nature portfolio

Corresponding author(s): Simon L. Girard

Last updated by author(s): Feb 17, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Python 3.11.5; matplotlib 3.10.1; numpy 1.26.4; pandas 2.2.3; scipy 1.10.1; seaborn 0.13.2; umap_learn 0.5.9; geneakit 0.1.0 available under MIT licence at https://github.com/Genopop/geneakit; scripts for the figures available at https://github.com/Genopop/slsj_structure; msprime 1.3.4; stdpopsim 0.3.0; scikit-allel 1.3.7; scikit-learn 1.8.0; scikit-bio 0.6.3; geopandas 1.1.2; requests 2.32.5

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Quebec genotype data is available via an independent data access committee by the CARTaGENE cohort (https://cartagene.qc.ca/en/researchers/access-request.html). Quebec genealogical data is available upon request to BALSAC (https://balsac.uqac.ca). These datasets are under restricted access due to the informed consent given by study participants.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Neither sex nor gender were considered during the analyses, as the scope of the study is to characterize the population structure of a whole population, regardless of the individuals' sex or gender.
Reporting on race, ethnicity, or other socially relevant groupings	When known, the location of marriage (municipality and/or region) of the individuals' parents was used as a proxy for the individuals' place of origin. The BALSAC population register is most comprehensive for French Canadians, especially for the Saguenay–Lac-Saint-Jean region, which is our population of interest.
Population characteristics	Genotype data was drawn from the CARTaGENE cohort, which comprises 43,032 participants aged 40-69 (in 2005) randomly recruited in the main urban areas of the Quebec Province, notably in the city of Saguenay, regardless of their ethnic origin and sex or gender. Of those participants, 29,337 individuals were genotyped and 9,405 also have a genealogical link to BALSAC. The latter, mostly of French Canadian origin, were used in this analysis. A second genealogical dataset from BALSAC included 80,348 probands (non-parent individuals) who married in Saguenay–Lac-Saint-Jean between 1931 and 1960.
Recruitment	The BALSAC population register is constructed from civil records (birth, death, marriage certificates) from New France and contemporary Quebec which have been digitalized and reconstructed in genealogies. The CARTaGENE contemporary cohort recruited individuals who reside in one of six metropolitan areas of Quebec (Montreal, Quebec City, Sherbrooke, Saguenay, Gatineau, Trois-Rivières), regardless of their birthplace.
Ethics oversight	This study was approved by the ethics board of the Université du Québec à Chicoutimi (UQAC).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For BALSAC genealogical analyses, individuals were included if they married in Saguenay–Lac-Saint-Jean between 1931 and 1960. For UMAP visualisation, siblings were removed based on kinship >= 0.25. For CARTaGENE genetic and genealogical analyses, HDBSCAN was used alongside UMAP to select the individuals of most likely Saguenay–Lac-Saint-Jean origin on the whole cohort with both genotype and genealogical data.
Data exclusions	Individuals in BALSAC who have genealogical (expected) kinship with less than half the other individuals were excluded to account for completeness. Individuals from CARTaGENE with low quality genotyped data were removed from the analysis.
Replication	The whole genealogy of the region of Saguenay–Lac-Saint-Jean region was used to visualise the population structure. As such, sampling was not considered useful. Whenever a stochastic function was used, a random seed was provided.
Randomization	All available data was used without randomization or sampling, because the whole population structure was analysed.
Blinding	To ensure confidentiality for the BALSAC data, marriage years are rounded up to the nearest 5. For both BALSAC and CARTaGENE, individual anonymity is rigorously maintained through the use of unique, non-identifying codes. Otherwise, blinding was not relevant to the study as the whole population was included in the analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.