richelieu Valley

and the many mills in Vermont (Klyza and Trombulak 1999). Vicero (1968) estimated the French Canadian population of Vermont at 5,500 in 1840, 12,070 in 1850, 16,580 in 1860 and 29,000 in 1870. Meanwhile, the total population of Vermont grew from 85,000 in 1790 to around 314,000 in 1850 and then remained fairly stable (332,000 in 1880 and 350,000 in 1920). Despite the stabilization in population growth, the number of surnames tripled between 1800 and 1880 (4490 to 12,800). Continued growth in the number of names implies turnover which was associated with out-migration of many of the sons of land owners and in-migration of French Canadian and Irish workers into low-wage jobs that opened as a result of industrialization. For each decade the number of surnames gained was larger than the number of surnames remaining from the previous decade (Table 3.1).

The Vermont surname distribution matches well with other numerical distributions from studies in other areas (including Quebec), where there are a large number of unique surnames and a small number of high-frequency surnames. On a recent United Kingdom electoral roll, for example, 42% of surnames occur once, 16% occur twice and 7% three times. Studies in Switzerland, Austria and Italy also show similar distributions. If we examine the distribution between 1840 and 1880, despite the population and name size differences, we see the same frequency distribution (Figure 3.1).

Figure 3.1 summarizes the frequency distribution of Vermont surnames showing the similarities between 1840 and 1880. Such frequency curves are helpful in evaluating whether a sample is representative of the entire population, and in comparing population structures and estimating rates of extinction. Surname frequency curves, when examined at different time periods, are also used to estimate migration rates.

When considering the loss of surnames, we must take into account the large number of surnames that become extinct as a result of no male descendant. Surname extinction is a complex theme affected by in-and-out migration, birth and death rates, and cultural and religious attitudes toward intermarriage (Hull 1998). Galton-Watson (1875) propose some probabilistic scenarios for name extinction; and Kendall (1873) reviewed the role of mathematical branching theory in

estimating how many names will become extinct in a population where there is growth and decay. In a closed or isolated population, name extinction is most rapid in early generations, purely as a result of the laws of probability. Of 1000 surnames approximately 237 (or 24%) will become extinct in the first generation, 100 more by the second generation, and progressively fewer in each generation as the same names grow making extinction improbable.³ In a completely closed population, after many generations, all but one name would become extinct. We will see in the chapter on 1881 Quebec that the results from our sample between 1765 and 1881 indicate the extinction of names.

The methods presented for Vermont were calculated at four scales: at the county scale (14 counties), at a more refined scale of 38 hexagons, for a single county (Orange County), and at the very fine-grained scale of the 232 populated places as defined in the 1880 census. The 1840 data were used only to provide basic isonymy (same-surname) measures at the county scale: where the object is to show how population growth (1840-1880) changes the dynamics of isonymy results. To produce the hexagon map for 1880, I applied the same method as I did for 1881 Quebec (see chapter 2). In this case a uniform grid of hexagons was superimposed on the 232 'places' recorded in the 1880 census, and populations were aggregated into corresponding hexagons.

Measures of inbreeding, isonymy, kinship and genetic distances

Darwin (1875) and Arner (1908) were the first to use surnames as population markers in their studies of marriages between first cousins. The use of surnames as a means to estimate inbreeding, using the percent of same surname or cousin marriages, continued and is found occasionally in the literature, but was not defined mathematically nor applied in a systematic way until the work of Crow and Mange (1965) on the Canadian Hutterites, where they coined the term population isonymy which means 'same surname' and denotes marriages between persons of the same name.

After Crow and Mange put forth their surname model, isonymy became an unofficial branch of human biology, physical anthropology and human genetics. Researchers quickly realized the power of surnames as a surrogate genetic variable

because in many Western societies surnames are transmitted through the paternal line, and one can think of the family name as an allele carried on the Y chromosome. In men, the same Y chromosome is inherited along the male genealogical line in the same way as the surname. Lasker and later Relethford (1988) followed up the work of Crow and Mange with important research that defined ways of extending the concept and application of isonymy beyond same name marriages to lists of surnames.

Measures of inbreeding by same-surname marriages

Although I am inquiring into surname distributions as evidence of settlement processes, it is important to take this detour into the geneticist's interpretation and inferences, in order to see the way in which this cultural element corresponds to a biological phenomenon. We have to ignore some of the terminology disturbing to a social scientist ('breeding') and keep in mind the probabilistic rather than deterministic nature of the models.

Since the work of Crow and Mange, same surname studies have been used in a variety of applications, but they can be defined as two basic types. The first uses marriage isonymy to estimate the extent of random and non-random components of inbreeding. Marriage isonymy refers to a measure of how often individuals bearing the same surname marry or how often pairs of the same name-pairs occur. The second type of isonymy refers to the number of names shared in a single or multiple populations. Each method will be explained in greater detail.

What geneticists refer to as the 'coefficient of inbreeding' is an estimate of the probability that two individuals have identical alleles at a given locus as a result of receiving them by descent from a common ancestor.⁴ A consequence of inbreeding is to increase the frequency of homozygous genotypes in a population, relative to the number that would occur with random mating. Seldom does inbreeding occur at high enough levels to have any significant effect of increasing deleterious genes (or alleles) to a dangerous level because with increased inbreeding there is an associated increase in rates of mortality and morbidity in children, removing these individuals from the breeding pool. Three methods are commonly used for estimating inbreeding rates or consanguinity. Consanguinity

refers to a level of marriage between relatives. In the Catholic church, the clergy usually allowed marriages beyond one's third cousin, but marriages of second cousins were often accepted.

The first method and the most accurate, uses family pedigrees or genealogies and ancestor chains (paths) to calculate the probability of a child inheriting the same allele from both parents. This method uses the frequency of marriages between persons of a common ancestor in a defined population. The coefficient can be applied to a variety of same family relations like siblings, half-siblings, uncle niece, and cousins of any order. The longer the pedigree, the more complex the relations and calculations become. In each case the coefficient is exactly onefourth the proportion of isonymy in the parents (Cavalli-Sforza and Bodmer 1991; Cavalli-Sforza, Menozzi, and Piazza 1994). For example, the inbreeding coefficient would be 1/8 if an aunt and nephew married, 1/16 for first cousins, 1/64 for second cousins. The coefficient of inbreeding is calculated as $F = \sum P(1/2)^A$, where the number of ancestor paths (p) is multiplied by one-half the number of chains of descent (A) (Hartl 2000). In figure 3.2, the lines or chains represent gametes of the offspring inherited from their parents. The equation calculates the probability that an individual, in our example the child, carries genes identical by descent. In the example displayed there are five paths: GDACE.

A second method of estimating inbreeding in a population employs marriage records. Frequencies of same-surname marriages are extracted from marriage records (adequate sample size) for a particular place. The inbreeding coefficient (F) from same-surname marriages is calculated as F=P/4, where P is the proportion of marriages (of the total) having the same surname (Lasker 1985). Inbreeding rates approaching 1.0 would be considered high. This method is prone to error when there are multiple founders in the population. For example if a Smith marries a Smith with unrelated genealogies then this would cause miscalculations in the inbreeding calculation driving the isonymy values much higher making the population appear much more related.

A third method was presented in a 1965 paper by Crow and Mange.⁵ This method, often referred to as "marital isonymy," uses lists of women's and men's

surnames (and when available actual marriage records) to estimate the average proportion of genes at paired loci that are identical from the same ancestors through both parents (Lasker 1985). Crow and Mange devised the coefficient to take into account both random and non-random inbreeding components.

The random component refers to the inbreeding resulting from panmixia in a finite population where all possible pairings of surnames are presumed equally probable. It is calculated to approximate how much inbreeding has accumulated since the surnames from the founding population were first established, by summarizing the commonality of surnames in the population. A founding population refers to a newly isolated pioneer population, sometimes small, which has different allele frequencies and a higher frequency of mutations than the larger and more diverse population from which it derived. In a later study we will look at the Quebec population over a hundred year period and how modernization, inmigration from Britain, and urbanization serve as mechanism for increasing random mating.

The non-random component of inbreeding refers to an increase or decrease of inbreeding as a result of selective mating (Lasker 1985; Cavalli-Sforza, Menozzi, and Piazza 1994; Fix 1999). Because non-random breeding affects generational changes in inbreeding, it is helpful for interpreting population changes in spans of thirty years.

Calculations of the random and non-random components of inbreeding are based on the assumption that isonymy accurately reflects common ancestry. For example, where a Smith marries a Smith, we assume they are related although we do not know to what degree. The genealogical ambiguity is a good example of the risk of overestimating inbreeding. The coefficient of inbreeding from matched lists of surnames is calculated as follows:

$$F = F_n + F_r(1 - F_n)$$

where F represents 'total inbreeding' for an entire region. The formula contains a random component (Fr) for estimating the regional mixing of a population, and a

non-random component (Fn) for estimating group behaviour toward isonymous marriages. The Fr coefficient is calculated as follows:

$$F_r = (\sum_k pkqk) / 4$$

where pk is the frequency of the k-th surname in men and qk is the same for women; the summation is over all surnames. The Fn coefficient is calculated as follows:

$$F_n = \left(P - \sum_k pkqk\right) / 4\left(1 - \sum_k pkqk\right)$$

where P is the proportion of isonymous marriages in the entire population.⁶ In order to calculate P you need to know the number of isonymous marriages and total marriages for a specified period (Lasker 1985; Rodriguez and Barrai 1997; Roguljic, Rudan, and Rudan 1997; Relethford 1988; Cavalli-Sforza, Menozzi, and Piazza 1994).

Another method which uses marriages, although less popular than inbreeding coefficients, is the 'repeated-pairs approach' (RP). Instead of using the repetition of same surname marriages this method uses the repetition of same pairs of names. This might include, for example, the number of times a Bessette-Beauchamp marriage occurs in a defined population at a certain span of time. This method provides a measure of population subdivision or the tendency (probability) of the people to marry among themselves within their town or at a specified distance (Lasker and Kaplan 1985, Koertvelyessy et al. 1988). It might be used as a means to estimate at what rate members of different ethnic groups begin to integrate (inter-marry). This certainty has a cultural component as well. The non-random repetition of surname pairs is calculated as follows:

$$RP = \sum \left[S_{ij} (S_{ij} - 1) \right] / \left[N(N - 1) \right]$$

where Sij is the number of marriages with a husband of the ith surname and a wife of the jth surname, and $N = \sum Sij$.

To derive the random component of RP, the wives' maiden names from the same sample are randomized and re-matched with the list of men and the pairs recounted. This randomization should occur multiple times to insure accuracy.

Having information on the number of same name marriage pairs, birth, death and migration rates, and the surname frequency, we can begin to estimate future rates of inbreeding, genotype frequencies, kinship networks, and to what extent a member of the population might have alleles that are identical by descent. In a later chapter on same surname marriages, and repeated surname marriages, I will look at the inbreeding coefficients and population parameters which lead to increases in the probability of same surname marriages or the tendency of a surname to pair with or mate with names from a locally (defined) population.

Since many studies do not have full information about the founding population, or even the number of generations elapsed, the approximation is helpful. In the case of Quebec we have exceptional information about the founding population and its relation to the population under study (1880 or today). For Vermont we have much less control, even though we can assume several waves of arrivals between 1700 and 1880 (see Klyza and Trombulak 1999). Most of the methods just presented work best when there are two or more population periods to compare.

Same-surname quotients (isonymy equations)

According to Lasker the main point of surname studies "is to measure the different probabilities of finding the same surnames in different times, places, groups and especially in marital partners." The suite of isonymy or 'same surname quotient' can be calculated for a single population, between two populations, or for an entire region. These measures provide probability values which can be used to interpret structure from frequency distributions of surnames in the populations at a given time or for various times. The surname quotient (isonymy) provides values which are then used to calculate genetic distances and

kinship measures between populations. The basic idea behind these computations is merely to estimate a probability index that provides a statement about the average commonality of surnames in a sample. The greater the number of people who share the same surname, the greater the probability that any two people picked at random will share that name. A value of one would indicate that everybody in the population has the same surname, a value below one would indicates some shared names, while a zero value would occur if all names were different.

The localized or within-population surname quotient is calculated as follows:

$$I_{ii} = \sum_{k} nik(nik-1)/Ni(Ni-1)$$

$$I_{ii} = \left(\sum_{k} m_{ik} \mathbf{f}_{ik}\right) / (M_i F_i)$$

where nik is the numbers of individuals with surname k in population I (a given county or hexagon) and Ni is the total number of surnames in population i. The second equation is a variant which distinguishes men (mik) and women (fik) (Relethford 1988).⁷ Lasker (1985) has also discussed options for deriving variance measures for the coefficient of relationship, by dividing samples into two, or dividing samples into subgroups which might consider non-immigrants, migrants, occupations, economic status and people arriving at different time periods. Confidence intervals can also be calculated using randomization techniques including bootstrapping, jackknifing or numerical modelling (Manly 1991).⁸

The between-population quotient provides a method for measuring the degree of surname similarity between two populations, and it is this measure on which I rely on for many maps and observations in the Quebec study. The computation takes half the ratio of all possible pairs of isonymous surnames to all possible pairs of surnames. The measure (Iij) is calculated as follows:

$$I_{ij} = \sum_{k} (nikn_{jk}) / N_i N_j$$

where nik and njk are the numbers of individuals with surname k in locations i and j and Ni and Nj are the sample totals in locations i and j. This measure provides a probability of picking one surname from each population and the surnames being the same. If two populations have very similar surname structures, then the probability of picking two similar surnames will be high (Relethford 1988). Care must be taken because the results can be biased when samples are too small. Though Lasker (1981) has discussed creating confidence intervals when applying same-surname methods, it is a problem that has yet to be addressed in the literature.

Regional surname quotient provides the probability of selecting two persons of the same surname at random from the entire study region. This value is used in the calculation of regional a priori kinship or to compare temporal changes or regional differences. This regional measure is calculated as follows:

$$R = \{\sum_{k} \left[(\sum_{i} n_{ik}) (\sum_{i} n_{ik} - 1) \right] \} / \left[(\sum_{i} N_{i}) (\sum_{i} N_{i} - 1) \right]$$

and combines local measures to provide a normalized average over all subpopulations.

The three surname measures were applied to 1840 and 1880 Vermont at the county scale, and for the1880 sample I include results at the hexagon scale; I use Orange County as an example of yet a more refined treatment of 'places'.

The 1840 within county same-surname (isonymy) values are, as expected, higher than for 1880, indicating diversification over the 40 years. At both dates values are lower than observed in other North American studies, notably 1840 Massachusetts, 1800 Bedford County, Pennsylvania, 1810 Kings County, New York , and 1850 Antebellum Georgia (Christensen 2001, 2000, 1999; Relethford 1988). This implies a considerable diversity of names and population origins in Vermont. These local isonymy comparisons may be misleading because the measures is sensitive to sample size, sizes of populations and the number of populations considered. This is apparent in the analysis of 1880 Orange County, where the 17 locations show higher local isonymy values, closer to the studies referenced above.

As shown in table 3.2, local isonymy values decreased in all but two counties (Lamoille, Washington) between 1840 and 1880. The effect is most pronounced in Addison, Chittenden, Rutland, Caledonia and Essex counties. The low isonymy values (high diversity) for Addison and Chittenden counties are partially explained by the presence of the state's two urban areas (Chittenden and Burlington), which were attracting Irish and French Canadian immigrant workers. Vicero (1968) reported that these two cities in 1880 had the largest immigrant populations in Vermont and the largest French Canadian populations in the United States. For 1840 and 1880 the lowest isonymy values (most diverse areas) were found in counties bordering the western slopes of the Green Mountains (Figures 3.3, 3.4).

The population grew slowly between 1840 and 1880, but the name structure changed significantly. Many of the areas in southern Vermont which had the least surname diversity in 1840 remained in the same position into 1880. Vermont's population was undergoing rapid change, but the change is not evident from overall population figures which are stable after 1850. If we re-examine table 3.1, showing the numbers of surnames lost and gained during the century, it becomes apparent that there was significant turnover. These changes are a result of numerous factors. First, Vermont sent about 32,000 soldiers to the Civil War, and more than half never returned. Second, Vermont never became an industrial powerhouse like other New England states, and many residents opted to find jobs in other states. Finally, farm sons who did not receive land from their fathers or could not afford to buy land opted for opportunities in the mid-western 'frontier'. This included many French Canadians who after 1880 tended to move toward larger urban centers, notably Lowell and Boston, Massachusetts.

The hexagon map for 1880 shows a pattern of local isonymy much like 1840, with the lowest values (greatest diversity) located close to the urban areas of Burlington and Rutland and along the east side of the mountain backbone (Figure 3.5). State-wide isonymy at this scale shows more of the regional surname

variability. The county maps make it look as if each county has a homogenous distribution, while the hexagons allow us to see that the lowest isonymy values are in fact found close to Rutland and Burlington City, and as you move away from these locations the values increase. In other words, the hexagons provide a more unbiased display of local surname variability and allow you to pick up population hotspots with extreme values.

The between-population surname quotients (values) show to what degree a particular population resembles other places in its surname structure. As illustrated in figure 3.6, most counties in 1880 Vermont are most closely related with Windham. The arrows in the figure track the strongest relationship for each county. Essex County, for example, has one of the lowest diversity levels in Vermont and shares the most surnames with Windham and Windsor. The large number of counties sharing surnames with Windham is partially the result of its low diversity (a small number of names), and the fact that it was one of the earliest settled regions in the state. In figure 3.7, which shows the second most related areas by surname similarity, you notice the smaller distances. In other words, geographic proximity begins to display a pattern of diffusion or regionalization of surname structure. Neighbouring counties like Franklin, Lamoille and Orleans, for example, have similar surnames.

To simplify results with large similarity or dissimilarity matrices, multidimensional scaling (MDS) is often used. MDS is a data reduction procedure that can be used on any pairwise similarity matrix. The primary objective is to plot meaningful relationships between pairs of variables, allowing researchers to interpret dissimilar objects as being far apart in Euclidean space and similar objects as close to one another. The calculations are based on an n x n distance matrix calculated from the n x p-dimensional matrix. MDS uses an iterative search process for the ranking and placement of n entities on k dimensions so as to minimize the stress of the k-dimensional configuration. Stress is a measure of departure from the association between the dissimilarity in the original pdimension space and distance in the reduced k-dimension. where n is the number of rows and p is the number of columns in the matrix (Kruskal and Uslaner 1978;

McCune and Grace 2002; Cox and Cox. 1994; Systat 1998).⁹ What MDS does is best grasped in the context of a simple matrix of geographic locations with the distances between them used as the dissimilarity measure. If MDS is applied, the results will produce a geographic map with the places located at the correct distances from one another.

When MDS was applied to the Vermont county-level between-population isonymy values, it produced robust results accounting for 93% of the variance with a Kruskal stress of .150. Figure 3.8, showing the first two dimensions, displays the organization of counties based on these values and it reveals that the between population isonymy values are related to the geography of Vermont. Each location is situated in terms of its similarity to all other locations. Populations close together on the diagram are more similar in surname structure. The diagram does show consistency with actual geographic reality; with the counties more closely related in a north-south direction. This pattern is likely the result of the backbone mountain ranges running north to south, hampering interaction (migration) along the east-west axis (Figure 3.8).

Measures of kinship

Measures of genetic kinship and distance are important tools for interpreting structure of human populations and determining the similarity of two or more populations. The kinship coefficient can be defined as the probability that a gene taken at random from an individual, at a given locus, will be identical by descent to a gene taken at random from a second individual at the same locus. Converting the isonymy values to measure kinship provides a normalization which allows these measures to be approximately compared with kinship derived from other metrics or other studies. The purpose of this coefficient is to measure the loss of heterozygosity relative to a reference population, which in this case is overall regional population. Commonly, allele frequencies are used to calculate genetic kinship, but other methods exist including using surnames, blood types, metric traits (e.g. cranial size), and migration matrices (Relethford 1988; Cavalli-Sforza and Bodmer 1991; Cavalli-Sforza, Menozzi, and Piazza 1994).

The kinship coefficient is an average measure of how related a population is to

the set of individuals in another population. The coefficient of kinship is calculated by the following equation:

$$\Phi_{ij} = (H_o - H_{ij}) / H_o$$

where Φ_{ij} is the heterozygosity expected when a random allele from population i is paired with a random allele from population j, and H_o is the heterozgosity of the reference population (Relethford 1988, 476). Because this equation has been the subject of confusion, due to the difficulty of defining the reference population, Relethford (1988) converted it to incorporate isonymy using Morton's (1973) terms of 'a priori' and 'conditional' kinship. A priori kinship refers to kinship relative to a founding population and is a measure of the unreduced covariance between populations; while conditional kinship refers to kinship relative to the contemporary population and is a measure of reduced covariance between populations (Roguljic et. al, 1997; also see Rogers and Harpending 1986 and Wood 1996).

A priori kinship for each population is computed as follows:

$$\Phi_{ij} = r_{ij} + (1 - r_{ij})\Phi_R$$

where Φ_{ij} is the unreduced covariance of population i and j, r_{ij} is the reduced covariance of population i and j, and Φ_R is the unreduced covariance of the entire region. Φ_R represents the weighted mean value from the entire matrix of Φ values (reference population) and is equal to regional isonymy divided by 4.¹⁰

$\Phi_R = R/4$

Conditional kinship for each population is computed as follows:

$$r_{ij} = (I_{ij} - R) / [4(1 - R)]$$

where Iij is the random isonymy between populations i and j, and R is the regional isonymy.

Conditional kinship values for the 14 Vermont counties in 1880 are all negative, indicating that these communities have a rather low degree of surname similarity when compared to other North American studies (Table 3.3). Negative values at the county-wide scale were expected, but when conditional kinship was calculated on 17 towns in Orange County, Vermont (1880), the kinship values were mostly positive. We see here that the calculation is sensitive to population size and geographic scale, and more appropriate for examining sub-populations (Christensen 1999, 2000, 2001; Relethford 1988; Rodriguez and Barrai 1997; Roguljic, Rudan, and Rudan 1997).

For four counties I made a comparison between the a priori kinship patterns at the county and the hexagon scale. Figure 3.9 illustrates the a priori kinship relationships for Orange, Addison, Windsor and Washington County. In each case we see minor distance decay (isolation by distance), but it varies depending on direction and location (Malecot 1975). In most cases a bordering county has a stronger kinship relationship with a neighbouring county immediately to the north or south. At the county scale the mountains seem to have had substantial influence on migratory patterns. If we included bordering counties of New York and New Hampshire, we might well see distance decay which would extend from Western Vermont into Eastern New York, and from Eastern Vermont into Western New Hampshire, as a result of inter-state migration. East-west structure is evident in the percentage of Vermont residents who were born in New York and New

The hexagon maps offer greater precision of the geographic variability in name structure. Again the closest locations are most closely related (darker shading), and the strongest relationships are usually found north or south of the reference population, indicating mobility along the lowlands and valleys. Visualizing the between-population kinship values is helpful, but it is difficult to summarize whether or not there is any local spatial relationship and whether or not distance decay exists (Figure 3.10).

At the hexagon scale, multidimensional scaling for 'a priori' kinship provides a better description of population organization. The MDS results explain 64% of

the variance with a Kruskal stress of .28. The organization of the first two axes does not show such distinct geographic patterns as at the county scale. The diagram illustrates some general regional trends found at the county scale, splitting the state into East, West and Central Vermont (Figure 3.11). What is more apparent at the scale of hexagons is the localized surname structure. Many of the hexagons are most similar to a geographic neighbour, indicating a strong localized relationship among names. For example, locations 21, 22, 10, 15 and 19 are all found next to each other along the northwest side of the state; and the cluster of 34, 35, 37, and 28, the oldest regions in the state, are all located in Southern Vermont (Figure 3.11).

Measures of Genetic Distance from Surnames

A genetic distance is a measure of the genetic differences among populations.¹¹ Geneticists often use the frequency of a certain allele. For instance, if they had the frequencies for the same allele for different populations, then they could infer that the populations with the most similar allele frequencies had the shortest genetic distances between them, making them most related (Relethford 2001). When calculating genetic distances from surnames we are looking for populations that have the most similar names, analogous to the Y chromosome haplotype. For example, if the name structures (isonymy values) of three populations are compared (A, B, C), and the absolute difference in isonymy values between A and B is smaller than between A and C, and B and C, then we can infer that populations A and B are more genetically similar. Genetic distance takes into account only the two populations for which the measure is being calculated and not the regional same surname computations which is the case for kinship measures.

Relethford (1988) has identified a number of possible equations for calculating genetic distances from surnames, but he recommends using the following equation if only a relative measure of population dissimilarity is needed. This equation is defined as:

$$d^2 = I_{ii} + I_{jj} - 2I_{ij}$$

The d^2 is the distance between populations i and j based on random isonymy. In this equation the data are normalized by the sum of local isonymy values for the two populations minus the between population isonymy values. Because local isonymy is accounted for in this measure, the results will not necessarily correlate with the between population isonymy values.

For the Vermont sample, only the relative measure of genetic distance was used. These measures show some structure, suggesting that the largest genetic distances (least related) are generally found a few counties away. At the county scale, the genetic distances for Vermont do not follow a normal distance decay model: county boundaries are split by mountains dividing the state leading to stronger north-south migration flows. Stronger flows of east-west migration appear to be a result of interstate moves from New York and New Hampshire.

To evaluate the significance of a relationship between genetic and geographic distances, a Mantel test was applied. Mantel's test (1967) is a method for assessing the association between two independent dissimilarity matrices and determining whether the association is stronger than one would expect by chance alone (Mantel 1967; Manly 1986; Legendre and Legendre 1983; Sokal 1981; Manly 1991). The Mantel test measures the association between two triangular matrices with a suitable statistic (usually correlation) and evaluates the statistical significance by randomly reordering and recalculating the statistic. The null hypothesis is that there is no association between the two matrices. This technique has traditionally been used to compare genetic and geographic distances between populations to determine the effect of distance on determining species micro-differentiation (source).

Application of the Mantel test to the genetic distances between Vermont counties reveals no significant relationship between genetic distances and geographic distances, so we must accept the null hypothesis (correlation coefficient .1486, 1000 iterations, p=.0001). For Quebec where populations are more defined and there has been less mixing, we might expect to see a relationship

between genetic and geographic distances.

Micro-differentiation

We borrow from geneticists one further computation to measure the global (regional) level of genetic differentiation across all populations (Wright 1951). The Fst and Rst coefficients known as "micro-differentiation" are especially pertinent for comparisons at successive time periods. We calculate an average squared genetic distance of a population from a single central point or centroid. In one case (Fst) the central point refers to the founding population, and in the other (Rst) it refers to the contemporary population. Relethford (1988) refers to the quantity of Fst as the average within-group 'a priori' kinship and Rst as the average within-group conditional kinship (Workman et. al. 1973; Roger and Harpending 1986; Relethford 1988; Roguljic et. al 1997).¹² Estimates of Fst and Rst are influenced by sample size and number of places (populations) included in the calculation. [closed population, new immigration after a few generations will have high Fst].

Observing Fst values over time or comparing them between populations provides information about the differences of structure and rates of gene flow (Relethford 2001). Because in-migration (or population mixing) from outside tends to reduce Fst (or Rst), the measure tends to decrease over time; and regions of high population density or attraction to migrants (such as Montréal or Burlington), are likely to have lower coefficients of micro-differentiation. If however, we are looking at a rather isolated region (such as Charlevoix or the mountains of Vermont), we would expect a higher Fst (or Rst): the value is likely to increase for several decades after a founding population settles, and then decline again. A population which possessed only a small number of surnames at the start, and then underwent elevated inbreeding due to random mating, would show high coefficients of microdifferentiation. Increasing values of Fst (or Rst) are indicative of decreasing heterozygosity, a consequence of growth of the founding population with little influence from outside. Care should be taken when comparing Fst and Rst values with other studies since the results are highly sensitive to the number and size of geographic regions included which influences population size.¹³ The measure of the unreduced variance of population

microdiffentiation is calculated as:

$$F_{ST} = R_{ST} + (1 - R_{ST})\Phi_R$$

Overall we will generally find that a decrease in population size will lead to an increase in microdifferentiation, and an increase in population size will lead to a decrease in microdifferentiation (Relethford 1991). It is also important to keep in mind that Fst decreases regularly with increasing size of subpopulation (parish, counties, state) (Jorde 1980; Cavali-Sforza et al. 2004).

Vermont Fst and Rst values went from .00162 and .0132 in 1840 to .0010 and .00036 in 1880 (Table 3.4). These values provide a good example of how, over time, with increasing diversity, the regional Fst and Rst values will decrease. The 1840 values are indicative of some Massachusetts regions before midnineteenth century, and of regions today in England. The lower 1880 values are closest to those derived for Bedford County, Pennsylvania (1800), but the values are difficult to put into context because of the rarity of regional studies in the United States. Most of the studies have focused on measurements at the size of counties or villages (Relethford, 1988; Christensen, 1999).

Using the hexagons to calculate Fst and Rst values for 1880 Vermont puts the regional variability in surname structure into context. The 32 hexagons produced an Fst value of .002, lower than the values calculated at the county scale. We could infer that the stability of the population overall was associated with considerable turnover (net migration), rather than natural replacement (Table 3.4). This is what was suggested by the initial observations of turnover of surnames from one decade to the next (Table 3.1). The numbers who left were great, the numbers who entered from elsewhere were substantial, and some regions are characterized by greater isolation than others. ¹⁴

We might compare these values with a population with exceptionally high values of Fst and Rst: The Nebraska Amish in Pennsylvania have Fst values around .1014 and Rst values around .0107. Such high values are indicative of geographic and religious isolates (Relethford 1988).

Isolation by Distance and Surnames (Distance Decay Models)

The above surname equations assess differences among populations, but do not address geographic variation (isolation by distance). Genetic differences between places often decrease as geographic distance between them increases. Many geographic phenomena reflect this property and in geography we refer to this concept as Tobler's (1970) 'first law of geography,' which explains that everything is related to everything else, but near things are more related than distant things. Biologists refer to this concept as 'distance decay.'¹⁵

The simplest way to think about this concept is in terms of three theoretical populations (A, B, C). A and B are 50 miles apart; population C is 5 miles from A and 55 miles from B. If we know migrants have moved through this area, passing all three towns, we might expect A and C to be more genetically similar because they are close together, and most often we would be correct, but we must consider whether the places were settled by different populations or subject to racial discrimination, a legislative frontier, or even a cultural or linguistic barrier. In England, Europe and other older regions researchers have found extreme cases of genetic distance decay. Studies in North America, although few, have not revealed such a strong distance decay (Christensen 1999, 2000, 2001). For many present-day populations, technological innovations and North American 'mobility' have extended marriage fields and made distance decay much harder to detect. Regardless of technological innovation there are many historical populations, particularly along the Atlantic seaboard, where population differences can be detected.

Distance decay in isonymy studies is most often calculated using kinship values or genetic distances. The distance decay model provides a prediction of the relationship of kinship to geographic distance and attempts to determine whether the coefficient of kinship decreases exponentially with geographic distance. The isolation by distance model is calculated for 'a priori' kinship using the following equation:

$$\Phi_{ij} = ae^{-bd}$$

where a is an estimate of the unweighted mean kinship, b is a measure of distance
decay and d is the shortest Euclidean distance between places.

In theory, 'a' should be equal to Fst. In order to calculate for conditional kinship we have to adjust for the negative values. Conditional kinship is calculated using the following equation:

$$r_{ij} = (1 - L)ae^{-bd} + L$$

where L is the adjustment unit to remove the negative values. L is a function of R.

$$L = R/(4-R)$$

For 1880 Vermont, I observed no significant or organized distance structure. This is in part attributable to the scale of observation and the narrow shape of Vermont with the most extreme differences in population structure being the shortest distances apart, separated by the north-south ranging Green Mountains. In addition, at this time there was a tremendous amount of population mixing and out-migration. I suspect for Quebec that the distance decay will be more representative of some areas in England and Europe.

As I have indicated, same surname analysis (isonymy) can be very informative for a first inspection of population structure, but care must be taken and a number of assumptions acknowledged (or met when possible). First, it is important to remember that when we use isonymy methods, as we did here for Vermont, we will not always know the structure of the founder population (as we do for Quebec), and we are assuming that each unique name is received by a common ancestor. In the study of Quebec, I shall provide some clues and approximations of how much we overestimate when we make such an assumption. Second, when calculating inbreeding with isonymy, we are assuming that same surname marriages occur between first cousins and not some distant relative with the same name from a remote founder. Third, when inferring migration from men's surnames, we are assuming that men and women have the same propensity to migrate and choose the same destinations. In fact, the textile mill towns of

Vermont tended to attract female labour, however, while quarry and timber regions attracted a male labour force. Throughout the nineteenth century large cities were selective of young women and some immigrant streams, notably the Irish, contained much larger proportions of women, others (notably the Italian) of men. Fourth, with the standardization of names there will be a tendency to overestimate the number of unique names, leading to overestimation of diversity. This reminds of us of the advantage of making spelling corrections when sources are available. Finally, isonymy methods are best used as relative measures (not absolute) for comparing one study area. When used to compare the same region at different times, or the results of another researcher's study, care should be taken in how the results are interpreted. How the study region is defined, the sizes of the samples and whether or not names were standardized can greatly affect results.

Ordination and multivariate methods for surnames

We can advance a step further by subjecting the same basic idea (the frequency tables of surnames in places) to a type of statistical treatment often used to explore social and biological variance: ordination refers to an array of mathematical techniques which summarize large multivariate datasets and extract major gradients which can then be examined for their association with social or environmental variables (McCune and Grace 2002). The most popular ordination techniques-- principal components analysis (PCA), correspondence analysis (CA) and canonical correspondence analysis (CCA) summarize variability by replacing raw data with a simplified matrix. What makes ordination methods attractive, as compared to many unidimensional and cluster statistics, is the potential for accounting for covariation found in the multivariate data set. Ordination methods are excellent for distinguishing strong patterns from weak ones and for exploring factors ('environmental' variables) which we hypothesize as influencing or driving a data pattern, thereby revealing unexpected patterns and processes (Lepš and Šmilauer 2003).

Ordination techniques are based on matrix algebra where eigen-analysis is used to reduce the large number of intercorrelated variables into a more condensed

form where the first few eigenvalues explain the greatest fraction of variation. Eigen-analysis is performed on a square symmetric matrix where the number of rows is equal to the number of columns. In biogeography and ecology, where I have explored software, functions, documentation and examples on ordination, the matrix columns are most often sites, and the rows are associated with species, or in this study surnames. Each cell therefore represents the number of persons of a given surname at a specified location (county, hexagon, or parish).¹⁶

Ordination techniques generate four types of information for summarizing the data: eigenvalues, total inertia, scores for locations, and scores for surnames. The relative importance of each axis is measured by the eigenvalues, whose number is equal to the sum of the numbers of rows and columns in the matrix. The eigenvalues are ranked from highest to lowest and researchers usually interpret only the patterns observed in the first three or four axes, since the share of explained variance becomes progressively less.

Inertia is a measure of the variance in the data. Often when working with ecological data, inertia values will decrease if the rarest species are removed from the data. Removing rare species will likely reduce the sample size drastically because of their large number. Many argue against removal of rare species and prefer to apply other mathematical transformations to reduce data variance. But when there are many singletons (surnames unique to one person) their removal does not result in a serious loss of information unless we are trying to understand diversity. How much information exists in singletons to reveal population relatedness? A single surname provides little clue to the overall population structure when using ordination methods, nor to the reasons certain surnames are clustered in certain regions. In many cases, the removal of singletons reduces noise and uncovers more accurately the population structure. The only time singletons might be of interest is if they were concentrated in certain areas. A clumping of singleton surnames would make me want to look further at what might be special about that location. But such questions can be explored prior to ordination methods, using diversity indices or simple queries.

The place and surname scores are the eigenvectors and provide information

about where places and individual surnames are situated along a particular dimension of variation. Two sets of scores (for two axes) can be mapped as a biplot showing two dimensions of variation, or the results of each individual axis can be positioned (mapped) on top of the geographic variables.

Ordination techniques are of two types: direct and indirect gradient. In direct gradient techniques we possess information about important properties of the environment, and we are trying to discover how species are distributed in relation to target variables, such as tree species along a gradient of soil moisture, or growing days, or perhaps surnames along a gradient of census reported 'foreign born'. Direct gradient techniques can encompass canonical correlation analysis (C-CA) and canonical correspondence analysis (CCA), each of which uses multiple nonlinear regression to constrain the environmental variables with the surnames or location data. In C-CA we assume a linear response between species and any environmental variables, while CCA does not require linearity and assumes a more conservative unimodal response curve (McCune and Grace 2002).

Indirect gradient analysis is usually employed in an exploratory phase, where the gradients are unknown and have to be inferred from the results of the species ordination. Indirect techniques include PCA, CA, and Detrended Correspondence Analysis (DCA). Use of PCA imposes the assumption of linearity among the variables (Legendre and Legendre 1983; McCune and Grace 2002, Ter Braak, 1985, 1986, 1999).¹⁷

On the 1880 Vermont census sample, detrended correspondence analysis (DCA) was performed. The sample contained 5,646 surnames and 14 counties (excluding unique surnames with frequency of one).¹⁸ The percentage of the surname data accounted for in the first axis using DCA was 12, and the cumulative variance on the first three axes was 28 (Table 3.5).¹⁹ A county map of the first axis (eigenvectors) reveals an east-west pattern which matches rather well with the known settlement geography of Vermont (Figure 3.12). To test whether there was structure other than the spatial distribution in the Vermont data at the county level, I created a null model (no structure) by shuffling the columns and randomly reassigning the data to new positions in the same column. The randomization

results produced a null model having no structure with only a small proportion of the variance explained. From this I infer that in fact there is location structure in the surnames in Vermont at the county scale (McCune and Grace 2002).

Further experiments were done with the Vermont data to examine the effects of removing rare species. As in a tropical jungle where there are many rare species, in human populations there tends to be a great number of rare and unusual surnames. Surnames that are more common will cluster in certain areas, and these clusters often result from short-range migration of ancestors from a small number of founders. Of 12,560 different surnames in the 1880 census of Vermont, only 48 occurred in all 14 counties.

Before running my final DCA, I examined the effect of removing rare surnames.²⁰ Removing low frequency surnames generates a logical trend on percent variance explained. As more surnames are removed, the percentage of variance explained in the first two axes increases; a result of having less data and fewer zero values in the matrix.

There is no single way to interpret ordination results, and because DCA and CA (indirect methods) are exploratory, one must use other variables and other sources for interpreting the results. Interpretation becomes an iterative process. In the Vermont data, regardless of which size sample was used, the first axis identified the east-west contrast with higher scores for places in the northern and eastern parts of the state. The second axis showed a north-south pattern of variation with high scores in the north and extreme variability in the northeast and northwest corners.

What variables might be driving Vermont's surname structure? To test some of the possibilities, I applied CCA, a direct gradient ordination technique (Ter Braak 1996, 1987). CCA is a method for trying to relate variation in species structure (ordination results) to predictor variables. CCA integrates ordination techniques and non-linear regression.

The dataset I examined with CCA contained all of the surnames that occurred 5 or more times in the state: 1824 surnames, 52,000 individuals in 14 counties.²¹ I tested ten predictor variables extracted from the 1880 Vermont census: number of

manufacturing establishments, total acres of land in farms, total annual wages paid in manufacturing, average number of youths and children employed in manufacturing, average number of males 16 and older and females 15 and older employed in manufacturing, and percentages of the population born in Canada, Ireland and Vermont. Using these variables we would hypothesize, for example, that nation of birth would be associated with certain names and reflect migration history, and that manufacturing, farming and wage levels might reveal attractors of migrants, perhaps in ways reflective of nationality or recent arrival.

With these predictor variables I am trying to discover whether the surname structure is more strongly related to one of the predictor variables than expected by chance. I am approaching this analysis with caution since there are other variables influencing the movements of people. After checking them for colinearity, four variables were rotated in and out of the analysis so as to not bias the computation. The correlation matrix in table 3.6 shows that the four variables retained in the final ordination calculation were reasonably independent of one another.

The cumulative percentage of variation explained by the first two axes was only 18%, for the first four axes around 33% (Table 3.7). Of the predictor variables, the one best related to the first axis is percent Canadian born (P value .02), followed by annual wages paid in manufacturing (Table 3.8). To determine the significance of the variables, a Monte Carlo randomization test was performed (1000 runs). This test randomly shuffles the samples of surname data in the matrix and creates a new data set each time. During each run the environmental variables are tested against the surnames/places matrix. If the probability (P) value is significant at <.05, then we can reject the null hypothesis that surname data are unrelated to environmental variables.²²

The scores on the first axis, we recall, show high loadings in the northeast and northwest corners of the state which is precisely where Canadian-born are concentrated; they also show French surnames (Figure 3.13). The high correlations between manufacturing wages are also associated with the structure of the population. The two largest urban areas have attracted large number of migrants, raising the diversity in the western part of the state. This is consistent with the

interpretation I made from the isonymy equations, for higher diversity in Burlington and Rutland.

As we have seen from these examples, ordination methods offer a flexible alternative to same surname methods. While the ordination methods provide less specific quantitative information about relatedness of places, they provide values (eigenvalues) which can be interpreted as estimates of the relative relatedness of all of the names and locations. If the researcher requires a relative value for both names and locations, this method would provide a meaningful result. The true strength of the ordination method, however, is its ability to assess surname data at multiple spatial scales and to uncover several levels of population hierarchy which researchers have recently acknowledged as being absent from surname studies. In the next few sections we will examine further options for assessing the relatedness of names and exploring a hierarchical structure.

Cluster and Fuzzy Clustering

Cluster analysis is a general term for an assortment of statistical procedures for grouping entities in terms of specified variables (Aldenderfer and Blashfield 1984). There exist many definitions of clusters and clustering methods, and an explanation of them is beyond the scope of this thesis. But I will discuss the two methods applied in this research: 1) hierarchical agglomerative and 2) fuzzy clustering (Equihua 1991; Aldenderfer and Blashfield 1984).

The first method agglomerates the surnames step by step (hierarchically) into clusters, starting with pairs of the most closely related surnames (or places) and nesting each cluster into a larger group at each iteration of the algorithm. When the method finishes, every entity (surname or place) belongs to a cluster, and the resulting clusters and their subgroups can be viewed as a tree or dendrogram. Cluster analysis is helpful in developing taxonomy or grouping data into meaningful and significant structure when we have little a priori knowledge. The difficulty with cluster analysis is determining an appropriate number of clusters. Setting the number too low clusters can force data into groups in which they do not belong. Ordination and fuzzy cluster methods do not create unreal clusters, but instead provide relatively more natural relationships between variables and are therefore helpful in deciding on an appropriate number of clusters.²³

Fuzzy cluster analysis is based on the notion of fuzzy sets, where variables are not rigidly grouped, but are organized according to varying degrees of membership in more than one set. In hierarchical cluster analysis a place or surname either belongs or does not belong to a particular group (say a set of French, Chinese and Russian surnames). Where there are no sharp boundaries between clusters, fuzzy clustering may provide a better representation of relationships. It can also be helpful for determining how many clusters should be selected when using ordinary strict clustering techniques by determining how far apart variables are from each other.

The best-known fuzzy clustering algorithm is fuzzy c-means developed by Dunn (1974) and Bezdek (1974, 1981 & 1987). It is available from a number of sources, but the technique itself has been integrated into only a few statistical packages. For this project, I used a program called Fuzzy Grouping which provides information on the degree of membership of each surname, its location in relation to each cluster, and a dissimilarity matrix comparing surname and locations.

Using the two clustering methods provides a means to differentiate regions with similar profiles of surnames, but like the same-surname and ordination methods spatial scale must be considered. A surname is likely to cluster in one or two regions and diffuse outward, decreasing in frequency. If we add more surnames for examination at a regional scale, each local concentration begins to influence the overall regional trends. It is the array of all of these concentrations of high-frequency surnames that defines the overall structure. The rarer names add to regional diversity and to the dispersion of surnames and account for the randomness of the data, but these single names do little to influence the overall structure which is driven by the common names which occur in large numbers.

Cluster analysis allows us to explore the surname structure at several scales. Moving to a higher order (fewer clusters) adjusts the spatial scale of exploration. A small number of clusters may offer a regional perspective, while a great number

of clusters may provide a micro-regional perspective on surname variation.

For 1880 Vermont surnames, I attempted both hierarchical and fuzzy clustering. In each case the sample set consisted of 5,646 surnames and 38 hexagons, excluding unique surnames with a frequency of one. For hierarchical clustering I used Euclidean distances as the distance between objects (locations vs. surnames) and Ward's procedure for grouping the distances between clusters.²⁴ Based on their surname profile, places were grouped into clusters of 25, then 15, 10, 5, and 3. A 'preferred' number of clusters was determined after examining the dendrogram and the geographically mapped clusters.

The maps display the clusters in their associated geographic locations. In most cases a pair of adjoining hexagons belongs to the same cluster, indicating a considerable similarity of their surname profiles. Examining a sequence of maps (Figure 3.14) reveals the hierarchical structure of the regions and subregions. From a display of 15, 10, 5 and finally 3 clusters, we see similarity expressed at ever higher orders, from local to regional. Thus we can imagine the potential of this technique for conceiving a regionalization of a complex geography like that of Quebec.

I move now to the fuzzy clustering results. After examining the partition coefficient, which provides information on the optimal number of clusters, I chose three clusters. It is important to consider that fuzzy clustering does not provide hard or distinct clusters, and that this number just provides a metric to measure the similarity between each location. The results are in the form of a similarity matrix describing the degree of relatedness of each pair of places. It also provides also a numerical comparison of clusters versus locations. The results do show the east-west split demonstrated in the earlier same-surname, ordination and hierarchical clustering results, but overall the results are much more difficult to interpret.²⁵

Given the difficulty in interpreting the results of fuzzy clustering when compared with the coherent and consistent outcome of the hierarchical clustering, it is probably not appropriate to apply fuzzy clustering to the Quebec data. Overall, the clustering experiment suggests that the surname distributions possess strong geographical concentrations and that the structure is hierarchal. Each level

in the hierarchy of surname similarity corresponds to geography at a particular scale.

Relevance Networks

'Relevance Networks' was developed by Butte (1999) and the Children's Hospital Informatics Program at Harvard University as an integrated statistical method for analyzing correlations among variables in large data sets and for creating networks of strongly associated variables. The method was initially conceived as a hypothesis-generating tool to explore the large data sets used in genomics and patient medical databases when no prior model or knowledge of data structure is available. Very specialized, the method can currently be performed only using a program called RELNET.

As a first step, a pair-wise comparison is made for all pairs of locations (features or variables). This is accomplished by choosing a dissimilarity metric (in this case the Pearson correlation coefficient) and a maximum number of associations. Once the pair-wise correlations are completed, the networks are created, filtering out associations that do not meet a threshold. When using the Pearson correlation coefficient, the threshold is set in terms of the r^2 value. By adjusting this threshold upward we exclude the weakest networks because only the variables having the improved linear fit (r^2) are included. The final results are presented as interconnected networks. In other words, we are employing once again the same data set, a matrix of a relatedness measure much like the others, but the important difference is the application of a threshold for the strength of relationships which will be taken into consideration. Setting an appropriate or optimal threshold is important.

Here I further elaborate on how the method works using a sample from Vermont containing 5,646 surnames and 38 hexagons (excluding unique surnames with frequency of one). I created a set of relevance networks for locations and another set for surnames. After creating the correlation matrix for the locations, the number of networks is examined interactively at various thresholds (r^2 values). At a low threshold we obtain a large number of networks, and as threshold is

increased fewer networks result. Figure 3.15 illustrates the 34 relevance networks created when the correlation value is set at .4: each line connecting a pair of locations (e.g. hexagons) represents a connection, and the thicker lines in this dense network represent slightly stronger connections. The final network, illustrated in figure 3.16, was completed at a .6 correlation value. Two networks include 18 of the 38 total features. The large network is split in two, but connected at hexagons 11, 20 and 22. Examining both the hexagons and the relevance networks shows the relationship between geography and the areas in Vermont that have the strongest relationships. These relationships markedly extract the between population surname values calculated for the hexagons earlier.

The networks of surnames are more difficult to interpret than the network of places. We are of course dealing with a very large dataset (over 5000 names). Correlation values were set high to retain only the most related or relevant connections between names. At a correlation value of .9, the results produced 8 networks with just 18 names (Figure 3.17).

Bayesian Probability

Bayesian analysis is a suite of methods based on 'Bayes' theorem for making inferences from data using probability models. A Bayesian method for inferring the origins of migrants into a recipient area was first applied by Degioanni and Darlu (2002), who examined surname records from a receiving area (study area) at two periods of time: a surname which first appeared in period 2 is the in-migrant for whom we want to estimate a probability of origin from areas gk (referred to as area k). Using Bayes' theorem p(si|gk) is the probability of observing surname si within the gk'th area, which refers to the number of locations in the origin (emigrating, migrating) population. The method proceeds by successive approximations. During the first iteration, the 'a priori' emigration (Φ) should be set to .01, but this value is replaced with the value of pgok calculated in step two.

$$p(g_k|s_i) = \frac{\Phi(g_k)p(g_k|s_i)}{\sum_k \Phi(g_k)p(g_k|s_i)}$$

The coefficient calculates the probability of a surname migrating to the recipient area in proportion to its frequency. To derive a probability for the full set of surnames, we sum the probabilities for all the names and then calculate a weighted average. When the weight is set to 1, we obtain the probability of the origin of the set of names, and when we make this weight equal to the frequency of the surname within the recipient area, we are estimating the probability of the origin of population of migrants (bearing those surnames)

$$pgok \frac{1}{\sum_{i} w_i} \sum_{i} w_i p(g_k|s_i)$$

Conclusion

We have looked at the surname geography of Vermont in a naive and unbiased way, applying the whole panoply of methods available from human biology, population ecology, and the latest developments in clustering and relevance networks. From the resulting parameters, graphs and maps of surname relationships, we observed at a regional scale (county-wide and hexagon scale) differences in composition in the eastern and western counties (split by a mountain range) and between the far northern and southern counties. When we examined the possible influences on the distribution of names in Vermont, we found some of the data variance associated with the larger share of French Canadians in the northeastern counties and with annual wages paid in manufacturing, another indicator of an industrializing region attractive to French Canadian immigrants. The name geography revealed by the various methods discloses a history and generates a series of questions about population movements of the past. Prior studies of the history and geography of Vermont suggest some further reasons for certain surname patterns.

What we can infer from the Vermont exercise is constrained by spatial and temporal boundary problems and a want of information on the 'founder' population. In turn, the space under analysis, the state and county boundaries are arbitrary and porous to the population movements we are trying to observe, but the imposition of a hexagonal grid somewhat 'dissolved' these boundaries, making

display of relationships easier to read. Scale issues could have also been further improved by using places or towns and calculating isonymy between them. This technique would be only possible if we develop methods for calculating confidence intervals for calculating isonymy values for various population sizes. The analysis is also confined by available sources to certain moments in time: while trends are interesting and surname-relatedness at a moment in time reflects a prior history of population movement, we cannot under these constraints pinpoint when the various movements occurred. This problem can be partially resolved by examining earlier censuses, suites of marriage or land records. The absence of information on a 'founder population' would be exceedingly difficult to resolve for Vermont because of the high turnover and widespread diffusion of popular English names throughout the northeastern US. A possible solution might be to only focus on relatively rare English names or examine land ownership. There is some evidence that land owners were less likely to leave the state (McGrory Klyza and Trombulak 1999).

Although the exercise does provide some insight into the population history of Vermont, the prime purpose of the analysis of surname geography of Vermont 1880 (or 1840 and 1880) was to help choose among methods by identifying their relative efficiency and appropriateness for answering specific geographical questions. It is now possible to do some 'sorting' (accept, reject, confine, extend) of the various methods and to decide with which questions each of them deals best.

Initially, I conceived the Vermont case as a simplifying vehicle for summarizing the statistical methods, but the findings led me to treat it as a pilot program of analysis for the Quebec data. The historical geography of Quebec- the real target- is a much bigger problem and a better test of the uses of surname geography. Quebec has a longer history, a much larger population, a larger subset of population nodes, and is a vast space situated in a very complex topography which facilitates or inhibits movements of people. The onslaught called, therefore, for some efficiency in methods, some choices and some more explicit questions. We recall also the advantage of this case study: Quebec has an exceptionally welldefined and well-studied founder population, and my Quebec datasets, unlike the Vermont datasets, will allow us to frame questions about French Canadian population movements in a way which could be extended into Vermont, into the US and into the entire continent. Because the Quebec samples are well defined, the results can be used to refine the observations which have been derived for regions where little is known about the founder population.

The series of tests on surnames in Vermont assisted me to select a specific set of problems in the historical geography of Quebec, so that for each problem I can select from my toolbox just the right size screw driver or wrench. On the basis of the Vermont exercise I am made the following choices/decisions about the Quebec analysis. I apply the 40km hexagon grid, so that I can overcome boundary problems of counties (especially arbitrary in Quebec, and misleading with respect to limits of the ekoumène), and overcome the difficulties of progressive subdivision of parishes, complicated and erratic over the long history of settlement. The only test which I outright rejected for the Quebec analysis was fuzzy clustering because of the obscure and hard to interpret results. Hierarchical clustering seems appropriate since the surname geography itself is hierarchical. But it will be important to look at the number of total clusters iteratively and evaluate an appropriate number of clusters. The other methods will not be applied to every Quebec dataset, but will be invoked to answer specific historical questions at a particular spatial and temporal scale.

The backbone of the analysis is the measures of regional, local and between population similarities measures (isonymy). These measures provide an overview of name relatedness, zooming into some regions to answer specific questions. During the Quebec analysis I focus strictly on surname similarity measures, without delving into the derivative forms like kinship and genetic distance, since these forms are more appropriate at a fine spatial scale or for comparative studies. Associated measures like inbreeding and the 'repeated marriage pairs' are used in a case study to provide an indicator of marriage fields and determine the geographic overlap between surname groups. Using the same marriage methods I hypothesized that in more heterogeneous populations with greater name diversity the proportions of same surname or name marriage pairs are lower than in more homogenous populations with lower name diversity: a factor also affected by marriage field distances.

From the test in Vermont, Detrended Correspondence Analysis (DCA) seems to be the most economical approach to an overview of 'relatedness' at different scales. The method provides excellent insight into the organization of the locations (parishes or hexagons) although interpreting the organization of names can be difficult due to the variability in their numbers. This problem is not uncommon in community and forest ecology and is usually solved by removing species (or names) whose frequency is below a certain threshold (e.g. >5 occurrences). Despite the suitability of DCA to explore the population structure for the 1881 census where the size and complexity of the data structure call for a simplifying statistical tool, I do not include analysis using this method.

DCA and hierarchical clustering techniques are complementary, and one technique can be used to verify the results of the other. Clustering is a rigid method, useful for organizing names and mapping groups of surnames, but because the method is so rigid, it sometimes groups objects in the wrong clusters. To get a good sense of the number of clusters, we need to examine the results in the context of other measures such as same-surname (isonymy) or relevance networks. Having prior information (e.g isonymy) as a check for determining the number of clusters is important and can help improve the accuracy of the final results. Clustering will be used extensively with the eighteenth and nineteenth-century Quebec data sets. It is the tool best suited for uncovering the hierarchical structure so characteristic of the distance-decay phenomenon and the results can be easily mapped. We will see, as I apply clustering and map each lower order cluster, the hierarchical population structure.

Relevance networks will be applied to eighteenth-century marriage records for Quebec and to the nineteenth-century census data. As the name implies, the technique identifies the salient relationships, those which meet a pre-determined threshold. The threshold will be set fairly high for this exercise to reveal the strongest networks.

The final method I apply is that of Bayesian probabilities. It is used to appraise

the parts of Quebec from which individuals were emigrating to Vermont, New Hampshire and Maine. The method is not without faults as we are not tracking individual migrants in real-time, but are instead inferring movements based on the frequencies of names found in the origin and destination populations. Despite some flaws, we will see the utility of the method when it is used on a population where we know the single founders and have control over surname spellings.

While there are many more methods, I have covered the most relevant and easiest to apply. No single method provides everything one needs to know about the population, and each method is better suited for certain questions. While my approach has so far been exploratory, identifying the array of tools that can be applied to surnames and raising some precautions concerning the data, a better way to proceed (the strategy) with the Quebec work is to begin with a question (or questions) and select the most appropriate method.

During the course of the Quebec work it will also be important to consider spatial scale. As we will see at different spatial scales we get slightly different interpretations of population organization. Both interpretations are 'correct,' but the patterns may arise from distinct demographic or migration processes.

NOTE

⁵ This method is an expansion of G. Darwin's 1875 research on using marriage records to compare similar surnames to estimate inbreeding.

⁶ (number of isonymous marriages divided by total marriages).

⁷ The equation $I_{ii} = (\sum m_{ik} f_{ik}) / (M_i F_i)$ is twice the value of Lasker's (Ri) withingroups coefficient of relationship. For more information on the use of this equation see (Relethford 1988) (Lasker 1998; Lasker 1985).

Almost always population heterogeneity (diversity) increases with town size. Changes in population size, increased urbanization, deteriorating of family farming, and improved transportation has led to what Dahlberg (1943) refers to as the breakdown of population isolates—including decrease in consanguinity and distance between birth and place of marriage

⁸ The variance is a measure of how spread out a distribution is. It is computed as the average squared deviation of each number from its mean.

⁹ When calculating multi-dimensional scaling (MDS) the user specifies the number of dimensions, and the method minimizes the stress. The residuals measure the influence of each point on the stress statistic. Stress is the square root of the ratio of the sum of squared differences between the input distances and those of the configuration to the sum of configuration distances squared. The best way to understand multidimensional scaling is by examining a distance matrix between places. The first two axes would resemble a geographic map showing the euclidian distances between each place.

¹⁰ The covariance of two features measures their tendency to vary together (co-vary). The variance is the average of the squared deviation of a feature from its mean and the covariance is the average of the products of the deviations of feature values from their means.

¹¹ (The smaller the genetic distance between two populations the closer they are genetically (Relethford 2001))

¹² Rst is defined as: $_{Rst} = \sum (wr_i)$ where wi is equal to the number of people in population i divided by the total number of people in the entire region rii is equal to $_{W_i} = n_i / \sum n_i r_{ii} = (I_{ii} - R) / [4(1 - R)]$, Fst is computed as $F_{ST} = R_{ST} + (1 - R_{ST}) \Phi R$ Fst is a measure of unreduced variance and Rst is a measure reduced variance. FST value range between between 0 and 1. Values of -.5 - 0.15 is moderate genetic differentiation, and greater than 0.25 is very extensive genetic differentiation.

¹ Geospatial and Statistical Data Center, University of Virginia. Available data differ somewhat from decade to decade, according to what was collected in the census and the items chosen for transcription to electronic form (1790-1960).

² Gerald O. Lesperance, Vermont French-Canadian Genealogical Society

³ For an excellent review of surname extinction see: Hull, D. M. 1998; Lange, Kenneth. 1981

⁴ Inbreeding should not be confused with the "coefficient of consanguinity" which is the probability that two relatives will have identical alleles by descent from a specific allele in a common ancestor.

¹³ High values of Fst show that the region with all of its populations included in the study are more dispersed from the mean; representing that they are more different from each other. Low values of Fst are clustered close to the mean; representing that they are not that different from each other. A population with all different surnames would have a low Fst.

¹⁴ To put genetic micro- differentiation values into context, consider the high Fst and Rst values of the Nebraska Amish in Pennsylvania who have Fst values around .1014 and Rst values around .0107. Such high values are indicative of geographic and religious isolates (Relethford, 1988).

¹⁵ This section only covers a small proportion of the concepts of surname analysis and geographic/spatial structure. For a more rigorous review see Epperson's book *'Geographical Genetics' (2003)*

¹⁶ (See McCune and Grace 2002 for a review of a broad range ordination of techniques based on PcOrd Software; Mooney and Swift 1999 for a clear and concise explanation of matrix algebra and eigenanalysis, also Legendre and Legendre 1983).
¹⁷ see (Ordination Web Site)

http://www.okstate.edu/artsci/botany/ordinate/terminol.htm

¹⁸ Correspondence analysis is an eigenanalysis-based ordination method that is a variant of the linear method of principle components analysis (PCA). Different fields rediscovered this technique, which is very popular in France and Europe. It is most popular in plant geography and community and forest ecology where it is used to explore site-species relationships in samples of many species at many sites. Correspondence analysis is a tool for representing and exploring the association between two or more categorical variables by representing the categories (rowscolumns) surname/places) as points on an x/y axis. Categorical variables with similar distributions are represented as points close in space, while dissimilar or less similar variables are located further apart. For this study I used the program Canoco: Software for Canonical Community Ordination. DCA is used if there is an arch effect in the ordination diagram. When running CA the first axis will be accurate, but the second axis is distorted by becoming an arched function of the first axis. The arch is caused by a unimodel (non-linear) response of species along gradients and can be removed by detrending. Detrending involves dividing the axes into segments and then reassigning the high sample scores so they are centered on the centroid. The arch can also be avoided by reducing the variance in the data by removing the surnames that only occur once.

¹⁹ Percent of Variance of the species data explained by the axis is calculated as $100*(\lambda_1 + \lambda_2)/\sum$ eigenvalues Canoco will give you this in the results file for PCA, CA, CCA except for DCA. PCORD does not provide this information either

²⁰ [Removing rare species is often done in ecological research]. [There seem to be various opinions on whether or not rare species should be removed from the sample before an ordination method is used.]

²¹ I also ran CCA with all surnames (12,560) and with all of the surnames occurring

once removed. Each time the results were similar

²² See Bryan F. J. Manly (1996) '*Randomization, Bootstrap and Monte Carlo Methods in Biology*' for info on this topic.

²³ (Cluster sources: Sneath, P. H. A. and R. R. Sokal. 1973; (See Legendre and Legendre 1983; Numerical Taxonomy. San Francisco: W. H. Freeman; McCune and Grace 2002; McGarigal et al. 2000 Multivariate Statistics for Wildlife Ecology Research)

²⁴ Euclidean distance(x,y) = { $\sum_{i} (x_i - y_i)^2$ }

Ward's method is distinct from all other methods because it uses an analysis of variance approach to evaluate the distances between clusters. In short, this method attempts to minimize the Sum of Squares (SS) of any two (hypothetical) clusters that can be formed at each step. Refer to Ward (1963) for details concerning this method. In general, this method is regarded as very efficient; however, it tends to create clusters of small size" (Statistica 2004). Ward's method is recommended as a general-purpose linkage method that minimizes distortions in the underlying space.

²⁵ Another possibility for clustering surnames is through the use of Neural Networks, particularly the Self Organizing Maps (SOM). A neural network is a structured numerical learning algorithm designed for data classification, estimation, simulation, and prediction. Neural networks have their roots in our understanding of the human brain and the fundamental concepts are based on attempts to mimic the way the brain processes information. The many methods which fall into the category of neural networks involve the concept of information being moved along a network as in graph theory, with a set of synapses or connectors (edges, links and arcs) connecting the nodes or neurons. The network function is determined by the connections between nodes, the number of connections and the weighted adjustments made at the connecters or nodes as the data makes its way through the network. It is common for neural network models to be adjusted, or trained, using a collection of data as a source input.

Neural network methods which fall into two main categories: supervised and unsupervised. Supervised networks must be supplied with the true classes of the training data and the desired output when training the input data. An example of a supervised neural network is optical text recognition (OCR) or image recognition, where the letters or pictures are supplied and deciphered prior to trying to train other data to match the original inputs. Unsupervised neural networks are designed to look for unknown structures in data.

	1800	1810	1840	1880	1920
Total surnames 'unique'	4490	7600	10300	12800	21000
٠		1800-1810	1810-1840	1840-1880	
Surnames Lost		2056	4586	6550	
Surnames Gained		5218	7298	9023	

Table 3.1 Total surnames in Vermont from 1800-1920, and total names lost and gained by each decade



Figure 3.1 Log-log distribution of the frequency of occurrence of surnames in 1840 (red squares) and 1880 Vermont (blue diamond)

County	1840 Sample (Total Heads of Household)	1840 # of Unique Surnames	1880 Sample (Total Heads of Household)	1880 # of Unique Surnames	1840 Local Isonymy	1880 Local Isonymy
Addison Bennington Caledonia Chittenden Essex Franklin Gran Isle Lamoille Orlange Orleans Rutland Washington Windam	(10tal Heads of Housenoid) 4146 3303 3978 4010 767 4021 688 1945 5020 2506 6139 4470 5179 7325	# of Unique Surnames 1690 2161 1549 1801 1069 1247 396 1016 1690 1702 1971 1283 402 1897	(1014) Heads of Household) 5299 5100 6295 7356 2006 6600 855 3001 6161 5461 10066 6000 6425 8254	# of Omque Suffaitles 2225 1906 2174 2944 1056 2702 431 1230 1885 1959 3251 1986 1840 2271	0.00167 0.00167 0.00193 0.00135 0.00326 0.00124 0.00407 0.00187 0.00209 0.00177 0.00166 0.00164 0.00210 0.00174	0.00124 0.00136 0.00147 0.00094 0.00195 0.00100 0.00351 0.00193 0.00143 0.00141 0.00128 0.00166 0.00207 0.00168
Total	53497		78879			

Table 3.2. Total surnames and same-surname values (isonymy) for all heads of household in 1840 and 1880 Vermont

Source: 1840 and 1880 United States Census

Table 3.3 1840 and 1880 Vermont conditional kinship values determined from surnames. 1840 values are located in the upper diagonal and the 1880 values are located in the lower diagonal.

County	Addison	Bennington	Caledonia	Chittenden	Essex	Franklin	Grand Isle	Lamoille	Orange	Orleans	Rutland	Washington	Windham	Windsor
Addison	1	-0.000042	-0.000004	-0.000035	0.000021	-0.000043	-0.000086	-0.000021	0.000016	-0.000013	0.000018	-0.000016	0.000011	-0.000012
Bennington	-0.000516	1	-0.000055	-0.000057	-0.000022	~0.000076	-0.000107	-0.000056	-0.000053	-0.000053	-0.000028	-0.000050	-0.000011	-0.000034
Caledonia	-0.000495	-0.000505	1	-0.000182	0.000080	-0.000050	-0.000118	-0.000005	0.000042	0.000038	-0.000020	-0.000006	0.000017	-0.000006
Chittenden	-0.000517	-0.000530	-0.000527	1	-0.000017	-0.000072	-0.000096	-0.000245	-0.000043	-0.000064	-0.000032	-0.000052	-0.000035	-0.000041
Essex	-0,000465	-0.000474	-0.000408	-0.000502	1	-0.000045	-0.000111	0.000049	0.000061	0.000022	0.000022	0.000034	0.000073	0.000040
Franklin	-0.000528	-0.000537	-0.000519	-0.000541	-0.000494	1	-0.000117	-0.000038	-0.000035	-0.000064	-0.000036	-0.000055	-0.000031	-0.000041
Grand Isle	-0.000542	-0.000557	-0.000687	-0.000563	-0.000553	-0.000606	1	-0.000071	-0.000113	-0.000083	-0.000084	-0.000104	-0.000078	-0.000100
Lamoille	-0.000480	-0.000487	-0.000442	-0.000504	-0.000414	-0.000539	-0.000603	1	0.000016	-0.000018	-0.000009	-0.000002	0.000001	0.000018
Orange	-0.000478	-0.000492	-0.000421	-0.000508	-0.000397	-0.000553	-0.000605	-0.000423	1	0.000027	0.000021	0.000032	0.000020	0.000022
Orleans	-0.000494	-0.000507	-0.000430	-0.000528	-0,000431	-0.000558	-0.000593	-0.000447	-0.000423	1	-0.000022	-0.000016	-0.000278	0.000004
Rutland	-0.000488	-0.000480	-0.000485	-0.000512	-0.000456	-0.000574	-0.000607	-0.000460	-0.000465	-0.000493	1	-0.000004	0.000035	0.000018
Washington	-0.000479	-0.000484	-0.000428	-0.000503	-0.000400	-0.000551	-0.000584	-0.000398	-0.000391	-0.000432	-0.000461	1	0.000011	0.000002
Windham	-0.000447	-0.000431	-0.000417	-0.000547	-0.000384	-0.000531	-0.000594	-0.000386	-0.000337	-0.000414	-0.000415	-0.000379	1	0.000058
Windsor	-0.000481	-0.000431	-0.000456	-0.000505	-0.000426	-0.000559	-0.000596	-0.000396	-0.000399	-0.000444	-0.000453	-0.000415	-0.000453	1

Source: 1840 and 1880 United States Census



Figure 3.2 Ancestor paths (chains) showing the probability of homozygosity of a child between two cousins. Ancestor paths include F,D,A,C,E.

Table 3.4 Measures of micro differentiation determined from surnames for 1840 and 1880 Vermont.

 ΦR is equal to regional isonymy divided by 4.

Fst is genetic microdifferentiation and refers to the average within-group (populations) unreduced variance. Rst is genetic microdifferentiation and refers to the average within-group reduced variance.

		ΦR	Fst	Rst
	1840	0.00031	0.00162	0.00132
	1880	0.00073	0.00109	0.00036
Hexagons	1880	0.0031	0.00195	0.00120



Figure 3.3 Vermont map illustrating same-surname values (local isonymy) for 1840. Counties having the greatest diversity of surnames have the lowest values (lighter colors).



Figure 3.4.Vermont map illustrating same-surname values (local isonymy) for 1880. Counties having the greatest diversity of surnames have the lowest values (lighter colors).



Figure 3.5 Vermont hexagon map illustrating 1880 same-surname values (local isonymy). The hexagons are aggregates of the places as represented in the 1880 census. Hexagons having the greatest diversity of surnames have the lowest values (lighter colors).



Greatest between population surname similarity

Figure 3.6 Strongest between population surname similarity (isonymy) connections in Vermont. The arrows display the locations from which each county has its strongest surname similarity. In this example, most of the arrows connect with Windham County in Southern Vermont. Note the circles represent the size of the population of Men 21 years and older and the color of the circles provide a relative measure of name diversity.



Second-order between population isonymy

Figure 3.7 Second strongest between population surname similarity (isonymy) connections in Vermont. The arrows display the locations from which each county has its strongest surname similarity. In this example, most of the arrows connect with Windham County in Southern Vermont. Note the circles represent the size of the population of Men 21 years and older and the color of the circles provide a relative measure of name diversity.



Figure 3.8. Multidimensional Scaling results explaining the between population surname similarity (isonymy) values for 1880 Vermont. Proportion of the variance explained is .932. Note that the results show that there is a substantial east-west divide in population surname structure and Lamoille and Washington County have a very close relationship according to names.



Figure 3.9 Between population 'a priori' kinship values for Washington, Addison, Orange and Windsor counties. The gray (darkest) county is the reference population in which 'a priori' kinship values are compared to all other counties. Counties with the lightest shading have the lowest values of 'a priori kinship' when compared to the reference population.



Figure 3.10. Between population 'a priori' kinship values for small areas within Washington, Addison, Orange and Windsor counties. The gray (darkest) hexagon is the reference population in which 'a priori' kinship values are compared to all other hexagons. Hexagons with the lightest shading have the lowest (least related) values of 'a priori kinship' when compared to the reference population.



Figure 3.11. Multidimensional scaling results explaining the 1880 Vermont 'a priori'kinship values.

Results show that some of the differences in 'a priori kinship are a result of the distances (locations) between each location. If the pattern (organization) were purely random there would be no geographic structure and we would not be able to isolate geographic relationships as we do above. Proportion of total variance explained is .659

No. of places: 14 No. of surname: 4491			an an an ann an ann ann ann ann ann ann		
Axes	1	2	3	4	Total inertia
Eigenvalues	0.24	0.179	0.141	0.099	2.006
Lengths of gradient	2.314	1.913	1.62	1.889	
Cumulative percentage variance of species data	12	20.9	27.9	32.9	
Sum of all unconstrained eigenvalues	2.006				

Table 3.5. Detrended correspondence analysis of surnames for 1880 Vermont



Figure 3.12. First and second axis (dimensions) of the detrended correspondence analysis postioned onto map

1	No. of manufacturing establishments	Total acres of land in farms	Total annual wages paid in manufacturing	Percent Canadian Born	Average No, of males 16 and older employed in manufacturing	Average No. of females 15 and older employed in manufacturing
Manu	1					
FarmAc	0.5602	1				
anwages	0.5222	0.3322	1			
CAN	-0.5246	-0.5107	-0.2248	1		
Males16	-0.3781	-0.1751	-0.5021	0.2239	1	
Fems15	0.379	0.21	0.6143	-0.2853	-0.5453	1

Table 3.6. Correlation coefficients of environmental variables for 1880 Vermont

Table 3.7. Results of canonical correspondence analysis of surnames for 1880 Vermont

Axes	1	2	3	4	Total inertia
Eigenvalues	0.223	0.217	0.19	0.183	2.469
Surname-environment correlations	0.98	0.994	0.962	0.998	
Cumulative percentage variance					
of surname data	9.1	17.9	25.6	33	
of surname-environment relation	19.8	39.1	55.9	72.1	
Sum of all unconstrained eigenvalues					
Sum of all canonical eigenvalues	2.469				

Table 3.8. Vermont 1880 example:probability (p) values for CCA.

Conditional Effects	
P Value, 1000 Runs	
	0.0200
	0.1650
	0.5070
	0.6330
	0.6310
	0.6330
	Conditional Effects P Value, 1000 Runs



Figure 3.13. 1880 Vermont: ordination diagram (tri-plot) illustrating results of the CCA.



Figure 3.14 Vermont hexagons clustered at various degrees: 15, 10, 5, 3. The different clusters can be distinguished by examining the numbers. Each common number is representative of the hexagons in each cluster



Figure 3.15 Relevance Networks correlations set to 4.8 34 features, 1 network



Figure 3.16. Relevance Networks correlations set to 6.0 created 2 networks, one with 18 features and the other with 2 features. This example used the same data as above and we can see that as the correlation values increase only the most relevant networks remain.



Figure 3.17 Relevance Networks set a 8.5 for surnames, 8 networks representing a total of 18 surnames (features). The names here represent those which are most correlated, meaning they show up most often in the same locations.
Chapter 4

18th Century French-Canadian surname population structure

In this chapter I apply surname methods to Quebec before 1800. Interpretation focuses on the use of a surrogate census for the year 1765 and eighteenth-century marriages of men bearing the names of 832 single founders. A number of in-depth studies have examined demography and kinship in the population of New France from parish records and genealogical reconstitutions, but few have attempted a regional display of settlement across the territory which is now Quebec (Courville, 1996; Bouchard 1985). I shall concentrate on what we can infer from methods presented in the previous chapter, about name patterns and their geographical organization, the settlement of regions, name diversity, and local and regional migrations.

We know from the previous work of Gagnon and Toupance (2002) and Bouchard (1985) that by the beginning of the eighteenth century surnames were polarized in the three governments of Quebec City, Trois-Rivières and Montréal. To what extent does this early population organization represent a hierarchical structure and how will this early organization affect the future population organization? There is also evidence for local and kinship networks and marriage fields of short radius. Were these elements already strong enough to shape a local structure of names and if so, to what extent ? Does the St Lawrence River serve as a barrier to movement or a network channel assisting settlement? I examine the development of the name structure and comment briefly on how its organization was shaped by migration, land concessions and cultural barriers.

The chapter begins with a brief review of nineteenth-century settlement history. I follow with an interpretation of the name structure using various methods, including same-surname methods (isonymy), cluster analysis and relevance networks. Because I use only single founders in the samples, we recall that the presence of a surname in several different parishes implies a degree of relatedness from which we can infer genealogical links and a history of migration.

Historical Geography

Immigration to New France was initially rather slow and by 1650, only 700 settlers were living permanently in the colony, with most located close to Québec City, Montréal or Trois Riviéres, along the St Lawrence river. Slow population growth was a result of a policy of not populating the new colony at the expense of the métropole, and a desire to send only enough settlers to ensure that the population could expand through natural increase (Charbonneau et al. 1993; Kuczynski 1930; Henripin 1972). The small size, combined with a moderate death rate and a shortage of women, kept the growth rate low in the first few decades.¹ Fearing the colony would fail, the French Crown began increasing the numbers of immigrants, and particularly women immigrants (Landry and Légaré 1987; Landry 1988). Larger numbers arrived between 1663 and 1772, and three quarters of the entire founder population arrived before 1800. In all, 10,000 colonists married in the colony, and approximately 6500 of them had at least one male son who married. With this boost, by the mid eighteenth-century the population had begun to grow exclusively by natural increase (Charbonneau et al.1993; Courville 1996).

By 1760, the population reached 80,000, by 1800, 200,000, growing at an annual rate of 2.5%, sufficient without immigration to double the population about every 25 years (Charbonneau et al.1993; Courville 1996). This growth prevailed until the late-nineteenth century when further modernization led to decreases in death rates and by the 20th century decreases in birth rates (Haines and Steckel 2000). Before 1800 the British accounted for only a small proportion of total population. In 1760 there were 500 British in Quebec, mostly artisans, merchants, soldiers, and bureaucrats. Their numbers would increase with the influx of British Loyalists after the American Revolution, and by 1815 there were around 50,000 British in Quebec. Other populations living in Quebec at this time were the Acadians whose numbers increased in Quebec after their deportation from Nova Scotia in 1755, and the Native Peoples located primarily on the periphery of the seigniorial lands.

The seigniorial system provided a social and economic organization to facilitate settlement. The company of *One Hundred Associates*, under obligation to the French government (King), was in charge of parceling out seigneurial lands (large tracts of land) to individuals. A seigneur was obliged by the king to cede rotures (concessions of land) to colonists (Harris 1969; Courville 1996). Seigneuries were conceded along the river in long narrow trapezoids fronting the river so they could maximize the number of families who would have access to the river. The morphology of the seigniories is strongly influenced by the need for access to the waterway for communication and movement of goods.²

Settlement progressed from the nodes of Montréal and Quebec City along the banks of the St Lawrence towards Trois-Rivières . By 1780 there were two ribbons of settlement, extending along the north and south shores of the St Lawrence between Quebec City and Montréal. By this time the population had expanded to 80,000 and some older seigneuries like Beaupré and Ile d'Orléans were already losing young settlers to newly opened lands behind the older seigneuries or upstream along nearby tributaries (Harris 1969, 1987). Backfill of the seigneurial lands was very typical and this had an impact on the movement of families (names) who often chose to live close-by.

When we look at the geographic distribution of names in the eighteenth century we must consider the influence of land inheritance on settlement and migration patterns.³ Continuous subdivision and crowding of seigniorial lands were a result of both the increasing number of rotures being ceded and the inheritance of property within large families. When a censitaire (landholder) died, his land was divided among his wife and children. The spouse would receive half, and the other half was divided equally among the children (Harris 1969, Trudel 1971). Because families were large, it did not take long before land could no longer be subdivided and individuals would have to consider other options which might prompt them to move further away.⁴

As a typical settlement process, an individual (head of household) might acquire a single roture from a seigneur, begin a family and farm the land. If no more land was available, he might sell his strip of land and move to a new region

where land was plentiful. With such a settlement process in an agrarian society, there would certainly be a period of equilibrium, when land could no longer be subdivided and density in these areas remained stable. This of course would change with modernization when higher-density populations could be sustained.

Rapid population growth and settlement under the seigneurial system had an impact on patterns of surnames. Clusters of names were more likely to occur where families could divide land and give portions to their children, but once land was filled, we see the same name pop up some distance away where land was available. The opening of prime farm lands on the largest and most valuable tributaries perhaps influenced settlers to move further away from the original farm rather than to remain on less valuable and smaller lots (Courville 1996; Bouchard et al. 1995).

The desire of families to stick together seems to be a dominant factor in the diffusion of names. 'Family' was the foundation of work groups, the sole form of 'insurance' for old age, and an institution of intense sociability and socialization (Collard 1990; Bouchard 1996). These choices are not limited to Quebec, but are well documented and were favoured by the system of inheritance under the Coutume de Paris, by the system of land grants and the high natural increase (Harris 1969). If we look at the distribution of five names at this time we can visualize the tendency for them to remain in relatively close proximity, and we will see later that these names will continue to cluster. Figure 4.1 shows the distribution of marriages for each of the five names between settlement and 1800. Each name is regionally concentrated and disperses outward or away from a central point. We will see the results of this dispersion in a later chapter. Though these are only five of thousands of surnames in the French Canadian population, they exemplify the overwhelming pattern which characterizes most of the names in Quebec. For example, looking at the Beauchamp surname (Blue) one can see that most of the marriages occurred close to Montreal. We will see, a century later, that the names, although dispersed, are still found in and around this region.

Eighteenth century same-surname results (isonymy)

Genetic differentiation across the total population was estimated in terms of Fst and Rst values (Wright 1951) for four different time periods, using marriage records (1621-1799) and for 1765 using the surrogate census. Marriage records show the lowest Fst and Rst values in the first 30 years of the eighteenth century, followed by a modest increase during the remainder of the century. Knowing that we are grounded in a founder population, we can be confident that the initial low values are indicative of a genealogically young population with a large number of unique names distributed randomly along the shores of the St Lawrence. After 1740 the Fst values begin to increase, precisely what we would anticipate since migration was slowing down, population size was increasing, and we are beginning to the see the domination of certain names through probabilistic advantage. Overall the micro-differentiation values for the eighteenth century are fairly high, and in a later chapter we will see a further moderate reduction as the population increases and some French Canadian names begin to emerge in larger numbers and the new non-French Canadian immigrants begin to integrate into parts of the population, increasing the overall diversity of names.

The values derived from the 1765 surrogate census give results very similar to those derived from marriage records covering the last thirty years of the eighteenth-century. These results are a good indicator of the reliability of the 'census' I created from baptisms. Overall, the differentiation values for the single founder marriage and census sources ranged from .0098 to .0129 (Table 4.1). Despite the variability, values are still rather high when compared with 1800-09 Massachusetts, seventeenth and eighteenth -century Kings County, New York, 1936 Western Ireland, 1753-1850 Otmoor England, and the measures I presented earlier for 1840 Vermont (.0016)⁵. Explanation of this lies in three features of the study: (1) We are already dealing with a large number of parishes scattered along a considerable distance. (2) This is still a 'young' population, within a few generations of founders' arrival. (3) Most important, the use of single founders

keeps the number of names to a small sample. Most studies have not been able to identify the single founders. I compared these results with the entire population from the 1765 surrogate census. Here, the Fst value was .0034 and the Rst value was .0032: values much lower and closer to the above cited studies, all of which used data from an entire population and assumed the monolithic origins of names.⁶

The range of micro-differentiation values between the sample in this study and those found in other studies provides evidence of the need to resolve the problem of determining the variability as a function of population size and number of parishes. This could be done through a method of data normalization, or surname simulation models which could be set up to observe how genetic differentiation values change under different demographic parameters. In theoretical population genetics, simulation models are often used to estimate changes in Fst under different levels of gene flow (migration) between populations. Until such a technique is carried out for surname studies, we must exercise caution in our comparisons.

Local same-surname (isonymy) values for the 91 parishes ranged from .005 in Montréal, settled in 1642 to .55 in Les Éboulements and settled 91 years after the settlement of Montréal. The low value for Montréal is indicative of a population with great name diversity. Other populations with high diversity of names include Ouebec City settled in 1621, Trois-Rivières settled in 1634, L'Assomption settled in 1724, and Chambly settled in 1706 (Table 4.2). This is consistent with the findings of Tremblay et al. (2001), Vézina (2001) and others who have demonstrated from descending genealogies that Montréal, Quebec City and the Richelieu Valley are the most diversified. The early implantation of diversity in western Quebec is still apparent in presentday genetics and the Richelieu area remains one of the most connected (and diverse) to other areas of Quebec. We also find a high diversity of names north of Montréal along the St Lawrence River at L'Assomption and Berthier-en-Haut in the Lanaudière. All of these places with high name diversity are well situated in the river system on which communication depended and which we find the earliest systems of mail, courier and reporting through the hierarchy of colonial and ecclesiastical governments. We will see later

that well-situated parishes over time will continue to have a high diversity of names and a high level of connectivity with other parishes.

While the earliest-settled locations had more time to grow their populations, a plot of date of settlement versus local isonymy values shows no relationship, a likely result of the constraints on population density and sustainability. Figure 4.2 shows that many parishes settled after 1700 have populations just as diverse as earlier-settled parishes and most have values within one standard deviation from the average of .07. Exceptions, or outliers, include St-Joachim (settled in 1725), Les Éboulements (settled in 1733), and L'Isle-aux-Coudres (settled in 1741). The majority of parishes showing such low diversity are found north of Quebec City 7 (Figure 4.3). Besides Montréal and Québec City, seats of trade and government where there was higher population density, most rural populations at this time had limited carrying capacity, and when farm land was no longer available, families moved, and sometimes large kin groups moved to newly settled land (Bouchard 1996). The availability of new lands permitted out-migration, with loss of names, and diverted migrant streams from elsewhere: thus the older parishes diminished in surname diversity. Another consideration, of course, with the same-surname values, is the meaning that can be attached to the date of the opening of parish registers. Often parishes were settled well before they were officially recognized. **Between population same-surnames values**

As we saw with Vermont, the between-population same-surname values offer inconvenient numbers of comparisons. With 94 parishes, we are considering 4371 pairs of relationships. To simplify the results, I will focus on the extreme cases, the earliest-settled parishes, and the parishes with largest populations, and I will use multidimensional scaling to assess and reduce the dimensionality of the numerous relationships. Tables 4.3 and 4.4 present results for the 30 largest parishes in 1765 and for the first 30 parishes settled.

Between population same surname values average .005 in the 94 parishes. This means that if we take two parishes at random, and choose a single individual from each, the likelihood of their bearing the same surname is very small, about one-half of one percent. If, however, we choose the nearby parishes of Rivière-du-

Loup (Louiseville) and Nicolet with a between-population value .891, the chances of selecting two of the same names is high, and if we picked Champlain and Longueuil (relatively distant) we would find that they shared no names. Figure 4.4 illustrates the geographic context of Rivière-du-Loup (Louiseville), highlighting the locations with name associations as well as the four strongest name affiliates, all located between Quebec City and Montréal. We notice that Rivière-du-Loup, located on the western side of the St Lawrence, (shows) strong affinities with parishes on both sides of the river, providing our first evidence of kinships extending across the waterway.

The Mantel test revealed no obvious relationship between the array of geographic distances and same surname values (isonymy). Multidimensional scaling did, however, reveal some structure and some salient relationships (proportion of variance (RSQ) 63% in the first two dimensions. A biplot of the first two dimensions provides evidence of the connectedness of parishes. The outermost circles (orange), illustrated in figure 4.5, represent parishes that share fewest surnames with other parishes, and the innermost circles (blue) are those which share the most surnames. An arrow indicates the link between a highly integrated parish and its nearest geographic neighbour. From these observations, it appears that there is a weak relationship between geographic distance and samesurname values, particularly with the most integrated parishes. This partially explains why we did not find a significant relationship between geographic and same-surname distance with a Mantel test. The exceptions are interesting: the majority of highly integrated parishes have a geographic neighbour in the set of least or moderately integrated parishes. Figure 4.5, showing the nearest neighbours for the most integrated parishes by date, suggests that this relationship may result from moves out of the most integrated parishes, that is the overflow from older parishes to nearby 'frontier' parishes. This is what is meant by 'le monde plein' and 'le monde vide' (Bouchard 1996).

One might expect a relationship between time of settlement and interconnectedness, with the earlier settled parishes being the most integrated. The data reveals some evidence of this, but the effect is weak. Figure 4.6, a plot of

parish opening date versus parish interconnectedness and size of parish, shows no relationship. The biplot roughly divides the province into three 'gouvernements' of the French régime, that is Quebec City, Trois-Rivières and Montréal. The majority of the Trois-Rivières parishes are located in the upper right quadrant of figure 4.7; the parishes of Montréal, which are among the most integrated and largest, are mostly found in the center, and the parishes of the Quebec region are located in the upper left (Figure 4.7). In some parishes great variability accounts for their placement outside their 'statistical' regions. In each of these cases distance plays a role in structuring the geographic organization of the parishes, but it is not as linear and systematic and thus the results do not appear when trying to examine distance versus same-surname relatedness.

To demonstrate the utility of between population same-surname calculations, I selected three cases for mapping the geographic distributions. In figure 4.8, displaying the relatedness of Chambly to the other parishes, we see connections among parishes in the Richelieu Valley with Quebec City and with the parishes just south of Montréal. Chambly is very integrated, and it shares at least one surname with all but 18 parishes. In the second display, Quebec City shows strongest surname relatedness with its immediate neighbours. Although it is the most integrated parish of all, sharing at least one surname with all other parishes, the relationships are relatively weak beyond its immediate region. The final case, Trois-Rivières, is a moderately integrated parish sharing a surname with 58 other parishes; it has its strongest relationships with parishes in the immediate vicinity, around the rim of Lac Saint Pierre. In all three cases a zone of intense interaction extends across the river: rivers function as connectors rather than barriers in the conditions of the eighteenth century.

As we saw in Vermont, same-surname methods are an important tool for understanding the geographic organization of names at both the local and regional scale. Local isonymy provides a means to appraise and calculate the number of names per thousands of people. Diversity of names has yet to become an important topic among historical geographers, but the number of names (families) and the size of the population have important consequences for 'marriage markets'

demographic characteristics, and of course genetic consequences. As we will see in a later case study, the diversity of names, especially where the marriage fields are of a short radius, affects the probability of marrying someone of the same name or of the same genealogical line. Between population relatedness based on samesurnames provides for various interpretations. In this example, it demonstrated the interconnectedness and relative relatedness of certain populations and provides insight into the settlement process. Geographic surname relationships between old and new parishes provide evidence of population movements. Multidimensional scaling provided further evidence of the same-surname relationships and a way to summarize the large number of relationships. MDS revealed that regional relationships were intact in the eighteenth-century Quebec population. The Mantel test confirmed a lack of distance-decay which I believe is an accurate interpretation, but the between population values suggest a moderate affinity between local neighbouring places. The Mantel test calculated a regional value and failed to capture local and moderate relationships. It cannot deal with directional components or topographic which moderate the effect of sheer distance. To examine to what extent there may have been local name affinities, we move now to Relevance Networks, a method which will extract the most important and strongest connections between places.

Eighteenth-Century Relevance Networks

Using the same 1765 census data, we turn now to relevance networks. For the parishes, I set the correlation threshold to .5, which created 23 networks including 69 of the 117 total parishes. As described earlier, this algorithm connects only those parishes that meet a threshold of similarity in their surname profiles, as defined by a correlation value. The correlation value was tested at various thresholds, and I chose a value which created the most networks. Figure 4.9 illustrates the relationship between the correlation values and the resultant number of networks and parishes. At the lowest threshold only a few networks appear with a large number of parishes. As the threshold correlation value is increased, we get more networks, but the number of parishes included in the networks declines. This implies that many parishes have some level (low) of name relatedness, but only a select set share a strong and meaningful relationship.

To create the networks of surnames, we make a separate analysis. Among the 832 single-founder names the largest possible number of relevance networks was 41 (when the correlation threshold was set to .6). Lowering the threshold below .5 groups three-quarters of the names into a single network. To simplify the explanation and focus on the most related sets of names, I set the correlation value high (at .8), generating 23 networks which included only 63 surnames.

The question of greatest interest to the geographer is: How well does the evidence of surname relatedness reflect geographical options? We shall see that relevance networks do reflect distance effects: members of the strongest networks are often found within short geographic ranges.

A Montréal network (15) of parishes is the largest, the most integrated, and has the longest geographic reach. It consists of 11 parishes split into two groups connected by Saint Nicolas and Sainte-Geneviève (Pierrefonds). The first group has a mix of long and short distances, notably Saint Nicolas (200 km apart) which has affinities with Montréal and Chambly. Montréal's strong connection with a distant parish is understandable, since its large population and high surname diversity made it more integrated with other parishes (Figure 4.10, 4.11, 4.12, 4.13). If we examine figure 4.12 we can see that the Montreal (15) network is mostly associated with parishes south and east of the Island of Montreal. We also find Montreal to be very integrated from the results of the between population same-surname values and multidimensional scaling. The second group connected to this network is most concentrated west of Montréal with the exception of Pointe du Lac. The anomalies raise intriguing questions for the local historian.

The next two largest networks include 7 parishes each. The most integrated of the two is the Trois-Rivières network (2), all of whose parishes are located between Montréal and Quebec City, all but Trois-Rivières on the south side (right bank) of the St. Lawrence. As mentioned earlier, cross-river name similarities suggest cross-river ties of kinship. The second of our seven-parish networks is the

Longue-Pointe network (13) concentrated in two areas within 60 km of one another: One set southeast of Montréal, with the strongest link between Repentigny and L'Assomption, which are geographic neighbours; the second northwest of Montréal (Figure 4.10, 4.11, 4.12, 4.13).

In the Ile-Jésus network (12) five out of six parishes form a tight cluster north of Montréal, with a short range. The longest geographic distances connect St-Vincent-de-Paul and Sault-au-Récollet and the strongest connection in this group (highly correlated) is between St-Vincent-de-Paul and St-Martin-de-Île-Jésus (Figure 4.10). Given the early settlement and easy communication on the lowland waterways, one can suppose that marriage distances extended beyond the adjoining parish. These parishes are known for the mobility of youthful voyageurs, and for interaction with the markets and logistics of defence at Montréal (1770s, 1812).

Two other networks that show strong geographic organization include the St-Joachim/Beauce network (14) and the Côte-du-Sud/ Baie Saint Laurent network (Figure 4.10). Even in these small networks we note the short geographic distances. An excellent example is the three-parish network of Ste-Foy, St-Augustin and L'Ancienne-Lorette (18), all close neighbours. An example of longer geographic linkages, suggesting some longer paths of riverine migration, is the network that includes Yamachiche, Grondines and Berthier-en-Bas (5) (Figure 4.11).

The networks of surnames are an inverse of the network of places, and we are seeing the names that were driving the place networks. The largest of the 23 is a chain of nine names concentrated along the St Lawrence River in the Lanaudière region (Figure 4.14, 4.15, 4.16). The next two largest networks were network (F) associated with Ile Jésus, and network (B) consisting of 5 surnames concentrated in Montréal. Mapping the 23 networks does not produce perfect associations, but we get rather close fits, with the networks overlapping. For example, if we look at the geographical dispersion of marriages associated with three of the networks, we see that the names associated with each network lie almost on top of each other in the same locations. I have summarized the locations of each of the networks in

Figure 4.16.

Although relevance networks neglect weak relationships, the technique is straightforward for identifying the most intensely related. It is more sensitive to strong relationships and I would argue that it more effectively pinpoints relative relationships, than does the multidimensional scaling of the between population same-surname results, because it is not a global interpolated summary. For Quebec, the relevance network method demonstrates that certain names are grouped in certain areas of the province, and this reinforces the values observed using same-surname methods.

Eighteenth-century Cluster Analysis

To explore the hierarchical nature of surname structure, we turn now to cluster analysis. The sample employed consists of all marriages of men from the 832 single-founder sample, recorded between 1771 and 1799. Two items stand out. First, Quebec City, Montréal, L'Assomption, St-Denis-sur-Richelieu, St-Martinde-Île-Jésus, and St-Eustache have the most diverse array of surnames. Second, Baie-Saint-Paul, Les Éboulements, L'Isle-aux-Coudres, La Pocatière and Charlesbourg have unusual concentrations of a single surname and therefore lower diversity. In Baie-Saint-Paul, for example, the Tremblay surname occurs 46 times and Simard 37 times in a population of only 128 (Table 4.5). This anomaly results in part from early settlement of the names in this region as well as specific events of seigneurs allocating land to several members of a family (Courville 1996).

It is well known in Quebec that some names are more clustered or populous in certain regions. We observed name clustering in our previous maps of the geographical organization of names, as well as the maps of relevance networks (see Figure 4.10). This regional clustering and overlapping of names is the result of the regional hierarchical organization of names caused by overlapping migration and marriage fields. To examine the effect of extent regional name clustering on the early development of hierarchical name structure, I use clustering methods to summarize the 832 single-founder sample of marriages.

Hierarchical agglomerative clustering was performed on surnames and places separately, using relative Euclidean distances and Ward's group linkage method. I examined 50 clusters, and the patterns which emerged at each succeeding step or cluster below 50. To determine the significance of the clusters, I calculated a participation rate to tally the percentage of all names in a cluster belonging to a particular set of place clusters. Good clusters produce participation rates that are either very high or very low. Based on the participation rates, I concluded that at a regional scale, 15 clusters provided the best fit for both names and places (Table 4.6).⁸

The great number of surnames and the wider range of values make them more difficult to cluster than places. Among the 15 surname clusters, the number of names ranged from a low of 13, with 125 individuals in the Favreau cluster, to 284 names and 3369 people in the Drouin cluster (Table 4.5). The distinctive geographic concentrations do not become immediately apparent until we cumulate all marriage events for each parish, and identify the parish with the greatest number of marriages. This step is necessary because we are interested in the aggregate of all the surnames in a cluster. In this way, we can map the locations where the majority of surnames in a cluster reside. The Pinsonneault (26) cluster, for example, consists of 15 different surnames exhibiting 171 marriage events in 25 different parishes. Among the 25, just three parishes account for three-quarters of the 171 marriages, all in a single region just east of Montréal Island: St. Constant, Saint-Philippe-de-Laprairie (St-Jean-François-Régis) and Laprairie. The Pinsonneault cluster, therefore, is mapped as belonging to these three locations (Figure 4.17). In the same manner the Choquet (6) cluster consists of 70 names with 741 marriages in 70 different parishes, but sixty-two percent of the marriages took place in close proximity to one another in the Richelieu Valley northeast of Montréal (Table 4.7). Like the two clusters just described, eight others show the same kind of isolation with the majority of marriage events occurring in neighbouring parishes: Archambault (2), Meloche (8), Favreau (15) Vezina (16), Massicotte (24), Boissonneau (32), Bessette (36), and Lheros (44) (Figure 4.17).

Because the clusters have such specific geographic concentrations, they can be

interpreted as contingent probabilities, suggesting that a Bessette marriage, for example, is more likely to be found with a Sainte Benne marriage in Iberville (Richelieu Valley) than with a name from the Vezina or Lheros clusters concentrated north of Quebec City (agglomeration of Quebec). The remaining clusters are found in very specific regions. The Archambault cluster on the islands of Montréal and Ile Jésus; the Meloche cluster at Pointe-aux-Trembles and in the Lanaudière; the Favreau cluster in the Richelieu valley; the Massicotte cluster in the Saint-Maurice Valley; and Boissonneau in the Côte-du-Sud.

Of the five remaining name clusters, four are more dispersed along the St. Lawrence [Ossan (73), Valiquette (3), Poitras (1), and Dionne (7)]. The Dionne and Poitras clusters are in Quebec City as well as regions of Côte-du-sud and Bas Saint-Laurent. Ossan is found in Montréal and the Bois-Francs regions, and the Valiquette cluster is located in a region just northwest of Montréal and the Bois-Francs. From cues such as these, the local historian may be able to identify the likely stream of settlers and look for evidence of the founder couples.

The last, the Drouin cluster, does not possess a discernible geographical organization. It consists of a very large number of marriages (3369) and 284 different names. It is the most variable and most geographically integrated, sharing at least one name with every parish. The majority of the parishes represented in this cluster are located in the Trois Riviéres region and north, but it cannot be identified with a single region like the others. This behaviour leads me to believe that the names included are diversified and spread more evenly throughout the province. The fact that many surnames cluster in certain regions does not mean that every name behaves in this way. Some names, like Laberge, concentrate in more than one area, while others show little geographic structure.

The Drouin cluster, with its great geographic dispersion, raises two important ideas. Our clusters of names tend to represent: names that cluster and names that have no intense geographic concentration. Can we conclude that less clustered names have more varied names on their family tree? I would argue yes, and that there appear to be three groupings of type of names: 1) names that are highly clustered (i.e Beauchamp); 2) names that are clustered, but split between two

regions; 3) names that cannot be identified as belonging to any specified region based on their occurrences.

To achieve a better perspective on the geographical organization and the resulting population hierarchy created by the layering and clustering of names, the parishes (130) were mapped into 50 clusters each at a higher level grouping (descending from 50 clusters) (Figure 4.18). Figure 4.18, showing the structure at six hierarchical levels, illustrates the regionalization of 30 clusters. As we reduce the number of clusters, the lower-order clusters (lower number) collapse into higher order clusters. Cluster 101, for example, (maps with 20 and 15 clusters) found in the Bas Saint-Laurent region, is amalgamated into cluster 63 already present in the Côte-du-Sud region. In each case the collapsing clusters demonstrate the overlapping and hierarchical organization of the surnames in Quebec. In other words, clustering reveals regions, sub regions and 'mini' regions (Figure 4.18, 4.19).

A few clusters are notable because of their stability. Cluster 15 in the Charlevoix region remains intact through all iterations. This stability is certainly a reflection of the isolation and uniqueness of this population. Cluster 14 in the Côte-du-Sud also remains stable, but is eventually merged into Cluster 11 of the Quebec City region. This relationship is most likely the result of short-distance migrations between Quebec City and Côte-du-Sud, just across the river. Another significant group includes cluster 10 found in and around the Montréal region; it contains many of the names that would settle west of Montréal and in the Ottawa valley (e.g. Beauchamp and Labelle). We also find evidence in the cluster maps of sets of names moving together up the river valleys. In 1881, this type of chainmigration up the river valleys will become more apparent.

When examining the tree of 50 cluster amalgamations, I paused at 15 clusters which appear to be the most significant. If we reduce further than 15 clusters, we find locations which share few names agglomerating together. Table 4.4 shows the breakdown and the parishes present in each of the 15 clusters. We find that the numerical and geographic breakdown of the 15 parish clusters match very well with the 15 clusters of names.

Using cluster analysis and mapping the clusters reveal the hierarchical and overlapping population structure that have been described by geneticists (Hartl 2000), whereby large populations split into smaller subpopulations through a process of stepwise (overlapping) chain migration. It is these short distance movements of individuals that served to hold the eighteenth century local populations together and to limit the regional level of genetic divergence. What becomes obvious using cluster analysis- and not with the other methods- is an overlap in the relatedness of most places with neighbouring places. The Beauchamp surname, for example, highly concentrated in the Montréal region, is nevertheless present in much lower numbers as one moves outward from the area of highest concentration. The clustering of a name in one location and its random occurrence in others is a characteristic of many surnames in the eighteenth century. In aggregate and with a large enough sample, this structure creates the hierarchical structure of names and the multiple levels of name relatedness throughout the province (overlapping webs).

Conclusion: Eighteenth-Century Surnames

This chapter on eighteenth-century surname structure revealed that some of the movements of settlers and the early development of the name regionalization were driven by the seigniorial system of granting land and by the limited number of opportunities which may have compelled larger numbers of individuals to move greater distances from the areas where their families were initially concentrated. What is evident when viewing the geographic name structure at this time is the filling of the seigniorial lands; first, along the Saint Lawrence, followed by smaller tributaries 'behind' the earliest seigneuries. It took several decades (two or three generations) to begin to see the concentration of names in certain areas. Instances are known where people of a particular surname (like Tremblay) obtained rights to large pieces of land. Many names diffused by leaps to several locations, which is why a name like Laberge eventually became clustered in more than one location at substantial distances apart. We also see at this time a strong affinity of name

relatedness (kinship) in local areas and across the St Lawrence River which illustrates the importance of the river as a communication and travel network.

By 1800, the earliest names to arrive had as many as 5 generations of growth, and we begin to see clustering in certain regions. In the next chapter, which describes the population 80 years later, we will see that initial settlement and the name structure by 1800 has a significant impact on future organization, with indications of settlement radiating from the early-established parishes in a process of short step-wise chain migrations, mostly up river valleys toward the new frontiers.

The importance of control over a founding population has been pointed out repeatedly. The selection of a single founder sample gives us the assurance that we are observing diffusion, that the relatedness of names reflects the links of descent. We have also seen that it affects the absolute values of global measures like Fst and Rst, and this introduces a note of caution into comparison with populations in other parts of the world.

The close attention to founders points to a key factor in any interpretation: all these phenomena - the numbers, the spatial concentration, and the dispersion - are time-dependent. As Desjardins et al. (1997) put it, 'tout s'est joué très tôt.' This is a property of genetic studies, not sufficiently acknowledged in geographic work: initial structure of a first generation channels development in successive generations.

NOTE

³ Land inheritance for the seigneur followed French laws and was a modified form of primogeniture, which included an appearement with the preconception that an individual had a 'natural right' to land held by the family. Most often the seigneur's widow would receive half of the seigneurial land, the eldest son $\frac{1}{4}$ and the remaining $\frac{1}{4}$ would be divided by all other children (Harris 1968)

⁴ Multiple seigneurs owned pieces of the original seigneuries and some seigneurs controlled more than one seigneurie

⁵ See Christensen, 2000, 1999; Relethford, 1988, Lasker, 1985; and Roguljic, 1997

⁶ It is important to acknowledge the lower values calculated when using the single founder sample because future results will also be affected by the sample.

⁷ No relationship between date and population size; Spearmans rank correlation indicates a relationship between size of population and number of surnames. (Spearmans Rank).

⁸ Important considerations when clustering include: progressive aggregation, stability of results, alternate starting points, shuffle of starting points and use of resampling and bootstrapping techniques. New methods have recently been developed to test the significance of each cluster under randomization conditions, with a focus on how clusters vary when data is resampled or removed.

¹ Death rate of New France was high by today's standard, but low at the time when compared to France and other old world and southern North American colonies.

 $^{^{2}}$ A roture was a concession of land within a seigneurie which would be taxed (cen), and could not be sub-conceded. It was considered a 'family-size' farm, often 60 to 120 arpents.



Percent of Total Marriages for selected names (approximate)

- 1.265823%
- 1.265824% 3.797468%
- **3.797469% 5.063291%**
- 5.063292% 11.392405%
 - 11.392406% 36.708861%



Figure 4.1. Geographic distribution of marriages for five surnames between 1621-1799 Source: PRDH

Table 4.1. 1765 Quebec measures of micro-differentiation determined from surnames. Two data sets are presented: the first results are from marriage records (1710-1799) divided into 30 year periods, and the second results are from the 1765 surrogate census created from baptisms.

	Date	Pop Total	Parishes	Surnames	Fst	Rst	Φ_R
Marriages	1710-39	2158	60	658	0.009833	0.009147	0.000693
Marriages	1740-69	5365	130	873	0.011397	0.010730	0.000673
Marriages	1770-99	9217	150	857	0.012702	0.011843	0.000870
Census Single Founders	1765	6980	125	1346	0.012974	0.012395	0.000585
Census Total Population	1765	25034	130	5239	0.003496	0.003260	0.000236

 ΦR is equal to regional isonymy divided by 4.

Fst is genetic microdifferentiation and refers to the average within-group (populations) unreduced variance. Rst is genetic microdifferentiation and refers to the average within-group reduced variance.



Figure 4.2 Parish opening dates prior to 1780 versus parish surname diversity. The greatest diversity values are below .05. Red line represents average isonymy (diversity) values and standard deviation.



Figure 4.3. Surname diversity for the 1765 Quebec surrogate census. The parishes with the lowest values (largest circles) have the greatest surname diversity



Figure 4.4 Between population surname similarity values (isonymy) between Rivière-du-Loup (Louiseville) and all other parishes. The star represents Rivière-du-Loup and largest circles (red) are the most closely related, followed by the less related parishes represented by decreasing circle size.



Figure 4.5 Multidimensional scaling of between parish surname similarity (isonymy) values for the surrogate 1765 census. Orange (<40 connections) circles represent least connected parishes and blue (>70 connections) circles represent the most connected parishes; and gray circles are moderately integrated (>40 & <70).



Figure 4.6. Parish opening date versus the number of parishes where at least one surname is shared (interconnectedness). In this example, we can see that the parishes Quebec City and Montréal have the greatest interconnectedness. The Y axis shows the total number of parishes which have some level of relatedness. For example, Montreal shares names with approximately 85 parishes. Red line represents average number of interconnected parishes.



Figure 4.7. Multidimensional scaling results of between parish surname similarity values (isonymy) for the 1765 surrogate census (sample, single founders). The parish opening dates are presented here as a color gradient: lightest colors represent the earliest opening date. The arrows provide the nearest neighbors for the selected parishes. See map key for parish names.



Figure 4.8 Between population surname similarity values (isonymy) between selected parish (star) and all other parishes. The most closely related parishes have the largest circles



Figure 4.9 Relationship between the correlation values and the number of networks and parishes



Figure 4.10 Relevance Networks illustrating the (correlation) connections between 1765 parishes. For this example the correlation threshold was set to to .5, which created 23 networks including 69 of the 117 parishes



Figure 4.11 Relevance Networks illustrating the (correlation) connections between 1765 parishes. For this example the correlation threshold was set to to .5, which created 23 networks including 69 of the 117 parishes



Figure 4.12 Relevance Networks illustrating the (correlation) connections between 1765 parishes. For this example the correlation threshold was set to to .5, which created 23 networks including 69 of the 117 parishes (see figure X for network descriptions)



Figure 4.13 Relevance Networks illustrating the (correlation) connections between the 832 single-founder surnames. For this example the correlation threshold was set at .8 creating 23 networks containing 63 names



Figure 4.14 Relevance Networks illustrating the (correlation) connections between the 832 single-founder surnames. For this example the correlation threshold was set at .8 creating 23 networks containing 63 names



Figure 4.15 Geographic distribution of the Relevance Networks of the 832 single-founder surnames(see figure X). Letters represent the approximate regions where the majority of marriages for the selected network occured



Figure 4.16. Map illustrating the name clusters. The location represent the top locations where the majority of marriages within that cluster occurred. Clusters were created using groom from the set of 832 single-founder names or the years 1771-1799.



Figure 4.17. Sequential maps illustrating the clusters of places at different hierarchies (30,20 & 15 clusters). Clusters were created using groom from the set of 832 single-founder names for the years 1771-1799.


Figure 4.18. Sequential maps illustrating the clusters of places at different hierarchies (10,5 & 3 clusters). Clusters were created using groom from the set of 832 single-founder names for the years 1771-1799.

Table 4.2. Quebec surrogate 1765 census sample

Parish ID	Parish	Opening Date	Population Size	Number of Surnames	Local Isonymy	Parish	Parish ID	Opening Date	Population Size	Number of Surnames	Local Isonymy
3901	Montréal	1642	414	235	0.00496	6001	Trois-Rivières	1634	54	38	0.01468
4501	Québec	1621	394	235	0.00430	1301	Châteaunuav	1738	54	23	0.11111
2804	L'Assomption	1724	310	107	0.01457	4602	Paruport	1673	53	25	0.04717
1003	Chambly	1706	189	87	0.01497	2004	Ot Martin de Île Herve	1774	52	30	0.01367
6203	Verchères	1724	180	61	0.02638	2904	St-Martin-de-IIE-Jesus	4750	52	25	0.04374
2902	St-Vincent-de-Paul	1743	167	61	0.03001	0001	Les Cedres	1752	52	25	0.04374
703	Berthier-en-Haut	1751	148	703	0.01361	4604	Registre des Hurons de Loretteville	1775	52	21	0.05053
2305	Ste-Geneviève (Pierrefonds)	1741	148	53	0.03181	3801	Ste-Anne-de-Beaupré	1657	52	13	0.16440
5901	Terrebonne	1727	137	62	0.02018	3001	Pointe-de-Lévy (Lauzon)	1679	51	24	0.07843
6202	Varennes	1693	136	56	0.03856	3203	St-Antoine-de-Tilly	1702	51	14	0.18980
1001	Boucherville	1668	122	49	0.03455	3101	L'Islet	1679	48	21	0.07270
2502	La Pocatière	1715	118	38	0.06475	3705	St-Pierre-de-la-Rivière-du-Sud	1727	48	22	0.08599
2701	Laprairie	1670	115	58	0.02304	6205	Beloeil	1772	46	26	0.02005
3301	Rivière-du-Loup (Louiseville)	1714	110	50	0.02118	1202	i es Éboulements	1733	45	5	0.55657
1905	Sault-au-Récoilet	1736	102	45	0.03164	402	Dia Maria da Daguna	1745	40	36	0.00000
602	La Durantaye	1693	96	49	0.02215	402	Ste-Marte-de-Beauce	1740	45	20	0.02990
5101	St-Charles-sur-Richelieu	1740	93	45	0.03343	3803	Ste-Familie I.O.	1000	43	21	0.06534
1002	Lonqueuii	1701	92	30	0.11945	3802	Château-Richer	1661	43	15	0.11849
5102	St-Denis-sur-Richelieu	1740	87	49	0.02138	4101	Bécancour	1716	42	23	0.03252
6303	Yamaska	1727	85	32	0.04202	3702	Cap-St-Ignace	1679	41	19	0.04878
4605	Charlesbourg	1679	85	36	0.10812	2301	Lachine	1676	40	25	0.02564
2303	Pointe-Claire	1713	84	39	0.04274	2803	St-Sulpice	1706	39	25	0.02834
603	St-Vallier	1713	83	33	0.06553	1102	Champlain	1679	38	26	0.01991
2304	St.J aurent	1720	82	46	0.02770	1204	i leie nu Coudres	1741	38	11	0 19203
604	St-Charles (Bellenhasse)	1749	80	40	0.02627	3704	St Emposie de la Pivière du Sud	1733	37	20	0.05556
3201	otbinière	1697	80	18	0 16139	4400	St-François-de-la-roviele-du-Sdd	1710	27	20	0.00000
6204	St-Antoine-sur-Richelieu	1741	79	36	0.04609	4102	Nicolar	4745	37	10	0.00757
4702	St-Ours	1750	79	30	0.06751	6302	La-Baie-du-Febvie	1/15	37	14	0.08258
2805	Mascouche	1750	77	36	0.02597	4405	Deschambault	1705	37	12	0.10360
2704	St. Jean-Francoic-Bégis (St. Dhilippe)	1753	72	41	0.02152	3103	Saint-Jean-Port-Joli	1767	37	13	0.18769
2503	Kamouraeka	1700	71	25	0.07203	1501	Oka	1727	36	18	0.13016
5302	Vamachicha	1727	70	30	0.05714	3002	St-Nicolas	1694	34	10	0.28877
702	i evoltrie	1732	69	38	0.02984	2703	St-Constant	1752	33	21	0.03598
4404	St-Augustin	1693	68	28	0.03951	1104	La Pérade	1693	33	20	0.06439
2501	Biolare Quelle	1695	64	26	0.06944	3703	Berthier-en-Bas	1710	33	17	0.08902
2801	Repetion	1679	63	37	0.03328	2901	St-Francois-de-Sales (Ile-Jésus)	1702	32	23	0.02218
1502	St-Eustache	1769	62	36	0.02380	1903	I oncue Printe	1724	32	15	0.08468
4603	L'Ancienne-i orette	1676	62	26	0.07245	3806	St-Bierre I O	1679	32	16	0.08669
2802	L'achensie	1697	61	35	0.05355	3904	Liange Cardian	1660	30	16	0.06667
1105	Ste-Geneviève-de-Batiscap	1728	60	27	0.04011	3004	C Alge-Galgeri	1766	20	15	0.00001
4701	Sorel	1675	60	27	0.04746	3003	SUTIENN	1700	29	10	0.03430
3701	Montmany	1679	60	35	0.04859	705	Saint-Cuthbert	1//0	28	19	0.03439
1201	Roje St Boul	1681	60	16	0.20226	6301	St-François-du-Lac	1687	28	15	0.06349
1001	Bainta aux Trambles	1674	58	29	0.03993	4402	Neuville	1679	27	15	0.04558
401	Stalosenh-de-Resuce	1738	56	25	0.04935	4401	Cap-Santé	1679	27	14	0.07407
-101	or or other in the second	1100		20	0.04000	3202	Ste-Croix (Lotbinière)	1727	27	9	0.13390
						3808	St-Jean I.O.	1680	26	13	0.09846
						3809	St-Joachim	1725	26	8	0.47385
						3807	St-Francois I.O.	1679	23	14	0.05138
						4606	Ste Fou	1699	22	14	0.04762
					1	-000				••	

Table 4.3. Between population surname similarity values for the 30 largest parishes in Quebec as of 1765 and the four most related parishes (based on between population isonymy values)

Parish ID	Parish Name	First Parish	lsonymy	Second Parish	lsonymy	Third Parish	isonymy	Fourth Parish	lsonymy
3901	Montréal	3002	0.01698	3804	0.00322	2305	0.00281	1903	0.00279
4501	Québec	3804	0.00821	4603	0.00720	3802	0.00655	4606	0.00542
2804	L'Assomption	5102	0.03579	2801	0.01639	1901	0.01046	1903	0.00938
1003	Chambly	3002	0.02272	1002	0.01024	6205	0,00713	1903	0.00413
6203	Verchères	5102	0.02219	1002	0.01492	5101	0.01314	2805	0.00771
2902	St-Vincent-de-Paul	2904	0.01889	3802	0.01824	1905	0.01673	1502	0.01304
703	Lachine	703	0.02027	705	0.01713	702	0.00969	3203	0.00927
2305	Ste-Geneviève (Pierrefonds)	3301	0.20270	3002	0.03219	5601	0.01515	2303	0.01504
5901	Terrebonne	3301	0.09489	2805	0.01512	2902	0.01128	2904	0.00997
6202	Varennes	5102	0.02681	1001	0.00790	6303	0.00761	1002	0,00687
1001	Boucherville	1002	0.01345	3002	0.00868	6202	0.00790	5102	0.00377
2502	La Pocatière	2503	0.04536	2501	0.02979	4701	0.00847	3705	0,00565
2701	St-André (Kamouraska)	1202	0.03188	2704	0.01534	1204	0.01487	1201	0.01232
3301	Rivière-du-Loup (Louiseville)	4102	0.89189	3202	0.88889	3201	0.85000	705	0.71429
1905	Sault-au-Récollet	3201	0.01801	2902	0.01673	3801	0.01414	3202	0.01162
602	La Durantaye	3809	0.04447	603	0.02523	402	0.00993	604	0.00951
5101	St-Charles-sur-Richelieu	5102	0.02216	6203	0.01314	4605	0.01025	6204	0.00667
1002	Longueuil	2703	0.02108	6203	0.01492	1903	0.01359	1001	0.01345
5102	St-Denis-sur-Richelieu	3801	0.03667	2804	0.03579	6202	0.02681	703	0.02227
4605	Charlesbourg	4604	0.04457	4606	0.01497	3803	0.01067	5101	0.01025
6303	La-Baie-du-Febvre	5102	0.00837	6202	0.00761	4702	0.00655	2301	0.00529
2303	Pointe-Claire	1501	0.02414	5601	0.01740	2305	0.01504	2803	0.01221
603	St-Vallier	3301	0.24096	3809	0.13346	602	0.02523	3703	0.02300
2304	St-Laurent	3201	0.01204	1905	0.01136	3202	0.00858	2904	0.00727
604	St-Charles (Bellechasse)	3301	0.17500	3703	0.01515	602	0.00951	3705	0.00938
3201	Lotbinière	3301	0.85000	3202	0.11250	3203	0.10809	6302	0.09088
4702	St-Ours	6204	0.00977	2304	0.00710	6303	0.00655	1901	0.00633
6204	St-Antoine-sur-Richelieu	5102	0.02131	2801	0.01567	1901	0.01550	4702	0.00977
2805	Mascouche	2802	0.05082	4404	0.01998	5901	0.01512	6203	0.00771
2704	St-Jean-François-Régis (St-Philippe)	1202	0.03056	2703	0.01810	2701	0.01534	1204	0.01425

Table 4.4. Between population surname similarity values for the first 30 parishes opened in Quebec and the four most related parishes (based on between population isonymy values)

Parish ID	Parish Name	First Parish	lsonymy	Second Parish	lsonymy	Third Parish	Isonymy	Fourth Parish	lsonymy
4501	Québec	3804	0.00821	4603	0.00720	3802	0.00655	4606	0.00542
6001	Trois-Rivières	3301	0.29630	3202	0.02606	3203	0.02288	6302	0.01752
3901	Montréal	3002	0.01698	3804	0.00322	2305	0.00281	1903	0.00279
3801	Ste-Anne-de-Beaupré	1201	0.06154	5102	0.03667	3803	0.03444	1202	0.02692
3802	Château-Richer	3101	0.05184	3705	0.03828	3804	0.03566	401	0.02824
3803	Ste-Famille I.O.	3801	0.03444	3807	0.02427	402	0.01568	3808	0.01521
1001	Boucherville	1002	0.01345	3002	0.00868	6202	0.00790	5102	0.00377
3804	L'Ange-Gardien	3301	0.30000	3802	0.03566	1903	0,03457	1202	0.02444
2701	Laprairie	1202	0.03188	2704	0.01534	1204	0.01487	1201	0.01232
4602	Beauport	3301	0.09434	3804	0.01572	4604	0.00689	4605	0.00644
1901	Pointe-aux-Trembles	1903	0.03125	2801	0.01752	6204	0.01550	6205	0.01499
4701	Sorel	1202	0.01222	2503	0.01080	2502	0.00847	1204	0.00570
4603	L'Ancienne-Lorette	3301	0.48387	4606	0.04106	4404	0.02657	3804	0.01774
2301	Lachine	3301	0.12500	2303	0.00923	1105	0.00667	2305	0.00659
4401	Cap-Santé	3301	0.14815	4405	0.04905	5302	0.01058	6001	0.00480
3101	L'Islet	3301	0.12500	3802	0.05184	3705	0.03212	401	0.01897
3701	Montmagny	3705	0.01528	3703	0.01364	3702	0.00854	3704	0.00676
4402	Neuville	1903	0.02431	3202	0.01920	3804	0.01481	6301	0.01058
2801	Repentigny	1901	0.01752	2804	0.01639	6204	0.01567	6205	0.00552
3001	Pointe-de-Lévy (Lauzon)	3003	0.03786	1202	0.03498	3002	0.02745	1204	0.02241
4605	Charlesbourg	4604	0.04457	4606	0.01497	3803	0.01067	5101	0.01025
1102	Champlain	3201	0.01217	1105	0.00965	4101	0.00877	4101	0.00877
3807	St-François I.O.	3301	0.21739	3808	0.02508	3803	0.02427	3801	0.01421
3702	Cap-St-Ignace	3301	0.12195	3804	0.01951	3802	0.01815	3101	0.01778
3806	St-Pierre I.O.	3301	0.15625	2503	0.03169	3003	0.01940	1502	0.00857
3808	St-Jean I.O.	3301	0.15385	1202	0.02821	3807	0.02508	3003	0.02387
1201	Baie-St-Paul	1202	0.26296	1204	0.09693	3801	0.06154	3809	0.04615
6301	St-François-du-Lac	3301	0.46429	3202	0.08201	3203	0.07283	3201	0.05580
1104	La Pérade	2802	0.02633	3002	0.01872	2901	0.01136	1105	0.01061
4404	St-Augustin	4603	0.02657	3804	0.02402	2805	0.01998	4606	0.01872

Places	Totals	Total Places	Surnames	Totals	Total Surnames	Cluster Name
1a	945	8	А	629	88	Poitras
2	1273	20	В	694	45	Archambault
3	488	11	С	574	51	Valiquette
4	241	6	D	741	70	Choquet
5	311	7	E	442	46	Meloche
6	824	8	F	3369	284	Drouin
7	397	5	G	408	24	Dionne
8	733	9	Н	120	13	Favreau
9	288	5	I	455	60	Vezina
10	765	11	- J	125	14	Massicotte
11	917	8	К	171	15	Pinsonneault
12	341	7	L	713	38	Boissonneau
13	880	13	М	255	38	Bessette
14	438	5	N	268	19	Lheros
15	305	6	0	182	23	Ossan
	9146	129		9146	828	

Table 4.5 Summary of total names and places in each of their corresponding clusters. Clusters were created using groom from the set of 832 single-founder names for the years 1771-1799.

Table 4.6. Summary of the parishes in the 15 clusters

Clusters were created using groom from the set of 832 single-founder names for the years 1771-1799. See figure X for the corresponding maps showing the geographic distribution of the 15 clusters

	1A	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cluster	1003	701	1103	3101	1101	1901	5101	1501	702	1902	1301	401	601	2501	1201
	2701	705	1104	3103	3201	1903	5103	1502	703	1905	3804	402	602	2502	1202
	2703	1001	1105	3702	3202	2801	6203	2301	704	2802	4404	403	603	2503	1203
	2704	1002	3301	3802	3203	2804	6205	2303	2302	2805	4501	3801	604	2504	1204
	3002	1102	4401	4801	3204	4702	6206	2304	2803	2901	4602	3803	605	3102	1205
	3901	2401	4402	5802	4104	5102		2305		2902	4603	3807	3001		7805
	4103	2806	4403		6302	6201		5601		2903	4605	3809	3003		
	4901	3302	4405			6204		6101		2904	4606		3701		
		3601	4406					6102		5901			3703		
		3808	7804							5902			3704		
		4101	7810							5903			3705		
		4102											3805		
		4701											3806		
		5202													
		5301													
		5302													
		6001													
		6202													
		6301													
		6303													

Table 4.7. Summary of the participation rate of the percentage of names in a cluster belonging to a set of place clusters. Clusters were created using groom from the set of 832 single-founder names for the years 1771-1799. See figure X for the corresponding maps showing the geographic distribution of the 15 name clusters

	А	В	С	D	Е	F	G	н	1	J	ĸ	L	M	N	0
Clust	r Poitras	Archambault	Valiquette	Choquet	Meloche	Drouin	Dionne	Favreau	Vezina	Massicotte	Pinsonneault	Boissonneau	Bessette	Lheros	Ossan
1a	1	62.5	50	62.5	25	87.5	37.5	50	25	0	50	62.5	62.5	0	25
2	33.3	42.8	14.2	38	4.7	100	9.5	14.2	42.8	4.7	4.7	28.5	28.5	0	19
3	18	0.36	9	18	9	100	18	0	27.27	45	0	18	9	54.5	9
4	50	0	33	0	0	83	50	0	33	16.6	0	66,6	0	16.6	0
5	28	14	14	28	14	100	42.8	0	14	14	14	0	14	42.8	0
6	87.5	100	62.5	87.5	12.5	100	0	0	37.5	0	0	50	12	12	0
7	80	80	40	100	20	100	20	20	60	0	0	60	60	20	20
8	55.5	44.4	55.5	11.1	100	100	44	11.1	22.2	0	11.1	22.2	11.1	11.1	11
9	60	80	80	80	20	100	60	0	40	40	0	40	0	0	20
10	54	100	100	54	36	100	45	0	27	0.	0	9	9	18	0
11	37.5	50	12.5	12.5	12.5	100	25	0	100	12.5	12.5	25	12.5	0.75	12.5
12	0	16	16.6	0	0	100	16.6	0	50	0	33.3	100	0	16.6	0
13	15.3	38.4	7.6	7.6	0	100	38.4	0	53.8	0	7.6	100	15.3	23	7.6
14	20	0	0	20	0	100	100	0	0	20	0	40	0	0	0
15	20	33	33	20	20	100	0.2	0	40	0	0		80	80	40

Table 5.1	Measures	of micro	differentiation	determined f	from surnames f	or
1765 and	1881 Queb	bec.				
			.			

Sample	Φ_R	Fst	Rst
1765 Single Founders	0.000585	0.012974	0.012395
1765 Full Sample	0.000236	0.003496	0.003260
1881 Single Founders	0.001065	0.006874	0.005815
1881 Full Sample	0.000219	0.001089	0.000869

 ΦR is equal to regional isonymy divided by 4. Fst is genetic microdifferentiation and refers to the average within-group (populations) unreduced variance. Rst is genetic microdifferentiation and refers to the average within-group reduced variance.





Figure 5.2. Multidimensional-scaling results of same-surname analysis (isonymy) for the 1881 census single-founder sample. See figure 2.4 for location names



Figure 5.3. Same-surname (isonymy) values for La-Grande-Riviere (72) and all other locations (1881). Locations with the lightest shading have the least surname similarity



Figure 5.4. Same-surname (isonymy) values for Baie-St-Paul (51) and all other locations (1881). Locations with the lightest shading have the least surname similarity



Figure 5.5. Same-surname (isonymy) values for Yamachiche (37) and all other locations (1881). Locations with the lightest shading have the least surname similarity



Figure 5.6. Same-surname (isonymy) values for Buckland (42) and all other locations (1881). Locations with the lightest shading have the least surname similarity



Figure 5.7. Same-surname (isonymy) values for Ste-Adele (16) and all other locations (1881). Locations with the lightest shading have the least surname similarity



Figure 5.8. Same-surname (isonymy) values for Stanstead (21) and all other locations (1881). Locations with the lightest shading have the least surname similarity



Figure 5.9 Relevance network. Total networks versus total number of hexagons included at different correlation values. Final network was set at .6. See figure 5.11 for the mapped network.



Figure 5.10 Relevance network. Total networks versus total number of parishes included at different correlation values. Final network was set at .8. See figure 5.14 for the mapped network.



Figure 5.11 Relevance networks determined from the 1881 census single-founder surname sample



gure 5.12. Relevance networks determined from the 1881 census single-founder surname sample transposed into the results of the between population same-surname (isonymy) analysis. (See figure X, above).



Figure 5.13. Close-up view of hexagon 33 (Halifax) and surrounding between-population surname values (isonymy)and relevance networks

Table 6.5. Spearman's rank correlation of the probability values for possible migrants from Quebec. 673 parishes * each of the three data sets (Maine, New Hampshire and Vermont)

	MAINE	NH
ME	1	
NH	0.475	1
VT	0.303	0.412



Figure 5.15. Seventeen clusters extracted from the 1881 census single-founder surname sample. Each number represents a cluster where we find similar name relatedness.



Figure 5.16. Seventeen name clusters mapped onto the results of the same-surname (isonymy) multidimensional-scaling results. The number on each hexagon corresponds to the clusters from figure 15.15. There is concordance between the cluster analysis results and same-surname multidimensional-scaling.



Figure 5.17. Thirty clusters defined from the single-founders surname sample extracted from the Quebec census, 1881 The larger numbers represent the regionalization of certain clusters.



Figure 6.1 Growth rate of marriage sample of 32 names versus total population

Table 6.1. Clusters for the sample of 32 surnames

Cluster Name Beauchamp	Total Marriages 5676	Cluster Name Bessette	Total Marriages 2045	Cluster Name Laberge	Total Marriages 6859	Cluster Name Gravel	Total Marriages 1899	Cluster Name Vachon
Names Included		Names Included		Names Included		Names Included		Names Included
BEAUCHAMP		BÉRARD		BUSSIERE		BUTEAU		LEGENDRE
LABELLE		BRODEUR		COTÉ		CARBONNEAU		SÉVIGNY
CONSTANTINEAU		CHICOINE		RÉAUME		GÉNÉREUX		BÉRUBÉ
BRISEBOIS		CHAPDELAINE		CADORET		CAMERON		CAOUETTE
CHARLEBOIS		BESSETTE		CHARLAND		GRAVEL		BILODEAU
				CUSSON				BOLDUC
				VAILLANCOURT				VACHON
				GAULIN				GONTHIER
				LABERGE				







Figure 6.2c. Geographic distribution of marriages from the 32 single-founder sample.

Chapter 5

The 1881 surname population structure of Quebec

The regional distributions of surnames by 1800 had a strong impact on subsequent geographical patterns. To examine the changes and the surname structure in place in 1881, I build upon the analysis of a century earlier. The methods from the previous chapter is repeated and when possible comparisons are made. Comparison, however, is difficult because settlement geography changed drastically between the two periods: Montreal had grown to a few hundred thousand; and settlement extended way beyond the bounds of the river valley. The analysis will focus, by inference, on migratory patterns which lodged certain surname groups (subpopulations) into newly opened parishes and regions. To put the geographic expansion and changes in name structure into context, the chapter begins with a brief review of Quebec's nineteenth-century historical geography and demography. I follow with an interpretation of the geographical name structure from the 1881 Quebec census. Using the same surname methods (isonymy), relevance networks and cluster analysis, I draw comparisons with the eighteenth-century results and describe the changes associated with new frontier settlement, urbanization and new opportunities presented to the population.

Nineteenth Century Historical Geography

During the first half of the nineteenth century Quebec began a transition from an agrarian to an industrial society. Factory production began replacing family and local artisans, mechanization of agriculture reduced demand for labour and manufacture's concentrated skilled and unskilled labourers in urban centers, often drawing them from rural regions. With railway, after 1850, many small Quebec towns became important industrial centers, and the labour opportunities in these locations provided vital job opportunities for struggling farmers and newly arriving immigrants (Courville 1996; Young and Dickinson 2003). The industrializing towns were fed with people, raw materials, timber and food from mostly the rural population that continued its extraordinary population growth despite the growth poles of Montréal and Quebec City. Between 1760 and 1825 the French Canadian population doubled approximately every 25 years, and from the late seventeenth century until 1875; it experienced high rates of marriage and fertility. During the last few decades of the nineteenth-century Quebec grew more slowly as birth and death rates declined. With a high rate of natural increase the rural population constantly jockeyed for land and job opportunities. The range of alternative responses by individuals, families and regions can be seen through their migration paths (Haines and Steckel 2000).¹

By 1880 new territories were already being settled outside the seigniorial lands in the Saguenay, Saint-Maurice watershed and the Ottawa valley. Exploiting the best lands first, settlers tended to move up the river valleys and combined the working for timber company's with the clearing of small farms to raise families (Hardy and Seguin, 1980, 1984). The economies in these areas were based mostly on agro-forestry and were affected by severe market fluctuations. While the new areas provided economic opportunity, urban areas still attracted a majority of the rural French Canadians. Between 1815 and 1881 the urban population grew from 30,000 to 378,512. A prominent and unusual feature was the massive growth of Montréal (from the 1850s) and the stagnation of Quebec City. In a later case study we will examine the extent to which rural French Canadians moved to Quebec's two largest urban centers (Young and Dickinson 2003; Linteau et al. 1983)

By providing markets for produce, railways facilitated rural settlement but also aided migrations to Montréal and New England. Many French Canadians moved to unsettled parts of the Eastern townships. French Canadian settlement in the townships was initially blocked by large land companies which promoted Protestant settlers or held land for speculation, but demand for unskilled labour in these areas discarded the policy and a variety of colonization schemes were implemented by the government, church missions and local societies (Young and Dickinson 2003).

Another significant movement was emigration to the United States, primarily

New England. Quebec has received the most attention in regards to the French Canadian exodus to New England because of the sheer numbers who left, and because at the time it was viewed as a threat to French Canadian cultural and political identity (Thornton 1998). Vicero (1968) who produced one of the first origin and destination studies of French Canadians in New England, estimated that between 1840 and 1860 about 22,000 left for New England, most moving just south of the border to Vermont. Lavoie's (1973) estimates for the same period were around 105 000 individuals. The pace of migration grew in the 1870s and 1880s as industrial opportunities began to push and pull people to new locations. Vigorous parish activities in the 1870s in mill towns like Lowell, Massachusetts, and Manchester, New Hampshire, provided institutional structures for 'Little Canadas' (Roby 1996; Thornton 1998). It has been estimated that between 1840 and 1930 approximately 900,000 French Canadians moved to the United States, about half to New England and half to New York and Michigan (Lamarre 2003; Lavoie 1973; Dulong, 2001). We will explore more on the French Canadian movements to New England in a later case study.

In this chapter we delve into the population structure in 1881, to examine whether or not the eighteenth-century surname structure plays an important role in what we see at this time. Large-scale population movements in Quebec occurred after the 1830s as a result of economic difficulties, industrialization, bad harvests, and fluctuating wheat prices; despite the difficulties new regions were still settled. Will the surname structure in 1880 tell us anything about these population movements and whether new places or older ones like St-Hyacinthe or St-Jérome, as they industrialize, draw settlers from local populations? Will between region migration 'mix-up' the population and generate a more homogenous structure? Will the hierarchical structure we observed in the 1760s remain apparent as settlement spreads inland from the older seigneuries along the river? Will the highlevel contrast between eastern and western Quebec persist despite such massive mobility? Will we still have networks of short radius as displayed by the relevance network technique? And finally, will we still find strong name allegiances across the St. Lawrence?

Quebec 1881: Isonymy

If we re-examine genetic differentiation in the single-founder sample of 832 names, the 1881 Fst value is .0068 and the Rst value .0058. These values, are lower than those calculated for the eighteenth century and are somewhat unexpected because we would assume that the higher concentration of certain names would increase the probability of name similarity among the total regional population. Theoretically, in a closed population with no in-migration, certain names will begin to dominate and values of Fst (and Rst) will increase. One issue that must be considered when interpreting these values is the extinction of names. Of the 2300 single-founder names (reported by the PRDH) I found only 1300 in the census of 1881, and only 650 of my 832. It is also possible that we are losing names because of spelling mistakes in the suite of transcriptions, (census takers, LDS, and spellcheck process). Though I did spell corrections, it does not rule out mistakes and a slightly skewed sample.

For the 1881 full census sample, comprised of all male heads of household of all national origins (232,000 men older than 16), a similar decrease is also observed (Table 5.1). This decrease is not a surprise considering the higher rate of flow between populations (nameflow and geneflow), and the great diversity of names introduced in the nineteenth century by immigrants from Britain and Ireland who settled heavily in the vicinity of Montréal and the Eastern Townships.²

The single-founder sample regional (R) isonymy value for 1765 was lower than for 1881 (.0005 1765 vs..001 1881) (Figure 5.1). The differences in isonymy are indicative of a trend towards the regional population growth and the clustering of names by 1881. In 1765 names had not yet had enough time to grow, and there was still little chance of selecting two similar names at random from two regions.

In 1765 the regions with the highest name diversity were the Richelieu Valley, Lanaudière, Trois-Rivières, Quebec City, and Montréal and its immediate surroundings; and they remain the most diversified regions in 1881. The high diversity of French Canadian names in these areas influenced the diversity of regions to which they 'overflowed'. As populations moved up the river valleys

away from the Saint Lawrence, the newly opened regions often absorbed populations and surnames closest to them. For this reason, if a population was settled from a region of high name diversity, it too would have high diversity. Examples include the Beauce, a region settled by a diverse set of names from the Agglomeration of Quebec and Quebec City; the Richelieu Valley and Estrie (Eastern Townships), predominately colonized by surnames from the Montréal and Quebec City regions (Figure 5.1). In larger urban places, which attracted skilled and unskilled migrants, we also see a substantial increase in diversity, while rural places maintain stability in name diversity.

Most regions with exceptionally low name diversity in 1765 had moderately increased by 1881, but the regions pioneered by settlers from these places would also have low diversity. An obvious example is the Saguenay, settled after 1840 primarily from parishes of Côte-de-Beaupré and Charlevoix, both of which were characterized by low diversity of surnames (Figure 5.1).

Comparisons of 'between population' same-surname values at the two dates suggest that the initial pattern played an important role in subsequent organization and observable relatedness. As late as 1881, we see the powerful constraints of geography on the movement of pioneers along river valleys. The Richelieu Valley, for example, having drawn settlers from all directions, ends up with the greatest variety of surnames and emerges as one the most integrated regions in the province, with genealogical connections to most other regions. Settlers had moved south from Quebec City and Trois Riviéres, and east from Montréal. In contrast, initial settlement of the Saguenay was directed along one river valley which funnelled people from only a few population nodes.

The between population same-surname (isonymy) values provide further evidence of migration flows and population relatedness. When the 1881 census 'places' are re-cast into 91 hexagons, the average between population value is 0.006; slightly higher than the .005 calculated for the 117 parishes in 1765. The increase implies that in the 1881 sample there is a slightly greater probability of randomly drawing the same name in two locations (hexagon). This slight increase may be the result of a decrease in the number of total unique surnames as a result

of surname extinction and emigration, and the greater number of some surnames like Bessette or Beauchamp.

Multidimensional scaling based on the table of between population samesurname values reveals distinct population relationships (proportion of variance explained 60.0%, for axes 1 and 2). The biplot of the first two axes, illustrated in figure 5.2, reveals a regional structure of surname relatedness. Hexagons in area C (Saguenay and Charlevoix in gray) with the most strongly related values show a distinctive structure when compared with their geographic neighbors. The regionalization of names points out the distinct movement of settlers from Charlevoix to the Saguenay (Gauvreau et al. 1987).³ Places in area C (Saguenay) are tightly clustered in the upper right quadrant adjacent to places in the Bas-Saint-Laurent, another area reported to have sent migrants to this region. Other areas showing strong relatedness are sets B and D, which make up the geographic regions of Côte-Nord, Côte-du-Sud, Bas-Saint-Laurent and Beauce (Figure 5.2). At the opposite pole, set J is mostly made up of regions which show no specific relatedness to other places, with values for these hexagons being loosely dispersed in quadrants 1 and 3.

From the Mantel test we cannot argue a significant relationship between geographic distances between places and similarity of their surname profiles. How do we explain the absence of a distance decay effect, despite the apparent 'geography' of the map and what we know about the short-distance marriage fields? First, creation of hexagons has already absorbed the most significant component of distance decay, the marriages within a single parish or the neighbouring parish. If we did the same isonymy calculations at the parish level without aggregation, we would see distance decay similar to what St Hilaire (2002) has found in his research on local marriage fields. Second, the topography of the province, coupled with importance (down to the 1850s) of travel by water introduce strong directional bias, channelling movement along certain paths. It is noticeable along the rim of the Gaspé, for example, and of course the people who lived in fishing villages were exceptionally mobile. Penetration of settlers into new regions followed stream valleys, sometimes several small ones as was the case of the movement of settlers into the Saguenay region. Another reason we do not see a distance effect is because the Mantel test is a regional (global) summary and does not account for any local or directional effects, making it a technique with little flexibility to pick out the hotspots of strong and localized relationships. To pick out directional components of isonymy we might choose to perform a directional correlogram or geostatistical method.

Despite the multidimensional scaling results of the between population values with a high percent variance explained (60% on the first two axes), it is still difficult to summarize the extent to which locations are geographically related. Like the Mantel statistic the method provides a global summary of all the relationships.

To illustrate some of the salient and meaningful relationships, I provide the between population values for six locations, five of which were sparsely populated or 'empty' areas in the late eighteenth century. The first, illustrated in figure 5.3, is La-Grande-Riviere (hexagon 72) located on the southern shore of the Gaspé peninsula. It has 15 different surnames (of the 832), a local isonymy of .08, slightly higher than the regional average and is strongly related to Les Chlorydormes (89), Cherbourg (80), and St-Féréol (50). The pattern shows relatedness with the Saguenay, the northern tip of Bas-Saint-Laurent, and the northern coastline of the Gaspé, a recognized path of communication and settlement (Remiggi 1983). The overall pattern, suggesting weak relationships with its immediate neighbours, arises from seasonal mobility, a fishery economy, extensive out-migration. As Remiggi (1983) reported, the people of the Gaspé region were extremely mobile, engaging in leap-frog settlement.

The second example, the Baie-St-Paul hexagon (51), includes the parishes of L'Islet and Charlevoix, with 73 different surnames and a local same-surname value (isonymy) of .17, higher than the regional average (Figure 5.4). As we might expect from St-Hilaire's account of movement of settlers from this region (Charlevoix) to the Saguenay between 1840 and 1881, we see strong relationships along the Saguenay river and Lac Saint Jean, as well as the northern coast of the Gaspé peninsula and Beauce (Figure 5.4). At closer inspection we also see at the

boundaries a moderate relatedness with the Mauricie and Bois-Francs, and still weaker relatedness with the Outaouais and Estrie regions.

The impressive feature described in the between population analysis is the relatedness to parishes of recent settlement and the absence of relatedness to any other seigniorial lands settled earlier. This exemplifies the whole argument of this chapter: that surnames in place by 1760 'unfold' into new areas of settlement, creating a structure which persists to the present. It is likely that there were successive waves of settlement, associated with the great forest clearance operations oriented to somewhat different resource markets. These are better known in the US, as the white pine industry transferred operations from Massachusetts to New York to Michigan to Minnesota. The picture is more complicated for Quebec, but there is also a succession of timbering 'frontiers' (William 1989).

The Yamachiche hexagon (37, figure 5.5) well developed by 1800, has a strong relatedness with locations nearby, notably Trois-Rivières and the parishes of the Mauricie, with 206 different surnames (a large number), and a local isonymy of .024 (moderate diversity). This population in 1881 shows very strong name relatedness to the frontier region to the north, which includes the newer parishes in the watershed of the upper Saint-Maurice River, as well as older regions of Côte-du-Sud and Côte-de-Beaupré, and with the growing Bas-Saint-Laurent region (Figure 5.5). TheYamachiche hexagon shows a curious relatedness with hexagon St-André-Avelin (10) in the Outaouais. (Yamachiche and St-André-Avelin share 27 single-founder names).⁴ Clues to the population relationships, resulting from movements in the timber industry, have been described by Hardy and Séguin (1984), and will be further explored as a case study in chapter 6.

Located beside the Côte-du-Sud and Beauce regions and along the New Brunswick border, the Buckland (42) hexagon has a local isonymy value of .02 and 67 surnames. Buckland has its closest relationships with its immediate geographic neighbors, followed by Quebec City and the regions of Bas-Saint-Laurent, Charlevoix and the Saguenay (Figure 5.6). This hexagon has a more diversified population than those with which it is most closely related, and the multidimensional scaling results placed it in group D, made up of locations with little specific relatedness to other locations. Surnames in Buckland may be residuals of settlement by many populations. This would be consistent with studies which report Loyalists and missions from New England into the Eastern Townships, followed by a gradual re-settlement by French Canadians as branch railways were built in the 1860s.

The Ste-Adèle hexagon (16) is not closely associated with many other locations and is clearly more closely related with western Quebec and the Montréal region. It has a same-surname (isonymy) of .02 and 80 unique surnames. Located northwest of Montréal, Ste-Adèle is most strongly related to its immediate geographic neighbors in the Laurentides and Lanaudière, and a few locations in Côte-Nord and Côte-du-Sud (Figure 5.7). The region was settled mostly by individuals from Montréal and areas along the St Lawrence northwest of Montréal; among them names from the Beauchamp-Labelle cluster, one of the most diverse population groups in all of Quebec.

The final example, the Stanstead hexagon (23) with a local isonymy of .02 and 49 unique surnames, is also closely related with its immediate neighbors, as well with locations in the Outaouais and Rive Sud de Montréal (Figure 5.8).⁵ Initially settled by the English, the region was repopulated in the mid 1800s by French-Canadians. This region's situation on the US border and its settlement from surrounding areas of high diversity during the nineteenth century account for the high diversity of names.

The between population same-surname results demonstrate the emergence of new 'name regions' since 1800 and show the impact of earlier movements of settlers on the regional surname structure. As settlement progressed out of the original seigniorial lands, many moved to the nearest 'open frontier' or moved with relatives from the homeland, re-generating their original levels of name diversity and name affinity. Therefore newly settled lands often had similar levels of diversity and shared many surnames of their origin populations. We could argue here that the surname structure or name regionalization as revealed from same-name analysis is more concentrated at the end of the eighteenth century

because of the extinction, growth and overrepresentation of certain French Canadian names (e.g Tremblay, Beauchamp).

1881 Relevance Networks

Using Relevance Networks at the scale of the hexagons and at a more refined scale of subdistricts, I will expand the story of surname structure and confirm our previous results by using a method which groups places with the most similar sets of surnames.

Figure 5.9 shows the correlation values for the Relevance Networks and the associated numbers of networks. At a low threshold many hexagons show a moderate connectivity, and as we impose a more stringent correlation value only the most salient connections remain. The number of relevance networks reaches a plateau at .7, including smaller numbers of hexagons. For the following discussion, I set the correlation value at .6, a level which identified 6 networks and included 31 of the 89 hexagons (Figure 5.9). To handle the 670 subdistricts, I again explore alternative thresholds, as shown in figure 5.10; and for purposes of division, I extracted the most significant relationships by setting the correlation to .82, which creates 25 networks and includes 109 of 670 subdistricts.

For the hexagons, the six relevance networks all match the most significant between population isonymy values. The map (figure 5.11) shows that the six relevance networks are organized geographically. In terms of the multidimensional scaling results, the only discrepancy is network D covering the Beauce and northern Estrie, not so apparent in the multidimensional scaling biplot (Figure 5.12). Inspection of the between isonymy values for these places argues that they are strongly related. It appears that multidimensional scaling is pulling these locations apart because of their mutual relatedness to the Quebec City region and a rather high overall integration, sharing some names with many other regions.

As with the other same surname measures, the Saguenay and Charlevoix network is the strongest (most relevant) network and remains intact even when a correlation threshold is set as high as .9. The second strongest network stretches

through the Côte-du-Sud and Bas-Saint-Laurent. Another network stretches southwest from Quebec City into the Bois-Francs and Estrie. Figure 5.13 makes it possible to compare this network with between-population isonymy for the embedded hexagon Halifax (33). The values show it is most strongly related to the adjoining hexagon St-Sylvestre (40) and its southern neighbour Ascot (27), with strong ties to neighbors in the Beauce region. The other two networks are found along the river, within the old seigniorial lands: one in the Lanaudière (Sorel hexagon), the other southwest of Montréal.

Even at the scale of hexagons, the relevance networks provide an unbiased strategy for uncovering a regional surname structure. Let us now zoom in at a finer scale to examine surname similarities among the 670 subdistricts, the smallest 'local areas' provided by the 1881 census. The 25 relevance networks (A-Y) are organized geographically. Most often the subdistricts making up the individual networks are near neighbours. The largest sets are situated in the Saguenay/Charlevoix and Côte-du-Sud regions (Figure 5.14). These two include 51 of the 107 parishes making up all relevant networks. Embracing 28 parishes, the Saguenay/Charlevoix network (C) is the strongest, and parts of this network remain intact even when a correlation threshold is set as high as .9. The only difference between what we find at this fine resolution and the hexagonal aggregates is the connection of Sainte-Félicité in the Bas-Saint-Laurent. The Côte-du-Sud network (F) also resembles the distribution observed for the hexagons, with the major difference being the appearance of a distinct subset comprised of Trois-Pistoles, Sainte-Françoise and Saint-Jean-de-Dieu (K).

The appearance of the two largest networks is no surprise, considering that both have been evident in all of the isonymy measures. The remaining networks containing from 2 to 5 parishes are much better delineated when they are aggregated into hexagons. It seems that some of the strong connections of surname similarity were masked by aggregation. In other words, much of the surname differentiation is of a very local nature, and intense social relations (intermarriage) occur over short distances. This confirms Saint Hilaire's (2002) observations, on which I based my choice of a 'radius' for the grid.

The remaining networks, illustrated in figure 5.14, can be divided into three types. First, there are networks made up of parishes in the old seigniorial lands (early settlement) clumped along the St Lawrence (networks A, L, N, O, S). Second, several networks combine older parishes and newer parishes at moderate distances; these, for the most part, have an intensity and a distinctive 'geography' tracking the St Lawrence and its tributaries upstream or into later-settled ranges behind the original seigniorial lands (networks E, J, K, M, Q, R, U and W). Finally, a handful of networks seem to represent longer leaps of interregional migration (networks G, V, Y, T, and X). Network P, for example, includes St-Pierre-de-la-Rivière-du-Sud settled in 1727, Buckland, Rolette, and Newport on the southern coast of Gaspé peninsula.

The precision and ease of interpretation of relevance networks make this method very helpful for interpreting the complexity of relationships in large populations. Because the method can dial up or down the level of precision (correlation) of networks, it allows one to examine the hierarchy of places which display varying degrees of relatedness. The method is also useful because it removes pseudo relationships that have no meaning. We will see in the next section that when using cluster analysis we have to be aware of false connections or clusters.

1881 Cluster Analysis

Cluster analysis was also carried out at hexagon and sub-district scales. To determine an appropriate number of clusters, I used results of the previous analyses as guides. Without this prior information it would be much more difficult to uncover convergence. With clustering, we have to try to avoid grouping locations (or surnames) with low or moderate similarity, since the rigidity of the algorithm sometimes 'forces' items into a cluster. For the hexagons, I began with 40 clusters and mapped successive higher-order outcomes, each time checking for convergence. The map presents 17 clusters (Figure 5.15). For the subdistricts I had only the relevance network results as an aid. I began with 150 clusters and, after mapping and comparing various higher-order clusters with the relevance networks, I adopted a final mapping of 30 clusters.

The hexagons are color-coded with the cluster number displayed. The regional groupings match our previous results. The relevance networks of the Saguenay, the Beauce and the one running southwest from Quebec City into the Bois-Francs and Estrie are well reproduced here by clusters 22, 17 and 13. To the south a unique cluster covers Montréal, Mauricie, the Outaouais and the Eastern Townships. If we compare clusters with the isonymy results (by mapping the clusters onto the multidimensional scaling diagram) we see a good match between their geographic structures (Figure 5.16).

Once again as we zoom in at a finer resolution, we are able to differentiate further, and to pick off, among the 670 subdistricts, some local high-intensity relationships which are masked by the aggregation into 89 hexagons. The 30 clusters of subdistricts uncover unique regions with similar surname profiles, but further differentiation occurs within a cluster or between some clusters. We know from our hexagon clusters that there are distinct regions, but the arbitrary nature of hexagons does not specify 'real' boundaries nor indicate the degree of admixture at these boundaries. Here we will begin by looking at areas with unique clusters and at some places in-between where there appear to be a greater admixture of names.

In figure 5.17 the clusters (tagged by a number) reveal regionalization even at this fine geographic scale. To make interpretation easier, I have traced a line around areas with large concentrations of a particular cluster. Each of the clusters has strong affinity with a certain region except Cluster 6. This cluster, for example, is concentrated in Lanaudière, but is also scattered throughout the province: these are orphan clusters or clusters with no distinctive characteristics and with considerable diversity of surnames. After examining the location of parishes in cluster 6 in more detail, it appears that it is likely a result of low numbers of total individuals but elevated name diversity. Most of the locations in this cluster have moderate population sizes and higher than normal diversity. In our previous analysis for 1765 we also found that a similar cluster (the Drouin cluster); one which was widely dispersed with no discrete geographical organization.

The geographic clusters that I have described from the original 30 suggest paths of settlement governed by lines of communication including roads and rivers. Clusters starting along the St Lawrence most often follow river valleys upstream, as illustrated in figure 5.17, for clusters 28, 98, 193. 199, 23, 18 and 52 (name these villages). The best example of the ability of the cluster analysis to pick up the movements of families (similar names) up the river valleys is found in clusters 193 and 28.

The Saguenay cluster (52) as always provides evidence of regional uniqueness and its original connection to Charlevoix. Here we see, at fine resolution, that the parishes along the northern coast of the Gaspé are associated with migrations from the Charlevoix and the Saguenay. These clusters (55) were causing the variability in surname structure apparent in the earlier analysis, and here we can more precisely see the exact parishes with strongest ties. On the Gaspé peninsula five clusters dominate, cluster 116 being tightly structured with the other four irregularly placed around the peninsula. An explanation for this would require a more elaborate historical geography. One might propose, for example, hypotheses of distinctive economic base (different fisheries), waves of settlement generated from other older parishes as they 'filled', or perhaps institutional networks like the schools managed by one or another of the religious orders, or under jurisdiction of a different diocese.

Overall, the cluster analysis has shown that the organization of French Canadian surnames is hierarchical with intense local relatedness (fine-grained differentiation), and well-defined overarching regional structure.

Discussion/Conclusion: Quebec 1881 Surnames

The distinct regional name patterns and the hierarchy of relatedness revealed by the 1881 census sample are the result of a number of factors. The French Canadian population of Quebec has a fairly high level of genetic relatedness (Jompe et al 2001). This is a consequence of a population expanding from a small set of founders. In terms of surname, their regional concentrations are more intense by 1881 due to the tendency of individuals marry someone close-by, and name extinction or emigration of rarer names. The effects of extinction and emigration of names was limited by the fact that during sample selection a majority of the names were found in the 1881 census. When we began the eighteenth-century analysis, there were around 2300 single founders, and in our 1881 census sample I found only 1376 of those single-founder names. For the selected sample of 832 names 650 were found in the 1881 census. The literature of surname extinction suggests that around 41% of the total names should be gone. We might make some allowances for nicknames, double surnames, and spelling problems, but this lower number is still suggestive of extinction, either by emigration or by a lack of male descendants. As stated earlier, surname extinction is usually most rapid in the early generation. Of 1000 surnames, approximately 237 will become extinct in the first generation, 100 more by the second generation and eventually 55% of the total will become extinct. When I collected phone users from the 2001 Bell Canada records, I found only 1230 (46.5%) of the 2300 singlefounders and 1289 in all of Canada. Extinction seems to have slowed, as theory suggests, or has continued with sons names having taken all the growth outside Canada.⁶ As a result of extinction and the dominance of certain names (e.g. Tremblay, Beauchamp, Labelle), the regional relatedness of names appears much stronger than it would if all surnames had survived. This is directly related to diversity levels and the likelihood of two surnames marrying. We will examine this in more detail in a case study, whereby we examine the change in same surname and same name pairs during the nineteenth century.

The various analyses revealed between 14 and 30 regions with strong surname relatedness. Similar regions have been defined in studies by Bouchard et al. (1985), Vezina et al (1998), and Tremblay et al. (2001).⁷ Like others, I have found greatest diversity of names in the Richelieu valley and Outaouais.

Results from the 1760s, confined to the seigniorial lands along the St Lawrence, suggest that the surname structure at this time governed to some extent the distribution a century later. In a majority of cases, frontiers were settled by residents from a nearby source. From the case of the Saguenay, Bouchard (1991, 1996) has argued that frontier parish expansion and migration were a maturation phase, whereby young married couples arrived with children and availed themselves of the youthful labour would clear land for farming: it took about a generation to clear fifty acres. After about a generation, some families would move to another new parish close by, where they could obtain a larger tract and where their children could settle down as neighbours. Saint-Hilaire (1988) explains that French Catholics had strong nuclear families and tended to migrate together and to establish farms; and Bouchard (1991, 1996) describes the process as a "reallocation model" which involves a ten to thirty-year process initiated by marriage of the first son and completed with the death of the surviving parent. Cluster analysis at the subdistrict level, showing the name clusters extending up river valleys, suggests that we are seeing this style of frontier settlement.

A distinctive feature of the eighteenth-century structure as compared with 1881 is the importance of clusters which include parishes on either side of the St Lawrence. Early family allegiances seem to result from marriages and social networks spanning the river. As settlement extended away from the St Lawrence, socialization across the river seems to have declined. In 1881 we still see riparian clusters, but as we see in figure 5.17, the cluster is likely to include neighboring parishes upstream on either side of the St Lawrence, not across the river. The Saint Lawrence in 1881 had few bridges (only at Québec and Montréal), and movement was now facilitated by better roads and by railways parallel to the rivers. Although steamboat service remained a factor on the St Lawrence, it favored the large towns, and for the new modes of travel by rail, the great river was an obstacle.

Important influences on regional surname structure included topography and hydrography. These two factors eased or inhibited communication, and despite the advent of modern mode of transports, their impact is still apparent in 1880. We find regions and small pockets of diversity, associated with corridors of transport, notably the Richelieu valley. With the influence of industrialization in the midnineteenth century, individuals had new choices regarding their lives. The growth of urban centers certainly pulled individuals into these locations, increasing the diversity of names in the new locations and stabilizing the diversity in many rural source regions.

NOTE

³ Also see Gagnon and Heyer 2001; De Braekeleer, 1995; 1993; Scriver, 2001; De Braekeleer and Gauthier 1996

⁴ This includes the parishes of Notre-Dame-de-Bonsecours, St-André-Avelin and Sainte-Angélique.

⁵ I am aware that the R(2) only represents R-square represents 12% of the variation in the data but when compared to the other locations (hexagons) not in group j as defined by the mds results, this result is strong. The locations that are clustered with their immediate neighbors show no distance decay.

The sample used for the same marriage (inbreeding) calculation included a subset of single founder marriages (832 of 2300 surnames). It included all marriages for each of the 832 surnames from their first arrival until 1799. The single founder sample accounts for 30 percent of all provincial marriages and shares similar statistical properties with the marriages for the entire population (Figure 1).

⁶ These numbers do not count variants of names, just the standardized single- founder names compiled from the PRDH database.

⁷ See Gagnon (2001) for a review on surname methods and a general geographic distribution of Quebec eighteenth century surnames.

¹ See Bouchard 1996 for an explanation of the settlement process; he uses the Saguenay as an example.

² For the subset of all males >15, there are 25,101 different names, and 13, 000 of these are unique (occur only one time).
Chapter 6

Case Studies

Methods employed in the previous chapters provide a regional interpretation of patterns and trends in surname structure and illustrate the types of information that can be extracted from list of names. ave observed the need for flexibility to evaluate surname similarity at different spatial and temporal scales, and we have seen that each technique, despite comparable results, is better suited for answering certain questions. In this chapter, I employ four case studies to 'bore' deeper into the application of certain methods and show how they can be combined to answer specific questions such as: What is the volume, concentration and direction of migrant flow? Which regions have the greatest surname affinity? How did changes in surname diversity (over time) and urbanization influence the rate of same-surname marriages and same surname pairs? Did diversity measures vary spatially? Each study will show a strategy for using name methods to provide specific kinds of evidence of migration histories as well as add depth to current historical studies

For each study, I identify a specific hypothesis, select an appropriate scale of analysis (spatial and temporal), and then pick from my toolbox methods, which I have already presented, to answer the question. None of the case studies is intended as a complete historical investigation and as we will see many questions and investigations will require more than one tool. Each study is intended as an exposition of methods for responding to specific historical questions and to provide specific examples of how these methods can be used. Unfortunately most isonymy (surname) methods have been used for regional interpretations and not for specific questions. I argue here that they can also be helpful in answering a range of historical migration questions at different spatial scales.

Case Study 1

Urban Choices: Montréal, Quebec City and inter-provincial movements.

In the nineteenth century, the movement of people from smaller places to the two largest population nodes in the province was in large part a result of both skilled and unskilled job opportunities. While a portion of the Montréal and Québec City population growth is accounted for by foreign immigration from England, Ireland and Scotland, the largest share was generated by rural to urban moves of French Canadians. We do not know very much about the impact of foreign immigrants on urbanization, but the decennial census provides some information on places of birth. Grace's (1999) work is one of the few attempts to differentiate arrival dates from estimates of settlement. Ramirez (1991) has described the concentration of Italian immigrants in Montréal in the 1890s and early 1900s. Surprisingly little has been written on the influence of the movements of rural French Canadians to urban areas, when in fact they were the prime sources of urban growth and the means of meeting the demand for 'cheap labour.' One reason for a lack of precision on the topic is that there was no specific counting method for tracking the moves. We have more specific information on immigrants, but nothing documenting internal migration. Information can be squeezed from the Census of 1901 which distinguishes 'urban' or 'rural' place of birth, but this source, recently released as a digital 5% sample, has yet to be exploited.

With limited sources for French Canadian moves to Montréal and Québec City, it becomes difficult to assess rates of urbanization. From which parts of the province did people arrive? Were certain places more likely to 'push' French Canadian migrants toward urban areas? At what rates during the nineteenthcentury were French Canadians moving to Montréal and Québec City? Did most choose the nearest of the two cities? or did they by-pass Quebec City, where employment opportunities were not increasing. By using a sample of singlefounder surnames, I estimate the rates at which migrants were moving to urban areas and from what regions they were coming. To explore urbanization rates and their regionalization, I track marriage records for 32 single-founder surnames, decade by decade, and for each surname, I tabulate the percentage of marriages celebrated in Montréal and Québec City. The logic here is that as the Montréal and Québec City populations increased and the number of French Canadian marriages became greater in these two areas than in rural places, we get an estimate of the relative rate of urbanization.

To take the inquiry of moves one step further in both space and time, I observe the long distance inter-provincial migrations of the same set of 32 single-founder names. The intention here is to use descriptive measures to observe the extent to which names have dispersed across Canada. The source is a 2001 Bell Canada phone directory, and I employ the set of 32 single-founder names.¹ The purpose is threefold. First, I want to highlight the results of surname movements (migration) in Quebec during the last 375 years. We will see that the structure in place in 1881 dictated much of the presentday distribution of names. We saw earlier that the dispersion of names in 1765 influenced where people would migrate or settle during the nineteenth century. While we anticipate a pattern of migration to large cities like Winnipeg, Edmonton, Toronto and Vancouver, I am especially interested in observing the rates of movement to other provinces? The final purpose of this study is to round off the story of the migration history of the French-Canadians by looking at more recent moves and the overall distribution

Demographers have shown that many young people moved before marriage (Thornton and Olson 2002), and there is some indication that young women took advantage of the demand for single female labour in domestic service to explore a 'marriage market' that might offer them upward mobility or a greater choice of lifestyles. They have also obtained evidence, from rather small samples, that children of rural in-migrants did achieve higher living standards than their parents.

Thirty-two single-founder surnames were selected from the sample of 832 discussed earlier, and marriages were extracted from the Drouin index (see chapter 2). The digital subset embraces 32,530 marriages over the period 1621 through 1900 and represents about 5 percent of all marriages in what is now the Province of Québec. Figure 6.1 tracks the curve for the 32 surname sample against the

counts of all known provincial marriages between 1700 and 1900. Both the sample and total provincial French Canadian marriages grow exponentially (straight line on semi-log plot), and the two curves illustrated in figure 6.1 suggest that this is a representative sample in terms of its temporal growth behaviour.

In the first part of the analysis I aggregated the 32 names into a single group and mapped total marriages by parish and by decade. The suite of maps provided a moving picture of how settlement proceeded. In figure 6.3, I provide a plot of the number of parishes present in each decade. In the second part of the analysis I use the 32 names to estimate the rates at which people were moving to Montréal and Québec City. To calculate the rates, I use the percentage of all marriages which were recorded from parishes on the Island of Montréal or in Quebec City between 1800 and 1900.²

For the first part of the analysis I do not differentiate the geographic concentration of each surname, and I derive a regional rate.³ In the second part, I use hierarchical cluster analysis to group names with similar geographic profiles, and I use these groupings to derive region-specific urbanization rates (Table 6.1). The clustering technique is the same used in the earlier chapter. This approach provides a way to answer the question of whether or not individuals from certain regions tended to move to the nearer of the two cities.

We have seen, in earlier chapters, that in the eighteenth century, the population hugged the St. Lawrence, and that settlement followed river valleys and smaller tributaries. The sequence of maps in figure 6.2, showing marriage locations for our 32 surnames, reveals the continued progression of settlement to the end of the nineteenth century. On the 1800 map, parishes remained close to the St. Lawrence with the exception of the parishes spreading into the Richelieu Valley, up the Chaudiére River, and into the Beauce region. Between 1700 and 1800, there was each year (on average) one new parish where a marriage occurred, but due to political reasons there were few new parishes before 1800. The number of parishes expanded from 17 in 1700 to 117 by 1800. Between 1800 and 1840, the number of parishes grew slowly, with further concentration in the Beauce and Richelieu valley, but the population in the 1870s moved deeper into the

Laurentides and regions northwest of Montréal, as well as along the St Lawrence in the Côte-du-Sud and Bas Saint Laurent regions. The exponential growth of the population during this period parallels the rate of erection of new parishes (Figure 6.2a).

After 1840, we see the start of significant expansion of new settlements away from the St. Lawrence. The number of parishes jumps from 204 in 1840 to 304 by 1850 (Figure 6.3). It is important to recognize here that we are seeing the combined impact of heavy in-migration (notably Irish Catholics), of high fertility stimulating demand for 'new' land, and of a strenuous ecclesiastical effort to expand vocations, missions and parish structure. New parish openings are most apparent in the large river valleys along the Ottawa, St Francis, Chaudière, and Richelieu. There is also substantial growth away from the banks of the St. Lawrence on the backsides of the original seigneurial lands in the Lanaudière, Laurentides and Mauricie, and settlement is initiated from Charlevoix into the Saguenay (Figure 6.2a).

Despite an increase in the number of rural inhabitants moving into urban centers like Montréal between 1850 and 1880, the number of rural parishes continued to grow. Settlement continued along the large rivers and eventually spilled up the smaller tributaries. We see increasing numbers of settlements, for example, in the Saguenay, Ottawa Valley, Bas Saint Laurent and the Gaspésie, as well as the agriculture lands of the Bois Francs and Estrie. ⁴ The ability of the French Canadian population to fill the new frontiers and move to urban areas like Montréal and Quebec City can be attributed to their high natural increase (6.2b, 6.2c).

The nineteenth-century urbanization rates, as revealed from my sample of 32 names, show an astonishing number of rural French Canadians seeking opportunities in Montréal. The French Canadian share of the Montréal population increased after 1850, from about half to two-thirds in 1900. In 1871, 60% of the total population of Montréal (144,044) was of French Canadian ethnic origin (Linteau et al. 1983; Young and Dickinson 2003). The city was characterized by 'mushroom growth': it doubled in size every 20 years; a phenomenon we see today

in 'developing areas' such as Nairobi or Manila.

Figure 6.4 shows that the percentage who married in Québec City remained stable (around 5 percent) from 1800 to 1890, but jumped to 10 percent by 1900. In the 1800s decade, 10 percent of marriages took place in Montréal; in the 1810s the rate increased to about 15 percent, fell to 11.5 percent during the next decade and remained at this rate for the next thirty years (Figure 6.4). During the 1860s we see the greatest increase in the Montréal share of marriages, and by 1900 around 25% of all the marriages were taking place in Montréal. The dip between 1810 and 1860 coincides closely with the number of new parishes and show that many were choosing to settle new lands instead of moving to Montréal or were leaving the province for New England or Ontario. The public works of the late 1840s promoted lumbering and forest clearance for settlement. These events had an influence on the decisions people made regarding where to move or to search for work. If we examine the last map (1880) in the series, we notice that much of the lowlands have been settled, and it is around this time that a greater number of individuals are moving to urban centers. These moves are a result of the 'industrial revolution' of steam power in Montréal beginning in the late 1850s and the mechanization of industry in the 1860s with the opening of numerous factories and foundries (Lewis 2000).

Despite the large number of intra-provincial moves, we have little information on the influence of location on the choices people made. Drawing from the classic literature of migration, we might postulate 1) moves to the city in response to industrialization, 2) a distance decay effect applicable to any urban 'migration shed', and 3) an effect of intervening opportunities: a preference for the nearest growth pole. From the subsample of 32 names we would therefore expect to find: a) the highest percentage increases for Montréal and Quebec City; b) for names whose eighteenth-century distributions are closer to Montréal, progressively higher percentages of marriages in Montréal; c) similar effect for Quebec City, but less dramatic; and d) lower percentages along the Vermont border where residents had alternative options in US towns and cities.

For Montréal and Quebec City, the relative distance from each location does

influence the proportion of individuals moving to these areas. The Beauchamp/Labelle cluster described earlier (six of the thirty-two names), for example, shows the greatest flow to Montréal from 1800 through 1900 (Figure 6.5). This cluster of names concentrates in the Rive Nord-Ouest de Montréal and in the Laurentians. Rather stable between 1800 and 1860, the rates climb rapidly after this time reaching 60 percent 'urban' in 1900. The growth and movement of individuals in this cluster to Montréal tracks very well with the observation of Olson (1996) who mapped the urbanization of the Beauchamp surname into Montréal. The second highest cluster in terms of net flow to Montréal was the Bessette, cluster drawing most heavily from Estrie and Richelieu regions in both 1765 and 1881. The urban share of this cluster increased steadily from 1800, but much more slowly than the Labelle cluster. It also showed a jump in the 1860s, topping off at 20 percent 'urban' by 1900. The other three sets of names, concentrated closer to Quebec City and the Saguenay, had the lowest percentages of marriages in Montréal (Table 6.2).

For Quebéc City, the Montréal results are inverted. The Laberge name cluster has the highest overall net flows from 1800 through 1900. It was originally concentrated in the Côte-du-Sud, close to Quebec City, but there was also a small concentration in the Rive Sud de Montréal region as early as 1725. The highest net flow to Quebec City was around 11 percent in 1900. The Gravel and Vachon clusters, close to Quebec City, also behaved in much the same way. The clusters with the lowest rates of marriages in Quebéc City during the nineteenth century were the Bessette and Beauchamp clusters, both as we have seen originating in the Montréal region (Figure 6.5, Table 6.2).

To interpret these results, we have to consider the many other opportunities people had. The Drouin marriage records provide information for some parishes in eastern Ontario, New Brunswick and New England. We can think of these as 'samples', but they are untested and perhaps erratic. The surname clusters also respond differently in terms of the percentages of marriages at each of these locations. Names belonging to the Labelle cluster, for example, report more marriages in Ontario; whereas the Bessette cluster reports more from New England, and the other clusters have low numbers in Ontario, higher numbers in New England and New Brunswick.

The results exemplify migration variability and confirm that people tended to seek new opportunities at the nearest growth pole. St-Hilaire, in his studies on Quebec City, has demonstrated from surname turnover that stagnation of overall numbers did not mean no migration. Rather, people were moving out of Quebec City, and others were replacing them, coming from the surrounding rural villages.

Throughout this study we find individuals, as well as the local regions where they concentrate, following paths postulated in the migration literature: 1) Name clusters tended to have the highest percentages of marriages at the nearest urban place (e.g. Montréal or Quebec City); 2) movement toward the largest growth poles displayed a distance decay effect, where residents who lived further from Montréal or Quebec City were less likely to migrate and more likely to seek other opportunities in nearby communities, in Western Canada or New England; and finally 3) the industrialization of the 1850s attracted higher proportions of marriages in Montréal and Quebec City. In the next case study we will employ the strategy for appraising region specific migration using surnames, by pinpointing the regions which sent more individuals to Maine, New Hampshire and Vermont.

For the exploration of regional migrations using phone records for the 32 names, we turn to the five name clusters presented earlier. In 1881 each cluster of names had a specific profile of regional frequencies in the various parts of Quebec. What does this regional distribution look like? Table 6.3 reports the distribution by cultural regions, of the total marriages for each of the five name clusters between 1870 and 1899. As reported in the table, each of the clusters had a specific location of the greatest concentration. The Beauchamp-Labelle cluster, for example, is mostly distributed in and around Montréal (as observed in 1881); the Bessette-Chicoine cluster had more than 70 percent of marriages in the Richelieu and Bois Francs; the Laberge-Cusson cluster had a great number of marriage in both the Quebec and Montréal regions; the Buteau-Gravel cluster had marriage concentrations in the Mauricie and Bois-Francs regions; and the Vachon-Bolduc cluster of names had more than 60 percent north of Quebec City.

To get a sense of the extent of migrations of the French Canadian population, we now examine the presentday distribution across Canada of the same five name clusters. When we examine the movements of French Canadians to Vermont, New Hampshire and Maine we noted that migration to these locations came from certain parts of the province, a possible result of proximity and processes of chain migration. We found a similar trend for movement to Montréal and Quebec City.

Table 6.4 profiles the distributions of the five clusters by province. For Beauchamp-Labelle cluster, we find a distinctively high percentage of residents in Ontario (21%). This is not surprising since the names in this cluster had always been widely dispersed west of Montréal and in the Outaouais region. Aside from this feature, there are few other patterns that stand out. The Laberge-Cusson cluster which had a large number of marriages in Montréal in 1881, has the second highest percent of residents in Ontario; and the Bessette-Chicoine cluster has relatively high percentages in the western provinces. In the future it would be worthwhile to explore these trends using a larger sample of names putting greater emphasis on quantifying the patterns and exploring the origins and destinations associated with distinctively French Canadian settlements of Manitoba, Alberta and eastern and northern Ontario.

Finally, I examine three names at a refined scale of postal codes: Beauchamp, Laberge, and Vachon. Figures 6.6, 6.7, 6.8 show the 2002 distributions of these three names. We do see geographic differences, but only between the numbers who went to Ontario. Here again Beauchamp has the greatest number.

The results, although by no means conclusive, do provide some evidence of trends and movements that involve different push and pull factors and chain migrations. There are avenues one might pursue in the future. I intend to test these ideas using larger samples, the 1881 and 1901 censuses, and today's phone records.

Case Study 2

French-Canadian Migration to Vermont, New Hampshire and Maine While the Quebec population continued to grow rapidly and the economy fluctuated, French Canadians began to look for new job opportunities. Some purchased land as the population continued to expand away from the St Lawrence, some moved to growing urban centers, and still while others opted to leave the province for New England and western Canada. The next question to answer is: Did Quebecers in search of survival opportunities go 'randomly' to the mill towns of New England, or do we observe different paths? Observations of Roby (1990), Boisvert (1996) Frenette (1989) and Vicero (1968) suggest highly specific, highly localized chain migrations, in several instances traceable to recruiting in one's 'home village'. From classic migration theory we might expect: 1) migrants would move to locations closest to their homes; 2) longer- distance moves were more likely a result of more selective chain migration; and 3) migrants living closer to Montréal would choose this location over New England.

To appraise the origins of migrants moving to Vermont, New Hampshire and Vermont, I use a Bayesian probability method, based on the model of Degioanni and Darlu (2002), who raised a similar question for twentieth-century villages of the Pyrenees (France).

The 1881 Quebec Census, and the 1870 and 1880 censuses for Vermont, New Hampshire and Maine were used in the analysis. In each case French Canadian single-founder names are used exclusively so that we are reasonably confident that they are moves of related families, however distant the relationship. According to Vicero (1968), Vermont in 1870 had the largest French Canadian population in New England, but it was quickly supplanted as a destination by Massachusetts and the other New England states with larger, faster-growing manufacturing centres.

After correcting the data for surname variants using the name dictionaries described earlier, each sample was queried for the 832 single founder names. The query identified 132 of the names in Vermont, 164 in Maine and 90 in New Hampshire (3 states, 3 lists of names). Using the names present in each state, I then collected (queried) their frequency distributions from the 1881 Quebec census. The computation was then performed independently for each of the three lists of names.

The objective is to specify the probable geographic origins of the group of surnames moving to a particular state. As discussed in an earlier chapter, the

Bayesian technique works on the assumption that for each name we find in the destination, the probability of its originating from a certain area in Quebec is a function of its frequency in that area. Because frequency of the selected names is an important component in the equation, extremely large populations can bias the results, making it look as if large numbers are migrating from that area. To minimize any such bias, Montréal was removed from the calculation. In the case of Montréal we can assume from classical migration theory and studies on Montréal that it was primarily an attraction pole, sending few French Canadian immigrants to the states.

The Bayesian analysis reveals distinctive geographic patterns for the likely origins of French Canadian migrants into Maine, Vermont and New Hampshire, with a stronger representation of names from regions in close proximity. Figure 6.9, showing the likely source regions (home parishes) for the Maine sample suggests networks with Côte-du-Sud, Bas-Saint-Laurent, Beauce and the northern Estrie region. In contrast, in figures 6.10 and 6.11 showing the possible origins for migrants to Vermont and New Hampshire, fewer names are associated with these regions and more from the Richelieu valley and along the Vermont and New Hampshire border. Frequency profiles of the surnames in the various parishes provide further evidence of distinctive pools of names moving to each state. Of the 282 names found in the three states, Vermont and Maine share only 26 (9%), New Hampshire and Maine share 30 (10%) and New Hampshire and Vermont share only 8 names (4%). A Spearman rank correlation was performed for these three samples, confirming significant differences among the three samples. It confirmed also that New Hampshire and Maine samples were more closely related (Table 6.5).

Each of the three maps shows a unique geographic pattern. Source populations for Vermont are clustered along the Yamaska and Richelieu River Valleys, and push furthest into the Lanaudière region. This map, illustrating the paths of possible migrants to Vermont, appears to stretch and reach toward the Vermont-Quebec border. The map of New Hampshire shows a pattern similar to Vermont, but less dense, with more locations represented in the adjacent zone of Estrie. Finally the Maine map illustrates a likely concentration of possible migrants from Saint-Hyacinthe (City) and surrounding locations as well as the Mauricie.

Conclusion/Discussion

Despite the idiosyncrasies of individual moves, and all the ambiguities of spellings in the nominal census, these experiments show that a rather small sample of surnames, controlled by the identification of the founder population, suffices to visualize rather complex movements and to fix them in the time frame. The method works well because we are aware of the regionalization of names. When specific names show up in New England, we can be fairly confident that they come from certain regions. The method fails to link specific migrants with their exact place of origin, but it does provide a good first attempt at trying to interpret the movements of the thousands who migrated to New England.

This case study used Bayesian probability to appraise the possible origins of migrants moving into Vermont, New Hampshire and Maine. We have accounts of the hundreds of thousands of French Canadians moving to New England, and we have a number of studies examining specific communities and individuals who moved to New England. Vicero (1969) provided the most detailed geographic depiction. Although his study details the locations where French Canadians concentrated in New England, it provides little information on their origins in Quebec. He attributes a majority of origins to the Eastern Townships, but does so with little specificity. Although many did come from the townships, we know from other studies, including the work of Boisvert (1996), that there were other regions significantly contributing migrants. Boisvert's maps of origins and destinations of the Boisvert clan into New England. With a great degree of specificity Boivert illustrates that this clan (surname) comes from the Mauricie, Nicolet and Trois-Rivières regions. By using dated lists of names the study could easily be expanded to provide a refined temporal analysis, highlighting when the different regions sent migrants. It might also be insightful to expand the study to include the other New England states which were also popular destinations of French Canadians. The surname strategy maneuvers around the imperfections of matching datasets from different sources.

To further expand this study and improve the specificity of the moves it would be worthwhile to use between population isonymy to get a better idea of the distribution of the origin and destination parish and towns.

Case Study 3 Lumberjacks and the settlement of the Mauricie: A possible link to a strain of Tuberculosis

This case study uses isonymy measures and Bayesian probability techniques to examine the genealogical links and migration paths of a population of lumberjacks into the Mauricie region in 1860s. A lumberjack population listed in the 1861 census consists of approximately 1126 men with 600 different surnames. This population has already been 'under the microscope' in regards to their daily lives as lumberjacks and their sharing of common experiences, including their mobile lifestyles bringing them each winter to the forests, and their daily lives among hundreds of others lumberjacks in lumber camps (Hardy and Séguin 1980, 1984; Fortin 1981, 1983). These men helped open the rugged tree-shrouded lands along the Saint Maurice and Shawinigan Rivers at a time when the flourishing forestry economy was essential to their economic survival and provided an alternative to seeking work in urban centers or mill towns in Western Canada (Ontario) or the United States.

We are interested in this population because of their possible link with an historical outbreak of a particular rare strain of tuberculosis. Routine clinical health surveillance of tuberculosis in Canada uncovered an unexpected concentration in Quebec of a variant strain of tuberculosis that is resistant to the antibiotic pyrazinamide (pza-r). Molecular epidemiological techniques have revealed that in Quebéc 6.2 % of Mycobacterium tuberculosis (MTB) isolates in the Canadian-born population are resistant to this drug, a very unusual rare phenotype. In a low-incidence setting, such as Quebec (1.9 cases per 100,000), it is remarkable that a highly unusual, drug-resistant strain would expand unrecognized over the last few decades and account for so many cases. It is common for unusual phenotypes to be associated with newly immigrated foreign born. For Quebec, it

has been determined through genetic fingerprinting techniques that the pzaresistant cases are not associated with the foreign-born cases concentrated in Montréal, and that 90% of the pza- r cases are of French-Canadian ancestry. As for the geographic distribution of TB and pza-r TB, a chi-square revealed that the in the eighteen Quebec health districts there was an overrepresentation of TB cases in the Mauricie and Bas St Laurent, and an overrepresentation of pza-r cases in the Mauricie.

Attempts at trying to solve the puzzle of the population origins of this strain of bacteria invited a combination of new and old methods in historical geography as well as methods in epidemiology and genealogy. We used case (patients) and control (pza-r) genealogies to test whether or not individuals within the French Canadian population were genetically more susceptible to this particular strain. Genealogical reconstructions for 40% of the pza-r patients were completed and the results revealed no concentration of descent from a particular founder or lineage, ruling out human genetic susceptibility.

The genealogical reconstruction did reveal a higher concentration of ancestral marriages of the cases in Bas-St-Laurent, Beauce and the Mauricie regions. If the ancestral marriages are grouped by generations or by 20-year blocks, we observe in the genealogies an overrepresentation in certain areas, and these can be tracked from one decade to the next. Prior to 1860 the Cote du Sud region was overrepresented; between 1860 and 1930 the Mauricie is overrepresented, and finally in the interval 1960-1979 both the Outaouais and the Mauricie are overrepresented.

Various spatial statistic techniques were also used to determine clustering of pza-r cases. The first technique used was nearest neighbor analysis. We compared the 76 cases (TB pza-r) with the 256 controls (TB). Using a specially designed resampling technique to adjust for uneven sample sizes we found that the geographic distances between the two groups were not different. To determine whether or not there was evidence of any clustering we used Cuzick & Edwards' method. The method uses a search radius between cases and controls to determine if cases are close to control. A null hypothesis of no clustering would be evident if

cases and clusters were found in the same locations. Our results showed that there was evidence of clustering. But because this is a global method it did not provide a location. To determine location of the potential clusters was we performed a cluster spatial scan statistic. Using Satscan (Version 5) we applied a Bernoulli model which compares the cases (pza-r) and controls (TB) and determines if the expected number of cases is greater than expected at various locations (Kulldorff and Nagarwalla 1995). We found significant clustering the Mauricie region.

Most of our evidence suggests some clustering in the Mauricie region; the region where there were a large number of lumberjacks. Did our lumberjack population of January 1861 have regional ties (surname associations) with these three regions? If there was a connection, we might suspect that the logging labour from Cote du Sud carried the bacillus into major timbering regions of the Mauricie and Outaouais, perhaps in the peak years of the timber trade in the 1830s and 1840s.

Though our inference of an outbreak of the disease in the nineteenth century remains speculative, the combination of evidence for present-day localization and historical population movements point to the timber industry as a possible vector in the diffusion of disease, in conjunction with its role as an instrument of settlement in the watersheds of the Saint-Maurice and Ottawa rivers. We are suggesting that the geographic distribution of the current cases follow historical migration and settlement paths. I will attempt in this section, again using samesurname and Bayesian methods, to track the origins or connections of the lumberjacks. My goal is two-fold. First, by taking an unbiased sample (where incidence of tuberculosis does not enter into the picture), somewhat larger in size (n=1100), I can make an independent appraisal from surnames of the migrant paths. Because of the single-founder constraint on selection of names, I have reasonable assurance of the family links. Using this sample we can identify the regions that have the greatest numbers of names associated with the lumberjacks. Second, I intend to compare the isonymy and Bayesian results with that of the genealogical reconstructions to argue the value of using surnames.

Data/Methods

Three different datasets are used in this analysis. All three are restricted to the surnames we know to be associated with a unique founder (see earlier discussion of confirmed single-founders). The first is the nominal census manuscript from January 1861, which provides the names of the lumberjack population in the 'unorganized territory' of Champlain County and the recently opened village of St Stanislas, situated in the upper reaches of the Saint-Maurice. This population has 1490 individuals with 640 unique surnames. Of the 640 names, 590 were identified as being French Canadian and 221 as single founders. Using only the 221 single founders, the base population used throughout the analysis consisted of 390 individuals. The second data set employed is the set of marriages of men of 832 single-founder surnames for the years 1740-1769 and 1770-1799. The final dataset used is that of married men of the 832 single-founder surnames (approximately 43,000 individuals) extracted from the 1881 Census.

Using the between population same-surname (isonymy) method, I compare the frequency of single-founder surnames in the population of 1861 lumberjacks with Quebec-wide populations for two periods in the eighteenth century and one in the nineteenth. Using this method, we obtain values representing the similarities between a subset of names to the timber population. As with the other cases, we are operating on the assumption that a greater similarity of surname structure reflects a relatedness of families, a history of proximity in the past, and past migrations. Available samples are not ideal: census of 1842 or 1851 is not fully digital for the whole province, and the leap from 1800 to 1860 is two generations at a time of intense mobility.

The Bayesian method is applied only to the 1881 subset. The same two subsets of data are used, but the 1881 dataset includes only names matching the 221 found in the list of lumberjacks. I do not include the entire sample because I am interested only in the probability of finding names from our timber population in other populations. For example, if a parish had a similar set of names it would have stronger connections and migrations could be inferred.

Results

The results are consistent with the hypothesised population movement from Quebéc City and Côte-du-Sud to the Mauricie . The results for the two 30 year periods in the eighteenth century show highest between population isonymy values in the parishes of Chambly, and Sainte-Geneviève-de-Batiscan, and La Pérade along the banks of the St Lawrence (Figure 6.12). When you compare the two lists, a few new parishes appear in the later period: Les Écureuils, Saint-Cuthbert, La-Baie-du-Febvre, and Rivière-du-Loup (Louiseville) in the vicinity of Trois Riverieres or upstream of it. Figure 6.12 [the map] shows that at end of the eighteenth century, populations most related to the lumberjacks concentrated in the Quebec, Trois-Rivières and Côte-du-Sud regions. The highest same-surname values are found in the Trois-Rivières, while the greatest number of marriages of a 'timberman's cousins' were concentrated in the Côte-du-Sud and Québec region. These two values do not coincide because of the small sizes of the populations. Remember that we are comparing very small populations with a total of 832 surnames from our single founder sample. The smaller populations around what would eventually be the Mauricie region have similar surname profiles, while the larger populations of the Côte-du-Sud Québec have a greater diversity of names.

For 1881, figure 6.13, displaying only the same-surname values greater than .002, shows the continued concentration of timber-population affinities close to the Mauricie region. The Côte-du-Sud population region has a smaller concentration of them than in the eighteenth-century and we now see high values in a few locations in the Outaouais, Bas-Saint-Laurent, Saguenay and Charlevoix regions. The high values in the Saguenay and Charlevoix regions contradict those of the Bayesian analysis, which shows a low probability of there being a surname from the timber population (Figure 6.13). Figure 6.12, based on the eighteenth century and figure 6.13 and 6.14, based on the late nineteenth century data, suggest continued subsequent diffusion, perhaps of their rather distant cousins.

Discussion

Surname evidence documents an unrecognized salience of movement between Cote-du-Sud and Mauricie 1835-1860, and the pattern of movement is consistent

with the hypothesis of an epidemic outbreak in the Mauricie prior to 1860. In addition, the results show a strong linkage between patient genealogies and the surname distributions, however using these two methods we can not be certain of exact moves and a possible link with the pza-r TB strain.

The pattern of movement detected from surname evidence suggests that members of this population (or their relatives) were underrepresented in Montréal. This is consistent with the curious under representation of PZA-R in Montréal today, and might explain why the numbers are proportionally very small when compared to the total Montréal French Canadian population. If the lack of sheer numbers of infected or latent cases moving into Montréal does not explain the low numbers of PZA-R cases, then one might propose as an alternative hypothesis differential success of the strain in a large urban area. Is there any means by which the strain would be filtered out in an urban environment? Might the strain be thought to have differential 'fitness' in urban vs rural habitat?

Case Study 4 Repeated-name and same-surname marriages

The purpose of this study is to examine the likelihood that individuals within certain areas tend to marry others in an immediately surrounding region. It is of some methodological importance to understand the way the regional name diversity affects the frequency of repeated-name marriages. As I described earlier, repeated-name marriage refers to how often we come across pairs of names (husband's name and wife's maiden name) (Lasker and Kaplan 1985; Koertvelyessy et al. 1988). How often, for example, do we find Bessette-Beauchamp marriages or Bessette-Jetté marriages in a specified region? I will use marriage records from settlement to 1900 to compute the 'repeated pairs of names method' and the rate of same surname marriages (inbreeding coefficient).

The analytical strategy is based on the standard assumption that biases in selection of mates will yield a higher coefficient of repetitive marriages (more repeated groom-bride pairs) than we would expect if all individuals in this same defined population had a random chance (same probability) of marrying some other individual (less repeated groom-bride pairs). We can refer to this as discriminatory marriage selection (mating) based on personal choice as well as other population related restraints such as distance, geographical barriers and local and regional migration. It might involve personal choice, linguistic habits, ethnic preferences or religion, religious roles, or a sufficient level of consanguinity as determined by the Catholic Church. Here I am interested in testing for a distancedecay effect of the phenomenon of repeated-name marriages, and a movement from a high rate of non-random mating to an increase towards random mating levels, especially in the urban places like Montréal and Quebec City.

We would expect, from what we now know about the regionalization of names in Quebec, that we will have a higher rate of repeated marriage pairs during the eighteenth-century: a rate of matching names that steadily decreases as population size increases and the influx of English and Irish in the nineteenth-century helps increase the number of names in the pool and helps increase probability of non-French Canadian groom-bride name pairs. Locations in urbanized areas like Montréal, where there is a great diversity of names, will have a more diverse set of repeated groom-bride name pairs; values much closer to random, and more diverse than isolated rural regions, like the Saguenay with fewer migrants and more stable population.

The study is broken into two periods to accommodate the available marriage data. The first source is the set of marriages for the 832 single-founder names 1700 to 1799, and the second is the set of marriages for the 32 single-founder names 1800 to 1900. The eighteenth-century data were aggregated into 11 hexagons, and the nineteenth-century data were aggregated into 21 cultural regions defined by IREP (see chapter 2). Estimates of random-mating were created by randomizing the marriage partners. I shuffled the original data 100 times for each region and matched brides and grooms at each iteration. The median value for the region is used as the test value for random mating. The assumption of same-surname pairs is only partially met because the brides in my sample come from the larger pool of non-monolithic surnames (non-single founder populations).

Results/Discussion

The results of the eighteenth-century same-surname marriage and repeated-pair marriage analyses are consistent with the proposed hypothesis: we do see an increase in the overall regional same surname, and repeated-pair coefficients, but we also find geographic variability. Table 6.6 summarizes the coefficient (P) of same-surname marriages (inbreeding, i.e Beauchamp-Beauchamp) for the decades of 1750 through 1790. This value begins low and increases over the decades, indicating an increase in the probability of same surname marriages. This is a result of the increase in the relative population size, as well as of the laws of probability significantly increasing the number of some surnames and causing others to become extinct. During this time (1750-1799) we also find the overall calculation of same-surname marriages (F), the random and non random components the lowest of the decades calculated. These coefficients are indicative of a small population with a limited name regionalization. Some individuals were still arriving from France, and it would be some time before their male descendants would enter into the surname pool. The pattern appears to indicate near-random selection of mates, approaching panmixia. From the close networks and tight clustering of the population, I would argue that people were still marrying individuals from the same region (marriage field), but because there had been so little time for the population to produce large numbers of certain names, the chances of marrying an individual of same name remained small. There were not yet many people who shared the same name (Table 6.6).

In subsequent decades, we see a steady decrease in all the coefficients. Interestingly, we see a greater gap developing between the random and nonrandom coefficients of same-surname marriages. Again, as discussed above, rapid population growth due to natural increase, the regionalization of names, and the increase of certain names increased the likelihood of marrying an individual with the same name.

In terms of the random component, it may be misleading to examine regional patterns because there may be geographic variability. By calculating the random

component for each region separately, we are presuming a local marriage market (as demographers do in using sex ratios) and the size of the local pool of names of 'eligible' partners would vary from place to place. Same surname marriages were found throughout the province, but we observe the greatest proportion north of Trois-Rivières. The highest percentages are found in Lotbinière settled in 1697, Les Éboulements settled in 1733 and St-Nicolas settled in 1694. Gagnon also identified this region as having a high number of same surname marriages.⁵ The obvious conclusion is that parishes with greater diversity of names had fewer same-surname marriages.

We also find geographic patterns in the analysis of same-surname marriage pairs. Table 6.7 shows the random versus non-random coefficients for samesurname marriage pairs for 11 geographic regions (Figure 6.14). The regions having the greatest number of same-surname marriage pairs are Côte-de-Beaupré, Beauce and Bois Francs, while Montréal and Quebec City have the fewest. Are these numbers what we would expect from the sizes of the populations and the levels of diversity of name pools in the rural regions? To answer this question, I used the names from all of the marriages to simulate a set of marriage pairs at random in each of the 11 regions. Table 6.6 shows that ten of 11 locations have observed frequencies lower than the randomly generated values. The exception, case 52, is located north of Quebec City and is one of the regions with the low name diversity. These results confirm the notion of regionalization described by the same-surname marriage coefficient, and characterize the very local shortradius marriage fields at this time.

In the nineteenth century, the proportion of same surname marriages increases slightly between 1800 and 1860, and declines slightly thereafter (Table 6.8.). This trend seems to result from the growth and concentration of names in certain areas. The LaBelle surname, for example, concentrated in the Laurentides and Outaouais has two LaBelle-Labelle marriages between 1800 and 1830, and 11 in the next two periods (1830-1860, 1860-1890).

All 19 regions show decreasing rates of same-surname pairs. Table 6.9 shows the values for the 19 regions for three 30 year periods (about 3 generations) For

the first (1800-30) we find Montréal and Quebec City have the lowest proportions of same-surname marriages, a reflection of the great diversity of available names. Except in Côte-du-Sud, all 17 regions have values that approach random mating. For the second period (marriages 1830-1860) overall values have decreased, but several regions have values lower than the randomizations imply: Laurentides, Outaouais, Estrie, Bois-Francs, Côte-du-Sud, Charlevoix, Région de Québec. In the last period (marriages 1860-1890), values decline further, and almost all the locations have same-surname marriage pair coefficients close to the randomized calculations. These trends make sense in a population that is growing rapidly through both natural increase and immigration. The increasing mobility of individuals and extension of their marriage fields causes a trend toward random matches of surname pairs within the region.

Overall, this case study has demonstrated how the use of marriage records and same-surname methods can provide evidence of marriage patterns and inferred marriage fields; it has shown the tendency of a population to mate randomly or amongst those in closest proximity; and hdiversity and the regionalization of names plays in kin selection. When marriages are used to interpret population structure, it is important to utilize the time and geographic elements usually available in such a source to delve further into population trends as well as regional patterns of variation. As this case study demonstrates, in the development of surname analysis it is important to recognize that structural properties and stochastic processes operate in spite of all the personal, whimsical or idealized 'choices' we make as individuals. These underlying properties can be discovered from probability analyses despite the capricious nature of individual spellings and transcriptions.

NOTE

¹ This dataset was collected with only the standardized surnames. I did experiment with a few of the variants, but they most often did not produce large numbers of individuals. To complete this dataset, in the future, it would be possible to collect the variants of the single-founder names.

² The parishes which were included for Montréal and Quebéc City were determined from the *Inventaire* des registres paroissiaux catholiques du Québec 1621-1876. ³ For the City of Montréal, parishes within the counties (regions) of Hochelaga, Jacques-Cartier, and

Montréal were included and for Quebec City only the parishes within Quebec City.

As mentioned earlier there is often a lag between the time people settle a region and when the parish is officially opened.

⁵ For a discussion of Lotbinière and same-surname see Gagnon 2001.



Figure 6.3 Number of parishes represented in sample by decade (31 single founder surnames)

Table 6.2. Rank order of percent of each of the five surname clusters in Montréal and Quebec City

Rank/Surname Cluster	Montreal	Quebec City
1	Beauchamp	Laberge
2	Bessette	Gravel
3	Laberge	Vachon
4	Gravel	Bessette
5	Vachon	Beauchamp



Figure 6.4. Percent of total marriages from the 32 name sample occuring in Montréal and Quebec City



Cluster Names District Name	Beauchamp-Labelle Marriages per district	% Total	Bessette-Chicoine Marriages per district	% Total	Laberge-Cusson Marriages per district	% Total	Buteau-Gravel Marriages per district	% Total	Vachon-Bolduc Marriages per district	% Total
Abitibi	1	0.0373	1	0.1044	0	0.0000	0	0.0000	0	0.0000
Témiscaminque	10	0.3726	2	0.2088	0	0.0000	0	0.0000	1	0.0770
Outaouais	82	3.0551	4	0.4175	21	1,9535	25	8.2781	6	0.4622
Laurentides	274	10,2086	1	0.1044	19	1.7674	0	0.0000	2	0.1541
Rive Nord-Ouest de Montréal	383	14.2697	2	0.2088	25	2.3256	4	1.3245	6	0.4622
Lanaudière	228	8.4948	26	2.7140	86	8.0000	46	15.2318	18	1.3867
Montréal	1025	38,1893	100	10.4384	225	20.9302	32	10.5960	88	6.7797
Rive Sud de Montréal	38	1.4158	4	0.4175	96	8.9302	5	1.6556	3	0.2311
Richelieu	237	8.8301	479	50.0000	61	5.6744	13	4.3046	26	2.0031
Estrie	162	6.0358	73	7.6200	69	6.4186	28	9.2715	151	11.6333
Mauricie	16	0.5961	43	4.4885	17	1.5814	27	8.9404	16	1.2327
Bois-Francs	104	3.8748	205	21.3987	63	5.8605	22	7.2848	1	0.0770
Côte-de-Beaupré	12	0.4471	10	1.0438	30	2,7907	1	0.3311	17	1.3097
Beauce	5	0.1863	1	0.1044	61	5.6744	11	3.6424	348	26.8105
Côte-du-Sud	7	0.2608	1	0.1044	52	4.8372	21	6.9536	242	18.6441
Charlevoix	0	0.0000	4	0.4175	8	0.7442	1	0.3311	12	0.9245
Saguenay	19	0.7079	0	0.0000	31	2.8837	15	4.9669	89	6.8567
Côte-Nord	2	0.0745	0	0.0000	2	0.1860	0	0.0000	2	0.1541
Bas-Saint-Laurent	32	1.1923	0	0.0000	69	6.4186	6	1.9868	143	11.0169
Gaspésie	9	0.3353	0	0.0000	16	1.4884	4	1.3245	20	1.5408
Îles-de-la-Madeleine	0	0.0000	0	0.0000	0	0.0000	1	0.3311	0	0.0000
Région de Québec	25	0.9314	1	0.1044	76	7.0698	36	11.9205	95	7.3190
Agglomération de Québec	13	0.4844	1	0.1044	48	4.4651	4	1.3245	12	0.9245
Total Marriages	2684		958		1075		302		1298	

Table 6.3. Percent of the total marriages for each of the 5 clusters of single-founder surnames by region. (32 single-founder surnames, 1870-1900)

Table 6.4. Percent of the total phone users for each of the 5 clusters of single-founder surname 32 single-founder surnames 2002)

Cluster Name	%QC	%MB	%ON	%AB	%BC	%NF	%SK	%NB	%NS	%PE
Beauchamp-Labelle	72.00	1.15	20.95	2.70	1.50	0.02	0.45	0.13	0.16	0.02
Bessette-Chicoine	84.73	3.19	5.79	3.30	1.58	0.01	1.20	0.08	0.13	0.00
Laberge-Cusson	81.47	0.44	13.89	1.82	1.42	0.04	0.67	0.13	0.12	0.00
Buteau-Gravel	88.92	0.29	7.00	2.00	2.00	0.01	0.31	0.07	0.22	0.01
Vachon-Bolduc	87.12	0.49	7.02	2.76	1.46	0.01	0.39	0.60	0.00	0.00



Figure 6.6. Distribution of phone users with surname Vachon (Bell Canada, 2002)



Figure 6.7. Distribution of phone users with surname Beauchamp (Bell Canada, 2002)



Figure 6.8. Distribution of phone users with surname Laberge (Bell Canada, 2002)



Figure 6.9. Probable origins of migrants who settled in Maine before 1880



Figure 6.10. Probable origins of migrants who settled in New Hampshire before 1880



Figure 6.11. Probable origins of migrants who settled in Vermont before 1880



Figure 6.12. Geographic distribution of between population same-surname values (isonymy) for the population of single-founders (131 of 832) associated with the single-founder surnames of the 1861 timber population (832 surnames, marriage records: 1770-1799)



Figure 6.13. Geographic distribution of between population same-surname values (isonymy) for the population of all 1881 single-founders associated with the single-founder surnames of the 1861 timber population (221 of the total 1881 single-founder population).



Figure 6.14. Probability of finding a surname matching the 221 single-founder surnames defined in the 1861 Mauricie lumberjack population

Chapter 7

Conclusion

In this thesis, my aim has been to explore the use of surnames as markers for examining population structure and for analysing and mapping migration. As I come to the end of this review of surname methods, and their application to the French Canadian population, I leave you with commentary on the sources and methods employed and a summary of what we have discovered through their application. I will conclude with suggestions for applying surnames in other areas of historical migration research or as a surrogate variable for modelling changes in surname population structure.

This work was intended as an exploration of the various surname methods available and not as a condemnation of other methods by which we can also interpret migration. Methods such as linked family histories, for example, or longitudinally linked migration studies are superior and more accurate at pinpointing individual moves, but they require large amounts of genealogical data that do not exist for most places, and seldom can they get at regional patterns. Surname methods offer an option for inquiries at multiple spatial or temporal scales or for various geographic regions, as I have shown in this work. I have tested a variety of statistical methods, and from the chockful toolbox of methods, I have chosen the best tools for to confront different questions, datasets and models.

Quebec was used as the case study because its high-quality sources provide a near 'gold standard' for discovering what a variety of multivariate methods tell us about population thus surname structure and how we can use them to infer migration. Methods like the ones presented throughout have often times been applied in regions where little is known about the founding population or the data quality or have been tested on simulated data. The results achieved from the methods, I believe, have provided a fairly accurate representation of what we should expect from them and a measure of the amount of noise in data when we know little about the sample (e.g. multiple versus single-founders). For example,
the results of between population values were calculated with both a singlefounder population and a total population thus providing estimates of the extent to which other studies may have overestimated relatedness due to over counting popular names like Smith or Jones. I will eventually be able to employ comparative studies to determine the sensitivity and specificity of surname methods versus the extremely accurate linked family histories available for the French Canadian population.

The large number of primary sources that contain lists of surnames summarizing or defining populations makes surname methods attractive for biologists and anthropologists. Surname methods were originally conceptualized by biologists as a surrogate variable to imply transmission of the Y and infer family genealogies and the genetic structure of populations (Crow 1965; Lasker 1980, 1985). When the idea was formulated, there were few methods for inferring differences in genetic structures of populations, and names were a logical and powerful alternative to family histories, serologies (blood typing), protein electrophoresis, or anthropometric differences (e.g. skull size, body size). As exemplified by recent publication trends, surnames are still used as a variable in population studies, but the arrival of polymerase chain reaction (PCR) techniques (1985) and high-throughput genetics are helping biologists isolate genetic markers that are more accurate and specific than surnames. With cheaper access to genetic data by biologists, the number of surname studies has diminished somewhat, but names are often still used as a comparative measure alongside genetic findings (Balanovsky 2001; Roguljic, 1997; Koertvelyessy 1988). Cavalli-Sforza et al. (2004), in a recent book on the population structure of Italy, describes the utility of surnames in genetic studies. He uses surnames to describe random genetic drift and migration, and in a sampling strategy, he uses geographic clusters of surnames to pinpoint subpopulations from which his research group extracts genetic data (blood samples). Surnames still remain the quickest and most flexible alternative for investigating the historical trends in population structure.

The basic assumption of the isonymy method implies that men having the same name also carry a similar Y chromosome and are therefore related by

descent. This assumption has been the subject of controversy, with researchers deploring its tendency to overestimate population relatedness. Because of the importance of this assumption, scholars began to study the relationship between surnames and the molecular characteristics of the Y chromosome. Even newer methods have been developed which use polymorphic markers (on the nonrecombining portion of the Y chromosome and multiallelic microsatellites) to identify a lineage that can be associated with a surname (Jobling 2001). Many of these studies are working from the assumption that genetic variability within the Y chromosome will also show geographic structure; meaning, certain markers on the Y chromosome should be unique to a time and place. These results suggest that the basic assumption of isonymy may be valid or at least more accurate than many have suspected (Colantonio et al. 2003).

The late Gabriel Lasker and his colleagues did a thorough job describing the many other ways surnames can be interpreted to describe migration histories, population and distance models, and the regionalization of names. It was his book 'Surnames and Genetic Structure' and his atlas of British surnames that brought attention to the geography of name distributions, and to the fact that names which most of us think of as ubiquitous, found in equal proportions in all places, are in fact not. In studies of Great Britain, Lasker has shown that even common names like Jones and Smith have unique distributions (Lasker 1980, 1983, 1985). For Quebec, this study has shown that French Canadian names are organized in three distinctive ways. First, there are names like Beauchamp, Bessette or Tremblay which dispersed around one or two close-by locations and their distributions resemble that of a dispersion of a disease outbreak or diffusion of culture; with the greatest name concentrations in a central location and a decreasing gradient with an increasing distance from that central node. Second, there are names like Laberge that in the seventeenth century separated some distance and dispersed around two or more growth poles. Finally, there are names that tended to have no uniquely defined structure and appeared in random clumps throughout the province. This group appeared to be related to the cluster that had little affinity with any one area (e.g. see Drouin cluster, chapter 4).

Care should be taken when selecting a source and an investigative method. This is important because of the sheer number of sources that are becoming available through electronic means. A few decades ago, it would have been unthinkable, in a short period, to collect all the individuals and their names from the Quebec 1881 census, but today we have an electronic version of this census, as well as samples of others, and access to censuses and marriage records for other countries. Integrating nominal sources with statistical computer languages like R (Splus) provides a vehicle to complete surname analyses like those I have presented, zooming to various scales (or level of aggregate) in a modest amount of time. With access to such large amounts of data, new information can be derived about migration paths and historical trends in a large number of locations. Yet care needs to be taken throughout the process of sampling and analysis in order to ensure meaningful results. In this study, I have shown the advantages to be derived from careful selection of a handful of sources and a selection of names.

An important consideration when collecting surname data for a study of population structure is to have knowledge of when the list was collected, how well the name list represents the region or parish under investigation, and whether the names are spelled correctly. If a random sample of names is used, care should be taken to ensure the geographic relevance, and tests should be performed to test whether a second random sample produces similar results. It might also be appropriate to remove singletons, since the majority of names are very common. In temporal studies, one must ensure that boundaries remain the same so that results can be accurately compared. If the study, for example, involves using lists of names, it is possible to compensate for disparities in the sources. I have experimented with spell checking rules, and the application of hexagon aggregation resolved the problem of changing administrative boundaries, so that I was able to substitute one source for another. In particular, I created a surrogate 1765 census from baptismal records because the original was flawed (Landry 1975).

The multiple origins of names and the complication of spellings have drawn much criticism of surname methods because of implicit assumptions that each

name arose from a single origin or genealogical line. The name Smith or Jones for example, frequent in many English speaking populations, might be treated as monolithic in a 'mindless' application of surname equations. Such popular names can bias results and overestimate relatedness of populations. In my studies of Quebec, therefore I specifically selected single-founders (see chapter 2). This ensures that name relatedness between two locations can be reasonably attributed to genealogical links. In some instances it is impossible to meet this assumption, but this study provides a framework from which I can begin to develop and test the extent to which we overestimate when we assume monolithic (multiple) name origins. For 1765 and 1881, if I compare my samples of single-founders with the full populations (all surnames), I obtain a strong correlation in the local and between population similarity values, but higher overall values for the singlefounder samples, indicating the populations are more closely related. This suggests that we can employ the method, but we need to recognize the weakness of the absolute values obtained, and we need to know something generally about the 'pioneering' phases of the region under study, as indeed we do for much of North America. Generally, researchers have recognized the problems of absolute values, particularly when comparing regions with no known historical links.

Attempts have been made to estimate the extent to which name variants and spelling mistakes can skew results of same-surname analysis. Souden and Lasker (1978), for example, in a study of East Kent, England, examined the effects of merging surname variants (e.g. Read or Reid) concluded that the absolute values will differ, but the overall rank order and geographic concentrations remain much the same. These results raise issues in regards to the comparisons researchers have made from studies conducted in different regions. I did not have to worry about correcting surname spellings for the pre-eighteenth century Quebec sample because this was already completed by the PRDH, but for the 1881 census sample all of the French Canadian names had to be standardized to match the list of single-founders. During this process the non-French Canadian names were also corrected using various North American and Canadian name databases. These corrections favour good results, but it can be argued that two researchers would

occasionally generate somewhat different 'solutions'.

Recent work has focused on providing algorithms or automated solutions for correcting variants which result from spelling errors, or aggregating together names having a common lineage that had changed at some point in time (e.g. bessette, besette). Palgrave (2004) examined the linguistic aspects that result in name variants, and provided a methodological strategy for recognizing them. Christian (2002) examined name variants caused by speech style, dialect variation, and phoneme rules, with a better way to match variants of the same name. Research in this area will obviously help in the process of dealing with name variants and will result in more accurate and meaningful results in surname studies.

Sample size is an important consideration in surname studies. Throughout this study a single-founder French Canadian population was used, which amounted to a sub-sample of the total population of the order of 26 % of the unique surnames or 30% of the population of French Canadians in 1765. When tracked spatially and numerically against the full population, they match well. Lasker (1980) argued a preference for large samples, but there is little information on what exactly the size should be. I can suggest that a random sample of names from older populations will not approach a normal or Gaussian distribution, but most often will have a distribution curve similar to those presented in chapter 3 (Figure 3.1). This curve summarizes a distribution made up of a few common names like Smith or Tremblay and a very large number of rare names.

There are also important considerations in preparing and organizing a geographical analysis of name samples (or populations). It is important to check the scales at which the data are available. It is not always ideal to study name structure at the finest scale available because some populations may be too small for statistical reliability. Population totals and surname distributions should be examined for each location, and populations that are too small should either be aggregated, removed from the analysis, or examined non-statistically. Of course the scale of inquiry will be greatly influenced by the questions being pursued. If a researcher, for example, is inquiring about marriage fields or the effect of distance

on surname similarity, data aggregated at the county scale would mask these movements. In the 1881 Quebec data, for example, we found little evidence of distance decay of name similarity at the hexagon scale (40km); yet we know from the work of St Hilaire (2002) that these short-distance marriage fields were important.

The aggregate level in which analysis is performed is another important concern in isonymy and genetic studies which have important implications on numerical results. Jorde (1980) explained that values of Fst (or Rst) decrease as the size of the geographic aggregate containing the subpopulation increases in size (e.g. state, county, subarea) resulting in a larger population used in the calculation. We saw a similar decrease of Fst values for 1881 Vermont when we moved from the hexagon to county aggregate. This variability in Fst values can have important implications in comparative studies. It might be helpful to have relative Fst estimates of what we should expect using different size population. Simulated population data could be used in the calculation.

Throughout this work, I reviewed and applied five families of methods: 1) same-surname measures or isonymy, kinship and genetic distance, 2) ordination methods such as correspondence analysis and canonical correspondence analysis, 3) hierarchical and fuzzy clustering, 4) relevance networks, and 5) Bayesian probability. Each method is better at answering certain questions, and we often need to confirm findings by using more than one method, by applying the technique at several spatial scales, or by using samples of different sizes. In the following paragraphs, I will describe the pros and cons of each method noting considerations that may improve results in the future.

Same-surname methods (isonymy) have been applied extensively worldwide, and studies range from local inquiries to large provincial or regional studies (Colantonio et al. 2003). In the methods section, I presented them as a suite of mathematical equations which measure micro-differentiation, inbreeding, kinship and genetic distances within a population or between populations. Though each offers a different measure of relatedness, all are based on the same concept of establishing probabilities of surnames being shared within or between populations.

In other words, surname similarity is a probabilistic method, where if we pick two individuals at random (one from each population) the chances that they will share the same name are contingent on the numbers of people of each name found in the two populations. If, for example, there are many Tremblays in both populations, then we have a greater chance of picking two Tremblays than two of some rare name.

This probabilistic concept of randomness is the theoretical foundation of how researchers conceptualize marriages (mate selection) and name distributions in population structure studies. The assumption of randomness, as we learned from the Quebec study, is almost never fully met due to demographic, cultural, religious and administrative influences, as well as distance and geographic barriers. The funnelling of marriages (names) and migrations in certain directions and towards certain places is part of the process which has generated and continues to change the structure of the Quebec population.

Same-surname (isonymy) methods were used for the eighteenth-century and 1881 census data and provided the basis for comparing and calibrating the other methods. Using this method with my specifically selected sample of singlefounders, I am fully convinced of its value in determining surname relatedness within and between populations. Though I took a broad approach looking only at major patterns, the method is capable of specifying the relatedness among all the locations, and it could be easily completed at the finest scale (parishes), aggregated into hexagons of larger sizes, or applied (as in the case of the lumberjacks) to test for the relatedness of one population with a number of others.

I found the same-surname method very flexible, and the distance matrix between parishes provides a useful metric either for visual comparisons using maps or statistical comparisons using multidimensional scaling, regression or the Mantel test. A possible approach that would have added more specificity to migrant flows is the tracking of parish opening dates and the examination which names and how many showed up in those parishes.

The multivariate methods used throughout provided less specific and less valuable information for estimating population relatedness, but a superior capacity

to get at the overall geographic population structure. These methods refer to a variety of (eigenanalysis) techniques that reduce the dimensionality of data and describe the proportion of variance associated with each orthogonal component of the input data matrix. In the methods chapter, I reviewed statistical considerations, demonstrating the use of detrended correspondence analysis and canonical correspondence analysis.

Multivariate methods have been little used for exploring surname population structure. An exception is Degionni et al. (2001) who used correspondence analysis to compare the extent to which Flemish patronyms are distributed in various French départements. It is surprising that multivariate methods have not been used extensively because they are extremely popular in ecology and biology for studying the structure and organization of plant and animal populations.

Although I did not use the method in the Quebec study, my tests on Vermont and earlier tests conducted on Quebec data (conference paper), convinced me of the value of using this and other multivariate methods as tools to initiate exploration. In the Vermont case, the first and second axes of the correspondence analysis revealed the same overall east-west regional name trends found with same-surname values (isonymy) and cluster analysis. When using correspondence analysis, it is important to remove singletons (rare names occurring once) and other names occurring only a few times because of the tendency of these low numbers to distort results. In the classic frequency distributions of names, the driving force which determines population structure (not diversity) is the set of names that occur in moderate numbers (the middle of the distribution). Rare names do little to uncover the overall hierarchical name structure.

Future endeavours using multivariate statistics might include tests of specificity and sensitivity of these techniques in determining the geographic organization of names and population relatedness. Eigenanalysis provides dimension scores for each name and place, and the first few dimensions account for most of the variation. If these values were standardized, would the relative statistical distances between pairs of the scores correlate with the distances obtained from between population same-surname techniques? When the first two

axes are plotted, is the number of names that hover around a place indicative of its diversity? Because I have same-surname (isonymy) results for a 'gold standard' dataset for Quebec, I will be able to test these questions and look for further applications of correspondence beyond its use as an exploratory tool.

Cluster analysis has been surprisingly little used in the examination of the geographic structure of names. Schurer (2004) in his article "Surnames and the search for regions," showed the utility of cluster analysis for uncovering similar regions in England. After its use in the Vermont and Quebec study, I am convinced of the strength of this method for assessing distributions of names, but I have some reservations and some suggestions to maximize the amount of information extracted. Cluster analysis has the advantage of producing very simplified results as compared with same-surname (isonymy) methods and correspondence analysis, but the simplification can lead to places and names being placed into clusters which are not the optimal partners. Deciding on the number of clusters for the names and places can be difficult to determine without priori knowledge about the name structure. For the Quebec analysis, I did have prior knowledge from other studies, as well as from the same-surname analysis and correspondence analysis. For the eighteenth-century Quebec names, I calculated a participation rate for each cluster, to tally the percentages of all names in a cluster belonging to a set of place clusters. This works on the assumption that good clusters produce participation rates that are either very high or very low. The method worked satisfactorily, but did not provide a way to test the statistical significance of the clusters, a major complaint by researchers. Re-sampling (bootstrapping) cluster methods has recently been developed in bioinformatics to remove faulty clusters and optimize cluster selection. I will consider these new techniques in future work.

Clustering was most informative when I observed the suite of clusters (or stages in a clustering process) and mapped the successive patterns. This technique was performed on the eighteenth-century marriage data, and it provided a view of the hierarchical structure that is so hard to obtain with the other methods. I began with a large number of clusters and as the number was reduced we could see

certain places with a strong name affinity 'joining' the cluster. This application provided a visual display of what population geneticists and biologists refer to as the stepping-stone model. This model reveals the leaps or moves from one population to another, and the hierarchical clustering method provided a simple and fairly dependable way to probe the hierarchical structure, an important property recognized in the literatures of genetic and population biology.

Relevance networks is a method developed to analyze correlations among variables in large data sets and to create networks of related variables. Little has been published so far, and the method has never been applied in surname studies. My applications to eighteenth-century and 1881 single-founder Quebec data extracted almost exactly the strongest set of connections between places as determined from the same-surname methods. Names were also analyzed, and the connections revealed by relevance networks proved to have well-defined geographical or spatial properties. The strength of this method is that which its name implies: locating the most relevant connections in large datasets. We are able to change the sensitivity of the connections (the threshold of networks we want to connect), but we do not yet have a way to summarize numerically the relative strengths of connections. If this were possible, the method might provide a broader and more informative interpretation of the population like that of same-surname methods. Therefore, the method at best provides a way to 'mine' an array of data for the strongest connections between names and places.

The final method used in this study was Bayesian probability for inferring the origins of migrants into a recipient area from a list of surnames (Degioanni and Darlu 2002). The method requires extensive knowledge of both the receiving and originating populations so that linkages and migrations are correctly defined. I used the Bayesian method to answer two questions. The first involved trying to evaluate the extent of the movements of French Canadians from various part of Quebec to Vermont, New Hampshire and Maine (see Chapter 5). In the second case I used the Bayesian method to identify origins of single-founder names present in the 1861 Mauricie timber shanties. In the first case, the method provided compelling evidence of different origins of migrants. In the case of the

lumberjack population, it also identified regions where the names were most populous, but the same-surname method gave better results because of its ability to deal with large places, whereas the Bayesian method overestimated relationships to large urban places with greater diversity and numbers of names. To avoid such estimation in the analysis of French Canadian migration to New England, I removed Montréal from the calculation.

Each of the methods was useful, but they are not equally useful for answering certain questions. Before a surname study is conducted, it is important to define the question and have a broad understanding of the region's population history, to consider the options of spatial and temporal scales and to understand how each method is affected by sample size and the monolithic origins of names. I intend to continue testing these methods as well to explore others. The French Canadian single-founder population dataset which accurately approximates true genealogical lineages provides an ideal 'gold standard' dataset to test the sensitivity and specificity of statistical methods. As Gagnon (2001) has done in his same-surname (isonymy) studies, it would be beneficial to test the methods against the fully linked genealogical databases available for the entire French Canadian population from settlement to today.

Two areas in which we might expand and apply surname methods are genetic studies and migration history. Genetic studies have already benefited from the use of surname methods to describe and isolate populations at a selected time. Roguljic et al. (1997), who used surnames to estimated inbreeding, kinship, and population structure on the island of Hvar, Croatia, found a moderately close relationship in the results of names and migration, sociocultural, anthropometric, physiological, dermatoglyphic, and genetic traits. For Russia, Balanovsky et al. (2001) found that the trend for 75 popular names coincided with the latitudinal trend of principal components for genetic and dermatoglyphical data. They concluded that "similarity between the main scenarios for the genetic markers and surnames demonstrates the effectiveness of the use of surnames for analysis of the gene pool." For Quebec there has been less emphasis on the comparison of surnames and genetic markers, and more on the use of the names as a descriptive measure.

As there are ongoing genetic studies in Quebec and excellent demographic and genealogical sources, it would be an ideal population on which to test some of the assumptions about genetics and surnames and calibrate their comparative potential (see Scriver, 2001).

A second avenue of research for the use of surname models is the study of historical migrations. Most migration is one of three types: local or regional, intrastate, or global migration. My work on the French Canadians has demonstrated the flexibility of using names to track migratory movements at different scales; in other words, we can build bridges between the conventional compartments or scale of analysis. For Quebec, using the surname methods (isonymy), I inferred local moves of the lumberjack population, and the probable locations of migrants moving to New England. Broader patterns were inferred by aggregating data into hexagons, and long- distance moves were inferred from modern-day phone records. Using the same sets, I was able to track the history of their movements from one century to the next and to observe trends and movements representative of the entire population descended from 'French Canadian' founders.

The simplest way to track migration is to sequentially document the appearance, disappearance and persistence of names. When following such a procedure, care must be taken to ensure that boundaries remain the same at areas of interest. Aggregating in this situation can help assure all of the areas of interest are included. GIS software is well-tested in this respect.

Names can also be applied to more specific questions. For simplicity let us pose a few typical questions that are of interest to historical geographers: Do most migrants move only short distances? Does the volume of migration increase with the expansion of industry and commerce? Is there a distance decay resulting from intervening locations or opportunities? Are the distances and directions that migrants travel a function of a hierarchy of central places? In the Quebec studies I was able to answer these questions using the surnames only. The relevance networks, for example, which connect the most strongly related places based on name affinity, show that a majority of migrants moved short distances. The set of marriage records for 32 names showed a steady increase in the percentage married

in Montréal after 1850. This set of names also provided evidence that the largest percentage of the population moving to Montréal in the nineteenth century came from a regional migration field to the north and west.

Surnames can be applied to a suite of other population models. The idea of using surname methods to provide evidence of central place theory is perhaps the most natural because of the geographic distribution of names which form different hierarchies. Measures of name relatedness over time can provide raw material for modelling the interactions between places. Greater surname relatedness can then be inferred as the flows between central places. This type of question could be better formalized in the context of graph theory where nodes and arcs provide the means by which we quantify the relationships (pushes and pulls) between places. The importance of connectivity is evident throughout the work on Quebec, where connectivity was associated with diversity and changes in access led to changes in name relatedness. The importance of connectivity suggests applicability of other types of population models such as gravity models, spatial interaction models, simulation models, diffusion models (Brownian movement), and dispersion models.

Although there are limitless areas of inquiry from which surnames can be applied, we need to be wary of the problems and assumptions when using surnames. A study must be carefully planned in terms of: sampling, regional aggregation and periodization. With careful planning surname studies will certainly provide a broad brush, labour-saving methodology that can overcome restrictions of spatial and temporal scale, to yield a broader, more systematic and balanced picture of migration. This will in turn provide a means to design more detailed inquires which might require and justify genealogies or origin and destination surveys.

Bibliography

- Abrahamse, A. F., P. A. Morrison, and N. M. Bolton. 1994. Surname Analysis For Estimating Local Concentration of Hispanics and Asians. *Population Research and Policy Review* 13 (4):383-398.
- Adams, John W., Alice Kasakoff and Jan Kok 2002 Migration over the life course in XIXth century Netherlands and the American North: A comparative analysis based on genealogies and population registers. *Annales de Démographie historique* no. 2, 5-27
- Agnar, Helgason, Eileen Hickey, Sara Goodacre, Vidar Bosnes, Kári Stefánsson, Ryk Ward, and Bryan Sykes. 2001. mtDNA and the Islands of the North Atlantic: Estimating the Proportions of Norse and Gaelic Ancestry. *American Journal of Human Genetics* 68:723-737.
- Albers, Jan. 2000. Hands on the Land: A History of the Vermont Landscape. Cambridge Massachusetts, MIT Press
- Aldenderfer, Mark S, and Roger K Blashfield. 1984. *Cluster Analysis, Quantatitative Applications in the Social Sciences*. London: Sage Publications.
- Allen, Rogers. R. 1988. Three Components of Genetic Drift in Subdivided Populations. American Journal of Physical Anthropology 77:435-449.
- Ashley, David J.B., and H Duncan Davies. 1966. The Use of the Surname as a Genetic Marker in Wales. *Journal of Medical Genetics* 3:203-211.
- Arner, G.B.L. 1908. Consanguineous Marriages in the American Population. Vol. 31, Columbia University Studies in History, Economics and Public Law. New York: Longman Grenn and Co.
- Balanovskaya, E. V., E. A. Pocheshkhova, O. P. Balanovskii, and E. K. Ginter. 2000.
 Gene-geographic analysis of a subdivided population: 2. Geography of random inbreeding based on surname frequencies in Adygs. *Russian Journal of Genetics* 36 (8):936-948.
- Balanovsky, O.P., A.P. Buzhilova, and E.V. Balanovskaya. 2001. The Russian Gene Pool: Gene Geography of Surnames. *Russian Journal of Genetics* 37 (7):974-990.
- Barbosa, A. A. L., I. J. Cavalli, K. Abe, M. G. Santos, and E. S. Azevedo. 1997. Family names and the length of the Y chromosome in Brazilian blacks. *Brazilian Journal of Genetics* 20 (1):93-96.
- Bardsley, Alan. "How many Smiths are there?" *Journal of One-Name Studies* 5, no. 10 (April 1996): 308-10.
- Barrai, I., A. Rodriguez-Larralde, E. Mamolini, F. Manni, and G. Scapoli. 2001. Isonymy structure of USA population. *American Journal of Physical Anthropology* 114 (2):109-123.
- Barrai, I, C Scapoli, M Beretta, C Nesti, E Mamolini, and A Rodriguez Larralde. 1996. Isonymy and the genetic structure of Switzerland .1. The distributions of surnames. *Annals of Human Biology* 23 (6):431-455.
- Barrai, I., A. Rodriguez-Larralde, E. Mamolini, F. Manni, and C. Scapoli. 2000. Elements of the surname structure of Austria. *Annals of Human Biology* 27 (6):607-622.

- Barrai, I., A. Rodriguez-Larralde, E. Mamolini, F. Manni, and G. Scapoli. 2001. Isonymy structure of USA population. *American Journal of Physical Anthropology* 114 (2):109-123.
- Barrai, I., A. Rodriguez-Larralde, E. Mamolini, and C. Scapoli. 1999. Isonymy and isolation by distance in Italy. *Human Biology* 71 (6):947-961.
- Beddoe. 1883. English Surnames, from an Ethnological Point of View. The Journal of the Anthropological Institute of Great Britain and Ireland 12:231-242.
- Beddoe, John. 1898. On Complexional Differences Between the Irish with Indigenous and Exotic Surnames Respectively. *Journal of the Anthropological Institute of Great Britain and Ireland* 27:164-170.
- Bélanger, Pauline et Yves Landry, avec la collaboration de René Jetté (1990), *Inventaire des registres paroissiaux catholiques du Québec 1621-1876*, Montréal, Presses de l'Université de Montréal, 354 p.
- Bideau, Alain. 1997. La reproduction de la population aux XVIIe, XVIIIe et XIX siécles: exemples français et québécois. *Annales de démographie historique*:137-148.
- Biondi, G., P. Raspe, and C. G. N. Mascie-Taylor. 2000. Genetic structure through surnames in Campobasso Province, Italy. *Journal of Biosocial Science* 32 (4):459-465.
- ———. 2001. Migration through surnames in Campobasso Province, Italy. *Journal of Biosocial Science* 33 (2):305-310.
- Biondi, G., P. Raspe, C. G. N. MascieTaylor, and G. W. Lasker. 1996. Repetition of the same pair of surnames in marriages in Albanian Italians, Greek Italians, and the Italian population of Campobasso province. *Human Biology* 68 (4):573-583.
- Blanchard, Raoul 1950 La Mauricie, Trois-Rivières, Editions du Bien Public, 154pp.
- Bittles, A. H., and J. J. McHugh. 1986. The Irish Famine and Its Sequel Population-Structure Changes in the Ards Peninsula, Co Down, 1841-1911. *Annals of Human Biology* 13 (5):473-487.
- Bittles, A. H., and M. T. Smith. 1994. Religious Differentials in Postfamine Marriage Patterns, Northern-Ireland, 1840-1915 .1. Demographic and Isonymy Analysis. *Human Biology* 66 (1):59-76.
- Boisvert, Michel 1996, « Migrations des francophones, 1652-1990: l'exemple des familles Boisvert », in *Population et Territoire*. Serge Courville et Normand Séguin (dirs.), Sainte-Foy, Presses de l'Université Laval, p. 136-137
- Boleda, Mario 1984, « Les migrations au Canada sous le régime français 1608-1760 », *Cahiers québécois de démographie* 131, p. 23-39.
- Boldsen, J., and G. W. Lasker. 1996. Relationship of people across an international border based on an isonymy analysis across the German-Danish frontier. *Journal of Biosocial Science* 28 (2):177-183.
- Bouchard, Gérard et al. 1985 La distribution des patronymes au Québec: témoins des dynamiques de population, *Anthropologie et sociétés* 9, 3, 197-218
- Bouchard, Gérard 1987, « Sur la reproduction familiale en milieu rural: systèmes ouverts et systèmes clos », *Recherches sociographiques* 28, 2-3, 229-251
- Bouchard, Gérard et al. 1987 La statistique agrégée des paronymes du Saguenay et de Charlevoix comme indicateurs de la structure de la population aux XIXe et XXe

siècles. Cahiers québécois de démographie 16, 1, (avril), 67-98

- Bouchard, Gérard, and Marc De Braekeleer. 1991. *Histoire d'un génôme : population et génétique dans l'est du Québec*. Sillery: Presses de l'Université du Québec.
- Bouchard, Gérard. 1992. Current Issues and New Prospects for Computerized Record Linkage in the Province of Quebec. *Historical Methods* 25 (2).
- Bouchard, Gérard 1996, Quelques arpents d'Amérique, Population, économie, famille au Saguenay, 1838-1971, Montréal, Boréal
- Bouchard, Gérard, et Richard Lalou 1993, « La surfécondité des couples québécois depuis le XVII^e siècle, essai de mesure d'interprétation », *Recherches sociographiques*, 34, 1, 9-44
- Bouchard, Gérard, Hubert Charbonneau, Bertrand Desjardins, Evelyne Heyer et Marc Tremblay (1995) « Mobilité géographique et stratification du pool génique canadien-français sous le Régime français », dans Yves Landry et al. (dirs.), *Les chemins de la migration en Belgique et au Québec du XVII^e au XX^e siècle*, Louvain, mnh / Academia, p. 51-59.
- Bouchard, Gérard, and M. DeBraekeleer 1990 Mouvements migratoires, effets fondateurs et homogénéisation génétique. In G. Bouchard and M. DeBraekeleer, eds. *Histoire d'un génôme, Population et génétique dans l'est du Québec*. Sillery, Presses de l'Université du Québec, 281-322
- Bouchard, Gérard and André Larose 1976 La réglementation du contenu des actes de baptême, mariage, sépulture, au Québec, des origines à nos jours. *Revue d'histoire de l'Amérique française* 30, 1 (juin), 11-20
- Boyce, A.J. 1984. *Migration and mobility: Biosocial aspects of human movement*. London: Taylor and Francis.
- Boyce, R.L. 1998. Fuzzy set ordination along an elevation gradient on a mountain in Vermont. *Journal of Vegetation Science* 9 (2):191-200.
- Brunet, Guy, Pierre Darlu, and Gianna Zei, eds. 2001. Le Patronyme : Histoire, anthropologie, société. Paris: CNRS EDITIONS.
- Butte, AJ, et al. 1999. Unsupervised Knowledge Discovery in Medical Databases Using Relevance Networks. Proceedings of the American Medical Informatics Association Symposium, 711-715
- Butte, AJ, et al. 2000. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proceedings of the National Academy of Science* USA, 97(22):12182-12186
- Caravello, G., and M. Tasso. 1999. An analysis of the spatial distribution of surnames in the Lecco area (Lombardy, Italy). *American Journal of Human Biology* 11 (3):305-315.
- Cavalli-Sforza, L L, and W.F. Bodmer. 1991. *The Genetics of Human Populations*. Mineola, New York: Dover Publications Inc.
- Cavalli-Sforza, L L, Paolo Menozzi, and Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton, New Jersey: Princeton University Press.
- Cavalli-Sforza, Luigi Luca, and Mark Seielstad. 2001. Genes, Peoples, and Languages: University of California Press.
- Cavalli-Sforza, L.L Antonio Moroni, Gianna Zei. 2004 Consanguinity, Inbreeding, and Genetic Drift in Italy (MPB-39) (Monographs in Population Biology) Princeton University Press.

- Charbonneau, Hubert, and N. Robert. 1987. Origines françaises de la population canadienne, 1608-1759. In *Atlas historique du Canada v. I, Des origines à 1800*, edited by R. C. Harris and L. Dechêne. Montréal: Presses de l'Université de Montréal.
- Charbonneau, Hubert and N. Robert 1987 Origines françaises de la population canadienne, 1608-1759. In *Atlas historique du Canada* v. I, Des origines à 1800, R.Cole Harris and L. Dechêne eds. Montréal, Presses de l'Université de Montréal, planche 45
- Charbonneau, Hubert, Bertrand Desjardins, Jacques Légaré et Hubert Denis (1996), « La population française de la vallée du Saint-Laurent avant 1760 », in Serge Courville et Normand Séguin (dirs.), *Population et Territoire*, Sainte-Foy, Presses de l'Université Laval, p. 31-43.
- Charbonneau, Hubert, Bertrand Desjardins, Andre Guillemette, Mario Boleda, and Real Bates. 1993. *The First French Canadians: Pioneers in the St. Lawrence Valley:* University of Delaware Press.
- Cheng, T. O. 1994. Use of Surnames to Identify Individuals of Chinese Ancestry. *American Journal of Epidemiology* 140 (1):76-76.
- Choquette, Leslie P. 1988. French Canadian Emigration to Canada in the 17th and 18th Centuries. Dissertation, History, Harvard, Boston.
- Choquette, Leslie P. 1997, Frenchmen into Peasants: Modernity and Tradition in the Peopling of Canada, Cambridge, Mass., Harvard University Press.
- Christensen, A. F. 1999. Population relationships by isonymy in frontier Pennsylvania. Human Biology 71 (5):859-873.
- ———. 2000. An isonymic study of the population structure of early Kings County, NY. *Human Biology* 72 (6):1017-1037.
- ———. 2000. Isonymy in Kings County, NY, 1698-1810. American Journal of Physical Anthropology:124-124.
 - ------. 2001. Calculation of isonymy from census and marriage records: A comparison from antebellum Georgia. *American Journal of Physical Anthropology*:50-50.
 - ———. 2001. Calculation of Isonymy from Census and Marriage Records: A Comparison from Antebellum Georgia. *American Journal of Physical Anthropology, Supplement* 32 (50).
- Christen, Peter, Tim Churches, and Justin Xi Zhu 2002 Probabilistic name and address cleaning and standardisation. Proceedings of the Australasian Data Mining Workshop, December, Canberra, Australia
- Christian, Peter. 2002. Surname variation and surname matching algorithms. In *Thesaurus* of British Surnames: INITIAL DRAFT.
- Christian, Peter. 2002. What is a surname variant? In Thesaurus of British Surnames.
- Clausen, Sten-Erik. 1998. Applied Correspondence Analysis: An Introduction, Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage Publications Inc.
- Clegg, E. J. 2001. Some factors affecting marital isonymy in three Outer Hebridean parishes. *Annals of Human Biology* 28 (4):455-466.
- Collard, Chantal 1990 Idéologie et pratique de la parenté, de la classification des parents aux stratégies familiales. In G. Bouchard and M. DeBraekeleer, eds. *Histoire d'un* génôme, Population et génétique dans l'est du Québec. Sillery, Presses de

l'Université du Québec, 121-144

- Collard, Chantal 1999, Une famille, un village, une nation : la parenté dans Charlevoix, 1900-1960, Montréal, Boréal.
- Cohen, R. ed. 1995 *The Cambridge survey of world migration*. Cambridge, Cambridge University Press
- Colantonio, Sonia, Gabriel Ward Lasker, Bernice A. Kaplan, and Vicente Fuster. 2003. Use of Surname Models in Human Population Biology: A Review of Recent Developments. *Human Biology* 75 (6):785-806.
- Cox, Trevor F., and Michael A. A. Cox. 1994. *Multidimensional scaling*. London: Chapman & Hall.
- Cox, Thomas R, Robert S Maxqell, Phillip D Thomas, and Joseph J. Malone. 1985. This Well-Wooded Land, Americans and Their Forests from Colonial Times to the Present. University of Nebraska Press.
- Courville, Serge 1990, Entre ville et campagne, L'essor du village dans les seigneuries du Bas-Canada, Sainte-Foy, Presses de l'Université Laval
- Courville, Serge, ed. 1996. Atlas Historique du Quebec: Population et Territoire. Laval: Les Press de l'Universite Laval
- Crow, J.F. 1983. Discussion (Isonymy: Special Issue). Human Biology 55:383-397.
- Crow, J.F., and A.P. Mange. 1965. Measurement of Inbreeding from the Frequency of Marriages between Persons of the Same Surname. *Eugenics Quarterly* 12:199-203.
- Cuzick, J, and R. Edwards. 1990. Spatial clustering for inhomogeneous populations. Journal of the Royal Statistical Society 52:73-104.
- Dale, Mark R.T. 1999. Spatial Pattern Analysis in Plant Ecology, Cambridge Studies in Ecology. Cambridge: Cambridge University Press.
- Danubio, M. E., and E. Amicone. 2001. Biodemographic study of a central Apennine area (Italy) in the 19th and 20th centuries: Marriage seasonality and reproductive isolation. *Journal of Biosocial Science* 33 (3):427-449.
- Danubio, M. E., and D. Pettener. 1997. Marital structure of the Italian community of Boston, Massachusetts, 1880-1920. *Journal of Biosocial Science* 29 (3):257-269.
- Danubio, M. E., A. Piro, and A. Tagarelli. 1999. Endogamy and inbreeding since the 17th century in past malarial communities in the Province of Cosenza (Calabria, Southern Italy). *Annals of Human Biology* 26 (5):473-488.
- Darwin, George H. 1875. Marriages Between First Cousins in England and Their Effects. Journal of the Statistical Society of London 38 (2):153-184.
- Darlu, P., A. Degioanni, and J. Ruffie. 1997. Some statistics on the distribution of surnames. *Population* 52 (3):607-634.
- Dean, K. C. 2000. In search of your Asian roots: Genealogical research on Chinese surnames. *Library Journal* 125 (9):80-80.
- DeBraekeleer, M. 1990 Consanguinité et parenté au Saguenay. In G. Bouchard and M. DeBraekeleer, eds. *Histoire d'un génôme, Population et génétique dans l'est du Québec*. Sillery, Presses de l'Université du Québec, 323-342
- DeBraekeleer, Marc, et To-nga Dao (1994), Hereditary disorders in the French Canadian population of Quebec, *Human Biology*, 66, 2, 205-249.
- De Guerra, D. C., H. Arvelo, and J. Pinto-Cisternas. 1999. Population structure of two black Venezuelan populations studied through their mating structure and other related variables. *Annals of Human Biology* 26 (2):141-150.

Degioanni, A., A. Lisa, G. Zei, and P. Darlu. 1996. Italian surnames and Italian migration to France 1891-1940. *Population* 51 (6):1153-1180.

Dulong, John P. 2001. French Canadians in Michigan: Michigan State Univ Press.

De Silvestri, A., and C. R. Guglielmino. 2000. Ethnicity and malaria affect surname distribution in Cosenza Province (Italy). *Human Biology* 72 (4):573-583.

Desjardins, Bertrand 1998, « Le registre de la population du Québec ancien », Annales de démographie historique, 2, p. 215-226.

Desjardins, Bertrand, Alain Bideau, Guy Brunet, Hubert Charbonneau et Jacques Légaré 2001, « De France en Nouvelle-France : les patronymes québécois hier et aujourd'hui », in Guy Brunet, Pierre Darlu et Gianna Zei (dirs.), *Le patronyme : histoire, anthropologie, société,* Paris, CNRS, p. 203-216.

Desjardins, Bertrand, A. Bideau, et al. 2000 From France to New France: Quebec Family Names, Past and Present. *The History of the Family* 5(2): 215-226

Devor, E.J. 1983. Matrix Methods for the Analysis of Isonymous and Nonisonymous Surname Pairs. *Human Biology* 55:277-288.

Dickinson, B. 2000. What's in a name? Using the Web to study Quebec place names and surnames. *Canadian Modern Language Review-Revue Canadienne Des Langues Vivantes* 56 (4):697-701.

Drouilly, P. 1989. Atlas des élections fédérales au Québec, 1867-1988. Outremont Ouébec, VLB

Equihua, M. 1991. Fuzzy clustering of ecological data. *Journal of Ecology* 78 (2):519-534.

FamilySearch. 1880 United States Census and National Index, Vermont [CD-Rom]. The Church of Jesus Christ of Latter-Day Saints 2001 [cited.

Everitt, Brian S., and Graham Dunn. 2001. *Applied Multivariate Data Analysis*. London: Arnold, Co-published by Oxford University Press.

Faucher, Albert 1975 Explication socio-économque des migrations dans l'histoire du Québec. *Mémoires de la société royale du Canada*, série IV, tome XIII, 91-107 (reprinted in Séguin 1980)

Ferrie, Joseph 1996 A new sample of males linked from the Public-Use-Microdata-Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules, *Historical Methods* 29, 141-156

Fix, A. G. 1979. Anthropological Genetics of Small Populations. *Annual Review of Anthropology* 8:207-230.

Fix, A.G. 1978. The Role of Kin-structured Migration in Genetic Micro-differentiation. Annals of Human Genetics 41:329-339.

Fix, G. Alan. 1978. The role of kin-structured migration in genetic microdifferentiation. Annals of Human Genetics 41:329-339.

Fortin, Claire-Andrée 1981 Profil de la main-d'oeuvre forestière en Mauricie d'après le recensement de 1861. *Bulletin d'histoire de la culture matérielle* 13 (automne), 75-76

Fortin, Claire-Andrée 1983 Les travailleurs forestiers en Mauricie au 19e siècle. Mémoire de MA (Études québécoises) UQTR

Fournier, Marcel 1997, Origine des familles pionnières du Québec ancien, Montréal,

Société généalogique canadienne-française.

- Frenette, Yves 1989, « La genèse d'une communauté canadienne-française en Nouvelle-Angleterre, Lewiston, Maine, 1800-1880 », *Historical Papers / Communications historiques*, p. 75-99.
- Frenette, Yves, avec la collaboration de Martin Pâquet (1998), Brève histoire des Canadiens français, Montréal, Boréal, 211 p.
- Galton, Francis and H. W. Watson. 1974. "On the probability of the extinction of families." Journal of the Royal Anthropological Institute 4 138-44
- Gagnon, Alain and Bruno Toupance. 2002. Testing isonymy with paternal and maternal lineage: the impact of polyphyletism and demographic differential. American Journal of Physical Anthropology 117: 334 341.
- Gagnon, Alain. 2001. Patronymes, lignées et généalogies, Trois méthodes d'estimation de la parenté au Québec ancien (1608-1800). In *Le Patronyme, Histoire, anthropologie, société*, edited by G. B. e. al. Paris: CNRS.
- Gagnon, Alain, H. Charbonneau, and B. Desjardins. 1998. Apparentement des conjoints par région au Québec ancien. In *Le choix du conjoint*, edited by G. Brunet, Fauve-Chamoux and M. Oris. Lyon: BOSC.
- Gagnon, Alain, and Evelyne Heyer. 2001. Fragmentation of the Quebec Population Genetic Pool (Canada): Evidence from the Genetic Contribution of Founders per Region in the 17th and 18th Centuries. *American Journal of American Biology* 114:31-41.
- ———. 2001. Intergenerational Correlation of Effective Family Size in Early Quebec (Canada). *American Journal of American Biology* 13:645-659.
- Gagnon, France. 1988. Parenté et migration: le cas des Canadiens français à Montréal entre 1845 et 1875. In *Historical Papers / Communications historiques*: Société historique du Canada.
- Gagnon, Alain, H. Charbonneau and B. Desjardins 1998 Apparentement des conjoints par région au Québec ancien. In G. Brunet, Fauve-Chamoux and M. Oris eds. *Le choix du conjoint*, Les Chemins de la recherche 43, Lyon, BOSC, 133-150
- Gagnon, France 1988, « Parenté et migration: le cas des Canadiens français à Montréal entre 1845 et 1875 », *Historical Papers / Communications historiques*. Société historique du Canada, p. 63-85.
- Gauthier, Benôit 1981 La sous-traitance et l'exploitation forestière en Mauricie, 1850-1875. Bulletin d'histoire de la culture matérielle 13 (automne), 59-67
- Gauthier, Benoît 1984 La sous-traitance et l'exploitation forestière en Mauricie, 1850-1880, Mémoire de MA, Études québécoises, UQTR.
- Gehring, Jake. 2001. Social Security Death Master File: A Much Misunderstood Index [Internet]. Ancestory.com 2001 [cited October 28, 2001 2001]. Available from http://www.ancestry.com/search/rectype/vital/ssdi/article.htm.
- Grace, Robert J. 1999 The Irish in mid-nineteenth-century Canada and the case of Quebec: Immigration and settlement in a Catholic city Ph.D. thesis, Université Laval, 2 vol., 660pp
- Gradie, I. Margaret, Lynn B. Jorde, and Gérard Bouchard. 1988. Genetic Structure of the Saguenay, 1852-1911: Evidence from Migration and Isonymy Matrices. *American Journal of Physical Anthropology* 77:321-333.
- Gradie, M. 1986. Spatial and Temporal Trends in Inbreeding in the Saguenay Population

of Quebec. American Journal Physical Anthropology 69:207.

- —. 1986. Spatial-Temporal Trends in Inbreeding in the Saguenay Population of Quebec. *American Journal of Physical Anthropology* 69:207.
- Gradie, Margaret I., and Danielle Gauvreau. 1987. Migration, Genes, and Hereditary Disease in Human Populations. *International Migration Review* 21 (3):592-608.
- Gauvreau and Margaret I. GRADIE, "Migration, Genes, and Hereditary Disease in Human Populations", International Migration Review, vol. 21, no 3 (1987), pp. 592-608.
- Gauvreau, D. 1991. Québec. Une ville et sa population au temps de la Nouvelle-France, Québec, Presses de l'Université du Québec
- Gueresi, P., D. Pettener, and F. M. Veronesi. 2001. Marriage behaviour in the Alpine Non Valley from 1825 to 1923. *Annals of Human Biology* 28 (2):157-171.
- Guglielmino, C. R., A. De Silvestri, A. Rossi, and V. De Micheli. 1998. Surnames, HLA genes and ancient migrations in the Po Valley (Italy). *Annals of Human Genetics* 62:261-269.
- Guglielmino, C. R., and A. Desilvestri. 1995. Surname Sampling For the Study of the Genetic-Structure of an Italian Province. *Human Biology* 67 (4):613-628.
- Guglielmino, C. R., A. deSilvestri, M. Martinetti, C. Daielli, L. Salvaneschi, and M. Cuccia. 1996. The genetic structure of a province as revealed by surnames and HLA genes: Potential utility in transplantation policy. *Annals of Human Genetics* 60:221-229.
- Haines, M.R. and R.H. Steckel. 2000. *A population history of North America*. Cambridge: Cambridge University Press
- Hagerstrand, T (1968): "A Monte Carlo Approach to Diffusion", pp 368-384 of Berry, B; Marble, D: Spatial Analysis, Prentice Hall, Englewood Cliffs.
- Hardy, René, and Normand Séguin 1984 Forêt et société en Mauricie. Montréal, Boréal Express /Musée national de l'Homme
- Hardy, René, N. Séguin et al., 1980. L'exploitation forestière en Mauricie: dossier statistique, 1850-1930 Trois-Rivières, Groupe de la recherche sur la Mauricie, UQTR Cahier no 4, décembre, 199pp
- Harris, R. Cole 1968, *The Seigneurial System in Early Canada, A Geographical Study*. Québec : Presses de l'Université Laval, 1966.
- Harris, Cole R., and Geoffrey J. Matthews, eds. 1987. *Historical Atlas of Canada: From the Beginning to 1800. Volume 1.* Toronto: University of Toronto Press
- Hartl, Daniel L. 2000. A primer of population genetics. 3rd ed. Sunderland, Mass. Sinauer Associates
- Hay, D. A., H. Hornsby, I. Jones, B. Mowry, and R. J. Mitchell. 1995. The Potential of Isonymy in Psychiatric Genetics. *Behavior Genetics* 25 (3):269-270. Québec, Presses de l'Université Laval
- Henripin, J. 1956. Les divisions de recensement au Canada de 1871 à 1951, méthode permettant d'en uniformiser les territoires. Montréal, Service de documentation économique, Ecole des hautes études commerciales de Montréal.
- Henry, Kevin, L Burge, and D Nguyen. 2003. Testing Differences between Case and Control Point Patterns Using Nearest Neighbour Distances and Bootstrapping.

Springer-Verlag: Lecture Notes in Computer Science 2669:33-42.

- Henry, Kevin, and Sherry Olson. 2004. Une descendance comme échantillon de la population. In *Culture française d'Amérique*, edited by Y. Frenette. Laval, Québec: Presses de l'Université Laval.
- Heyer, E. 1995. Genetic Consequences of Differential Demographic Behaviour in the Saguenay Region, Québec. *American Journal of Physical Anthropology* 98:1-11.
- Heyer, Evelyne, and Marc Tremblay. 1995. Variability of the Genetic Contribution of Quebec Population Founders Associated to Some Deleterious Genes. *American Journal of Human Genetics* 56 (970-978).
- Hill, Emmeline W, Mark A. Jobling, and Daniel G. Bradley. 2000. Y-Chromosome variation and Irish origins. *Nature* 404 (6776):351-352.
- Holloway, S. M. 1995. Coefficients of Relationship By Isonymy Among Scottish Males With Multiple Primary Cancers. *Disease Markers* 12 (3):205-214.
- Hull, D. M. 1998. A reconsideration of Galton's problem (using a two-sex population). *Theoretical Population Biology* 54 (2):105-116.
- Hurd, J.P. 1983. Comparison of Isonymy and Pedigree Analysis Measures in Estimating Relationships between Three 'Nebraska' Amish Churches. *Human Biology* 55:349-355.
- Ihaka R. and Robert Gentleman 1996. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics.Vol. 5 (3) 299–314.
- Imaizumi, Y. 1971. Estimation of Kinship Coefficient from Migration Matrix. Japanese Journal of Human Genetics 16 (2):100-&.
- Imaizumi, Y., and R. Kaneko. 1997. Surname and consanguineous marriages in Japan. Journal of Biosocial Science 29 (4):401-413
- Islam, M. N. 1995. A Stochastic-Model For Surname Evolution. *Biometrical Journal* 37 (1):119-126.
- Ivanov, V. P., M. I. Churnosov, and A. I. Kirilenko. 1998. Distribution of family names in rural populations of Kursk oblast. *Genetika* 34 (8):1114-1119.
- James, A.V. 1983. Isonymy and Mate Choice on St Bart, French West Indies: Computer Simulation of Random and Total Isonymy. *Human Biology* 55:297-303.
- Jetté, René 1983, *Dictionnaire généalogique des familles du Québec, des origines à 1730*, Montréal, Presses de l'Université de Montréal
- Jetté, René, Danielle Gauvreau and Michel Guérin 1990 Aux origines d'une région: le peuplement fondateur de Charlevoix avant 1850. In G. Bouchard and M. DeBraekeleer, eds. *Histoire d'un génôme, Population et génétique dans l'est du Québec*. Sillery, Presses de l'Université du Québec, 75-106
- Jobling, M. A. 2001. In the name of the father: surnames and genetics. *Trends in Genetics* 17 (6):353-357.
- Jongman, R.H.G., C.J. Ter Braak, and P Van Tongeren, eds. 1995. *Data Analysis in Community and Landscape Ecology*. New ed., with corr ed. New York: Cambridge University Press.
- Jorde, L. B. 1982. The Genetic-Structure of the Utah Mormons Migration Analysis. *Human Biology* 54 (3):583-597.
- ———. 2001. Consanguinity and prereproductive mortality in the Utah Mormon population. *Human Heredity* 52 (2):61-65.
- Jorde, L. B., A. W. Eriksson, K. Morgan, and P. L. Workman. 1982. The Genetic-

Structure of Iceland. Human Heredity 32 (1):1-7.

- Kaplan, B.A., W Lasker, and B. Chiarelli. 1978. Communality of Surnames of Biological Interrelationships Among Thirty-one Settlements in Upper Val Varaita in the Italian Alps. *Journal of Physical Anthropology* 49:251-256.
- Kendall, David G. "Branching processes since 1873." J. London. Math. Soc. 41: 385-406. Notes: Reprinted in: Studies in the history of statistics and probability. Vol 2. (1977) pp383-405
- Koertvelyessy, T, M.H. Crawford, R.G. Huntsman, M. Collins, D. Keeping, and M. Uttley. 1988. Repetition of the Same Pairs of Names in Marriages in Fogo Island, Newfoundland, and Genetic Variation. *American Journal of Physical Anthropology* 77:253-260.
- Koertvelyessy, T. A. 1995. Ethnicity, Isonymic Relationships, and Biological Distance in Northeastern Hungary. *Homo* 46 (1):1-9.
- Kohonen, T. 1995. Self-Organizing Maps. Vol. 30, Series in Information Sciences. Springer, Heidelberg.
- Kohonen, Teuvo, T. S. Huang, and M. R. Schroeder. 2000. Self-Organizing Maps. In *Springer Series in Information Sciences, 30*, edited by 3rd. New York: Springer Verlag.
- Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J. 1996. SOM_PAK: The selforganizing map program package. Report A31. Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland,. Also available in the Internet at the address http://www.cis.hut.fi/research/som lvq pak.shtml.
- Kuilikoff, Allan. 1986. "Migration and Cultural Diffusion in Early America 1600-1860: A Review Essay." *Historical Methods* 19, no. 4: 153-169.
- Kruskal, Joseph B., and Eric M. Uslaner. 1978. *Multidimensional Scaling, Sage university* papers series Quantitative applications in the social sciences. California: Sage Publications.
- Kuchermann, C.F., G.W. Lasker, and D.I Smith. 1979. Historical Changes in the Coefficient of Relationship by Isonymy Among Populations of the Otmoor Villages. *Human Biology* 51:63-77.
- Kulldorff M, Nagarwalla N. Spatial disease clusters: Detection and Inference. Statistics in Medicine, 14:799-810, 1995.
- Laberge, Alain, ed. 1993 *Histoire de la Côte-du-Sud*. Québec : Institut québécois de recherche sur la culture
- Laberge, Claude 1967 La consanguinité des Canadiens français, Population 22, 861-896
- Lamarre, Jean. 2003. The French Canadians of Michigan: Their Contribution to the Development of the Saginaw Valley and the Keweenaw Peninsula, 1840-1914: Great Lakes Books.
- Lasker, G. W. 1998. Calculating components of the coefficient of relationship. *Annals of Human Biology* 25 (5):501-502.
- Lasker, G.W. 1985. Surnames and Genetic Structure. Edited by G. W. Lasker, C. G. N. M. Taylor, D. F. Roberts and S. L. Washburn, Cambridge Studies in Biological Anthropology. Cambridge: Cambridge University Press
- Lasker, G. W. 1997. Census versus sample data in isonymy studies: Relationship at short

distances. Human Biology 69 (5):733-738.

- —. 1998. Calculating components of the coefficient of relationship. *Annals of Human Biology* 25 (5):501-502.
- Lasker, G. W., and B. A. Kaplan. 1995. Demography in Biological Anthropology -Human-Population Structure and Evolution. *American Journal of Human Biology* 7 (4):425-430.
- Lasker, G.W. 1977. A Coefficient of Relationship by Isonymy: A Method for Estimating the Genetic Relationship between Populations. *Human Biology* 49:489-493.
- Lasker, G.W. 1978. Increments through Migration to the Coeficient of Relationship of between Communities Estimated by Isonymy. *Human Biology* 53:235-240.
- - ——. 1985. Surnames and Genetic Structure. Edited by G. W. Lasker, C. G. N. M. Taylor, D. F. Roberts and S. L. Washburn, Cambridge Studies in Biological Anthropology. Cambridge: Cambridge University Press.
- Lasker, G.W., D.A. Coleman, N. Aldridge, and W.R. Fox. 1979. Ancestral Relationships within and between Districts in the Region of Reading, England as Estimated by Isonymy. *Human Biology* 51:445-460.
- Lauderdale, D. S., and B. Kestenbaum. 2000. Asian American ethnic identification by surname. *Population Research and Policy Review* 19 (3):283-300.
- Lavoie, Yolande 1973, L'émigration des Canadiens aux États-Unis avant 1930, Mesure du phénomène, Montréal, Presses de l'Université de Montréal.
- Lavallée, Louis 1992, *La Prairie en Nouvelle-France, 1647-1760*, Montréal, McGill-Queen's University Press
- Lavallée, Louis 1992, *La Prairie en Nouvelle-France, 1647-1760*, Montréal, McGill-Queen's University Press
- Lee, Evert. 1966. "A Theory of Migration." Demography 3: 47-57.
- Legay, J. M., and M. Vernay. 2000. The distribution and geographical origin of some French surnames. *Annals of Human Biology* 27 (6):587-605.
- Legendre, L., and P. Legendre. 1983. Numerical Ecology. New York: Elsevier Scientific
- Lewis, G.J. 1982. *Human Migration: A Geographical Perspective*. London: Billing and Sons Limited
- Linteau, Paul-André, René Durocher, and Jean-Claude Robert. 1983. Quebec A History: 1867-1927. Toronto: Lorimer.
- Louder, Dean, Christian Morissonneau et Eric Waddell (dirs.) (1979), « Du continent perdu à l'archipel retrouvé : le Québec et l'Amérique française », *Cahiers de géographie du Québec*, 23, 58
- Madrigal, L., and B. Ware. 1997. Inbreeding in Escazu, Costa Rica (1800-1840, 1850-1899): Isonymy and ecclesiastical dispensations. *Human Biology* 69 (5):703-714.

——. 1999. Mating pattern and population structure in Escazu Costa Rica: A study using marriage records. *Human Biology* 71 (6):963-975.

- Madrigal, L., B. Ware, R. Miller, G. Saenz, M. Chavez, and D. Dykes. 2001. Ethnicity, gene flow, and population subdivision in Limon, Costa Rica. *American Journal of Physical Anthropology* 114 (2):99-108.
- Mascie-Tasker, C.G.N and G.W. Lasker. 1995. Geographical Distribution of Common Surnames in England and Wales. *Annals of Human Biology* 12 (5):397-401.
- Mascie--Taylor, C.G.N., and G.W. Lasker. 1988. *Biological aspects of human migration*. New York: Cambridge University Press.
- Malcolm, M. Dow. 1987. Partial Correlation of Distance Matrices in Studies of Population Structure. *American Journal of Physical Anthropology* 72:343-352.
- Malecot, G. 1975. Isolation by Distance. In *Genetic Structure of Populations, Population Genetic Monographs*, edited by N. E. Morton. Honolulu: University of Hawaii Press.
- Manly, B. F. J. 1986. Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Researches on Population Ecology* 28:201-218.
 - ———. 1991. *Randomization and Monte Carlo Methods in Biology*. London: Chapman and Hall.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27 (209-220).
- Mathieu, Jacques 1987, « Mobilité et sédentarité: stratégies familiales en Nouvelle-France », *Recherches sociographiques*, 18, 2-3, p.211-227 p.211-227
- Mathias, R. A., C. A. Bickel, T. H. Beaty, G. M. Petersen, J. B. Hetmanski, K. Y. Liang, and K. C. Barnes. 2000. A study of contemporary levels and temporal trends in inbreeding in the Tangier Island, Virginia, population using pedigree data and isonymy. *American Journal of Physical Anthropology* 112 (1):29-38.
- PC-ORD for Windows 4.24. MjM Software, Gleneden Beach, Oregon.
- McGrory Klyza, Christopher, Stephen C. Trombulak. The Story of Vermont: A Natural and Cultural History (Middlebury Bicentennial Series in Environmental Studies)University Press of New England, 1999
- McCune, Bruce, and James B. Grace. 2002. *Analysis of Ecological Communities*. Gleneden Beach, OR: MjM Software Design.
- McInnis, R. M. 2000 The population of Canada in the nineteenth century. In *A population history of North America*, edited by M.R. Haines and R.H. Steckel, 371-432. Cambridge: Cambridge University Press
- Moch, Leslie 1992 Moving Europeans, Migration in western Europe since 1650. Bloomington, IN, Indiana University Press
- Mollié, A. 2000, « Bayesian mapping of Hodgkin's disease in France », in P. Elliott, J.C. Wakefield, N.G. Best, et D.J. Briggs (dirs.), *Spatial Epidemiology, Methods and Applications*, Oxford, Oxford University Press, p. 267-285
- Morissonneau, Christian 1978, « La colonisation équivoque: agriculture et développement », *Recherches sociographiques*, 19, 1, 33-54
- Morrissey, Charles T. 1984. Vermont A History. New York: W. W. Norton & Company.

- Mourrieras, B., P. Darlu, J. Hochez, and S. Hazout. 1995. Surname Distribution in France - a Distance Analysis By a Distorted Geographical Map. *Annals of Human Biology* 22 (3):183-198.
- Olson, Sherry, et Patricia Thornton. 2001, « La croissance naturelle des Montréalais au XIX^e siècle », *Cahiers de démographie du Québec*, 30, 2 (automne 2001), p. 191-230
- Olson, Sherry 1996, « Le peuplement de Montréal », in Serge Courville et Normand Séguin (dirs.), *Population et Territoire*, Sainte-Foy, Presses de l'Université Laval, p. 81-94 et 104-107
- Olsson, Gunnar. 1965. *Distance and Human Interaction: A Review and Bibliography*. Pennsylvania: Regional Science Research Institute

Paabo, S 1995. The Y chromosome and the origin of all of us (men). Science 268: 1141.

- Palgrave, Derek. 2004. Many surname variants are really misspelt deviants. *Journal of One-Name Studies* 8 (5):6-9.
- Petersen, W. 2001. Surnames in US population records. *Population and Development Review* 27 (2):315-+.
- Pettener, D., S. Pastor, and E. Tarazona-Santos. 1998. Surnames and genetic structure of a high-altitude Quechua community from the Ichu River valley, Peruvian Central Andes, 1825-1914. *Human Biology* 70 (5):865-887.
- Piazza, A, S. Rendine, G. Zei, A. Moroni, and L.L. Cavalli-Sforza. 1987. Migration rates of human populations from surname distributions. *Nature* 329 (22):714-716.
- Pooley, Colin, and Jean Turnbull. 1998. *Migration and mobility in Britain since the 18th century*. London: UCL Press Limited
- Poulain, Michel (1995), « En quoi l'anthroponymie peut-elle aider le démographe? » dans Yves Landry et al. (dirs.), *Les chemins de la migration en Belgique et au Québec du XVII^e au XX^e siècle*, Louvain, mnh / Academia, p.43-50
- Porteous, Douglas J. 1982. Surname geography: a study of the Mell family name c. 1538-1980. *Transactions of the Institute of British Geographers* 7:395-418.
- Porteous, John Douglas. 1987. Locating the place of origin of a surname. *The Local Historian* 17 (7).
- Ramirez, Bruno, avec Yves Otis (2001), Crossing the 49th Parallel : Migration from Canada to the United States, 1900-1930, Ithaca, N.Y., Cornell UniversityPress.
- Ramirez, Bruno 1991, On the Move. French-Canadian and Italian Migrants in the North Atlantic Economy, 1860-1914, Toronto, McClelland & Stewart
- Ramirez, Bruno. 1995. L'émigration des Canadiens-français aux Etats-Unis. In Le Québec contemporain. Entretiens avec l'histoire, edited by A. Champagne: Radio-Canada,.
- Raspe, P.D., and G.W Lasker. 1980. Biological Relationships by Isonymy Among the Scilly Isles with the Rest of England and Wales. *Annals of Human Biology* 7:401-410.
- Ravenstein, E.G. 1885. "The laws of migration." *Journal of the Statistical Society* 48: 167-235.
- Relethford, J. H. 1986. Density-Dependent Migration and Human-Population Structure in Historical Massachusetts. *American Journal of Physical Anthropology* 69 (3):377-

388.

- Relethford, John H. 1991. "Effect of Population on Marital Migration Distance." *Human Biology* 63, no. 1: 95-98.
- _____. 1991. Effect of Changes in Population-Size on Genetic Micro- Differentiation. Human Biology 63 (5):629-641.
- _____. 2001. Genetics and the Search for Modern Human Origins. New York: Wiley-Liss.
- Relethford, J. H., and J. Blangero. 1990. Detection of Differential Gene Flow from Patterns of Quantitative Variation. *Human Biology* 62 (1):5-25.
- Relethford, J. H., and M. H. Crawford. 1995. Anthropometric Variation and the Population History of Ireland. *American Journal of Physical Anthropology* 96 (1):25-38.
- Relethford, J. H., M. H. Crawford, and J. Blangero. 1997. Genetic drift and gene flow in post-famine Ireland. *Human Biology* 69 (4):443-465.
- Relethford, J. H., and J. H. Mielke. 1994. Marital Exogamy in the Aland Islands, Finland, 1750-1949. Annals of Human Biology 21 (1):13-21.
- Relethford, John. 1986. Density-Dependent Migration and Human Population Structure in Historical Massachusetts. *American Journal of Physical Anthropology* 69:377-388.
 - ———. 1986. Microdifferentiation in Historical Massachusetts: A Comparison of Migration Matrix and Isonymy Analyses. *American Journal of Physical Anthropology* 71:365-375.
- Relethford, John H. 1988. Estimation of Kinship and Genetic Distance from Surnames. *Human Biology* 60 (3):475-492.
 - ------. 1988. Isonymy, Inbreeding and Demographic Variation in Historical Massachusetts. *American Journal of Physical Anthropology* 77:243-252.
- Roby, Yves (1990), Les Franco-Américains de la Nouvelle-Angleterre, 1776-1930. Ouébec, Septentrion
- Roby, Yves (1996), « Partir pour les États », dans Serge Courville et Normand Séguin (dirs.), *Population et territoire*, Sainte-Foy, Presses de l'Université Laval, 121-131.
- Rogers, A.R. 1991. Doubts about isonymy. Human Biology 63:663-668.
- Rosenthal, Paul-André 1999, Les Sentiers invisibles, Espace, familles et migrations dans la France du 19^e siècle, Paris, EHESS
- Rodriguez, Larralde, A., and I. Barrai. 1997. Isonymy structure of Sucre and Tachira, two Venezuelan states. *Human Biology* 69 (5):715-731.
- Rodriguez-Larralde, A., I. Barrai, C. Nesti, E. Mamolini, and C. Scapoli. 1998. Isonymy and isolation by distance in Germany. *Human Biology* 70 (6):1041-1056
- Rodriguez-Larralde, A., D. C. de Guerra, M. Gonzales-Coira, and J. Morales. 2001. Gene frequencies and estimates of admixture in different regions of Venezuela based on the ABO and Rh blood groups. *Interciencia* 26 (1):8-+.
- Rodriguez-Larralde, A., J. Morales, and I. Barrai. 2000. Surname frequency and the

isonymy structure of Venezuela. *American Journal of Human Biology* 12 (3):352-362.

- Rogers, A. R., and H. C. Harpending. 1986. Migration and Genetic Drift in Human-Populations. *Evolution* 40 (6):1312-1327.
- Rogers, R. Allen, and Aldur W. Eriksson. 1988. Statistical Analysis of the Migration Component of Genetic Drift. American Journal of Physical Anthropology 71:451-457.
- Roguljic, Diana, Igor Rudan, and Pavao Rudan. 1997. Estimation of Inbreeding, Kinship, Genetic Distances and Population Structure from Surnames: The Island of Hvar, Croatia. *American Journal of Human Biology* 9:595-607.
- Rojasalvarado, M. D. A., and R. Garzachapa. 1994. Relationships By Isonymy Between Persons With Monophyletic and Polyphyletic Surnames From the Monterrey Metropolitan-Area, Mexico. *Human Biology* 66 (6):1021-1036.
- Rudan, D., I. Rudan, A. Sujoldzic, A. Chaventre, B. Janicijevic, T. Macan, and P. Rudan. 2000. Analysis of changes in genetic structure from surnames. A study of three generations of Pag Island isolates, Croatia. *Homo* 51 (2-3):110-131.
- Rudan, I., H. Campbell, and P. Rudan. 1999. Genetic epidemiological studies of Eastern Adriatic island isolates, Croatia: Objectives and strategies. *Collegium Antropologicum* 23 (2):531-546.
- Rudan, I., G. N. Ranzani, M. Strnad, A. Vorko-Jovic, V. John, J. Unusic, J. Kern, D. Ivankovic, R. Stevanovic, S. Vuletic, and P. Rudan. 1999. Surname as 'cancer risk' in extreme isolates: Example from the island of Lastovo, Croatia. *Collegium Antropologicum* 23 (2):557-569.
- Sattenspiel, L. 1983. A Migration Matrix Model of Day-Care Hepatitis. *American Journal* of *Physical Anthropology* 60 (2):249-249.
- Schiffman, Susan S., M. Lance Reynolds, and Forrest W. Young. 1981. Introduction to Mutlidimensional Scaling: Theory, Methods and Applications. New York: Academic Press.
- Schmidt, H. D., L. Efremovska, and Z. Handziski. 2000. Migration and isonymy of Aromun populations. *Homo* 51:S116-S116.
- Scriver, Charles R. 2001. Human Genetics: Lessons from Quebec Populations. Annual Review of Genomics and Human Genetics 2:69-101.
- Schurer, K. 2004. Surnames and the search for regions. Local Population Studies 72:50-76.
- Segel, Aaron. 1993. An Atlas of International Migration. London: Hans Zell Publishers
- Séguin, Normand, and René Hardy, Forêt et société en Mauricie, 1850-1930, Bulletin d'histoire de la culture matérielle 13 (automne), 51-57
- Siebert, Wilbur H 1913 The American Loyalists in the Eastern Seigniories and Townships of the Province of Quebec. Ottawa, Royal Society of Canada
- Smith, Daniel Scott (1989), « 'All in some degree related to each other', A demographic and comparative resolution of the anomaly of New England kinship », *American Historical Review*, 94, 1 (Supplément), 44-79
- Sokal, Robert R. 1981. *Biometry, the principles and practice of statistics in biological research*. 2nd ed. San Francisco: Freeman.
- Sokal, R. Robert, and Carlos A. Braumann. 1980. Significance Tests for Coefficients of

Variation and Variability Profiles. Systematic Zoology 29 (1):50-66.

- Sokal, R. Robert, Rosalind M. Harding, and Neal L. Oden. 1989. Spatial Patterns of Human Gene Frequencies in Europe. *American Journal of Physical Anthropology* 80:267-294.
- Sokal, R.R., R.M. Harding, G.W. Lasker, and C.G.N Mascie-Taylor. 1992. A Spatial Analysis of 100 surnames in England and Wales. *Annals of Human Biology* 19 (5):445-476.
- Sokal, Robert R. 1981. *Biometry, the principles and practice of statistics in biological research*. 2nd ed. San Francisco: Freeman.
- Southall, H. and Gregory B. White. 1997. Visualizing Past Geographies: The use of animated cartograms to represent long-run demographic change in Britain. *Graphics, Visualization and the Social Sciences* (Workshop Report: AGOCG Technical Report Series No. 33)
- Systat for Windows 9. SPSS Inc, Chicago.
- St-Hilaire, Marc (1996), Peuplement et dynamique migratoire au Saguenay, 1840-1960, Québec, Presses de l'Université Laval
- St-Hilaire, Marc. 1988. Origines et destins des familles pionieres d'une paroisse saguenayenne au XIXe siécle. *Cahiers de geographie du Québec* 32 (85):5-26.
- ------. 1996. La formation du territoire saguenayen. In *Atlas historique du Québec*, edited by S. Courville. Saint-Foy: Les Presses de l'Université Laval.
- Stephenson, Charles. 1974. Tracing Those Who Left: Mobility Studies and the Soundex Indexes to the U.S. Census. *Journal of Urban History* 1 (3):73-83.
- Stouffer, S.A. 1940. "Intervening opportunities: a theory relating mobility and distance." *American Sociological Review* 5: 845-67.
- Sykes, B., and C. Irven. 2000. Surnames and the Y chromosome. *American Journal of Human Genetics* 66 (4):1417-1419.
- Systat for Windows 9. SPSS Inc, Chicago.
- Tanguay, Cyprien (1887), Dictionnaire généalogique des familles canadiennes: depuis la fondation de la colonie jusqu'à nos jours, Montréal, E. Sénécal Thornton, Patricia, et Sherry Olson (2001), « A deadly discrimination among Montréal infants 1860-1900 », Continuity and Change 16, 1, p. 95-135.
- ter Braak, Cajo J.F. 1985. Canoco for Windows 4.02. Centre for Biometry Wapeningen, CPRO-DLO, Wageningen, Netherlands.
- ter Braak, Cajo J.F. 1985. Correspondence Analysis of Incidence and Abundance Data: Properties in Terms of a Unimodal Response Model. *Biometrics* 41:859-873.
- Thornton, Patricia A. 1998. "The Exodus: Leaving Home in the Age of Industry." . Draft of a paper for Canada: Confederation to Present. Department of Geography, Concordia University
- Tobler, Waldo R. 1987. Experiments In Migration Mapping By Computer." *The American Cartographer* 14, no. 2: 155-163.

Tremblay, Marc, Michèle Jomphe, and Helène Vézina (2001), « Comparaison de structures patronymiques et génétiques dans la population québécoise, dans G. Brunet, P.

Darlu et G. Zei (dirs.), *Le Patronyme : Histoire, anthropologie, société*, Paris, CNRS.

- Tremblay, Marc. 1997. "Transmission intergénérationnelle de la reproduction au Saguenay à la fin du XIXe siècle "Cahiers québécois de démographie (à paraître).
- Tremblay, Marc, Evelyne Heyer 1993. "Reproduction régionale différentielle des femmes nées entre 1860 et 1870 au Saguenay ". Cahiers québécois de démographie, 22-2 :263-283, automne
- Trépanier, Guy 1983 Économie, population et mobilité géographique en milieu rural: la paroisse Sainte-Flore en Mauricie, 1860-1901. Mémoire de maîtrise (Études québécoises) UQTR

Trudel, Marcel. 1971. The seigneurial regime. Ottawa: Canadian Historical Association.

- Underhill, P.A., G. Passarino, A.A. Lin, P. Shen, M. Mirazon Lahr, R.A. Foley, P.J. Oefner, and L.L. Cavalli-Sforza. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of Human Genetics* 65:43-62.
- Vernay, M. 2000. Trends in inbreeding, isonymy, and repeated pairs of surnames in the Valserine Valley (French Jura), 1763-1972. *Human Biology* 72 (4):675-692.
 - ———. 2001. Geographical repartition of surnames and genetic structure in the French county of Ardeche at the beginning of the XXth century. *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences* 324 (7):589-599.
- Vicero, Ralph (1968), Immigration of French Canadians to New England, 1840-1900: a Geographical Analysis (Ph.D. diss. (géographie), University of Wisconsin
- Vienna, A., and G. Biondi. 2001. Culture and biology: Surnames in evaluating genetic relationships among the ethnic minorities of Southern Italy and Sicily. *Collegium Antropologicum* 25 (1):189-193.
- Vienna, A., G. F. De Stefano, A. Bastianini, and G. Biondi. 1998. Biodemography in Siena, Italy. *Journal of Biosocial Science* 30 (4):521-528.
- Vienna, A., J. A. P. Garcia, C. G. N. Mascie-Taylor, and G. Biondi. 2001. The ethnic minorities of southern Italy and Sicily: Relationships through surnames. *Journal of Biosocial Science* 33 (1):25-31.
- Watson, H. W., Francis Galton. 1875. On the Probability of the Extinction of Families. *The Journal of the Anthropological Institute of Great Britain and Ireland* 4:138-144.
- Weaver, John C., James De Jonge, and Darrell Norris. 1993. "Trans-Atlantic Migration, 1831-1851."[map] In *Historical Atlas of Canada 1800-1891*, edited by R. Louis Gentilcore. Toronto: University of Toronto Press
- West, Robert C. 1996. An Atlas of Louisiana Surnames of French and Spanish Origin. Baton Rouge: Geoscience Publications, Department of Geography and Anthropology, Louisiana State University.
- White, P and Woods, R. 1980 The Geographical impact of migration. New York: Longman
- Widdis, Randy William. 1987. With scarcely a ripple: English Canadians in northern New York State at the beginning of the twentieth century. *Journal of Historical Geography* 13 (2):169-192.
- Williams, Michael. 1989. Americans and Their Forests, A historical geography

Cambridge University Press

- Withers, Charles W.J., and Alexander J. Watson. 1991. "Stepwise migration and Highland migration to Glasgow, 1852-1898." *Journal of Historical Geography* 17, no. 1: 35-55.
- Wood, J. W. 1977. Stability Test for Migration Matrix Models of Genetic Differentiation. *Human Biology* 49 (3):309-320.
 - ——. 1986. Convergence of Genetic Distances in a Migration Matrix Model. *American Journal of Physical Anthropology* 71 (2):209-219.
- Young, Brian J., and John A. Dickinson. 2003. A Short History of Quebec. Montréal: McGill-Queens University Press.
- Zelinsky, Wilbur. 1955. Some Problems in the Distribution of Generic Terms in the Place-Names of the Northeastern United States. *Annals of the Association of American Geographers* 45:319-49.
 - ———. 1970. Cultural Variation in Personal Name Patterns in the Eastern United States. Annals of the Association of American Geographers. 60:743-69.