



# The 20<sup>th</sup> century part of the BALSAC project and its use in health research

**UQAC**  
Université du Québec  
à Chicoutimi

**HÉLÈNE VÉZINA**

**Linking old and modern data on historical life courses for 20<sup>th</sup> century health research**

ROYAL DANISH ACADEMY OF SCIENCES AND LETTERS  
COPENHAGEN, DENMARK | 17-18 OCTOBER 2016



# Outline of the presentation

---



- The BALSAC database: content and structure
- Strategies to bridge the gap between 1965 and the present-day data
- Access and linkage to medical data
- Health-related research using BALSAC
  - Context and methodological approaches
  - Some examples of past and ongoing projects
- Access to data and privacy concerns
- Other issues and challenges
- Ongoing developments and plans for the future



## The BALSAC database: content and structure

## A brief history of the database

---



- Construction initiated in 1970 by Gérard Bouchard and his team at the Université du Québec à Chicoutimi
- 1972-86: Family reconstitution for the Saguenay region from 1842 (onset of French Canadian settlement) to 1971 (660 000 BMD records)
- 1989-2000: BALSAC undertakes to cover all of Quebec. Entry is restricted to marriages from 1800 onward while the database is being increasingly used in human genetics. The PRDH provides a copy of marriages from the beginning of French Regime (1608) to 1800. and these are integrated and linked to BALSAC
- 2001-11: Integration and linkage of more than 2 million marriage records for the period 1800-1965

# Map of Quebec



# BALSAC in 2016



- **2,2 M** marriage records  
Whole province (Catholic records: 80% population), complete up to 1965, partial for the rest of 20th century
- **550 000** birth and death records  
Saguenay-Lac-St-Jean region
- All records have been digitized and linked relying on linkage procedures based almost exclusively on nominative data.
- The database contains information on **2,5 M** families and **5 M** individuals of which **83%** are linked to parents (or previous spouse in case of remarriage)
- Various research projects in sociohistorical and biomedical sciences
- It is owned by a university consortium and is available to the whole scientific community



# 20th century records in BALSAC



	Births	Marriages	Deaths	Total
1900-1909	34 251	124 470	14 538	173 259
1910-1919	38 429	144 560	16 000	198 989
1920-1929	51 196	158 380	17 880	227 456
1930-1939	58 623	177 427	18 060	254 110
1940-1949	74 840	289 285	20 219	384 344
1950-1959	87 833	313 516	19 238	420 587
1960-1969	65 018	219 516	17 865	302 399
1970-1979	9 892	31 746	6 348	47 986
1980-1989	145	21 233	2 812	24 190
1990-1999		4 822	931	5 753
<b>Total</b>	<b>420 227</b>	<b>1 484 955</b>	<b>133 891</b>	<b>2 039 073</b>

Birth and death records for Saguenay-Lac-St-Jean region only  
Marriage records for all Quebec

# Catholic vs non Catholic records

---



- The Catholic church : very good and exhaustive recording of vital events in parish registers
- Information much less good for non-Catholics
  - Virtually 100% of Catholic marriage records contain the name of spouses' parents or previous spouse
  - For non Catholic records, name of the parents is known in about 23% of cases for husbands and about 30% for wives (improves with time)



# Data sources



## Quebec marriage records

**Before 1800**

- Registre de population du Québec ancien, Programme de recherche en démographie historique, Université de Montréal

**1800-1939**

- Civil records kept in law courts
- Microfilms of civil records from the Fonds Drouin

**1940-1965**

- Digitized images of marriage records from the Direction de l'état civil du Québec

## Saguenay birth, marriage, death records

**1840-1971**

- Civil records kept in law courts

# Content of marriage records

---



- Record number, date and place of registration, date of event, number of banns
- For each spouse:  
first and last name, residence, occupation, honorary title, ethnicity, origin, age, signature (Y|N)
- For the fathers and mothers of the spouses:  
first and last name, residence, occupation, honorary title, ethnicity, origin, signature (Y|N), presence (Y|N)
- For the ex-spouse:  
first and last name, occupation, residence, honorary title, ethnicity, origin
- Other:  
consanguinity (from dispenses), contract (Y|N) , notary, observation

# Content of baptism and burial records

---



## Variables in the database for baptisms:

- Record number, date and place of registration, date of event
- For the subject:  
first and last name, sex, residence, honorary title, ethnicity, origin
- For the father and mother:  
first and last name, residence, occupation, honorary title, ethnicity, origin, signature (Y|N), presence (Y|N)
- Observation

## Variables in the database for burials:

- Record number, date and place of registration, date of event
- For the subject:  
first and last name, sex, age at death, occupation, residence, honorary title, ethnicity, origin, civil status, status
- For the father, mother and spouse:  
first and last name, residence, occupation, honorary title, ethnicity, origin, signature ((Y|N), presence (Y|N))
- Observation

19. Date et heure du décès

[illegible]

☐ Masculin ☐ Féminin ☐ Indéterminé

21. Avez-vous coroné (voir l'aide-mémoire au verso de la copie 1) ☐ Oui ☒ Non

Intervalle approximatif entre le début étiologique et le décès ▼

## 22. Causes du décès

1. Maladie ou affection morbide ayant directement provoqué le décès\*

Antécédents. Affections morbides ayant éventuellement conduit à l'état prédit, l'affection morbide initiale étant indiquée en dernier lieu

a) \_\_\_\_\_  
due à son caractère de

b) \_\_\_\_\_  
dues à (ou consécutives à)

c) \_\_\_\_\_  
dues à (ou consécutives à)

d) \_\_\_\_\_  
(Accuse Article)

2. Autres états morbides importants ayant contribué au décès, mais sans rapport avec la maladie ou avec l'état morbide qui l'a provoquée

\* Il ne s'agit pas ici du mode de décès, par exemple: défillement cardiaque, syncope, etc., mais de la maladie, du traumatisme ou de la complication qui a entraîné la mort.

**23. Y a-t-il eu autopsie?**

1	Oui	2	Non
---	-----	---	-----

Si oui, la certification de la cause du décès tient-elle compte de l'information fournie par l'autopsie?

1 Oui 2 Non

#### 24. Présence de radio-isotopes

1 Oui 2 Non

**27. Personne décédée atteinte d'une maladie à déclaration obligatoire**

1 Oui 2 Non Préciser

[illegible]

25. S'il s'agit d'une femme, le décès est-il survenu au cours d'une grossesse ou dans les 42 jours?

1 Oui 2 Non

26. Si mort violente, cocher

À DES FINS STATISTIQUES SEULEMENT

Accident Suicide Homicide

28. Lieu (ferme, usine, etc.) et circonstances (novade, strangulation, etc.)

## 29. Qualité de l'auteur de la certification médicale

1	Médecin	4	Coroner	Autre
---	---------	---	---------	-------

30. Nom de famille et prénom usuel de l'auteur de la certification médicale

||31. N° de téléphone où l'auteur peut être rejoint

# Database Structure

---



- Relational Database (SQL Server)
- 3 Main tables
  - **Event** - one record for each person in a certificate
  - **Individual** - genealogical data
  - **Couple** - one record for each family
- Other tables
  - Codes for occupations, residences, etc...
  - Indexes for Family Headers, dates, etc...
- Names/Surnames
  - Original form
  - Phonetized (for linkage purpose)

# The geographical dictionary at BALSAC

Code de lieu	Code de variante	Nom de lieu	Valeur par défaut	Échelle de la resid	Lieu d'enregistrement	Année de fondation (paroisse)	Source de l'année de fondation	Indicateur d'URB	Regroupement pour le SLS	Divisions du Québec	Régions du Québec	Code de province, état ou département	Code de pays	Confession de la paroisse	Indicateur de saisie	Longitude	Latitude
resid	var	nom	top	entite	lieu	annee_fondation	mannee_fondation	indic_urb	regr	division	region	prov_etat_dept	pays	conf	val	longitude	latitude
1340	0	ROUYN NORANDA	1	1	0	0	0	1	0	70	24	6	1	0	0	-79,02201	48,23736
1340	1	AMULET	0	12	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	2	BOISVERT	0	12	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	3	DU CUIVRE	0	12	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	4	LAC DUFAULT	0	24	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	5	MERCIER	0	12	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	6	NORANDA	0	24	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	7	NORANDA NORD	0	24	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	8	OSISKO LAKE	0	32	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	9	PELLETIER	0	24	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	10	ROUYN	0	24	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	11	YOUVILLE	0	12	0	0	0	1	0	70	24	6	1	0	0	0	0
1340	12	BLESSED SACRAMENT	0	2	1010	1946	1	1	0	70	24	6	1	1	0	0	0
1340	13	NOTRE DAME DE PROTECTION	0	2	1166	1929	1	1	0	70	24	6	1	1	0	0	0
1340	14	SACRE COEUR	0	2	1234	1953	1	1	0	70	24	6	1	1	0	0	0
1340	15	ST JOSEPH	0	2	1662	1946	1	1	0	70	24	6	1	1	0	0	0
1340	16	IMMACULEE CONCEPTION	0	2	1045	1938	1	1	0	70	24	6	1	1	0	0	0
1340	17	STE BERNADETTE	0	2	2042	1957	1	1	0	70	24	6	1	1	0	0	0
1340	18	ST MICHEL ARCHANGE	0	2	1793	1925	1	1	0	70	24	6	1	1	0	0	0
1340	19	CENTRE HOSPITALIER ROUYN NORANDA	0	2	3161	1953	9	1	0	70	24	6	1	1	0	0	0
1340	20	UKRAINIAN CATHOLIC MISSION	0	2	3162	1953	9	1	0	70	24	6	1	2	0	0	0
1340	21	MARIAGES CIVILS	0	2	2291	1969	1	1	0	70	24	6	1	1	0	0	0

**22 027** different names of places in the dictionary  
corresponding to a total of **11 957** distinct localities

# Occupations in BALSAC



Distribution of spouses occupations  
in marriage records per period

Periods	Nb of records	Occupation of men		Occupation of women	
		n	%	n	%
< 1700	4 345	627	14,4	4	0,1
1700-1799	66 892	6 001	9,0	116	0,2
1800-1899	599 586	396 117	66,1	4 713	0,8
1900-1965	1 484 936	999 884	67,3	326 857	22,0
total	2 155 759	1 402 629	65,1	331 690	15,4

# Results of HISCO coding

---



## Distinct occupations

- 48 169 found in at least one record
  - 31 161 coded (64,7%)
    - 25 332 with an occupation (52,6%)
    - 3768 undetermined (7,8%)
    - 2020 not working (students, retired) (4,1%)
  - 17 008 not processed (35,3%)

## Occurrences

- 3 448 000 in BALSAC
  - 3 408 241 coded (98,8%)
    - 3 275 960 with an occupation (95%)

## HISCO codes

1123 out of 1675 codes were used for coding (67%)





---

Strategies to bridge the gap  
between 1965 and present-day data

# Selection of subjects in present-day population

---

## Population samples

- Genealogical or biodemographic study (without medical or genetic information):
  - BALSAC and marriage repertories
- Genealogical and genetic study:
  - Neutral DNA variants: direct recruitment in the population with consent form mentioning the BALSAC component of the project and REB approval
  - CARTaGENE project (participants who gave consent to genealogical option and filled the questionnaire)

# Selection of subjects in present-day population

---

## Patient samples

- recruitment through medical research project with consent form mentioning the BALSAC component of the project and REB approval
- CARTaGENE project (participants who gave consent to genealogical option and filled the questionnaire)

## Data requirements

- Collect family information to start genealogical reconstruction and locate parents or grandparents in the database

# Genealogical questionnaire (1)



## Guidelines for the genealogical questionnaire

### General instructions

- Answer to the best of your knowledge, exact information is very valuable but approximate information is also useful;
- If you wish, you may consult with members of your family to help you fill out this questionnaire;
- If you don't know, leave the space empty;

### Instructions concerning the PARTICIPANT (yourself):

- If you have been married more than once, your first marriage is the most informative;
- If you are not married, enter "not married" in the "Place" of marriage.

### Instructions concerning the PARENTS:

- Enter the names of the biological parents (natural father and mother);
- If you have to mention an adoptive parent, please check « Adoptive parent ».

### Instructions on DATES :

- Indicate the dates as precisely as possible (day/month/year), for example: 31/05/1945;
- If you only know the month and year, enter (month/year), for example: 05/1945;
- If you only know the year, enter (year), for example: 1945.

### Instructions on PLACES :

- Indicate the places as precisely as possible (parish, city, region), for example: *St-Dominique, Jonquière, Saguenay*;
- If you only know the city, indicate (city, region), for example: *Percé, Gaspésie*;
- If you only know the region, indicate it, for example: *Abitibi*;
- If the place is in Canada outside Quebec, indicate the province, for example: *Alberta*;
- If the place is outside Canada, indicate the country, for example: *Italy*.

Thank you for your participation

# Genealogical questionnaire (2)



We suggest that you complete this genealogical tree as a first step to facilitate the identification of individuals in the following pages.

1 : Participant (yourself)

Name : \_\_\_\_\_

First name : \_\_\_\_\_

SEX : ☐ Male ☐ Female

X : Spouse of 1

Name : \_\_\_\_\_

First name : \_\_\_\_\_

SEX : ☐ Male ☐ Female

2 : Father of 1

Name : \_\_\_\_\_

First name : \_\_\_\_\_

PARENT : ☐ Biological ☐ Adoptive

3 : Mother of 1

Name : \_\_\_\_\_

First name : \_\_\_\_\_

PARENT : ☐ Biological ☐ Adoptive

4 : Father of 2

Name : \_\_\_\_\_

First name : \_\_\_\_\_

PARENT : ☐ Biological ☐ Adoptive

5 : Mother of 2

Name : \_\_\_\_\_

First name : \_\_\_\_\_

PARENT : ☐ Biological ☐ Adoptive

6 : Father of 3

Name : \_\_\_\_\_

First name : \_\_\_\_\_

PARENT : ☐ Biological ☐ Adoptive

7 : Mother of 3

Name : \_\_\_\_\_

First name : \_\_\_\_\_

PARENT : ☐ Biological ☐ Adoptive

8 : Father of 4 - PARENT : ☐ Biological ☐ Adoptive

Name : \_\_\_\_\_

First name : \_\_\_\_\_

9 : Mother of 4 - PARENT : ☐ Biological ☐ Adoptive

Name : \_\_\_\_\_

First name : \_\_\_\_\_

10 : Father of 5 - PARENT : ☐ Biological ☐ Adoptive

Name : \_\_\_\_\_

First name : \_\_\_\_\_

11 : Mother of 5 - PARENT : ☐ Biological ☐ Adoptive

Name : \_\_\_\_\_

First name : \_\_\_\_\_

12 : Father of 6 - PARENT : ☐ Biological ☐ Adoptive

Name : \_\_\_\_\_

First name : \_\_\_\_\_

13 : Mother of 6 - PARENT : ☐ Biological ☐ Adoptive

Name : \_\_\_\_\_

First name : \_\_\_\_\_

14 : Father of 7 - PARENT : ☐ Biological ☐ Adoptive

Name : \_\_\_\_\_

First name : \_\_\_\_\_

15 : Mother of 7 - PARENT : ☐ Biological ☐ Adoptive

Name : \_\_\_\_\_

First name : \_\_\_\_\_

# Genealogical questionnaire (3)



#	INDIVIDUAL	Bio/Adoptive parents	SEX	NAME AT BIRTH	FIRST NAME	BIRTH		MARRIAGE (IF MARRIED)		DEATH (IF DECEASED)	
						Place <small>(parish, town, region if within Quebec) (province if outside Quebec, within Canada) (country if outside Canada)</small>	Date <small>(dd/mm/yyyy)</small>	Place <small>(parish, town, region if within Quebec) (province if outside Quebec, within Canada) (country if outside Canada)</small>	Date <small>(dd/mm/yyyy)</small>	Place <small>(parish, town, region if within Quebec) (province if outside Quebec, within Canada) (country if outside Canada)</small>	Date <small>(dd/mm/yyyy)</small>
	1 Participant						__/__/__		__/__/__		
	X Spouse								__/__/__		
PARENTS OF 1	2 Father of 1		M				__/__/__		__/__/__		__/__/__
	3 Mother of 1		F				__/__/__		__/__/__		__/__/__
GRANDPARENTS OF 1	4 Father of 2		M				__/__/__		__/__/__		__/__/__
	5 Mother of 2		F				__/__/__		__/__/__		__/__/__
	6 Father of 3		M				__/__/__		__/__/__		__/__/__
	7 Mother of 3		F				__/__/__		__/__/__		__/__/__
GREAT-GRANDPARENTS OF 1	8 Father of 4		M				__/__/__		__/__/__		__/__/__
	9 Mother of 4		F				__/__/__		__/__/__		__/__/__
	10 Father of 5		M				__/__/__		__/__/__		__/__/__
	11 Mother of 5		F				__/__/__		__/__/__		__/__/__
	12 Father of 6		M				__/__/__		__/__/__		__/__/__
	13 Mother of 6		F				__/__/__		__/__/__		__/__/__
	14 Father of 7		M				__/__/__		__/__/__		__/__/__
	15 Mother of 7		F				__/__/__		__/__/__		__/__/__



## Access and linkage to medical data

# Access to medical and genetic data

---



- No medical, genetic or epidemiological data in BALSAC (was very important in the 80s and 90s)
- Data come from researchers' projects and are linked to BALSAC for the duration of the project. This linked is destroyed at the end of the project (BALSAC keeps the genealogical data added in the course of the project).
- Molecular population geneticist: collection and analysis of genetic data
- Physician or health scientist for collection and analysis of clinical and genetic data linked to condition





- founded in 1983 and has provided financial and scientific support to BALSAC since the beginning
- the objective of the RMGA is to support and develop basic and applied human genetic research by creating, managing and promoting scientific infrastructures and knowledge transfer for the benefit of the Quebec population.
- the Network supports service platforms, develops Statements of Principles, offers scholarships to students and postdocs, and organizes workshops and meetings
- most Quebec scientists who do medical and human genetic research are members of the RMGA

## RMGA Strategic Theme in Population and Statistical Genetics : objectives 2014-17

---



- enhance and diversify BALSAC's services to the research community
- develop a protocol enabling the use of BALSAC for clinical purposes
- expand the range of analytical tools to support the joint analysis of genealogical and genetic data in genetic epidemiology, mapping studies and health care planning
- extend national and international collaborations for benchmarking and for the development of new analytical and statistical approaches.

# THE LARGEST POPULATION COHORT OF QUÉBEC

Supporting research efforts with high quality population data and biological samples from 43,000 individuals



## Recruitment

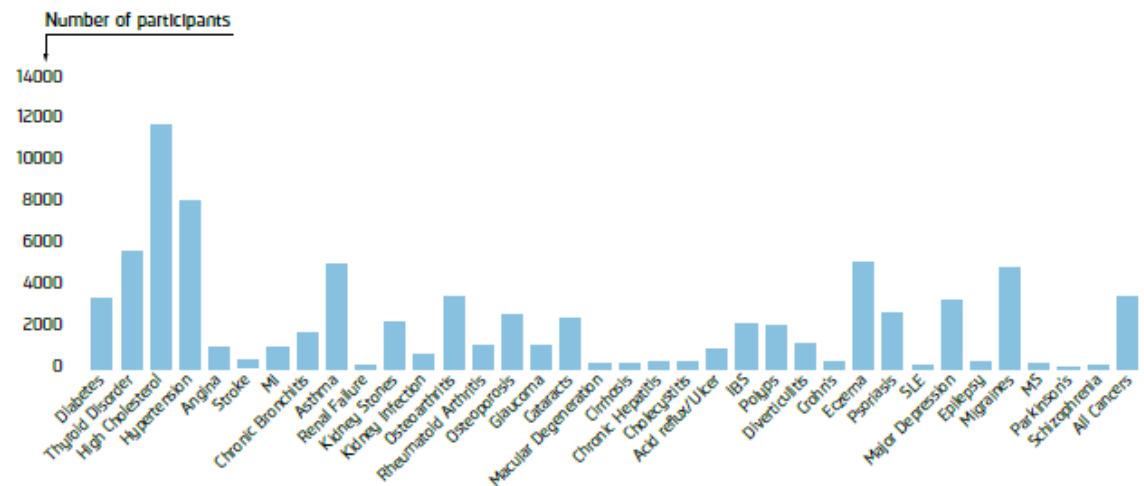
Phase 1: 2009-2011

Phase 2: 2012-2015

Participants aged 40-69 from 6 Quebec cities and regions

About 25% gave consent to genealogical option

Captures a broad and representative sampling of chronic diseases in the population.



MI: myocardial infarction; IBS: irritable bowel syndrome; MS: multiple sclerosis; SLE: systemic lupus erythematosus

43,000 participants recruited in two phases

One of the most deeply phenotyped population cohort in the world, with over 650 variables collected.

Focus on quantitative traits and endophenotypes, representing intermediate and heritable stages in the development of chronic diseases.

One of the largest collection of biomaterial from the general population in North America.

Participants signed a broad consent enabling longitudinal tracking (for > 50 years) using provincial health administrative databases.

Access to genealogical data through BALSAC.

Access to administrative health data (RAMQ) including: hospitalizations, diagnoses, interventions, prescription drug use.

Complete residential histories enabling exposure assessments through linkage to administrative databases.

Access to high density genotyping data.

Yearly follow-up questionnaire planned.

Ancillary research projects are also possible.



---

## Health-related research using BALSAC: context and methodological approaches

# BALSAC and human and medical research

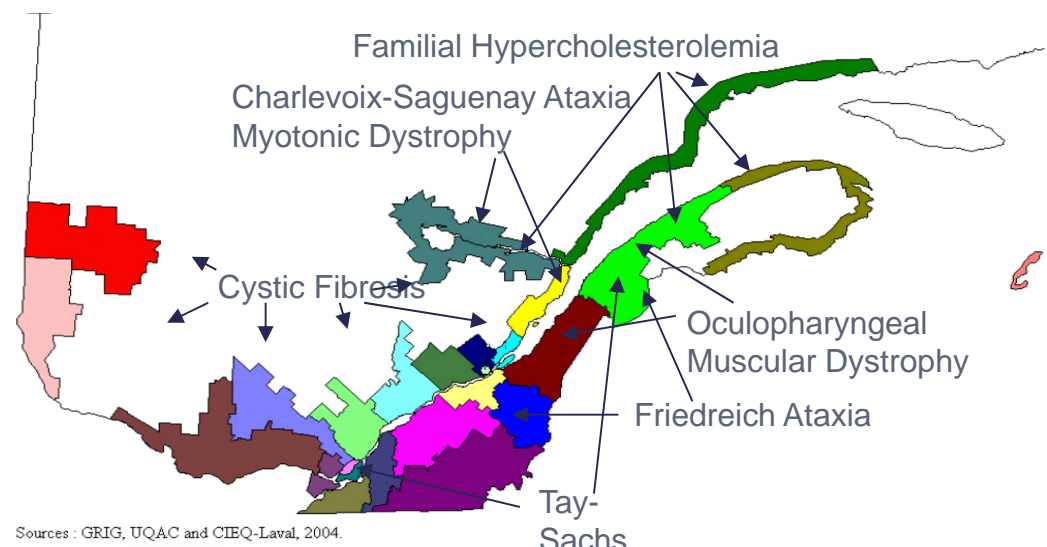
---

- BALSAC is a longitudinal database that allows automatic reconstruction of family histories (SLSJ), genealogical ties and kinship structure of the Quebec population over four centuries.
- It covers all regions of Quebec's territory since the beginning of European settlement in the 17th century to the contemporary period.
- Genetic and medical research started around 1980 when Gérard Bouchard initiated the development of a research program in medical and human genetics with Charles Scriver (McGill) and Claude Laberge (Laval University) and clinicians from the Saguenay-Lac-St-Jean region

# The Quebec context



- Geneticists have a special interest in founder populations and isolated populations because of their expected increased genetic homogeneity
- Quebec is a founder population of 6,5 million individuals
- We can study the characteristics and consequences of initial founder effect in a context of specific regional histories
- Higher frequency of inherited disorders in some regions



# Three approaches for genetic and medical studies using BALSAC

---



1) Individual biographies and 2) family characteristics and intergenerational approach

- The outcomes are variables related to death: (age at death, longevity) or a specific disease
- Complete family reconstitution is necessary (BMD) so studies are done with Saguenay-Lac-St-Jean data
- Use of bivariate or multivariate statistical methods for analyses
- Genetic analysis methods can be used for intergenerational transmission of traits (heritability calculations)

# Individual biography approach

## Once were farmers: Occupation, social mobility, and mortality during industrialization in Saguenay-Lac-Saint-Jean, Quebec 1840–1971

Alain Gagnon <sup>a,\*,1</sup>, Marc Tremblay <sup>b</sup>, H  l  ne V  zina <sup>b</sup>, Jamie A. Seabrook <sup>c</sup>

**Table 2**

Cox hazard models of mortality between marriage and age 60 per period.

Variable	Period					
	1840–1899		1900–1939		1940–1971	
	HR	p	HR	p	HR	p
Year of episode	0.979	0.000	0.982	0.000	0.987	0.000
Age at marriage	0.965	0.002	0.991	0.224	0.996	0.417
Occupation						
Farmer	REF					
White collar	1.502	0.078	1.567	0.000	1.416	0.000
Skilled worker	0.714	0.505	1.001	0.996	1.424	0.000
Lower skilled worker	1.098	0.633	0.978	0.835	1.456	0.000
Unskilled worker	1.312	0.022	1.296	0.000	1.311	0.000
Residence						
Rural	REF					
Semi-urban	1.315	0.043	0.936	0.383	1.130	0.142
Urban	2.700	0.000	1.341	0.002	1.382	0.000
Signed the register	0.834	0.206	0.768	0.000	0.209	0.000
Number of observations	44,279		119,118		186,962	
Number of subjects	5322		14,510		36,097	
Number of deaths	497		1361		2475	

Notes: HR: Hazard ratio; p: p-value.

Population sizes:

Rural: <1000 for the three periods.

Semi-urban: 1000–4999 (<1900); 1000–10,000 (1900–39), and 1000–15,000 (1940–71).

Urban: >5000 (<1900); >10,000 (1900–39); and >15,000 (1940–71).

Source: Gagnon et al. 2011



# Family and intergenerational approach



Hum Nat

DOI 10.1007/s12110-008-9031-7

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

Louis Houde • Marc Tremblay • H       V      

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

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

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

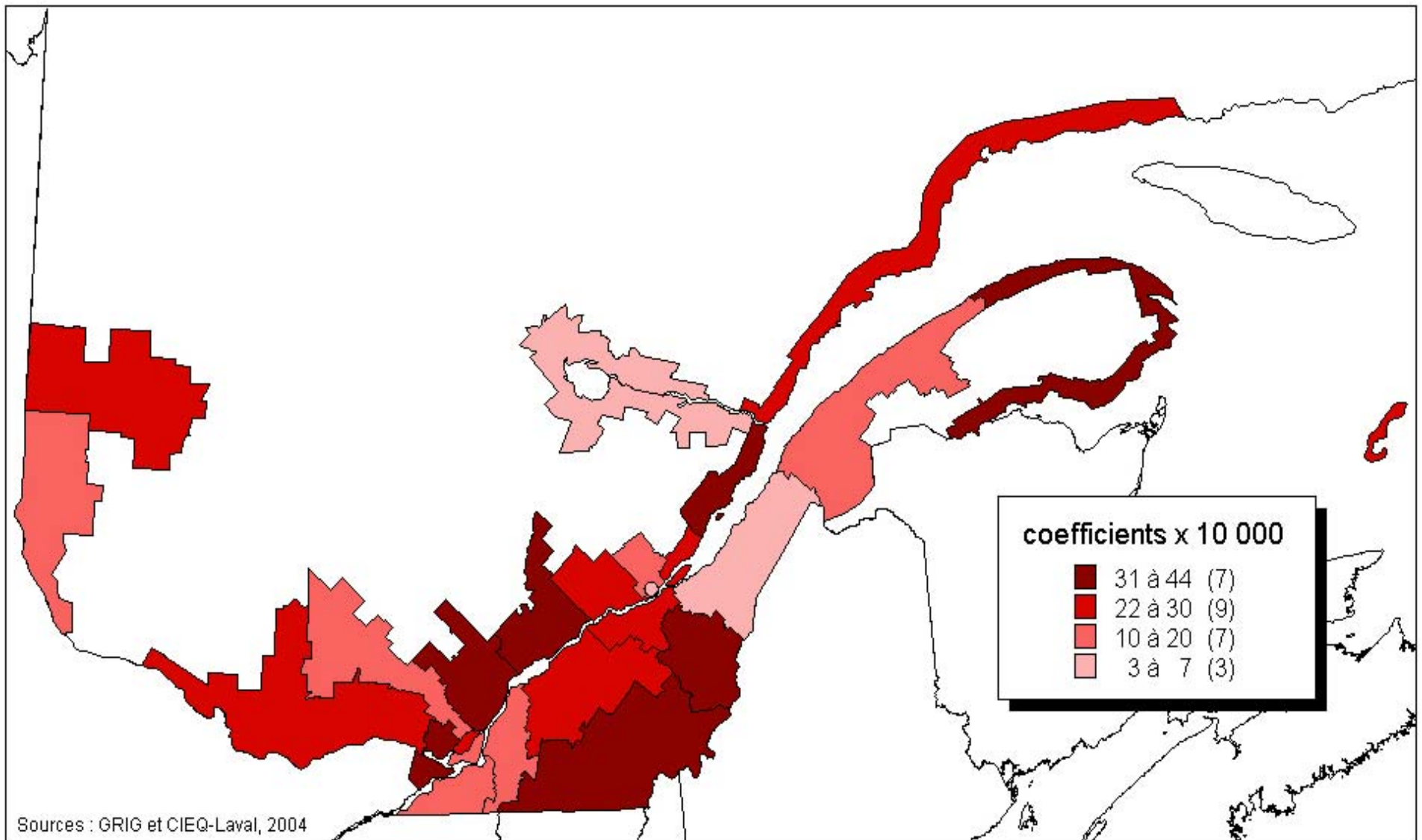
### 3) Extended family and genealogical approach

---



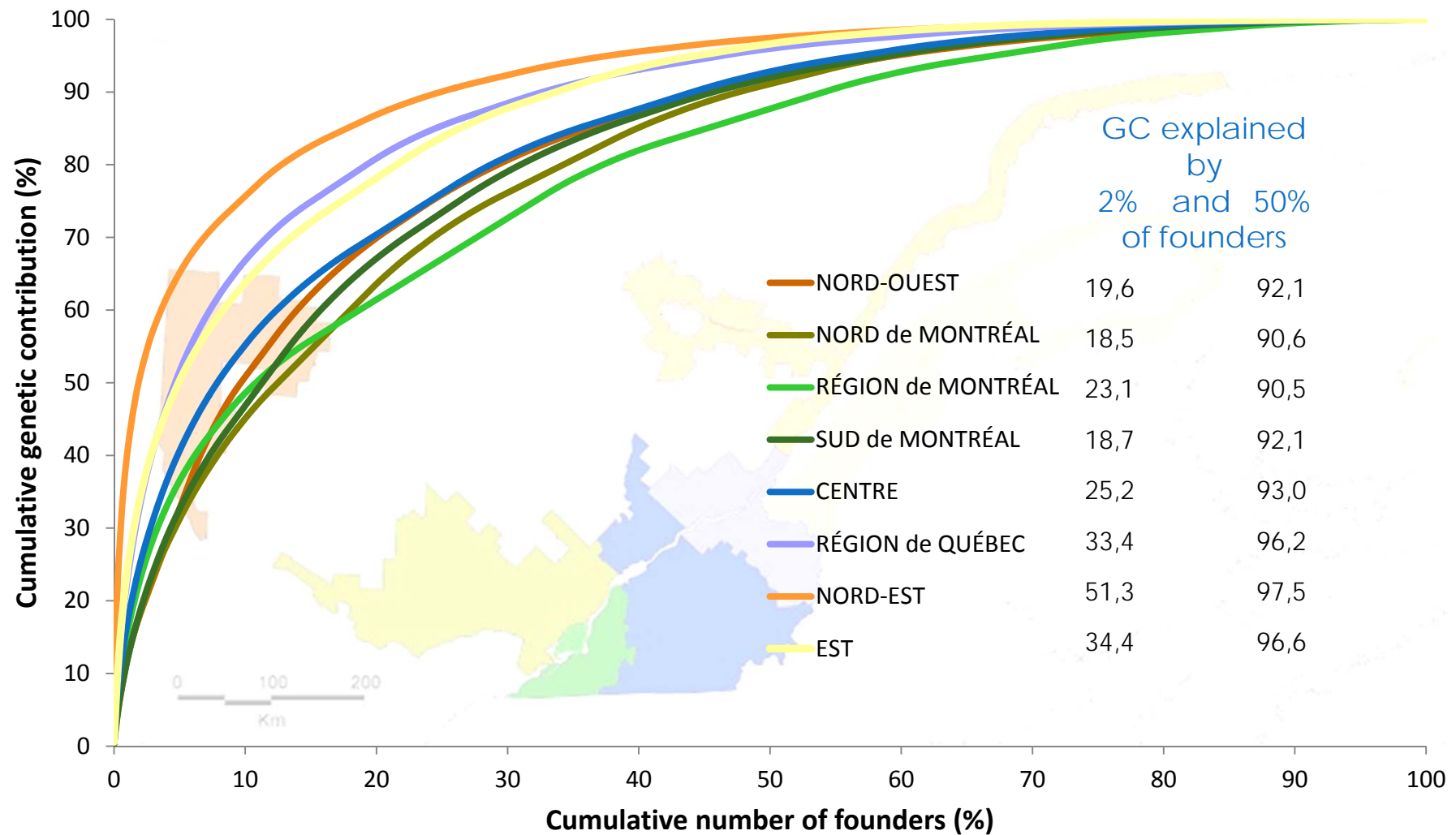
- The genealogy of a population contains information on people who participated to reproduction and therefore transmitted their genes
- It comprises ancestors and the links connecting them; these links are the transmission paths of genes from one generation to the next.
- Genealogical analyses rely on methods from population genetics (kinship and inbreeding calculations, genetic contribution of ancestors, gene dropping)
- Most medical and genetic research done at BALSAC has relied on this approach using genealogical data only or in combination with medical or genetic data.

## Mean inbreeding coefficients at the 4th generation

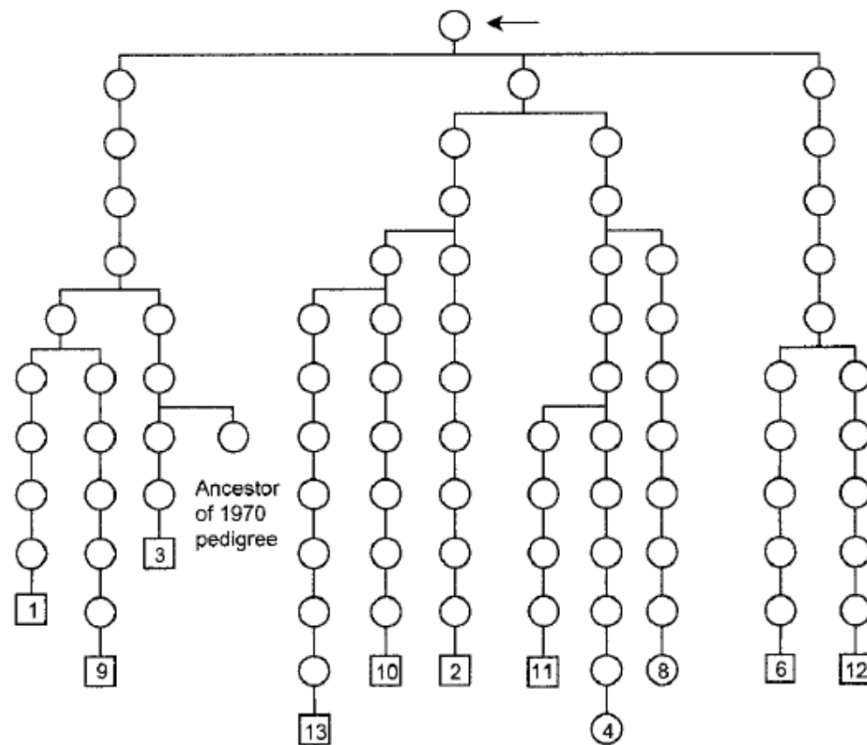


# Genealogical founders

## Genetic contribution (GC) per region

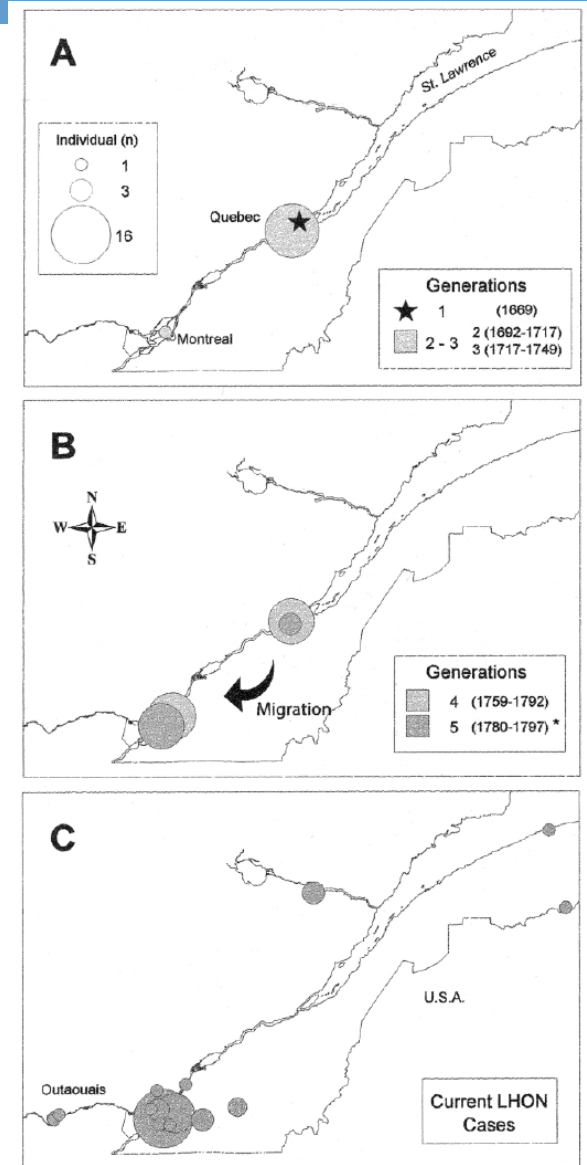


# Charting the introduction and diffusion of the T14484C Leber optic atrophy mutation in Quebec



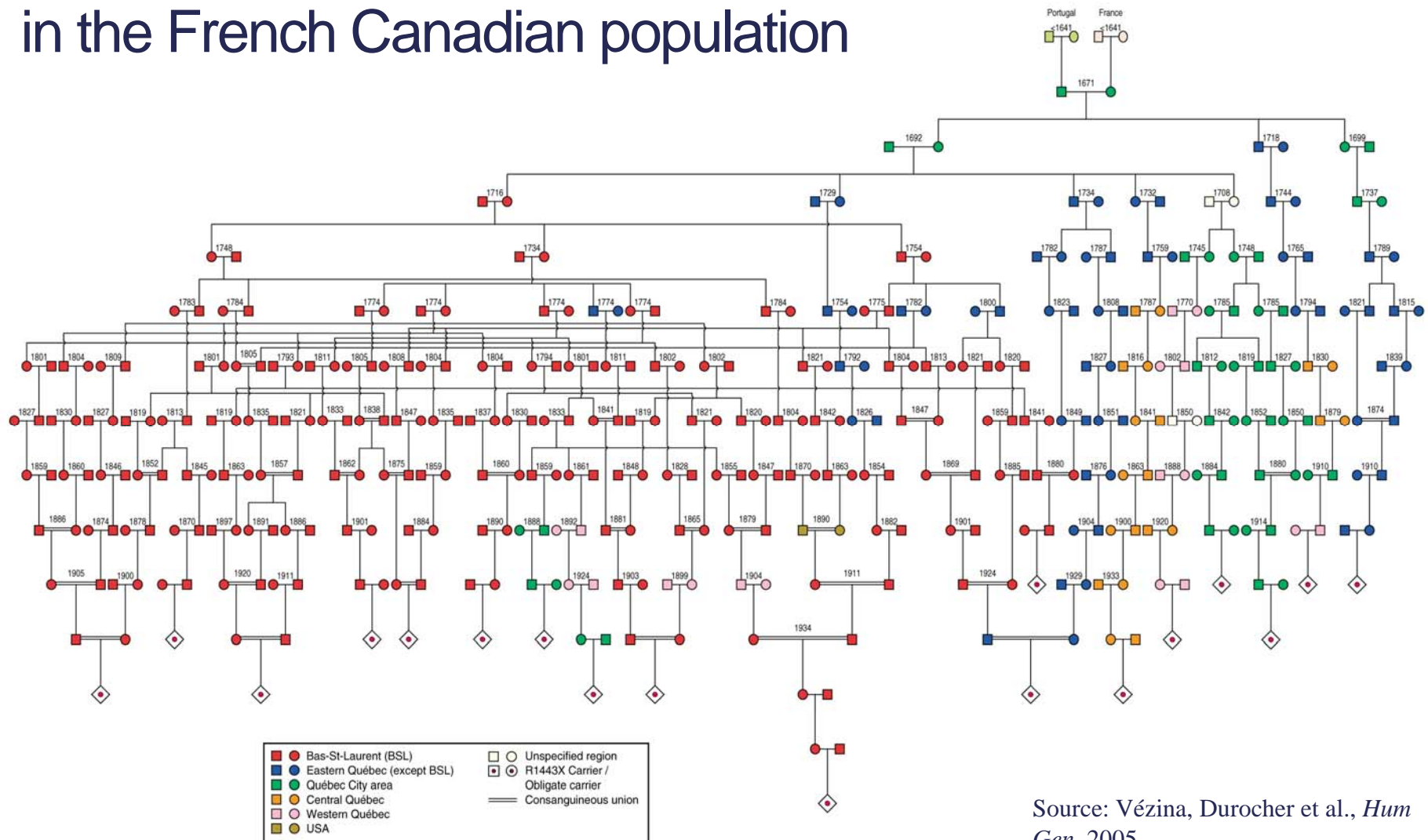
**Figure 1** Pedigree of the 11 probands sharing the same female ancestor and their relationship with the family reported in 1970 (Brunette and Bernier 1970).

Source: Laberge et al. *AJHG* 2005



**Figure 2** Geographic distribution of the marriage parishes of the single female ancestor's female descendants, by generation (1669–1800) (A and B) and geographic distribution of current cases that are due to T14484C mutation (C). An asterisk (\*) marks the 5th generation, which is incomplete, because data are available only through 1800.

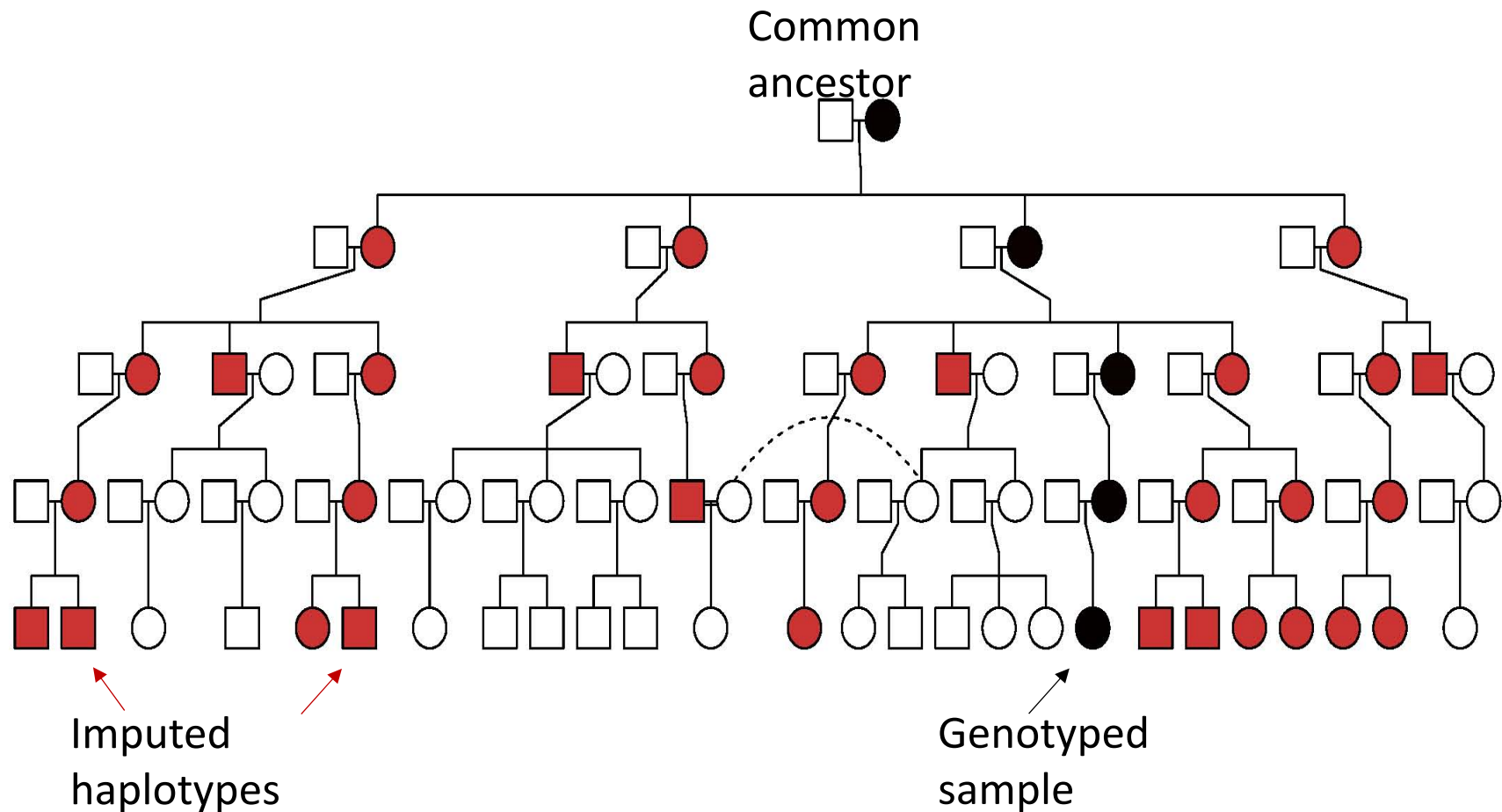
# Introduction and diffusion of the BRCA1 R1443X mutation in the French Canadian population



Source: Vézina, Durocher et al., *Hum Gen*, 2005



# Maternal haplotype imputation in genealogies

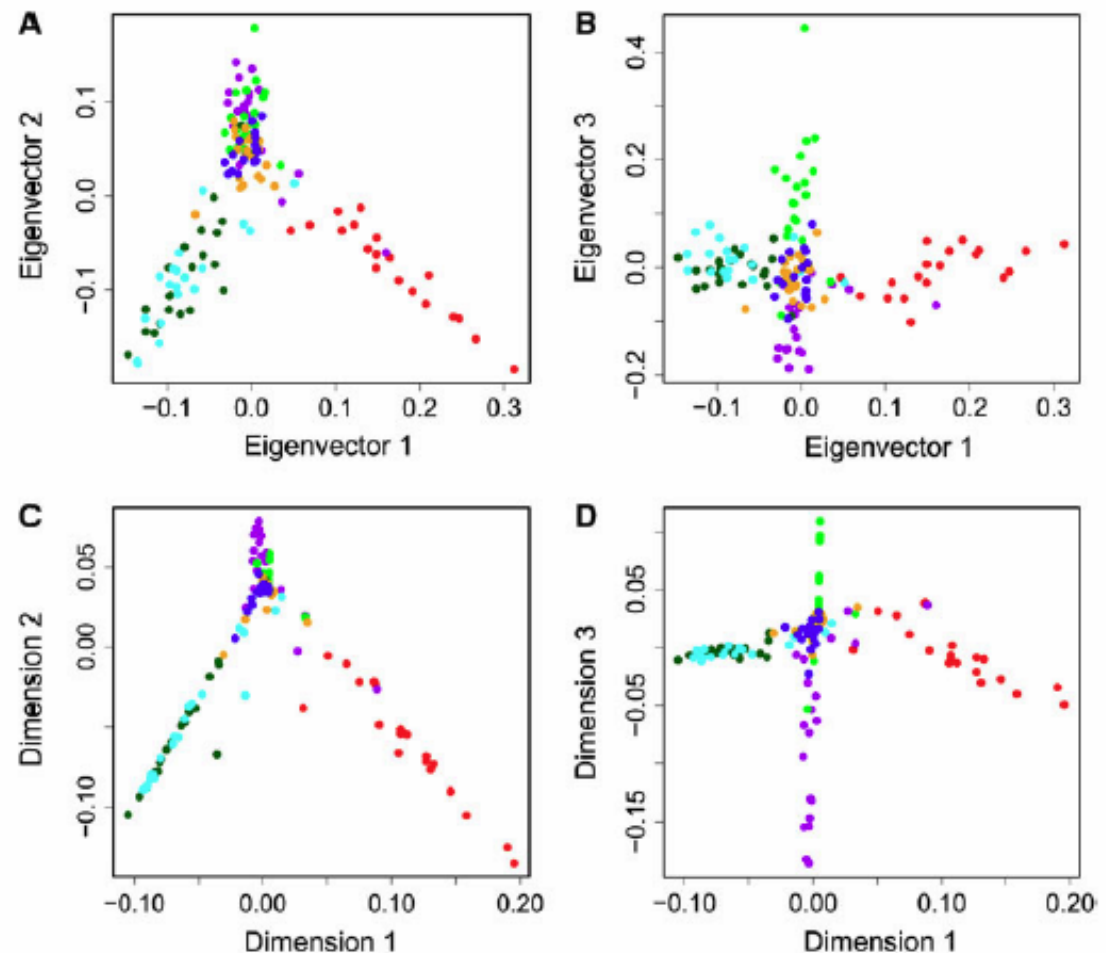


	Genotyped samples	Imputed 1930-60	Individuals 1930-60	Coverage 1930-60
Maternal	924	778 670	1 589 991	0.49
Paternal	414	182 567	789 372	0.23

## Genomic and genealogical investigation of the French Canadian founder population structure

Marie-Hélène Roy-Gagnon · Claudia Moreau · Claude Bherer · Pascal St-Onge ·  
Daniel Sinnett · Catherine Laprise · Hélène Vézina · Damian Labuda

Quebec population  
structure captured  
by genomic and  
genealogical data



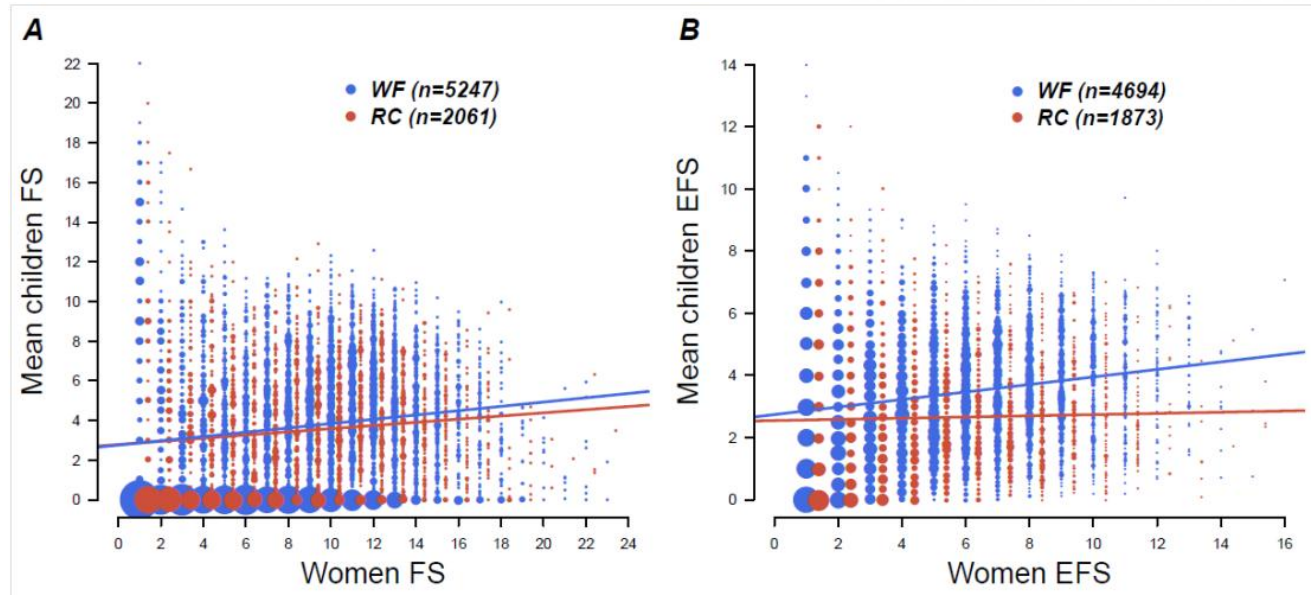
Source: Roy-Gagnon et al. 2011



# Deep Human Genealogies Reveal a Selective Advantage to Be on an Expanding Wave Front

Claudia Moreau,<sup>1</sup> Claude Bhérier,<sup>1</sup> Hélène Vézina,<sup>2</sup> Michèle Jomphe,<sup>2</sup>  
Damian Labuda,<sup>1,3\*</sup> Laurent Excoffier<sup>1,4,5\*</sup>

	No. of women	Mean no. of children (FS)	Mean no. of married children (EFS)	Mean age at marriage	FS ratio WF/RC	EFS ratio WF/RC	Marriage age ratio WF/RC
Wave front (WF)	2663	9.1	4.9	20.5	1.15***	1.20***	0.95 ***
Range core (RC)	1783	7.9	4.1	21.6			



Source: Moreau et al. 2011 Science

Individuals at the wave front of the range expansion have more chance to transmit their genes than people stayed behind. This trait is heritable.



---

## Health-related research using BALSAC : ongoing projects

## Research areas of ongoing projects (2015-16)

---



---

Genetic epidemiology	15
Evolutionary biology and biodemography	15
Social sciences	15
Methodological developments	4

---

49 projects

# Users' affiliation (2015-16)



	Researchers	Students and research assistants	
Québec	40	50	90
Rest of Canada	2	-	2
Outside of Canada	10	2	12
	52	52	104

# Genetic epidemiology of inherited disorders

---



- Genetics of spinocerebellar hereditary ataxias

Bernard Brais/McGill

- Characterization of founder effects in Morquio and Tay-Sachs in the French Canadian population

Hélène Vézina/UQAC, Gail Ouellette/Regroupement québécois des maladies orphelines

- Effects of maternal and paternal transmission in familial hypercholesterolemia

Patrick Couture/Université Laval, Daniel Gaudet/CSSS-Chicoutimi

- Eye Hereditary Diseases in Québec: study of mutations associated with glaucoma, dystrophies and retinal degeneration

Vincent Raymond/Université Laval

# Genetic epidemiology of complex traits

---



- Towards optimal use of Quebec genealogical resources: an integrated approach of genetic epidemiology and population genetics

Marie-Hélène Roy-Gagnon/U of Ottawa

- Personalized medicine of epilepsy: genealogical aspects

Simon Girard/McGill, Patrick Cossette/CHUM

- Genetic factors in congenital cardiopathies: towards an integral genomic analysis

Gregor Andelfinger/CHU-Ste-Justine

- Search for genes of hypertension and its associated factors

Pavel Hamet/CHUM

# Evolutionary biology and biodemography

---



- Dynamics of genes with sex-dependant transmission in the Quebec population

Alexandra Doyon/UQTR

- The role of contemporary evolution in modern human populations

Emmanuel Milot/UQTR

- Study of genetic surfing in the Charlevoix-Saguenay-Lac-St-Jean population

Damian Labuda/CHU-Ste-Justine, Laurent Excoffier/University of Bern, H          /UQAC

- Fecundity and Human Evolution

Marc Klemp & Oded Galor/Brown University, USA

# Evolutionary biology and biodemography

---



- Impact of early life conditions on longevity

Alain Gagnon/ Université de Montréal

- Accumulation of deleterious mutations during spatial expansion by sequencing exomes

Isabelle Duperret & Laurent Excoffier/ Bern University

- Genealogical characteristics of participants to the CARTaGENE project

Marc Tremblay, UQAC



# Overview of medical and genetic research with BALSAC since 1980

---



## 1980-2000:

- Studies of inherited disorders aimed at understanding and explaining the role of demographic dynamics and population history in the origin and spread of the mutant alleles involved in these disorders
- Studies of complex traits to gain a better understanding of the role of genetic factors in the aetiology of these disorders and to contribute to study design for gene mapping projects
- Genealogical studies in population genetics to investigate consequences of settlement history on population structure
- Few studies in biodemography mostly on intergenerational transmission of demographic traits (infant mortality, fertility)

# Overview of medical and research research with BALSAC since 1980

---



## 2000-2016:

- Studies of inherited disorders: use of new methods such as gene dropping and simulations
- Complex traits: genetic epidemiology projects aiming at improving study design (for instance reducing or taking into account heterogeneity of genetic and environmental factors)
- Populations genetics projects: joint analysis of molecular and genealogical data to investigate population structure,
- Biodemography and evolutionary biology projects : BALSAC as population laboratory, strong interest for birth and death data and contextual information
- From small to large datasets; from goals linked to Quebec population to larger scope; from analyses performed at Balsac to extraction of datasets sent to researchers; from simpler to more complex genealogical and statistical methods



Access to data and protection of privacy

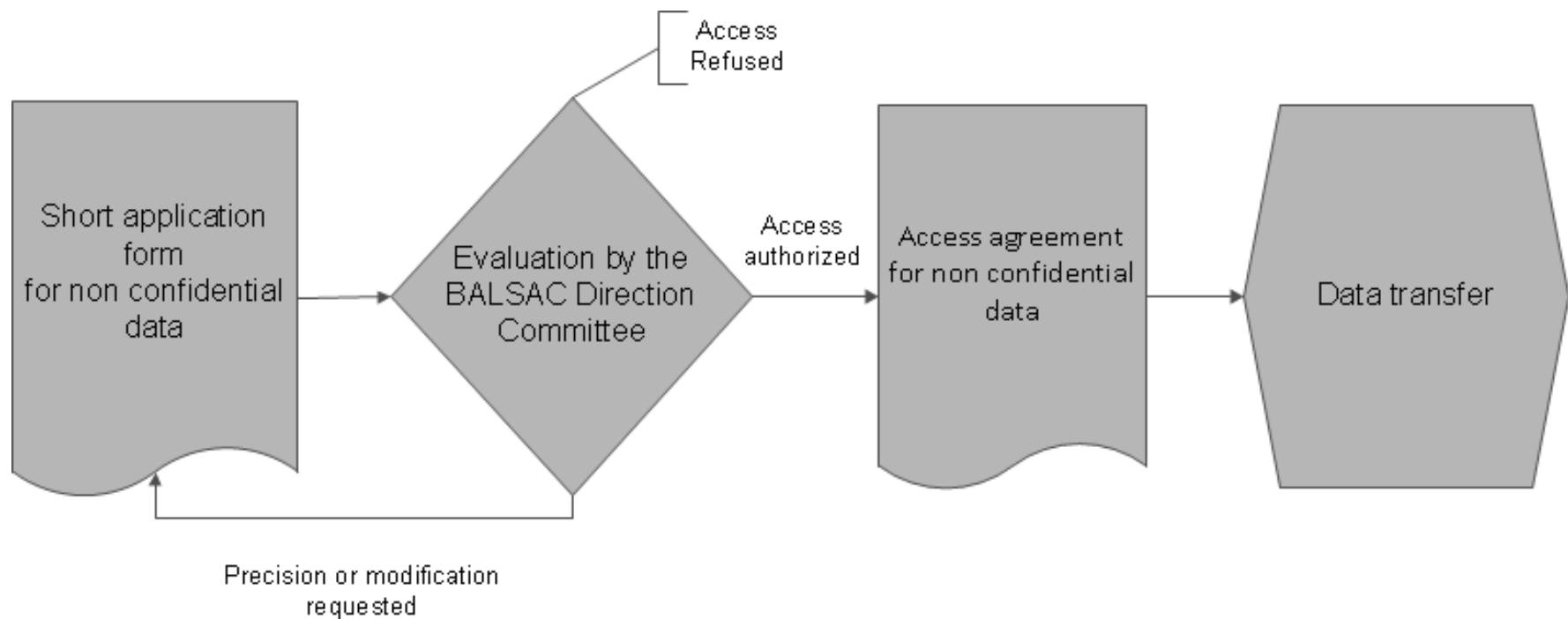
# Protection of privacy at BALSAC

---



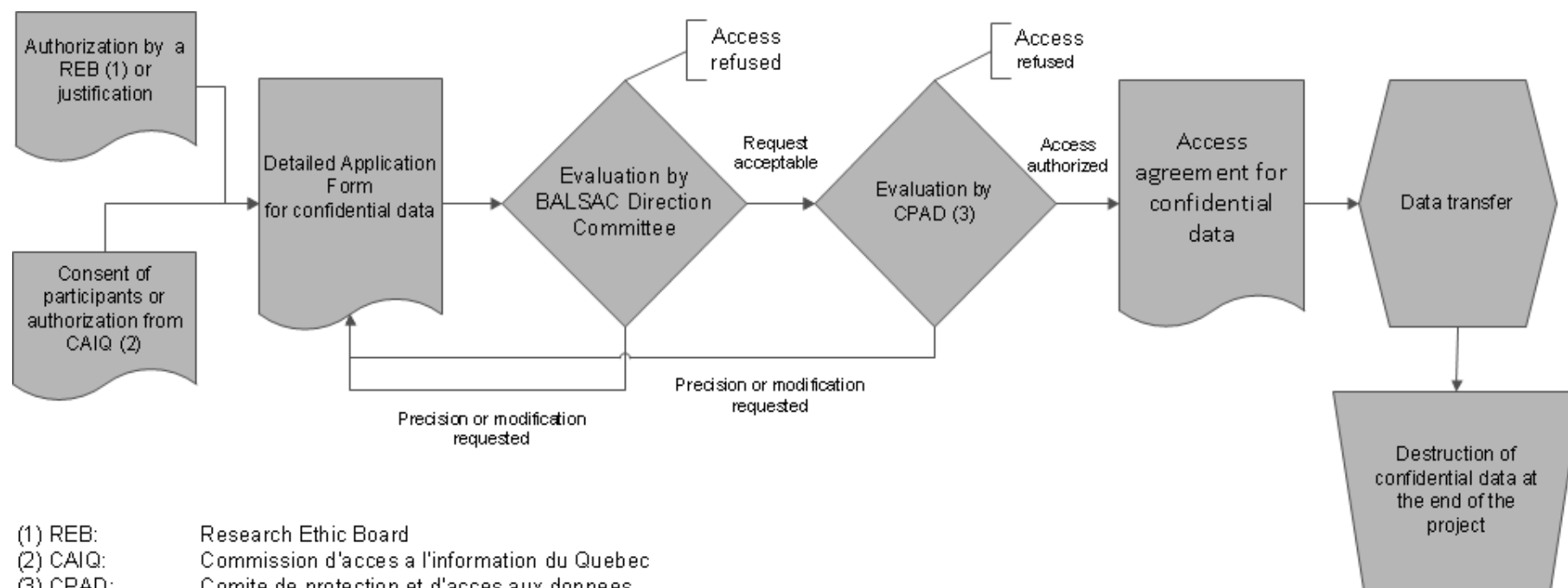
- Used only for scientific research projects
- Contains no medical or genetic data
- Anonymity preserved in the diffusion of datasets and study results
- No information transmitted to insurance companies, police, governments

# Access procedure for non confidential\* data



\*Anonymous or coded information that does not permit identification of individuals or personal information originating from documents considered public by law

# Access procedure for confidential\* data



\*Personnal information originating from non public documents



---

## Other issues and challenges

# Issues and challenges

---



- Funding for activities other than development and research: management of the infrastructure and services to researchers (free access to data – cost for data extraction and research assistance only): UQAC and its foundation, 3 partner universities, RMGA
- Critical mass of users (reach potential users (how? who?), train the next generation, initiate and participate to national and international collaborations)
- Longitudinal databases and health research: multidisciplinary research
- Language and geography



## Support to researchers

- ensures that data access requests are completed and sent in accordance with the terms of the BALSAC Data Access Policy.
- provides technical and professional assistance to those who wish to use BALSAC data for research
- services range from the development of custom data sets for research projects to more or less detailed data analysis, depending on needs.
- provides assistance to GENLIB users

# Development of genealogical analysis tools



## Handling genealogical data

Deep rooted genealogies because of their tree like structure as well as the complexity (dependence) of the data cannot be directly analyzed with standard statistical packages

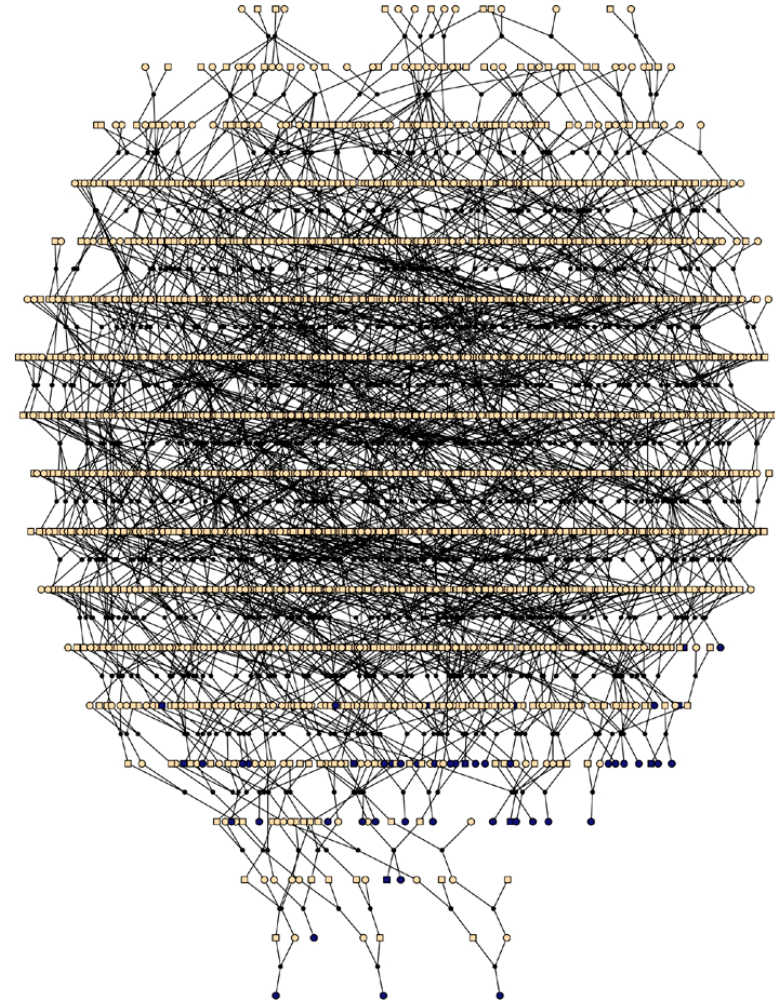
## GENLIB

A R package specifically developed for genealogical analyses

Available on <http://balsac.uqac.ca> :

GENLIB can be used for :

- descriptive analyses (completeness, average genealogical depth, etc.)
- genetic measures (genetic contributions, consanguinity, kinship, etc.)
- simulations (gene dropping) and statistical tests



<http://balsac.uqac.ca>

[HOME](#) [CONTACT US](#) [SITE MAP](#)



[FRANÇAIS](#) [Population Database](#) [Data Access](#) [Confidentiality](#) [Bibliography](#) [Did you know?](#)

## Welcome to BALSAC



The BALSAC database was created 40 years ago at Université du Québec à Chicoutimi (UQAC) in partnership with Université Laval, McGill University and Université de Montréal. The database is comprised of digitized Québec vital event records (mainly marriage certificates). These records are interconnected using a linkage method based on nominative information that allows for the automatic recreation of genealogical relations and kinship structures in the Quebec population over four centuries. BALSAC covers the entire territory of Quebec, from the beginning of European settlement in the 17th century to the contemporary period.

This is a remarkable piece of Quebec's scientific legacy and a unique infrastructure, both for its genealogical depth and the population it covers. BALSAC is used by researchers in Quebec, Canada and abroad who are interested in population genetics, genetic epidemiology,

### Partner Universities



### Information from BALSAC



### Funding Bodies

- Canada Foundation For Innovation
- Fondation de l'Université du Québec à Chicoutimi (FUQAC)
- Fondation J.-A. DeSève
- Ministère du Développement économique, Innovation et Exportation
- Réseau de médecine génétique appliquée (RMGA)

### Links



Ongoing developments and plans for the future



# The Integrated Infrastructure of Quebec Historical Microdata (IMPQ)

---



The IMPQ will integrate available historical microdata on the Quebec population dating back to the beginning of European settlement, into a set of relational databases

1. Fusion of the BALSAC and RPQA databases; integration and linkage of birth and death records for 1800-1849
2. Harmonizing existing census data series and expanding the geographical coverage
3. Linkage of census data to BALSAC families and across censuses

# French Canadian allele frequencies database

---



- We intend to integrate genetic information in the BALSAC infrastructure
- The goal is to set up a database that will include frequencies of genetic variants in the French Canadian population; this will be most useful to scientists who try to identify genetic risk factors to complex traits
- There will also be a connection between genealogical and genetic information for some individuals
- Major change for BALSAC, will include many steps including obtaining approval from UQAC authorities and partner universities

## Possibilities for future developments

---



- Birth and death records post 1850 (with cause of death?)
- Marriage records post 1965 (take into account the fact that less and less people get married)
- Integration of complementary data (environment, more socioeconomic information)?
- Pursue development on whole Quebec or concentrate on specific regions?
- Georeferencing?
- Translate data into IDS?
- Analytical tools for genealogical data (R platform)?





Mont-Valin, Saguenay  
photo: Réo Blackburn