

**UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES**

**ÉTUDE GÉNÉTIQUE QUANTITATIVE DE L'ÉVOLUTION  
DES TRAITS D'HISTOIRE DE VIE CHEZ DES POPULATIONS  
PRÉINDUSTRIELLES FRANCO-CANADIENNES**

**THÈSE PRÉSENTÉE  
COMME EXIGENCE PARTIELLE DU  
DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE**

**PAR  
WALID MAWASS**

**AOÛT 2023**

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

**UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES**

**DOCTORAT EN BIOLOGIE CELLULAIRE ET MOLÉCULAIRE (PH. D.)**

**Direction de recherche :**

---

Emmanuel Milot

Directeur de recherche

**Jury d'évaluation de la thèse :**

---

Emmanuel Milot

Directeur de recherche

---

Hugo Germain

Président de jury

---

Dany Garant

Évaluateur externe

---

Patrick Bergeron

Évaluateur externe

Thèse soutenue le 22/6/2023

## REMERCIEMENTS

Tout d'abord, je tiens à remercier le Professeur Emmanuel Milot, Ph.D., pour m'avoir donné l'opportunité d'entreprendre ce projet d'investigation en biologie et génétique évolutive. Le fait de m'avoir choisi pour faire partie de son laboratoire pour mon master et mon doctorat m'a aidé à me développer et à grandir pour atteindre mon objectif de devenir un scientifique et un académique, tout en m'inculquant les valeurs d'autonomie, d'indépendance et de confiance en soi. Son mentorat et ses conseils m'ont aidé à exprimer mes idées tout en m'apprenant comment mener une bonne science. Le contenu de cette thèse n'aurait pas été possible sans sa propre idée originale. Il m'a ouvert la voie pour explorer les questions intrigantes développées dans ce document.

Je tiens également à remercier un certain nombre de personnes extraordinaires que j'ai eu la chance de rencontrer, de fréquenter et de collaborer. Il s'agit notamment (sans ordre particulier) d'Alexandra Doyon, Léo Lavergne, Yannick Noël, Dr. Tommy Harding, Dr. Carla Aimé-Jubin, Camille Julien, Roxane Landry, Diane Martin-Moya, Jessie Beauchemin, Kariane LaRocque, Ariane Durand-Guévin, Audrey Gareau-Léonard.

Je remercie tout particulièrement les scientifiques et chercheurs extraordinaires qui m'ont aidé directement ou indirectement dans mon travail ou dans mon développement en tant que chercheur en biologie et en génétique de l'évolution. Je cite : Dr. Michael Morrissey, Dr. Jacob Moorad, Dr. Jarrod Hadfield, Dr. Alastair Wilson, Dr. Rebecca Sear, Dr. Pierre de Villemereuil. Je remercie très chaleureusement l'étonnante et instructive Dre. Francine M. Mayer qui m'a aidé à comprendre ce qu'il faut pour étudier une population et ses individus sans perdre de vue ce que cela signifie réellement en termes des expériences vécues par ces personnes passées. Son soutien et son aide ont été indispensables à la formulation d'un grand nombre d'idées présentées dans cette thèse.

Enfin, ce travail n'aurait pas été achevé sans le soutien inconditionnel de mon épouse, Fern, et l'amour et les soins infailibles qu'elle m'a prodigués pendant toutes ces années.

## AVANT-PROPOS

### CONTEXTE ET MOTIVATION

Pendant mon adolescence, les livres étaient les portes qui me plaçaient dans des mondes où je pouvais m'aventurer et apprendre des mots, des idées et des perspectives qui m'étaient inconnues à travers mon monde familial. Mon parcours de bibliophile m'a appris le scepticisme et la flexibilité pour regarder et analyser le monde qui m'entoure. La découverte du processus d'évolution constitue une étape importante dans ce parcours. Les esprits sont façonnés par des institutions locales dans leur environnement social. Le système d'enseignement secondaire de Tripoli, au Liban, avait un cursus de biologie sans aucune mention d'évolution. J'ai appris la génétique et l'ADN sans comprendre que c'était la base de l'hérédité de l'organisme et une grande partie du changement évolutif. J'étais complètement fasciné par le sujet de la génétique, et ma motivation pour devenir biologiste se développait lentement dans mon esprit. Cependant, il a fallu un livre commercial pour m'enseigner le concept de l'évolution et comment la génétique fait partie de ce processus biologique.

Mon esprit a été soufflé, hypnotisé par ce que signifiait comprendre que le changement évolutif était la source de la diversité dans la nature et le fil intemporel qui relie chaque être vivant. Ce voyage d'exploration de la nature était étroitement lié à mon voyage spirituel et religieux. Plus j'ai appris les explications naturelles du monde autour de moi, plus je me suis détaché de mon éducation religieuse et de ses explications métaphysiques. Il m'a fallu quelques années pour séparer les deux parties, et à cette époque, ma motivation a grandi exponentiellement. J'étais comblé par la passion et le désir de devenir un expert en biologie évolutive et en génétique. Après avoir abandonné la pression parentale infructueuse qui me poussait à chercher une carrière médicale, je suis entré dans le seul établissement d'enseignement supérieur public du Liban, l'Université Libanaise, pour obtenir un diplôme en sciences de la vie. Malheureusement, même au niveau de l'enseignement supérieur, les références à l'évolution étaient rares. L'apprentissage formel le plus proche du processus évolutif se trouve dans deux cours séparés sur la paléontologie et l'écologie. Néanmoins, j'ai rempli ce vide avec des livres

commerciaux écrits par des biologistes pour en savoir plus sur l'évolution et tous ses aspects.

Comme pour tout Libanais en quête d'éducation et de sécurité, le chemin de l'émigration est pavé par nos parents avant même que nous réalisions ce que signifie quitter notre pays d'origine. Mais j'étais déterminé à poursuivre mes études en biologie évolutive et je suis d'abord allé à Tours, dans la région centrale du Val de Loire, en France, puis à Trois-Rivières, ville riveraine du Québec, au Canada (une drôle de friandise et coïncidence, ces deux villes sont jumelées depuis 1987). Au cours de ces années, j'ai commencé à me plonger dans le côté académique de la biologie évolutive, en lisant des articles scientifiques, en discutant avec des chercheurs et, enfin, en menant mes propres recherches. Le premier modèle d'étude que j'ai étudié à Tours était la minuscule guêpe parasitaire *Nasonia vitripennis*. Ça faisait partie d'un projet visant à comprendre l'impact de l'exposition des mâles à des températures élevées sur la sex-ratio des femelles. Cela n'a duré que quelques mois, mais j'ai eu un aperçu du travail de recherche et je tenais à le faire. Ensuite, j'ai dû faire face à la décision administrative du département d'écologie et de biologie évolutive (EEB) de l'Université François-Rabelais d'effectuer des admissions sélectives des étudiants souhaitant poursuivre en deuxième année de master en EEB. Par conséquent, on ne m'a pas choisi et j'ai dû chercher d'autres solutions. L'une de celles-ci était l'offre du professeur Emmanuel Milot. Il m'a fait passer un entretien et m'a offert la possibilité de rejoindre son laboratoire à l'université du Québec à Trois-Rivières (UQTR), pour travailler sur un projet financé qui explore l'évolution des traits d'histoire de vie dans la population historique du Québec.

La décision a été monumentale à plusieurs égards. Non seulement étais-je maintenant un étudiant de recherche à temps plein travaillant sur la génétique évolutive des humains, mais je gagnais l'autonomie et le renforcement des relations. J'ai acquis de nombreuses compétences, notamment en informatique, en organisation et en communication, au laboratoire de génétique des populations d'Emmanuel Milot et j'ai développé mes capacités à mener des recherches scientifiques. Mon directeur a travaillé d'arrache-pied pour m'appuyer, comme l'ont fait mes collègues et mon département. La confiance que m'a insufflée mon directeur au cours des deux premières années a

catalysé mon entrée rapide dans le programme de doctorat, en plus de la croyance croissante en moi-même, en mon parcours universitaire et en mes objectifs de recherche. Le travail effectué pendant mon séjour au laboratoire de génétique des populations à l'UQTR ressemble à la fois à un aboutissement et à un point de départ : l'aboutissement d'un travail acharné, d'épreuves et de persévérance pour arriver à rédiger une thèse et à obtenir mon doctorat, et le point de départ de ma carrière universitaire et professionnelle en tant que scientifique désireux de trouver des questions plus intrigantes à partir de réponses possibles à des problèmes intrigants. En bref, j'ai enfin pu marier ma passion pour la biologie évolutive et ma curiosité pour la génétique dans ce travail, dans l'espoir d'être enfin reconnu comme un expert en génétique évolutive. Tout cela s'inscrit dans le cadre d'un grand principe qui m'a toujours guidé, depuis le jour où j'ai ouvert et lu un livre pour la première fois jusqu'au jour où j'ai pu rédiger mon travail de recherche, et c'est la curiosité : la curiosité d'apprendre, d'expliquer, de comprendre, d'échouer, de coopérer, de communiquer et de grandir.

## **MISE EN GARDE ET RECONNAISSANCES**

Avant de commencer le texte principal, je présente quelques mises en garde concernant les règles du texte et le contenu général qui, je l'espère, amélioreront la compréhension de cette thèse et la contextualiseront, suivis de reconnaissances importantes spécifiques au sujet principal de cette thèse :

- Mise en garde #1 : L'utilisation du pronom "je" dans le texte concerne mes propres décisions et les choix que j'ai pris. Dans ce cas, je suis le seul responsable de toute décision erronée en lien avec cette thèse. L'utilisation du pronom "nous" se produit quand le discours est explicatif, comme la présentation de preuves. Par "nous", on entend "moi et le lecteur". J'utilise également strictement le "nous" lorsque je parle d'une idée ou d'un résultat produit en collaboration, par exemple dans le contexte d'un travail de recherche conjoint avec quelqu'un d'autre. Cette voix est celle par défaut pour cette thèse.
- Mise en garde #2 : Les données explorées et analysées dans les chapitres de cette thèse concernent des individus humains qui ne sont plus en vie. Plus

précisément, les données décrivent des événements de vie d'individus passés. Il s'agit de descriptions réductrices d'expériences réelles, vécues par des êtres humains qui ont laissé leur empreinte sur ce monde d'une manière ou d'une autre. L'application de ce type de donnée aux fins de nos analyses se fait sans aucune connaissance des noms réels ou des détails personnels. Par conséquent, j'ai mené les recherches décrites dans ce mémoire sans aucun préjugé ou discrimination, au mieux de mes capacités.

- Mise en garde #3 : Conformément aux efforts récents visant à promouvoir la visibilité de la diversité des sexes dans les populations humaines, il est important de noter que les individus étudiés dans ce travail scientifique sont assignés ou considérés comme femelles d'après les actes utilisés pour reconstruire leur histoire de vie. Par conséquent, je ne prétends pas vérifier ou connaître leur identité sexuelle et leurs expériences réelles. Je note cependant que ces mêmes individus peuvent être désignés comme "femmes" dans les articles de recherche présentées ci-dessous. Ce choix reflète une convenance dans la description de ces personnes et ne constitue pas une détermination définitive de leur identité sexuelle qui m'est inconnue.
- Mise en garde #4 : Les déductions et interprétations présentées dans cette thèse sont basées largement sur des analyses statistiques. Il convient donc de reconnaître que toute déclaration biologique faite à la suite d'une telle analyse est limitée par les incertitudes inhérentes aux analyses statistiques. Pour être clair, je ne considère pas les différents résultats rapportés ici comme la représentation absolue de la réalité. Ils représentent plutôt les explications les plus probables détectés par les modèles construits et testés sur des données observées.

Maintenant, j'aimerais reconnaître que, en tant qu'immigrant au Québec, je ne peux pas comprendre les expériences réelles et vécues des êtres humains étudiés dans cette thèse, ainsi que les subtilités de leur culture et de leurs traditions. Je tiens à reconnaître que leurs vies et leurs histoires individuelles ont contribué à faire progresser notre compréhension des processus évolutifs, et je leur en serai toujours reconnaissant. Ma gratitude s'étend également aux nombreuses personnes et groupes qui ont travaillé



sans relâche pour documenter ces histoires de vie, les informatiser et les maintenir. Sans leur travail, cette entreprise scientifique serait impossible. Enfin, il est impératif pour moi de reconnaître, dans le cadre de mon principe anticolonial et anti-impérialiste, que l'établissement des premiers Français dans le Québec d'aujourd'hui n'aurait pas été possible sans le déplacement de nombreux peuples autochtones qui vivaient là depuis des milliers d'années et sans la destruction subséquente de leurs modes de vie, ce qui inclut (et ne se limite peut-être pas) aux Premières Nations suivantes : Abénakis, Algonquins, Atikamekw, Béothuks, Cris, Innus, Inuits, Iroquois, Malécites, Mi'kmaq, Mohawks, Naskapis et Hurons-Wendats.

## RÉSUMÉ

La mesure des changements génétiques associés à l'adaptation dans une population naturelle est un objectif permanent pour les biologistes évolutives. La sélection naturelle, un mécanisme de changement évolutif, peut conduire à une adaptation en un petit nombre de générations, c'est-à-dire évolution contemporaine. La génétique quantitative fournit une théorie et des méthodes permettant d'extraire des informations sur la matière première sur laquelle la sélection naturelle peut agir, la variance génétique additive. Jusqu'à présent, Milot *et al.* (2011) restent la seule étude chez l'humain à montrer de manière robuste l'évolution contemporaine en réponse à la sélection, en utilisant la population préindustrielle de l'île aux Coudres (IAC). Il a été déterminé par une analyse génétique quantitative combinée à des informations généalogiques que l'âge à la première reproduction (AFR) des femmes mariées, qui a évolué de 26 ans à 22 ans en 5 à 8 générations, était partiellement dû à un changement génétique en réponse à la sélection.

La généralité du changement microévolutif est difficile à démontrer du fait du manque de répliques naturels permettant de comparer les changements évolutifs dus à la sélection pour le même trait. Nous avons vérifié si le changement microévolutif de l'AFR sous forme d'une réponse à la sélection de l'AFR se produisait également dans d'autres populations canadienne-française préindustrielle considérées comme des semi-répliques naturelles. On s'attend à ce que le moment de la reproduction précoce soit sous sélection dans une population durant une expansion démographique. En utilisant des méthodes basées sur la régression et deux mesures différentes de fitness relatif, nous avons trouvé une forte sélection linéaire pour des valeurs plus précoces de l'AFR dans les trois populations étudiées. Une analyse génétique quantitative a confirmé que juste dans une seule population, l'AFR et le fitness relatif ont une variance génétique additive décente. Les prédictions indiquent une faible réponse génétique attendue à la sélection naturelle de l'AFR dans les trois populations. Le fait que le changement temporel détecté dans les valeurs moyennes de 'breeding values' de l'AFR soit tout aussi probable dans un scénario de dérive aléatoire seule constitue une démonstration robuste de la stase évolutive. Des informations généalogiques profondément enracinées sont disponibles pour la population historique canadienne-française, et il est vital de déterminer la fiabilité de cet ensemble de données en termes d'estimation des paramètres génétiques quantitatifs. Nous avons combiné différents scénarios de simulation avec les dernières méthodes d'estimation génétique quantitative basées sur les statistiques bayésiennes pour déduire le degré de fiabilité de plusieurs pedigrees canadien-français reconstruits. L'analyse de puissance et de précision effectuée sur les pedigrees reconstruits a révélé que la plupart des ensembles de données étaient suffisamment puissants pour détecter un effet génétique simulé. La précision et l'exactitude ont beaucoup souffert lorsque le modèle n'incluait pas toutes les sources de similarité entre les apparentés (par exemple, la similarité due à l'environnement familial), quelle que soit la taille de l'échantillon et la profondeur du pedigree. Finalement, nous avons émis l'hypothèse que les fluctuations des conditions environnementales pourraient induire des

changements dans l'architecture génétique de l'AFR et donc dans son taux de changement évolutif. Nous avons testé cette hypothèse en utilisant la population IAC en incorporant le taux de mortalité infantile comme indicateur de la difficulté des conditions au début de la vie. Notre approche était basée sur un test quantitatif de la présence d'interactions génotype-environnement (GxEs). Nous avons également testé si l'hétérogénéité environnementale avait un impact sur le taux d'évolution de l'AFR en appliquant une approche « état de caractère ». Nos résultats ont détecté la présence de GxEs sous-jacents à la variation de l'AFR et du fitness relatif dans cette population. Sur la base de nos prédictions utilisant la covariance Roberston-Price, les GxEs ont conduit à une augmentation dans le changement génétique attendue par génération de l'AFR dans des conditions difficiles au début de la vie par rapport à des conditions plus bénignes.

En conclusion, la preuve empirique du rôle des GxEs dans la contrainte ou la facilitation de la réponse à la sélection a un impact substantiel sur des sujets tels que la réponse au changement climatique et la gestion des espèces invasives. L'absence de généralisation du changement évolutif des traits d'histoire de vie par la sélection naturelle entre des populations apparentées en expansion révèle à la fois le rôle crucial des conditions locales et l'impact de facteurs non détectés qui limitent la réponse à la sélection et contraignent le changement phénotypique dans la nature.

**Mots-clés :** Évolution contemporaine, génétique quantitative, sélection naturelle, héritabilité, statistiques bayésiennes, stase évolutive, généalogies canadiennes-françaises, interactions génotype-environnements, âge à la première reproduction, fitness.

## TABLE DES MATIÈRES

<b>REMERCIEMENTS .....</b>	<b>ii</b>
<b>AVANT-PROPOS .....</b>	<b>iv</b>
<b>RÉSUMÉ.....</b>	<b>ix</b>
<b>LISTE DES TABLEAUX.....</b>	<b>xiii</b>
<b>LISTE DES FIGURES .....</b>	<b>xiv</b>
<b>LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES .....</b>	<b>xvi</b>
<b>LISTE DES SYMBOLES.....</b>	<b>xviii</b>
<b>CHAPITRE I</b>	
<b>INTRODUCTION.....</b>	<b>1</b>
<b>CHAPITRE II</b>	
<b>LACK OF GENETIC RESPONSE TO SELECTION IN A HUMAN LIFE HISTORY TRAIT IN NATURAL REPLICATES UNDERGOING EXPANSION .....</b>	<b>36</b>
2.1 Contribution des auteurs .....	36
2.2 Résumé de l'article .....	36
2.3 Lack of genetic response to selection in a human life history trait in natural replicates undergoing expansion.....	38
Abstract.....	38
Introduction.....	39
Materials and methods.....	44
Results .....	50
Discussion.....	53
References.....	58
Tables.....	66
Figures .....	70
<b>CHAPITRE III</b>	
<b>ASSESSING THE IMPACT OF PEDIGREE QUALITY ON THE VALIDITY OF QUANTITATIVE GENETIC PARAMETER ESTIMATES ....</b>	<b>73</b>
3.1 Contribution des auteurs .....	73
3.2 Résumé de l'article .....	73

3.3	Reliability of large-scale human genealogies for the estimation of quantitative genetic parameters .....	74
	Abstract.....	74
	Introduction.....	75
	Materials and methods.....	80
	Results .....	88
	Discussion.....	90
	References.....	95
	Tables.....	101
	Figures .....	104
<b>CHAPITRE IV</b>		
<b>GENOTYPE-BY-ENVIRONMENT INTERACTIONS MODULATE THE RATE OF MICROEVOLUTION IN REPRODUCTIVE TIMING IN HUMANS.....</b>		
	4.1 Contribution des auteurs.....	111
	4.2 Résumé de l'article .....	111
	4.3 Genotype-by-environment interactions modulate the rate of microevolution in reproductive timing in humans .....	112
	Abstract.....	112
	Introduction.....	110
	Materials and methods.....	114
	Results .....	124
	Discussion.....	126
	Conclusion.....	130
	References.....	131
	Tables.....	139
	Figures .....	142
<b>CHAPITRE V</b>		
<b>CONCLUSION.....</b>		
<b>RÉFÉRENCES BIBLIOGRAPHIQUES.....</b>		
		172

## LISTE DES TABLEAUX

Tableau	Page
2.1 Valeurs moyennes et écarts types des traits de l'histoire de vie des femmes mariées dans les trois populations étudiées.....	65
2.2 Résultats du modèle animal bivarié avec l'AFR et le fitness comme variables de réponse pour chaque population.....	66
2.3 Résultats du modèle animal bivarié continués.....	67
2.4 Indices bayésiens pour des paramètres du modèle animal.....	68
3.1 Tableau montrant les facteurs clés qui affectent la qualité et la performance du pedigree.....	100
3.2 Statistiques sommaires et propriétés généalogiques des pedigrees.....	101
4.1 Résultats du modèle animal de régression aléatoire bivariée ajustant la distribution IMR comme une fonction polynomiale de premier ordre.....	139
4.2 Indices Bayésiens des paramètres du modèle RRAM bivarié.....	139
4.3 Résultats du modèle animal catégorique bivarié.....	140
4.4 Indices Bayésiens des paramètres du modèle catégorique bivarié.....	140
4.5 Matrice de covariances et corrélations estimé du modèle RRAM bivarié.....	141

## LISTE DES FIGURES

Figure	Page
1.1 Paysage de fitness avec et sans contrainte.....	4
1.2 Pedigree: jeux de données et matrice d'apparentement.....	12
1.3 Les interactions génotype-environnements.....	18
1.4 Carte de Nouvelle-France.....	22
1.4 Carte des régions du Québec où des données de registres de baptêmes, de mariages et de sépultures sont disponibles.....	25
2.1 Sélection phénotypique dans AFR sous forme de régression de fitness relative sur AFR.....	69
2.2 Tendance temporelle de l'AFR moyen dans les trois populations étudiées.....	70
2.3 Le changement évolutif par génération attendu et observé dans AFR chez les trois populations étudiées.....	71
3.1 Schéma de la conception du flux de travail.....	102
3.2 Estimations de la variance génétique additive pour chacun des six pedigrees sur la base du scénario 1.....	103
3.3 Estimations de la variance génétique additive pour chacun des six pedigrees sur la base du scénario 1 avec taux d'erreur de 1%.....	104
3.4 Estimations de la variance génétique additive et de la variance génétique maternelle pour chacun des six pedigrees selon le scénario 2.....	105
3.5 Estimations de la variance génétique additive et de la variance environnementale maternelle pour chacun des six pedigrees sur la base du scénario 3.....	106
3.6 Distributions postérieures des paramètres estimés de génétique quantitative (avec l'intervalle HPD à 95%) pour chacun des six pedigrees sur la base du scénario 4 avec corrélation génétique de 0.1.....	107

3.7 Distributions postérieures des paramètres de génétique quantitative estimés (avec l'intervalle HPD à 95 %) pour chacun des six pedigrees sur la base du scénario 4 avec corrélation génétique de 0.5.....	108
4.1 Variation dans les traits d'histoire de vie des femmes mariés de l'IAC entre les environnements IMR.....	142
4.2 Covariance et corrélations génétiques dans et entre les environnements IMR pour AFR et fitness relatif.....	143
4.3 Effet attendu de l'environnement IMR sur la covariance RP.....	144



## LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

AFR	Age at first reproduction, âge à la première reproduction
ALR	Age at last reproduction, âge à la dernière reproduction
BE	Breeder's equation, équation de l'éleveur
BV	Breeding value, valeur génétique
EPP	Extra-pair paternity, paternité extra-paire
FC	French-Canadian, canadien-français
G×E	Genotype-by-environment interaction, interaction génotype-environnement
GWAS	genome-wide association study, étude d'association pangénomique
HPD	Highest posterior density, densité postérieure la plus élevée
IAC	île aux Coudres
IGE	Indirect genetic effects, effets génétiques indirects
IMR	Infant mortality rate, taux de mortalité infantile
LH	Life history, histoire de vie
LHT	Life history theory, théorie d'histoire de vie
LRS	Lifetime reproductive success, succès reproductif à vie
MCMC	Markov Chain Monte Carlo
pd	probability of direction, probabilité de direction

PRDH	Programme de recherche en démographie historique
QG	Quantitative genetics, génétique quantitative
REML	Restricted maximum likelihood, maximum de vraisemblance restreint
RMSE	Root mean squared error, erreur quadratique moyenne
RP	Robertson-Price covariance or identity, covariance ou identité Robertson-Price
ROPE	Region of practical equivalence, région d'équivalence pratique
RRAM	Random regression animal model, modèle animal de régression aléatoire
SE	Standard error, erreur standard
SI	Supplementary information, informations supplémentaires
STS	Secondary theorem of selection, théorème secondaire de sélection
GLMM	Generalized linear mixed-effects model, modèle linéaire généralisé à effets mixtes
wd	environmental window, fenêtre environnementale

## LISTE DES SYMBOLES

$f_a, f_m$	Fonction de régression aléatoire pour effets additifs $a$ , effets maternels $m$
$\sigma_P^2, V_P$	Variance phénotypique
$\sigma_A^2, V_A$	Variance génétique additive
$\sigma_M^2, V_M$	Variance maternelle ou environnement familiale
$\sigma_E^2, V_E$	Variance environnementale ou résiduelle
$\sigma_D^2, V_D$	Variance due à la dominance
$\sigma_I^2, V_I$	Variance due aux interactions entre gènes ou épistasie
$\sigma_A, Cov_A$	Covariance génétique additive
$h^2$	Héritabilité au sens étroit
$r_g$	Corrélation génétique
$CV_A$	Coefficient de variation génétique additive
$h^2$	Héritabilité au sens étroit
$R$	Réponse à la sélection dans une génération
$S, s$	Différentiel de sélection
$\Delta z$	Changement dans un phénotype $z$ en une génération
$G$	Matrice de variance-covariance génétique additive
$\beta$	Gradient de sélection linéaire ou coefficient de régression partielle
$\gamma$	Gradient de sélection quadratique

$a$	Effet génétique additive
$e$	Effet environnementale ou résiduelle
$\mu$	Moyenne de la population
$\theta_{x,y}$	Coefficient de consanguinité entre individu $x$ et $y$
$v_o$	Valeur reproductive à la naissance
$\lambda$	Taux de croissance de la population
$w$	Fitness relatif
$\bar{X}$	Valeur moyenne
$n$	Taille de l'échantillon

## CHAPITRE I

### INTRODUCTION

*« Considérée à la lumière de l'évolution, la biologie est peut-être intellectuellement la science la plus satisfaisante et la plus inspirante. Sans cette lumière, cela devient un tas de faits divers - certains d'entre eux intéressants ou curieux mais ne faisant aucune image significative dans son ensemble. » Théodose Dobzhansky (The American Biology Teacher, 1973)*

*« Je suis sûr que nos points de vue sur l'évolution auraient été très différents si les biologistes avaient étudié la génétique et la sélection naturelle avant et pas après que la plupart d'entre eux aient été convaincus que l'évolution s'était produite. » J.B.S. Haldane (Attribué.)*

Le fait de l'évolution explique l'immense diversité de la nature vivante et l'interconnexion entre les espèces (Mayr 2001). L'autre aspect de la vie qui n'avait pas d'explication satisfaisante jusqu'à la découverte par Darwin et Wallace du mécanisme de la sélection naturelle est l'incroyable adéquation des organismes à leurs environnements respectifs, bien adaptés dans leur morphologie, leur physiologie et leur comportement (Gardner 2009). La description de la sélection naturelle par Darwin dans son *Origine des espèces* a fait passer l'idée du processus évolutif du paradigme transformiste (notamment le lamarckisme) à celui variationnel (Darwin 1859). Pour qu'un changement évolutif se produise par sélection naturelle, trois conditions sont nécessaires : 1- l'existence d'une variation entre les individus dans leur phénotype ; 2- la transmission génétique héritable de cette variation de telle sorte que la progéniture, en moyenne, ressemble plus à ses parents qu'aux individus moins apparentés ; 3- la reproduction différentielle, à savoir les individus portant différents phénotypes laissent un nombre différent de descendants dans la génération suivante (Darwin 1859). Il est important de rappeler que le changement évolutif se produit aussi par la dérive génétique causée par le tirage aléatoire des gamètes d'une génération pour en produire une autre, ainsi que par d'autres mécanismes comme la mutation et le flux génétique (Falconer et

Mackay 1996). Cependant, pour rendre compte des changements évolutifs adaptatifs au sein d'une population<sup>1</sup>, il faut qu'il y ait une variation génétique héritable à sélectionner.

Néanmoins, la connaissance de ces trois conditions du changement évolutif par sélection naturelle ne dit rien, en elle-même, sur l'adaptation des organismes à leur environnement. Pour ce faire, Darwin a introduit un quatrième élément, soit les pressions environnementales qui sous-tendent la sélection naturelle. Même si Darwin a fait référence – métaphoriquement – à « la lutte pour l'existence », cette formule simple signifiait qu'en réalité, le changement adaptatif résulte de la relation entre la variation phénotypique et la survie/reproduction différentielles des organismes (Lewontin 1970). Ainsi, l'adaptation face aux pressions de l'environnement sur la survie et la reproduction des organismes se comprend comme étant relative aux autres individus. C'est de ce principe que découle le concept de « fitness »<sup>2</sup> dans la théorie de l'évolution (Ariew et Lewontin 2004), défini comme le nombre relatif de descendants produits par un individu (ou un génotype) par rapport à la moyenne de la population (Rosenberg 1983). Le fitness d'un individu peut se décomposer en plusieurs composantes, comme le fait de survivre jusqu'à un certain âge (p. ex. jusqu'à la maturité), le succès d'appariement ou encore le nombre de descendants produits (fécondité). Par conséquent, le fitness est lui-même étroitement lié aux traits déterminant ces composantes, tels que les traits d'histoire de vie (De Jong 1994).

Cette thèse porte sur l'évolution des traits d'histoire de vie dans une population humaine de l'époque contemporaine. Ce sujet est à l'intersection de plusieurs domaines, chacun nécessitant sa propre brève introduction. Ainsi, je vais donner une vue d'ensemble des différents éléments qui font partie de ce sujet complexe. Les sections qui suivent traitent de : i) l'évolution contemporaine par sélection naturelle ; ii) le cadre théorique de la génétique quantitative ; iii) les interactions génotype-environnement ; iv) l'évolution de l'histoire de vie humaine ; v) la population historique canadienne-française (en mettant l'emphase sur celle de l'île aux Coudres, dans Charlevoix).

---

<sup>1</sup> Nous utilisons ici la définition opérationnelle d'une population comme un ensemble d'individus qui partagent des ancêtres génétiques et vivent dans la même zone géographique

<sup>2</sup> Stearns (1976) propose une définition amusante du fitness qui rend compte de son statut dans les sciences de l'évolution : « *Fitness* : quelque chose que tout le monde comprend mais personne ne peut définir avec précision. »

## 1.1 Évolution contemporaine par sélection naturelle

Le matériel génétique transmis par les parents à leur progéniture se présente physiquement sous forme d'ADN (acide désoxyribonucléique), contenu et emballé dans les chromosomes. Chaque individu possède ce patrimoine génétique qui influence, d'une manière ou d'une autre et à différents degrés, son phénotype (Futuyma 2013). Selon le modèle le plus simple, un phénotype peut être réduit à une somme des effets génétiques et environnementaux :  $Phénotype (P) = Génotype (G) + Environnement (E)$ . Cependant, il convient de noter que le phénotype observé se construit par l'expression d'un génotype dans un environnement donné, qui peuvent interagir<sup>3</sup> (Falconer et Mackay 1996). Ce modèle conduit à une formulation centrée sur la variance requise pour que le changement évolutif se produise au sein d'une population : 1) une sélection qui fait que la distribution phénotypique des parents n'est pas la même que celle dans la population en général (incluant les individus non reproducteurs), et 2) une variance génétique héritable (donc additive) qui sous-tend cette distribution phénotypique, de sorte que les changements induits par la sélection soient transmis à la génération suivante (Lynch et Walsh 1998).

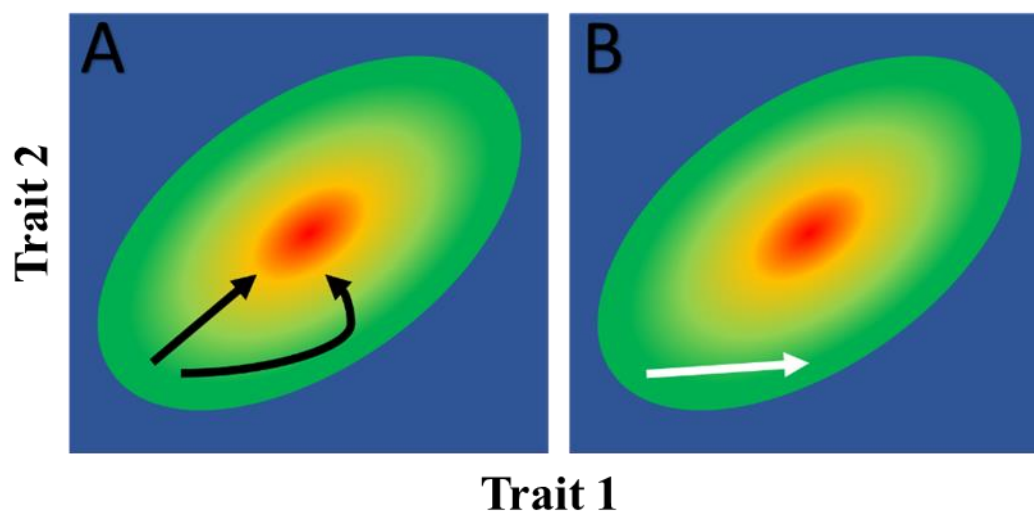
Pour illustrer comment la sélection naturelle conduit à l'adaptation, Sewall Wright (1932) a introduit la métaphore topographique d'un paysage accidenté où les vallées et les sommets représentent des valeurs de phénotypes qui sont, respectivement, moins ou mieux adaptées à l'environnement, et où la sélection opérant sur la variation génétique pousse les populations vers des « pics adaptatifs »<sup>4</sup> (voir la Figure 1.1). Cette « marche » vers un état plus adapté est toutefois difficile à démontrer dans les populations naturelles, car cela nécessite la détection de changements évolutifs se produisant « en temps réel » sous l'action de la sélection naturelle. Cela a sans doute contribué à répandre, même parmi les biologistes, l'idée que la sélection n'est pas fréquente dans les populations naturelles. Selon Endler (1986), cela n'est pas le cas et des compilations

---

<sup>3</sup> L'environnement peut également contenir d'autres génotypes. Cela donne lieu à des effets associatifs sur le phénotype dus aux interactions entre génotypes. Pour plus de détails, nous renvoyons les lecteurs au modèle proposé par Griffing (1967)

<sup>4</sup> Nous notons que le paysage de fitness est beaucoup plus complexe que celui présenté ici, mais peut toujours servir de représentation simplifiée de l'action adaptative de la sélection naturelle

d'estimations empiriques le montrent, tout en soulignant qu'une sélection forte (gradient de sélection standardisé  $> 0,5$ ) peut ne pas être particulièrement fréquente en nature (Kingsolver *et al.* 2001 ; Kinnison et Hendry 2001). Une méta-analyse de Kingsolver *et al.* (2012) révèle un portrait clair : i) la sélection directionnelle se produit régulièrement dans les populations naturelles, agissant sur des traits tels que la morphologie (p. ex. la taille du corps) et l'histoire de vie (p. ex. l'âge à la première reproduction); ii) la sélection stabilisatrice, qui tend à éliminer les phénotypes s'écartant trop d'une valeur moyenne adaptée, est rarement détectée empiriquement bien qu'elle devrait être fréquente en théorie (mais voir Sanjak *et al.* 2018); iii) la sélection sur la composante de fécondité du fitness est généralement aussi importante que la sélection sur la composante de survie (Corbett *et al.* 2018); iv) il existe une immense variation spatiale et temporelle dans la force et la direction de la sélection s'exerçant sur un trait donné dans la nature, bien qu'il soit parfois difficile de distinguer empiriquement cette variation réelle dans la sélection de l'erreur de mesure due à l'échantillonnage (Morrissey et Hadfield 2012; Kokko *et al.* 2017).



**Figure 1.1** Paysage adaptatif pour deux traits dans une population présentant un seul sommet de fitness (gradient du vert au rouge allant du fitness moyen le plus bas au plus élevé). (A) La sélection naturelle (flèches noires) déplace la population vers des combinaisons de traits la positionnant à des pics adaptatifs plus élevés (zone rouge), par le chemin linéaire le plus court ou par des chemins plus complexes ; (B) Des contraintes génétiques et écologiques peuvent empêcher la sélection d'amener la population tout en



haut du sommet adaptatif en canalisant la trajectoire qu'elle peut suivre (flèche blanche).

Une idée fausse veut que le changement adaptatif par sélection naturelle en réponse à un changement environnemental ne puisse qu'être lent et est donc impossible à observer directement, car il nécessite des milliers à des millions d'années pour s'accomplir. Elle repose sur le postulat (abusif) que tout changement dans la fréquence des génotypes en réponse à la sélection sera lui-même trop lent pour se traduire par un changement adaptatif du phénotype (microévolution) rapide. De tels changements concerneraient davantage les processus de spéciation et de divergence des lignées (c'est-à-dire la macroévolution; Gould 1994). On a longtemps pensé que la plasticité phénotypique était le seul mécanisme qui pouvait conduire à un changement à la fois détectable et rapide dans la distribution d'un phénotype dans une population, en réponse à des changements environnementaux rapides (West-Eberhard 1989). Bien que la plasticité soit un mécanisme majeur de réponse phénotypique à court terme (Ghalambor *et al.* 2007), on sait aussi que le processus de microévolution par sélection naturelle au sein d'une population peut se produire en quelques générations seulement, avec une ampleur assez grande pour être détectable par le suivi de populations en nature sur quelques années ou décennies seulement (Boag et Grant 1981, Coltman *et al.* 2003, Stuart *et al.* 2014).

Le changement évolutif par sélection naturelle peut être modélisé mathématiquement, dans un cadre de génétique quantitative, à l'aide de l'équation du « sélectionneur » (*breeder's equation*). Celle-ci permet de prédire, à l'intérieur de certaines limites, le changement attendu ( $R$ ), dans la génération des rejetons, dans la valeur moyenne d'un trait, en réponse à la sélection (directionnelle) s'opérant dans la génération parentale (Lush 1937 ; Falconer et Mackay 1996) :

$$R = h^2 S, \quad (1)$$

où  $S$  représente le différentiel de sélection et  $h^2$  l'héritabilité, soit la proportion de la variance phénotypique ( $V_P$ ) due aux effets additifs des allèles. L'héritabilité est relative à une population donnée ( $h^2 = \frac{V_A}{V_P}$ ; Wray et Visscher 2008). Par conséquent, si nous connaissons le déterminisme génétique d'un trait ainsi que la force et le signe (positif ou

négatif) de la sélection ( $S$ ) s'exerçant sur celui-ci dans la génération des parents, on doit pouvoir prédire, en principe, le changement phénotypique attendu dans la génération suivante par l'équation (1) (Lynch et Walsh 1998). Cependant, l'emploi de l'équation du sélectionneur n'est pas approprié lorsqu'il existe une covariance génétique entre plusieurs traits sous sélection (Morrissey *et al.* 2010). Dans ce cas, la réponse microévolutive du trait focal dépendra à la fois du signe et de l'ampleur de sa covariance génétique avec d'autres traits et de la sélection agissant sur ces derniers (Lande et Arnold 1983). Une forme multivariée de l'équation (1) permet de prendre cela en compte (Lande, 1979) :

$$\Delta \mathbf{z} = \mathbf{G}\boldsymbol{\beta} \quad (2)$$

où  $\Delta \mathbf{z}$  est le vecteur des réponses prédites à la sélection (c.-à-d. la forme multivariée de  $R$  dans l'équation (1)),  $\mathbf{G}$  la matrice de variance-covariance génétique additive et  $\boldsymbol{\beta}$  le vecteur des gradients de sélection.

En étudiant des populations de pinsons de Darwin aux îles Galápagos, Grant et Grant (1993, 1995) ont pu détecter des changements microévolutifs locaux, notamment dans la taille du bec des oiseaux. « L'observabilité » de cette microévolution tient à la mesure empirique, faite par les chercheurs sur le terrain, des pressions de sélection agissant sur la population, de l'héritabilité du trait en question et des changements phénotypiques observés de génération en génération (c.-à-d. les trois paramètres de l'équation du sélectionneur). Cependant, la prédiction de la microévolution sur plusieurs générations est compliquée par les fluctuations environnementales pouvant survenir à court terme et faire le varier le différentiel de sélection d'une génération à l'autre (voir p. ex. Gienapp *et al.* 2014). Dans de tels cas, il s'avère difficile, voire impossible, de prédire comment la distribution des phénotypes aura changé dans la population après plusieurs générations. Coltman *et al.* (2003) ont mis au jour une réponse évolutive sous la forme d'une réduction de la longueur des cornes et du poids de mouflons d'Amérique (*Ovis canadensis*) en Alberta, au Canada, dans une population qui avait été soumise à une intense chasse au « trophée ». Ainsi, en peu de générations (30 ans), la sélection a pu causer un changement génétique suffisant pour induire la tendance phénotypique observée dans les cornes (une conclusion vérifiée par de nouvelles analyses de Pigeon *et*

*al.* 2016). La démonstration empirique est désormais faite que la détection de changements microévolutifs locaux est possible « en temps réel », avec des méthodes appropriées de génétique quantitative (Kruuk *et al.* 2014). C'est ce que Hendry et Kinnison (1999) qualifient d'évolution contemporaine<sup>5</sup>.

Les démonstrations de l'évolution contemporaine proviennent de multiples taxons, y compris les humains, et couvrent diverses conditions environnementales (Postma 2014). Toutefois, chaque démonstration empirique se rapportant à évènement particulier de microévolution, un des défis de la biologie évolutive consiste à déterminer le degré de généralité des processus microévolutifs, et ce, pour des combinaisons population-trait contemporaines soumises à des conditions environnementales similaires. En effet, il est rare de trouver dans la nature des systèmes adéquats pour tester la répétabilité d'un changement microévolutif dans des environnements qui seraient spatialement indépendants (Bolnick *et al.* 2018). Autrement dit, les changements microévolutifs détectés sont associés à des combinaisons population-trait-environnement pour lesquelles il n'est pas facile d'identifier des « répliques naturels » à étudier. Cela s'applique également aux cas présumés « d'évolution parallèle » de populations ayant divergé à peu près en même temps d'un ancêtre commun et qui sont aujourd'hui isolées spatialement les unes d'autres (Schluter *et al.* 2004, Elmer *et al.* 2014). Schluter *et al.* (2004) ont démontré une hérédité parallèle de deux caractères, le nombre de plaques latérales et la forme du corps, dans deux lignées d'épinoches à trois épines qui se sont adaptées indépendamment aux ruisseaux d'eau douce sur les côtés opposés de l'océan Pacifique. Notamment, la réduction du nombre de plaques en eau douce implique une substitution au même locus majeur dans les deux lignées. Cela est l'objet de chapitre II de cette thèse.

## 1.2 Génétique quantitative évolutive

Le phénotype est le résultat des interactions, à de multiples niveaux, entre le génotype et l'environnement (Barton *et al.* 2007). La cartographie (mapping) de la relation entre le génotype et le phénotype reste un sujet d'étude très complexe. Même

---

<sup>5</sup> Selon Hendry & Kinnison (1999), ce terme est généralement utilisé pour désigner des changements phénotypiques récents ou en cours, d'origine génétique ou héréditaire, survenant dans les populations naturelles

avec l'avènement de la révolution génomique, il n'y a pas de description définitive du lien causal qui existe entre le génotype et le phénotype, qui est souvent non linéaire (Pigliucci 2010). Des études portant sur le repliement de l'ARN, fonction des protéines et simulations computationnelles ont montré que la relation  $G \rightarrow P$  difficile à décortiquer (Martin et Wagner 2007; Cowperthwaite et Meyers 2007). Un trait quantitatif en est un qui se mesure sur une échelle ordinale et présentant une variation phénotypique d'origine génétique et/ou environnementale (Lynch et Walsh 1998). Le trait peut se mesurer par des valeurs discrètes, comme le nombre de rejetons, ou continues, comme la taille d'un individu. Un trait quantitatif présente généralement une variation au sein et entre les populations (Barton et Keightley 2002), qu'il soit de nature morphologique, physiologique, comportementale, lié à l'histoire de vie, ou encore moléculaire, tel que le degré d'expression d'un gène mesuré par la production d'ARN messenger.

Il existe deux grandes catégories de méthodes pour étudier les changements évolutifs dans les populations : l'approche par génotype mesuré et l'approche par génotype non mesuré (voir Queller 2017 pour une discussion récente). De base, les modèles de génétique quantitative font partie de cette dernière approche, bien que des approches y intégrant les données génomiques ont commencé à émerger (Jensen *et al.* 2014). Le domaine de la génétique quantitative s'intéresse à l'étude de la variation interindividuelle au sein d'une population donnée et à la mesure dans laquelle des causes génétiques et non-génétiques la façonnent (Falconer et Mackay 1996; Walsh et Lynch 2018). La génétique quantitative classique repose sur une prémisse très intuitive : si des individus très étroitement apparentés, qui partagent donc beaucoup de gènes, sont phénotypiquement plus semblables les uns aux autres que des individus moins apparentés, et partageant donc moins de gènes, c'est que la variance phénotypique ( $V_P$ ) se compose en partie d'une composante génétique additive ( $V_A$ ) (Kruuk 2004; Roff 2012). Selon le modèle le plus simple,  $V_P = V_A + V_E$ , où  $V_E$  est la variance expliquée par les facteurs environnementaux. Les composantes de  $V_P$  peuvent être estimées avec plus de précision dans les expériences de croisements contrôlés, que ce soit chez des espèces d'élevage (animales ou végétales) ou des espèces modèles en laboratoire (Lynch et Walsh 1998; Wilson *et al.* 2010).

La génétique quantitative, débutée depuis la première moitié du XXe siècle, a démontré qu'il était possible d'étudier la variation génétique d'un trait quantitatif sans en connaître l'architecture génétique sous-jacente (Fisher 1930). Selon le modèle classique dit « infinitésimal », développé par Fisher (1918), la variation génétique qui sous-tend la variation phénotypique s'explique par l'action de nombreux gènes ayant chacun un faible effet sur le phénotypique (Bulmer 1971; Barton *et al.* 2017). D'ailleurs, comme l'ont démontré les récentes études d'association à l'échelle du génome, ou GWAS, le fait de disposer d'informations sur la génétique moléculaire ne se traduit pas automatiquement par une compréhension beaucoup plus complète de l'architecture génétique des traits (citons par exemple les problèmes de l'héritabilité manquante, de l'épistasie, la pléiotropie et interactions gène-environnements; Visscher *et al.* 2012; Korte et Farlow 2013). Cela est d'autant plus vrai pour les traits complexes dont l'architecture est très polygénique (Sella et Barton 2019). De plus, lorsqu'on considère les GWAS, il faut garder à l'esprit l'adage *cum hoc ergo propter hoc*, c'est-à-dire que la corrélation n'implique pas la causalité (Fischer 1970). Ainsi l'association statistique génotype-phénotype ne permet pas d'inférence directe sur les relations causales entre variation phénotypique et génétique, lesquelles requièrent d'être testées par des méthodes expérimentales.

### **1.3 Le modèle animal en génétique quantitative**

Une approche dont la popularité est montée en flèche, depuis deux décennies, pour l'estimation des paramètres génétiques dans les populations naturelles est celle dite du « modèle animal » (Kruuk 2004). Il s'agit ici d'une forme particulière de régression linéaire mixte en statistiques, à ne pas confondre donc avec les « modèles animaux » utilisés en laboratoire (p. ex. souris, rats). Son histoire est liée à la littérature sur l'élevage animal et la génétique statistique, développée par Henderson (1953, 1976, 1984). La force du modèle animal dont il est question ici, et qui l'a propulsé à l'avant-scène de la génétique quantitative « en nature » est sa capacité à utiliser tous les types de relations d'apparentement génétique que l'on peut retrouver entre individus au sein des pédigrées complexes et déséquilibrés dont on dispose habituellement pour les populations naturelles (Wilson *et al.* 2010).

Le modèle animal est très flexible dans la mesure où il peut tolérer les données manquantes et les pédigrées débalancés, et incorporer les facteurs tant génétiques que non génétiques modulant le phénotype (par exemple, les effets de l'âge, de cohorte, du sexe; Thompson 2008). Ces facteurs sont modélisés comme effets fixes ou aléatoires, dont nous pouvons résumer la différence de la manière suivante : les effets fixes affectent la moyenne de la distribution d'un trait phénotypique et les effets aléatoires influencent la variance du même trait.

Dans le cas du modèle animal, l'effet aléatoire d'intérêt premier est « la valeur génétique » ou « valeur de croisement » de l'individu (en anglais sa *breeding value*; Henderson 1977). Comme mentionné précédemment, la valeur génétique est l'effet additif, sur le phénotype  $z_i$ , des allèles portés par un individu  $i$  aux gènes affectant le trait  $z$  étudié. Cette valeur est en fait calculée comme une déviation du phénotype de  $i$  par rapport à la moyenne de la population. Dans le cas le plus simple, on modélise le trait d'un individu  $i$  comme suit :

$$z_i = \mu + a_i + e_i,$$

où  $\mu$  est la moyenne de la population,  $a_i$  est la valeur génétique (c'est-à-dire l'effet du génotype de  $i$  par rapport à  $\mu$ ) et  $e_i$  représente la partie de  $z_i$  qui ne s'explique ni par  $\mu$ , ni par  $a_i$  (=déviation résiduelle due à des facteurs environnementaux non mesurés ou à l'erreur de mesure). La modélisation de  $a_i$  en tant qu'effet aléatoire nous permet d'estimer la variance génétique additive  $\sigma_A^2$  du trait  $z$  au sein de la population étudiée. Il s'agit du paramètre qui détermine le potentiel évolutif dans le trait et son héritabilité (Lynch et Walsh 1998). En ajoutant, comme effets aléatoires, d'autres sources de non-indépendance entre les observations phénotypiques, on obtient des estimations d'autres composantes de la variance phénotypique totale (p. ex., Wolak 2012). En d'autres termes, comme tout modèle à effets mixtes, le modèle animal permet une analyse des composantes de la variance où chaque effet aléatoire suit d'une distribution présumée et spécifiée (gaussienne, de Poisson, *etc.*) avec une moyenne nulle et une variance inconnue qui doit être estimée (Zuur *et al.* 2009).

La covariance génétique additive entre une paire d'individus  $i, j$ , est  $2\theta_{ij}\sigma_A^2$ , où  $\theta_{ij}$  est le coefficient d'apparentement génétique, c'est-à-dire la probabilité qu'une des

deux copies d'un gène, tirée au hasard chez l'individu  $i$ , soit identique par descendance à une des deux copies du même gène tirée au hasard chez l'individu  $j$  (p. ex.  $\theta=0,25$  entre un parent et un rejeton, ce qui correspond à la moitié du degré de parenté; Walsh et Lynch 2018). En fournissant les liens généalogiques (connus) entre les  $n$  individus dont on a le phénotype, le pédigrée permet d'établir la matrice de d'apparentement génétique  $\mathbf{A}$  entre ceux-ci, pris par paires (cette matrice est donc de taille  $(n \times n)$ , tel qu'illustré à la Figure 1.2. Le modèle animal utilise l'inverse de cette matrice ( $\mathbf{A}^{-1}$ ) en combinaison avec des données phénotypiques individuelles pour estimer des paramètres génétiques quantitatifs, tels que la variance génétique additive  $\sigma_A^2$ . Dans le processus d'estimation de  $\sigma_A^2$ , les valeurs génétiques individuelles ( $a_i$ ) sont elles-mêmes estimées. Elles peuvent ainsi être extraites des résultats pour effectuer d'autres tests a posteriori, dont l'un s'avère essentiel pour détecter la microévolution. À savoir, la relation entre les valeurs  $a_i$  et une variable temporelle, comme la cohorte de naissance, permet de vérifier si la moyenne des  $a_i$  a changé dans le temps, selon la tendance prédite par la sélection naturelle (Hadfield 2010; pour un exemple empirique, voir Pigeon *et al.* 2016).

Le modèle animal peut être utilisé sous une forme univariée, soit avec une seule variable phénotypique réponse, ou multivariée, soit avec deux variables réponse ou plus (Meyer, 1991). Dans ce dernier cas, on estime alors la matrice génétique additive de variance-covariance ( $\mathbf{G}$ ) qui sous-tend tous les traits modélisés (Steppan *et al.* 2002). On peut donc estimer les covariances additives entre les traits et, partant de là, calculer les corrélations génétiques entre ceux-ci, déduits des résultats du modèle animal. Une forme matricielle, plus générale, du modèle animal est la suivante :

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

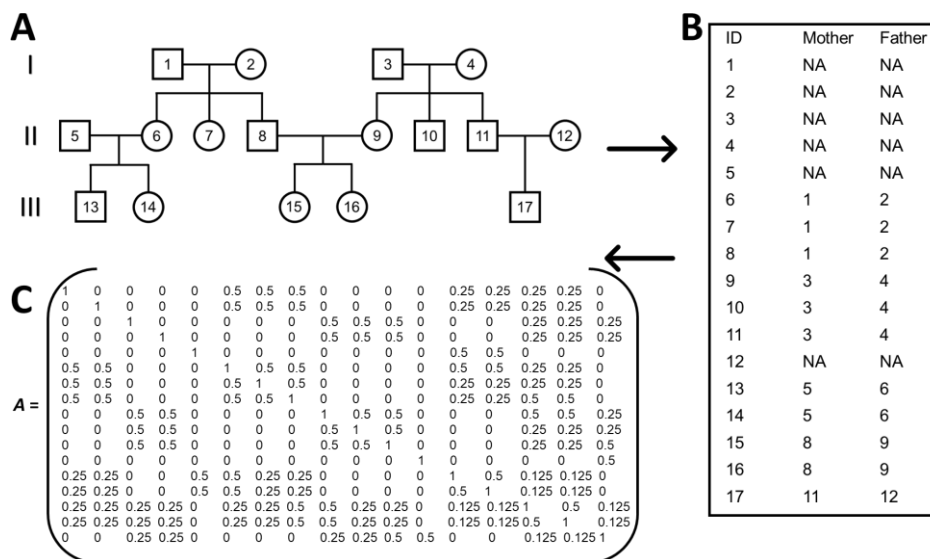
où  $\mathbf{z}$  est la matrice d'observations phénotypiques sur tous les individus (qui se réduit à un vecteur lorsqu'un seul trait est modélisé),  $\boldsymbol{\beta}$  est le vecteur des effets fixes,  $\mathbf{X}$  représente une matrice de conception (de 0 et de 1) reliant les effets fixes appropriés à chaque individu,  $\mathbf{u}$  est un vecteur d'effets aléatoires (dans le cas de ce modèle particulier, il définit la matrice de variance-covariance génétique additive  $\mathbf{G}$ ),  $\mathbf{Z}$  est une matrice de conception reliant les effets aléatoires appropriés à chaque individu et  $\mathbf{e}$  est le vecteur d'erreurs résiduelles. Des modèles plus complexes sont possibles, comme par

exemple l'introduction d'une fonction de régression aléatoire (Schaeffer 2004; Charmantier *et al.* 2014) pour élucider la manière dont les effets génétiques additifs d'un trait varient en fonction d'une variable environnementale (Brommer 2013).

#### **1.4 Défis de l'estimation de paramètres génétiques en nature**

L'un des objectifs de la biologie évolutive est de comprendre les effets de la sélection naturelle dans les populations naturelles (Futuyma 2009). L'approche discutée jusqu'ici consiste à reconstruire un pedigree pour la population étudiée et à mesurer les phénotypes individuels dans le but d'appliquer les méthodes classiques de la génétique quantitative (Kruuk et Hill 2008). Un pedigree décrit les relations de parenté entre les individus d'une population (voir la Figure 1.2; Pemberton 2008). Dans le passé, les généticiens ont utilisé les pedigrees pour étudier en laboratoire ou élevage l'hérédité des polymorphismes, la dépression de consanguinité et la variation génétique quantitative (p. ex., Leroy 2011). Ce n'est que dans les années 1970 et 1980 que les écologistes ont commencé à exploiter des données sur l'histoire de vie des individus et de liens de parenté pour reconstruire les pedigrees reliant des échantillons d'individus provenant de populations naturelles pour étudier ces types de question (Bulmer 1973; Boag et Grant 1978). Certes, ces méthodes de génétique quantitative, initialement développés dans le cadre de programmes de sélection pour l'amélioration génétique d'animaux d'élevage ou d'expériences de laboratoire (Hill 2010), permettent l'étude de l'évolution au sein de populations naturelles, mais au prix de certaines complications (p. ex. structure de données non contrôlée, facteurs de confusion non mesurés, difficile reconstruction de pedigrees; Teplitsky *et al.* 2014).





**Figure 1.2** (A) Exemple de pédigrée contenant 17 individus répartis en trois générations; (B) pédigrée transformé en jeu de données de 17 lignes, chacune correspondant à un individu indiquant l'identité de ses parents (lorsque connus); (C) matrice d'apparentement génétique entre chaque paire d'individus, prédit à partir des liens généalogiques fournis par le pédigrée.

Il convient de noter que les approches de génétique quantitative sont intrinsèquement mathématiques, statistiques, basées sur la modélisation et, selon Kruuk *et al.* (2014), il est plus exact de s'y référer en tant qu'analyse des composantes de la variance. Les analyses génétiques quantitatives basées sur les pédigrées dépendent fortement des propriétés de ceux disponibles (profondeur et complétude généalogiques, exactitudes des liens, *etc.*), qui peuvent varier beaucoup pour les populations naturelles (Pemberton 2008). Avant d'appliquer des méthodes de génétique quantitative sur de nouveaux jeux de données, ou encore de tester une nouvelle hypothèse sur des pédigrées utilisés précédemment, il est important d'explorer la validité de le faire, c'est-à-dire d'évaluer si le pédigrée est adéquat pour bien estimer les paramètres génétiques ciblés, tout en considérant les valeurs plausibles que peuvent prendre ces paramètres (Morrissey *et al.* 2007).

Inférer le potentiel évolutif d'une population seulement à partir de la quantité de variance génétique additive dans un seul trait est une simplification excessive (Arnold *et*

*al.* 2008; Teplitsky *et al.* 2014). La similarité entre les individus apparentés n'est pas seulement due au partage des gènes mais également à des sources non-génétiques de variance phénotypique, telles que l'environnement familial partagé par des frères et sœurs ou la transmission culturelle, en particulier chez l'humain (p. ex. le statut socio-économique; Meyer 1992; Cui *et al.* 2002; Wilson *et al.* 2010). Dans la mesure du possible, il est essentiel d'inclure ces sources de variation dans le modèle animal, car leur absence de l'analyse peut conduire à la surestimation de la variance génétique additive. Ainsi, le degré de ce biais d'estimation doit être déterminé pour les nouveaux pédigrées afin de mieux établir leur adéquation aux analyses génétiques quantitatives et d'éviter des conclusions erronées. Cela est l'objet de chapitre III de cette thèse.

### **1.5 Évolution de l'histoire de vie**

La connaissance du cycle de vie d'un organisme s'avère un aspect essentiel pour comprendre son évolution biologique : comment se développe-t-il, quand atteint-il la maturité et se reproduit-il, quelle est sa fécondité, son espérance de vie, *etc.* (Stearns 1992 ; Roff 2002). Ainsi, les traits d'histoires de vie décrivent les patrons de croissance, de maturation, de survie et de reproduction. Par conséquent, la composition génétique d'un organisme influence l'expression des traits d'histoire de vie (Flatt et Heyland 2011). Il existe une immense variation des cycles de vie dans la nature, entre et au sein des espèces et des populations, en partie le résultat de l'évolution par sélection naturelle (Promislow et Harvey 1990). Cette dernière agit pour maximiser le fitness, créant ainsi différentes « stratégies » d'histoire de vie dans la nature (Charnov et Schaffer 1973). Comprendre l'histoire de la vie implique donc de s'y intéresser sous la perspective de l'évolution. Selon Roff (2002), deux éléments sont à la base de l'analyse de l'histoire de vie. Tout d'abord, il y a la démographie (taux de natalité et de mortalité, structure d'âge de la population, *etc.*), qui est considérée comme la résultante de l'histoire de vie de l'ensemble des individus composant la population et qu'on peut relier au principe de variation de Darwin (p. ex. survie et reproduction différentielles; Cole 1954). Ensuite, il y a l'action de la sélection naturelle au niveau de la population lié à sa capacité à survivre dans un environnement changeant et en compétition avec d'autres espèces (lié au principe d'adaptation de Darwin).

Les facteurs environnementaux tels que la prédation, la disponibilité des ressources et le climat influencent l'évolution des traits d'histoire de vie (Partridge & Harvey 1988; Stearns 2000). Les populations peuvent adopter différentes stratégies en fonction de leur contexte écologique spécifique. Les organismes doivent allouer des ressources (p. ex. énergétiques) limitées entre des composantes concurrentes, tels que la croissance, la reproduction et le maintien de la santé cellulaire et physiologique (Roff 1993). Les recherches empiriques sur diverses espèces montrent que des changements dans la disponibilité de ressources, telles que la nourriture, peuvent entraîner des changements dans l'allocation énergétique relatif consenti à ces composantes (Koskela *et al.* 1998; Descamps *et al.* 2008; Rickard *et al.* 2010). Par exemple, chez certaines espèces d'oiseaux, la réduction de la disponibilité alimentaire peut causer une diminution de la taille des couvées ou un délai dans l'initiation de la reproduction (Monaghan *et al.* 1989; Siikamäki 1998; O'Brien & Dawson, 2011; Seress *et al.* 2020).

Moins les ressources environnementales sont abondantes et accessibles et plus le compromis entre l'allocation d'énergie immédiate et celle future, dans la reproduction ou la survie, risque d'être important et générer de la covariation négative entre certains traits d'histoire de vie (Stearns, 1976, 1983). C'est pourquoi existe dans la nature un continuum entre histoires de vie lentes et rapides, associées à des ensembles différents de traits (p. ex. courte longévité avec un taux de fécondité élevé pour les histoires de vie rapides). Cette relation est aussi connue sous l'hypothèse du « rythme de vie » (Ricklefs & Wikelski 2002), qui veut que les espèces ou populations, où leurs individus sont confrontées à une mortalité extrinsèque plus élevée selon leur âge (risque de décès dû à des facteurs tels que la prédation ou autres périls environnementaux), sont susceptibles d'évoluer vers des stratégies d'histoire de vie plus rapides (p. ex. Martin, 2015), caractérisées par une reproduction précoce et des taux de vieillissement plus rapides (Montiglio *et al.* 2018; Dammhahn *et al.* 2018). Cette influence génère principalement une variation interspécifique dans le rythme de vie and les stratégies d'histoire de vie. Dans les environnements où les taux de mortalité sont élevés et où la probabilité de survivre jusqu'à un âge avancé est faible, les individus peuvent avoir intérêt à se reproduire à un âge précoce pour maximiser leurs chances de transmettre leurs gènes (Promislow et Harvey 1990; Griskevicius *et al.* 2011). Une reproduction précoce

augmente la probabilité qu'au moins une partie de la progéniture survive jusqu'à l'âge de la reproduction (Quinlan 2010). Cette stratégie est souvent associée à la production d'un plus grand nombre de descendants, même si les chances de survie de chacun d'entre eux sont moindres (Nettle 2010; Caudell et Quinlan 2012).

Deux sujets de recherche sur l'évolution de l'histoire de vie nous intéresseront particulièrement ici. Il y a, d'une part, l'étude des changements génétiques (donc évolutifs) dans les traits d'histoire de vie. D'autre part, il y a la question de la proportion de ces changements qui s'explique par une réponse à la sélection naturelle induite par l'environnement. Avec un pedigree multigénérationnel, il est possible de mesurer le fitness des individus de différentes générations et de reconstituer leur histoire de vie.

### **1.6 L'étude de l'évolution d'histoire de la vie humaine**

L'histoire de vie humaine présente quelques différences par rapport à celle des autres primates : une puberté et un début de la reproduction tardifs, des intervalles relativement courts entre les naissances et, surtout, la ménopause suivie d'une période inhabituellement longue de vie post-reproductive des femelles (Mace, 2000). De plus, la culture au sein des populations humaines a un impact important sur l'expression des traits de l'histoire de vie, influençant la prise de décision et pouvant limiter la variation interindividuelle dans des directions particulières (Kaplan *et al.* 2000; Muthukrishna *et al.* 2018; Richerson & Boyd 2020). Néanmoins, les différences d'ordre génétique ne disparaissent pas, car ces traits présentent une héritabilité souvent non négligeable au sein des populations humaines, tel que démontré par les études en génétique quantitative et en génomique (p. ex. Bolund *et al.* 2016). Cela est particulièrement important pour les traits d'histoire de vie qui sont étroitement liés au fitness (Brommer 2000).

Ainsi, l'étude de l'évolution de l'histoire de la vie humaine est une entreprise d'investigation pertinente, surtout si on considère la disponibilité de méthodes adéquates et de vastes ensembles de données (Roff 2002). Les humains ont occupé un large éventail d'environnements, des tropiques à l'Arctique, et chaque environnement présente des défis et des opportunités différents. L'étude des stratégies d'histoire de vie permet aux chercheurs d'explorer la manière dont les humains se sont adaptés et continuent de s'adapter à différentes niches écologiques (Hill & Kaplan 1999). Par exemple, les colons français qui sont arrivés en Amérique du Nord au début du XVIIe siècle ont dû faire face

à des environnements écologiques différents et prospérer sous ces nouvelles conditions (Moogk 2000). Après la fondation de la Nouvelle France, les populations régionales, en expansion démographique et géographique, ont adopté une stratégie d'histoire de vie caractérisée par une fécondité plus élevée (Moreau *et al.* 2011) et un rythme de reproduction plus rapide (Milot *et al.* 2011). Ce changement a eu pour moteur des facteurs non seulement socio-démographiques et culturels, mais aussi biologiques dont, entre autres l'évolution. Ceux-ci ont eu un impact sur la composition génétique de la descendance des premiers Canadiens-français (Peischl *et al.* 2018). Les études en génétique évolutive sur les Canadiens-français démontrent la pertinence d'étudier l'évolution de l'histoire de vie dans les populations humaines en général.

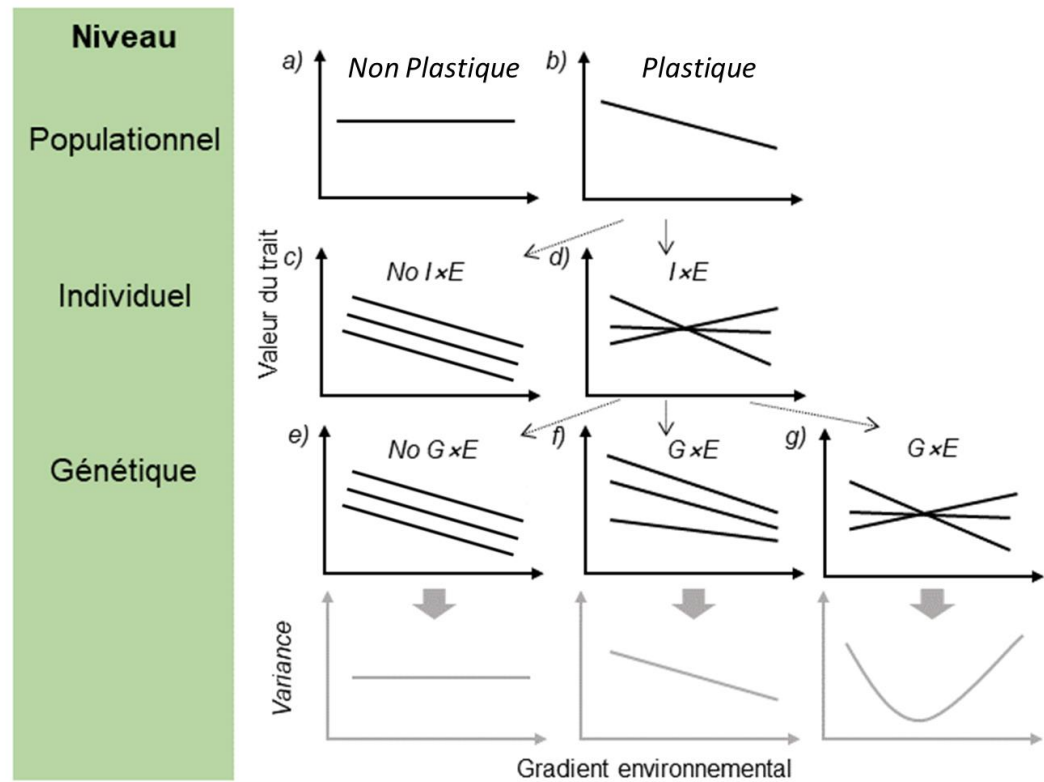
L'un des principaux compromis dans les stratégies d'histoire de vie est l'allocation des ressources énergétiques entre la reproduction et le maintien somatique (l'entretien et la réparation du corps; Stearns 1989). Les individus qui investissent davantage dans la reproduction allouent moins de ressources au maintien somatique, ce qui peut accélérer le vieillissement et accroître la susceptibilité des individus aux maladies liées à l'âge (Stearns 2012). Ces compromis de répartition entre les principaux aspects de l'organisme sont influencés par différents aspects génétiques (par exemple, pléiotropie antagoniste, Williams 1957). En plus d'influencer le compromis entre la survie et la reproduction, les facteurs génétiques jouent un rôle important dans la détermination des différences individuelles en termes de durée de vie et de susceptibilité aux maladies liées à l'âge (Finch et Tanzi 1997; Moorad et Promislow 2009). L'étude de l'évolution de l'histoire de vie chez l'humain s'est aussi particulièrement intéressée aux causes ultimes et aux fondements génétiques de la longévité, notamment à la manière dont les gènes ont évolué au fil du temps en réponse à différentes pressions de sélection naturelle (Lummaa 2007; Monaghan *et al.* 2008).

Les patrons d'histoire de vie varient énormément entre les populations humaines dans différents contextes culturels et écologiques (Hill & Kaplan 1999). Par exemple, chez les populations présentant un régime de fécondité dit « naturel » ou non-malthusien, c.-à-d. sans restriction culturelle imposée à la fécondité (p. ex. contraception, compromis entre carrière et taille de famille), l'âge moyen à la première reproduction peut varier de 10 ans et il en va de même pour l'âge à la dernière

reproduction ou la taille moyenne de la famille (voir tableau 2 dans Walker *et al.* 2006). Les études biologiques sur les stratégies d'histoire de vie humaines, notamment celles en écologie évolutive, se sont concentrées sur les sources de cette variation et sur le potentiel d'évolution que celle-ci représente, notamment en réponse aux changements environnementaux et écologiques rapides (Nettle *et al.* 2013; Bolund et Lummaa 2017; Evans *et al.* 2018). Elles ont montré que i) à nos jours, la sélection naturelle agit de plus sur les traits d'histoire de vie par le biais de la variation dans la composante de reproduction entre les individus, et ii) les différences génétiques peuvent influencer la variation des traits d'histoire de vie, ce qui signifie que, dans les populations humaines contemporaines, il existe un potentiel évolutif dans ces traits. Stearns *et al.* (2010) ont montré que l'opportunité et la force de sélection sont substantielles dans les populations industrialisées contemporaines (voir aussi Corbett *et al.* 2018), tandis que Moorad (2013) a révélé que les pressions de sélection changent rapidement dans le temps.

### **1.7 Les interactions entre le génotype et l'environnement**

Les organismes vivent dans des environnements hétérogènes, ce qui fait qu'il est courant que leurs traits soient modulés par les conditions environnementales. La capacité d'un même génotype à produire différents phénotypes selon les conditions environnementales est appelée « plasticité phénotypique » (Via et Lande 1985). Cette capacité d'ajustement phénotypique de l'organisme aux conditions ou changements de l'environnement est elle-même façonnée par la sélection. En effet, la variation interindividuelle dans la plasticité est en partie d'origine génétique et héritable. En conséquence, deux génotypes différents ne réagiront pas nécessairement de la même manière à un changement dans l'environnement, c'est-à-dire qu'ils n'induiront pas le même ajustement du phénotype. C'est ce qu'on appelle les « interactions génotype-environnement » ( $G \times E$ ; Figure 1.3). Autrement dit, l'effet du génotype d'un individu sur un trait particulier varie en fonction des conditions environnementales (Westcott 1986). Les  $G \times E$ s témoignent de la complexité des relations gène-environnement et du fait que les effets génétiques sur un trait ne sont pas fixes, mais peuvent être modulés par des facteurs externes.



**Figure 1.3** Illustration schématique de la plasticité phénotypique à différents niveaux d'organisation biologique. Le phénotype moyen dans la population peut s'avérer invariable peu importe l'environnement (a) ou, au contraire, varier suivant le gradient environnemental (b). S'il est invariable au niveau populationnel, alors il en sera de même au niveau des individus et des génotypes (c, e), ce qui signifie qu'il n'existe pas de variation interindividuelle dans la plasticité (pas d'interactions individu-environnement ( $I \times E$ ) ni génotype-environnement ( $G \times E$ )). En revanche, lorsque le phénotype moyen (b) varie selon l'environnement, alors cela peut correspondre à deux cas des figures : soit le phénotype de tous les individus change de la même manière (mêmes direction et ampleur) et il n'y a donc ni  $I \times E$  ni  $G \times E$  (non illustré), soit il existe une variation dans la plasticité individuelle (d). Dans le second cas, la plasticité individuelle peut avoir une base génétique ( $G \times E$ ), tel qu'illustré en f et g ou non illustré en e. Au niveau de la variance génétique, dans le cas où la plasticité individuelle ne présente pas une base génétique (e), alors la variance ne change pas avec le gradient environnemental. Dans le cas où la plasticité individuelle a une base génétique, la variance change avec le gradient environnemental et la forme de ce changement dépend de l'intersection (g) ou non (f) des normes de réaction

Imaginons un trait quantitatif comme la taille. Si la génétique joue un rôle important dans sa détermination, d'autres facteurs tels que la nutrition et l'état de santé général pendant la croissance l'influencent également. Des interactions G×E pourraient se manifester si l'effet de gènes liés à ce trait est plus prononcé sous certaines conditions nutritionnelles plutôt que d'autres (p. ex. environnement de qualité permettant une nutrition optimale ou, à l'inverse dans lequel les ressources alimentaires sont limitées). Il s'ensuit que certaines variantes génétiques peuvent offrir un avantage dans un environnement mais pas dans un autre, ce qui entraîne une sélection différentielle sur ces gènes (Romagosa et Fox 1993).

Une conséquence des interactions G×E est que la variance génétique additive peut changer à travers un gradient environnemental (dernière ligne de graphiques de la Figure 1.3). Ainsi, le potentiel microévolutif d'un trait ayant une certaine plasticité phénotypique peut être modulé par cette plasticité même, via les G×Es. Hoffmann et Merilä (1999) ont compilé et analysé de nombreuses études sur des invertébrés et des oiseaux montrant ce phénomène, où la capacité d'adaptation de certains traits (voir les tableaux 2 et 3 dans leur revue), par réponse génétique à la sélection, était limitée dans certaines conditions environnementales extrêmes (par exemple, basse température/nutrition). Les G×Es deviennent particulièrement importantes dans un contexte de changements environnementaux rapides causés les activités humaines ou leurs conséquences, comme le réchauffement climatique (Kang 2002). Les populations qui peuvent s'adapter rapidement à des conditions changeantes, notamment grâce aux G×Es, auront de meilleures chances de survie. En outre, les G×Es peuvent accélérer les réponses microévolutives dans certaines circonstances. Par exemple, si l'environnement change ou un nouvel habitat devient disponible, les individus possédant le génotype approprié permettant une meilleure réponse plastique peuvent avoir une plus grande probabilité de survie et de reproduction, ce qui entraîne la propagation rapide de traits adaptatifs (Charmantier & Garant 2005).

Pour étudier l'impact des G×Es sur la microévolution dans la nature, il faut idéalement partir d'un cas avéré et détectable de réponse génétique à la sélection sur un trait et pouvoir estimer des paramètres génétiques et environnementaux pertinents, ainsi que leur covariation, à différents points temporels (Wong et al. 2003). Les méthodes de



génétique quantitative basées sur les pédigrées peuvent incorporer des sources génétiques et non génétiques de variation phénotypique et leurs interactions (Wilson et al. 2010). À partir de données collectées sur le terrain en Estonie sur 37 ans, Brommer et al. (2008) ont ainsi étudié le rôle des G×Es dans la plasticité de la date de ponte chez les Goélands cendrés (*Larus canus*). Bien que cette étude n'ait pas clairement détecté de G×Es, elle fut l'une des premières à introduire de telles interactions dans un modèle de génétique quantitative. Ce type d'analyse n'a jamais été réalisé auparavant dans une population humaine.

L'un des traits d'histoire de vie les plus importants et les mieux étudiés chez les humains est l'âge à la première reproduction (AFR), soit celui qu'a un parent à la naissance de son premier enfant (p. ex. Bürkli et Postma 2014; Evans et al. 2018). Ce trait étroitement associé au fitness présente une variation substantielle au sein des populations et entre les populations (p. ex. Pebley et al. 1982). En particulier, la démographie influence la sélection s'exerçant sur l'AFR. En effet, une reproduction plus précoce représente un avantage sélectif dans une population en croissance, l'inverse étant vrai dans une population en décroissance (Caswell 1989). Par « avantage sélectif », on entend ici au plan de la représentation génétique d'un parent dans la prochaine génération (toutes choses étant par ailleurs égales). À partir des données de la Framingham Heart Study, une étude médicale multi-cohortes, Byars et al. (2010) ont mis au jour une sélection naturelle sur l'AFR, qui s'est maintenue dans le temps (une période de 60 ans), et démontré que ce trait était héritable dans cette population. De là, ils ont projeté la trajectoire évolutive future de l'AFR moyen et prédit sa diminution, à raison d'environ 15 jours par génération (si la sélection se maintient). En réalité, et pour de nombreuses populations, notamment celles industrialisées, c'est au contraire un repoussement de l'AFR moyen qu'on observe, qui se situe aujourd'hui à environ 30 ans dans la plupart des pays occidentaux (malgré des variations importantes entre les pays; [oecd.org/els/family/database.htm](http://oecd.org/els/family/database.htm)). Cette observation rappelle que les traits d'histoire de vie sont aussi modulés par la variabilité de l'environnement (Roff 1993).

La théorie de l'histoire de vie a permis de comprendre l'influence des facteurs environnementaux sur le rythme et le moment de la maturation et de la reproduction des organismes. Dans un ouvrage jalon, dans lequel ils synthétisent la théorie d'histoire de

vie et du développement comportemental, Chishlom et al. (1993) montrent comment l'expérience vécue au début de la vie peut façonner les stratégies de reproduction des individus. Au rôle des taux démographiques (p. ex. mortalité) déjà évoqué par les théoriciens de l'évolution, ces auteurs intègrent celui de l'environnement psychosocial dans lequel l'individu se développe, basé sur le modèle de Belsky et al. (1991). Ils font valoir que la variation dans ce dernier peut mener aussi bien à un avancement ou à un retard de la maturité et de la reproduction, pouvant ou non aller dans le même sens que l'effet précité induit par la démographie.

Dans l'étude d'une cohorte britanniques, Nettle et al. (2013) ont mis au jour une association entre les conditions défavorables vécues par certains bébés de sexe féminin, comme un allaitement écourté ou des soins parentaux limités, et le risque de grossesse précoce chez les femmes que ces bébés deviendront plus tard, et ce, même après avoir contrôlé pour le statut socioéconomique familial et l'âge de la mère à la naissance du bébé. Par conséquent, l'environnement psychosocial en début de vie peut influencer l'AFR chez les humains, ainsi d'ailleurs que d'autres aspects de l'histoire de vie, tels que la taille de famille ou le degré d'investissement parental (Nettle et al. 2011). Spécifions que cela n'exclut pas l'implication des gènes (West-Eberhard 2003). Selon Ellis (2004), les effets génétiques sur le développement pubertaire, qui se répercutent très probablement sur l'AFR, prennent la forme d'une norme de réaction<sup>6</sup>. À savoir que ces effets génétiques varient selon les conditions environnementales, incluant ici celles psychosociales. En d'autres termes, l'AFR est probablement plastique et sous influence des G×Es. Cependant, on ne sait pas dans quelle mesure le degré de plasticité phénotypique des traits d'histoire de vie humains est adaptatif, c'est-à-dire ajusté génétiquement par l'évolution passée. On ignore aussi si les G×Es sont à même d'avoir un impact sur le taux de microévolution lorsque ces traits sont soumis à une sélection naturelle.

## **1.8 La population canadienne-française comme système d'étude**

La population étudiée dans cette thèse est celle canadienne-française historique qui vivait dans ce qui est aujourd'hui le Québec, mais anciennement appelée Nouvelle-

---

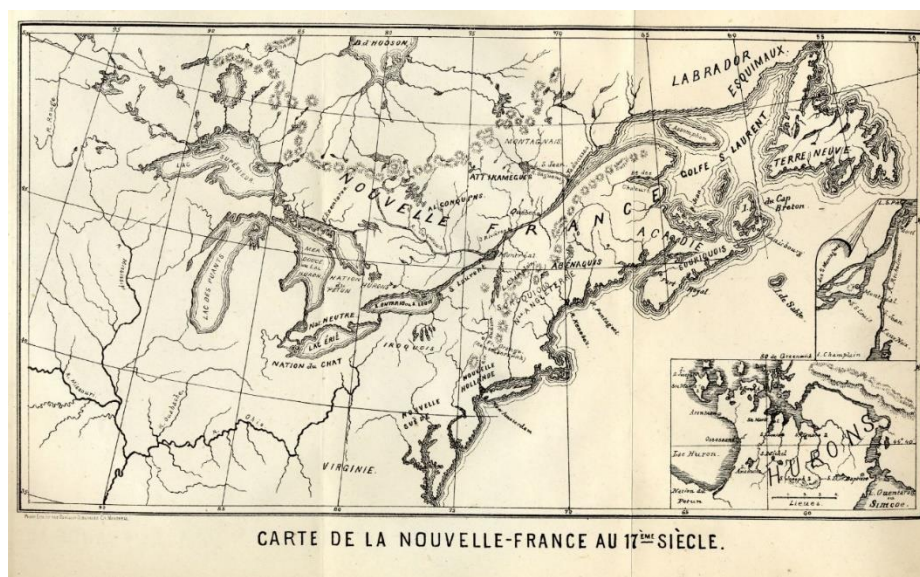
<sup>6</sup> Le concept de norme de réaction dans l'âge à maturité a été proposé pour la première fois par Stearns et Koella (1986)

France, puis Bas-Canada; Figure 1.4). Sur la base de nombreuses études démographiques et génétiques (p.ex. Bouchard et De Braekeleer 1991; Lavoie *et al.* 2005; Roy-Gagnon *et al.* 2011), deux caractéristiques de la population canadienne-française la rendent particulièrement intéressante pour l'étude génétique : i) les effets fondateurs, cinq à six millions d'individus de la population actuelle du Québec descendant d'un petit bassin de 8 000 à 10 000 colons et autres immigrants; ii) une homogénéité ethnique et culturelle relative dans la population, notamment dans les régions du Nord-Est, soit le Saguenay–Lac-Saint-Jean, Charlevoix, le Bas-Saint-Laurent, la Côte-Nord et la Gaspésie, où l'isolement et les normes sociales (Bouchard et De Braekeleer 1991) ont créé une restriction au « mélange », que ce soit par l'origine des immigrants ou dans les mariages qui y ont été célébrés<sup>7</sup>. Les effets fondateurs ont été étudiés de manière approfondie, particulièrement aux plans sociodémographiques et de l'épidémiologie génétique (Scriver 2001; Moreau *et al.* 2007). Ces effets expliquent notamment la prévalence élevée de maladies génétiques rares dans certaines régions, telles que la dystrophie myotonique (dominante) et l'ataxie spastique (récessive). La majorité de la population actuelle, environ 80%, est à la fois francophone et de confession catholique<sup>8</sup>. Selon Bouchard (1989), ce constat fait de la population canadienne-française une sorte de « macro-isolat » parmi les populations anglophones d'Amérique du Nord.

---

<sup>7</sup> Cette homogénéité contraste avec l'hétérogénéité et l'ascendance mixte d'une partie de la population des grandes régions métropolitaines du Québec (Montréal et Québec).

<sup>8</sup> Ce chiffre est tiré du recensement canadien de 2001. Des enquêtes plus récentes révèlent une diminution de la proportion de la population qui s'identifie comme catholique.



**Figure 1.4** Carte de la Nouvelle-France au XVII<sup>e</sup> siècle. Créée par la compagnie de lithographie Burland-Desbarats et tirée de l'ouvrage *Les Jésuites-martyrs du Canada* écrit par le père François-Joseph Bressani. Image provenant de la BANQ et relevant du domaine public au Canada.

La colonisation puis le peuplement de l'est du Canada ont commencé avec l'arrivée d'immigrants français au XVII<sup>e</sup> siècle. La majeure partie s'est faite le long des rives du fleuve Saint-Laurent, en amont et en aval de la région de Québec, servant de point de départ. Après la conquête anglaise de la colonie, l'immigration en provenance de France s'est pratiquement arrêtée. Cependant, la population a connu un accroissement naturel rapide et, au XIX<sup>e</sup> siècle, la pénurie de terres disponibles pour l'établissement des familles les a forcées à s'aventurer plus profondément dans les régions pour s'y établir. Cette donnée est pertinente d'un point de vue écologique et évolutif : les expansions dans de nouvelles zones signifient de nouveaux défis et des conditions locales pouvant être différentes pouvant se traduire par des pressions de sélection naturelle propres.

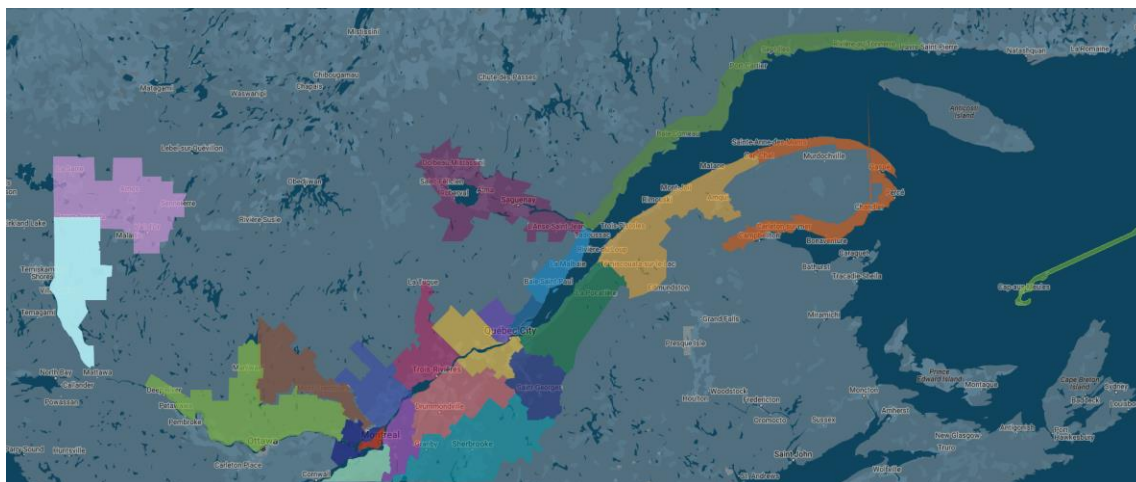
Comme mentionné précédemment, la plupart, sinon la totalité, des immigrants français et de leurs descendants étaient catholiques et l'Église faisait partie intégrante de leur vie sociale et politique. L'expansion géographique de la population fut accompagnée de l'établissement de nouvelles paroisses, desservies par des clercs tenant les registres locaux de baptêmes, mariages et sépultures. Combinés aux actes d'état civil et sur les recensements, ces registres paroissiaux ont permis de reconstituer les généalogies de la population canadienne-française depuis sa fondation (ou presque, le

premier acte de mariage connu datant de 1621). La qualité de ces généalogies se manifeste par le fait que de très nombreux individus de la population actuelle peuvent être reliés aux fondateurs (et au-delà, dans certains cas où un jumelage a été fait avec des ascendants européens). Regroupées dans le Registre de population du Québec ancien du PRDH et le registre BALSAC (Dillon *et al.* 2018; Vézina et Bournival 2020), ces généalogies couvrent l'ensemble des régions du Québec habitées par des Canadiens-français et concernent plus de cinq millions d'individus.

L'extensivité des généalogies canadiennes-françaises et le fait qu'elles couvrent autant d'individus et jusqu'à 10 générations en font un outil exceptionnel pour les études génétiques, de point de vue mondial même (p. ex. Anderson-Trocmé *et al.* 2023). On l'a vu, l'épidémiologie génétique s'y appuie largement pour étudier la prévalence de maladies génétiques rares dans les régions (Scriver 2001; Plante *et al.* 2008; Heyer 2009). De plus, et pour se rapprocher du sujet de la présente thèse, les généalogies servent de matériau pour comprendre les processus évolutifs et démographiques. Par exemple, Moreau *et al.* (2011) ont démontré que le fait, pour un individu, de vivre au front pionnier d'une vague d'expansion peut lui conférer un avantage sélectif, c'est-à-dire un meilleur fitness, tel que mesuré par sa contribution au pool génétique des générations suivantes; l'inverse est vrai pour ceux vivant au noyau, soit dans la zone géographique d'occupation plus ancienne. Plus généralement, en regroupant les généalogies descendantes des familles ayant habité une région, le pédigrée est porteur d'informations essentielles pour l'étude de l'évolution biologique de traits complexes, à savoir sur le fitness des individus, sur l'apparentement génétique entre eux et sur certains de leurs phénotypes, dont ceux relatifs à l'histoire de vie (taille de la famille, AFR, âge à la dernière reproduction, intervalles génésiques, etc.). Selon Wilson *et al.* (2010), la faisabilité ou la validité d'une analyse de génétique quantitative dépendra fortement de la qualité du jeu de données disponible. Un examen rapide des généalogies canadiennes-françaises révèle des propriétés recherchées pour l'analyse génétique quantitative, telles que les grandes profondeur généalogique et taille d'échantillon.

La population historique canadienne-française est l'objet d'étude de cette thèse. Pour les chapitres II et III, six populations régionales ont été analysées parmi celles délimitées par BALSAC (Figure 1.5) : Bois-Francs, Charlevoix et Côte-de-Beaupré pour

les deux chapitres, auxquelles s'ajoutent Gaspésie, Laurentides et Bas-Saint-Laurent au chapitre III. Le chapitre IV porte sur l'île aux Coudres, une sous-population de Charlevoix. Les données pour cette dernière proviennent d'un registre populationnel différent de BALSAC et décrit au chapitre IV.



**Figure 1.5** Carte des régions du Québec correspondant aux délimitations de BALSAC. Image tirée de BALSAC (<https://balsac.uqac.ca/apercu-des-donnees/>).

### 1.9 L'île aux Coudres, modèle d'étude de l'évolution biologique

Le déclencheur du travail de recherche présenté dans cette thèse trouve son origine dans l'étude de Milot *et al.* (2011) sur la génétique évolutive de l'âge à la première reproduction dans la population préindustrielle de l'île aux Coudres (IAC), une petite île de 34 km<sup>2</sup> gisant près de la rive nord du fleuve Saint-Laurent dans la région de Charlevoix. Cette étude rapporte la première la démonstration empirique d'un changement microévolutif dans une population humaine de l'époque contemporaine. Au cours d'une période de 140 ans (1800-1940), l'âge moyen à la première reproduction des femmes est passé de 26 à 22 ans. Les auteurs ont montré que ce changement phénotypique était dû, en partie, à un changement génétique en réponse à la sélection. À partir du pédigrée, qui avait été reconstitué à partir du Registre de population de l'île aux Coudres (Philippe et Gomila 1971; Philippe 1973; Mayer *et al.* 1988; Boisvert et Mayer, 1994), ils ont d'abord mis au jour une héritabilité non négligeable dans l'AFR et une sélection directionnelle, soutenue dans le temps, en faveur des valeurs plus précoces. Ensuite, les auteurs ont démontré qu'il y avait eu un changement, entre les générations, dans la moyenne des valeurs génétiques additives de l'AFR, que ce changement était

dans la direction prédite par la sélection (devancement de l'AFR), et qu'il était d'ampleur trop importante pour être expliqué par un simple scénario de dérive génétique. Aussi, Pelletier *et al.* (2017) ont démontré que ce changement génétique de l'AFR en réponse à la sélection a eu un impact sur la croissance de la population de l'IAC (la taille de la population aurait été inférieure d'environ 12 % durant la période d'étude sans l'évolution dans l'AFR), ce qui constitue un indice de dynamique éco-évolutive.

Bien que la population de l'IAC soit endogame et relativement isolée, son histoire est assez typique de celle canadienne-française en général, à certains égards (Martin, 1957). Notamment, le premier peuplement s'est fait par la migration familiale (30 familles installées entre 1720 et 1773) et l'attribution de terres aux paysans, suivant le régime seigneurial. Rapidement, soit avant la fin du XIII<sup>e</sup> siècle, la natalité pris le dessus sur l'immigration comme moteur d'accroissement. Cet évènement de fondation, suivi d'une expansion démographique, dans un contexte d'endogamie soutenue (mariages célébrés majoritairement entre époux natifs de l'île), ont créé une structure généalogique complexe et enchevêtrée, idéale pour les études de génétique quantitative, notamment pour distinguer les influences génétiques et environnementales sur les traits d'histoire de vie des individus. Dans la période préindustrielle, le degré élevé d'homogénéité entre les familles de l'île, aux plans de la classe sociale, de l'éducation et de la religion, s'est transformé en atout précieux pour la présente étude. En effet, ces facteurs étant connus pour être corrélés avec l'âge à la première reproduction (Pettay *et al.* 2007), leur variation limitée à l'IAC rend les effets génétiques sur ce trait plus faciles à détecter. Si l'on ajoute à cela l'uniformité relative de la répartition des terres entre les familles et le nombre limité de professions, du moins, jusque vers la fin du XIX<sup>e</sup> siècle, nous pouvons supposer sans trop de risque d'erreur que les effets de confusion dues aux facteurs socioéconomiques auront un impact modeste, peut-être même minime, dans les analyses génétiques quantitatives. Cela pourrait ne pas s'appliquer à d'autres régions du Canada français durant la même période. La région du Saguenay–Lac-St-Jean, par exemple, présentait une plus grande diversité de professions, ainsi qu'une hétérogénéité plus marquée dans la répartition des ressources entre les familles (Bouchard et De Braekeleer 1991; Gagnon *et al.* 2011). Cependant, l'homogénéité culturelle entre les familles par la

religion et les pratiques sociales est plus répandue dans la région (). Les régions proches de l'IAC, comme Charlevoix et Côte-de-Beaupré, en plus de partager des aspects environnementaux similaires (Sabourin 1962; Bird 1972; Raveneau 1977), présentent également des dynamiques démographiques comparables (c.-à-d. un taux de natalité élevé qui a ensuite diminué avec le temps alors que le taux de mortalité a augmenté; Gauvreau et Jette 1992; Moreau et al. 2011).

### 1.10 Questions de recherche

La question générale étudiée dans cette thèse concerne la dynamique évolutive des traits d'histoire de vie dans une population humaine de l'époque contemporaine, soit celle canadienne-française. Nous avons divisé cette enquête en trois questions spécifiques mais interreliées :

- a) Les études sont de plus en plus nombreuses qui appuient l'idée que l'évolution biologique est possible, peut-être même fréquente, dans les populations humaines contemporaines (p. ex. Corbett *et al.* 2018; Beauchamp 2016). À tout le moins, il y a ubiquité du potentiel évolutif dans ces populations, comme le révèle la démonstration répétée de la sélection naturelle agissant sur des traits héréditaires (p. ex. physiologiques, morphologiques et d'histoire de vie; Stearns et al., 2010). Cela est en phase avec le portrait plus général qui émerge des études sur diverses espèces sauvages (Ghalambor *et al.* 2015; Johnson et Munshi-South 2017), au point où les biologistes sont plusieurs à considérer que la sélection naturelle joue un rôle vraisemblablement beaucoup plus important qu'on ne le pensait dans la réponse adaptative des populations animales et végétales aux changements environnementaux rapides (Colautti et Barrett 2013). Mais les démonstrations de microévolution *réalisée*, et non seulement *potentielle*, restent difficiles à reproduire. Autrement dit, il s'avère hasardeux de vouloir trop généraliser à partir des cas particuliers, impliquant des traits spécifiques dans des populations ayant été exposées à des conditions environnementales les plus diverses (p. ex. DeFaveri et Merilä 2014).

Pour appréhender la généralité des changements microévolutifs, il serait utile de comparer le sort des mêmes traits dans des populations d'organismes



semblables et exposées à des conditions environnementales les plus similaires possible. Or, cela est envisageable avec certaines populations humaines. Nous nous intéresserons ici de celles issues de l'expansion coloniale européenne des derniers siècles. En particulier, les populations fondées en Amérique par des colons français furent soumises à des conditions environnementales nouvelles par rapport à leur mère patrie. Entre autres, nombre de peuplements jeunes et de faible densité, par rapport à l'abondance des ressources, ont connu une rapide croissance démographique. Dans ce contexte, la théorie d'histoire de vie prédit que peuvent survenir des changements phénotypiques adaptatifs rapides dans les traits d'histoire de vie, c'est-à-dire induits par une réponse à la sélection naturelle. Cette situation correspond au cas documenté de microévolution de l'AFR dans la population préindustrielle de l'île aux Coudres (Milot et al., 2011). Étant donné l'histoire de la population canadienne-française, plusieurs sous-populations qui partagent un bagage génétique avec la population de l'IAC et des conditions environnementales similaires ont connu une expansion démographique (Moreau et al., 2011). Cela conduit à la question suivante : **le cas documenté de microévolution chez IAC est-il un phénomène particulier et isolé ou général qui suit l'expansion des populations humaines dans de nouveaux environnements ?**

- b) L'étude et la réponse aux questions relatives à l'évolution des populations naturelles nécessitent une mesure précise de la base génétique sous-jacente des traits pertinents (Postma 2014). La génétique quantitative est un domaine d'étude qui se concentre sur cet objectif en décomposant la variation phénotypique observée en ses composantes génétiques et environnementales (Lynch et Walsh 1998). Les pedigrees, lorsqu'ils sont disponibles, fournissent des informations sur la parenté « moyenne » entre les individus d'une population donnée (Pemberton 2008). Avec les mesures phénotypiques individuelles, ces informations sont introduites dans le modèle génétique quantitatif, connu sous le nom de modèle animal, pour déduire la quantité de (co)variance génétique additive qui sous-tend le trait en question (Wilson *et*

*al.* 2010). Les pedigrees basés sur des observations sociales peuvent contenir un nombre non négligeable d'erreurs qui peuvent réduire la précision des informations fournies sur la parenté (Bérénois *et al.* 2014). Cette observation nécessite d'évaluer la fiabilité des pedigrees et des ensembles de données lors de l'estimation des paramètres génétiques quantitatifs (Morrissey *et al.* 2007).

Les généalogies canadiennes-françaises sont étendues, multigénérationnelles et couvrent plusieurs sous-régions (Vézina et Bournival 2020), ce qui en fait d'excellents candidats à utiliser dans les analyses génétiques quantitatives. Étant donné que ces généalogies profondément enracinées proviennent de données de recensement et les registres paroissiaux des mariages (et dans certains cas, baptêmes et sépultures), il est probable que les pedigrees sociaux basées sur celles-ci contiennent un nombre non-négligeable d'erreurs (Doyon 2018). Bien que quelques études de sensibilité aient montré que des taux raisonnables d'erreurs de pedigree n'affectent pas de manière majeure les résultats des analyses génétiques quantitatives (p. ex., Charmantier et Réale 2005), il n'est pas recommandé de généraliser cette attente aux pedigrees basés sur un système humain. Cela s'explique par le fait que les populations humaines présentent une paternité extra-pair relativement faible (Larmuseau *et al.* 2016) et un système d'accouplement particulier, par rapport à d'autres espèces de mammifères (Marlowe 2000). Étant donné que les pedigrees canadiens-français offrent la possibilité d'étudier la base génétique de traits complexes et d'enquêter sur des questions évolutives, **il faudrait procéder à une évaluation *a priori* adéquate et approfondie de leur fiabilité, en termes de puissance, d'exactitude et de précision, pour l'estimation des paramètres génétiques quantitatifs.**

- a) Les populations naturelles sont soumises à des conditions environnementales hétérogènes, tant dans le temps que dans l'espace (McDonald et Ayala 1974). Des études théoriques et empiriques ont montré que les fluctuations environnementales peuvent affecter l'architecture génétique de traits

complexes (tels que les traits d’histoire de vie) en un court laps de temps, affectant la quantité de variance génétique additive disponible pour la sélection (Robinson *et al.* 2009; Bolund *et al.* 2015). Cependant, l’impact de telles fluctuations sur le taux de microévolution dans la nature est encore peu exploré (p. ex., Huang *et al.* 2016), en raison de la rareté des systèmes naturels où des changements microévolutifs ont été démontrés et qui permettent de tester une telle hypothèse.

Il y a déjà eu une étude sur une population humaine contemporaine qui a mis en évidence un changement microévolutif dans un trait d’histoire de vie héritable (la population d’IAC; Milot *et al.* 2011). Si l’on ajoute à cela le fait que les traits d’histoire de vie sont étroitement liés aux composantes de fitness (Braendle *et al.* 2011), qu’ils sont sensibles à la variabilité de l’environnement (Roff 2002) et que l’on sait qu’ils présentent des interactions entre le génotype et l’environnement, **les changements rapides dans l’environnement pourraient moduler le taux de changement microévolutif du trait sous sélection.**

### 1.11 Objectifs et hypothèses de recherche

Après avoir présenté les différents sujets et domaines liés à cette thèse, ainsi qu’un bref aperçu historique de la population étudiée, suivi de l’énonciation des questions de recherche, j’exposerai ici succinctement les trois objectifs de ce doctorat, ainsi que les hypothèses qui découlaient de chacun d’eux (ces objectifs sont traités en détail dans les trois chapitres centraux [II à IV]).

- **Objectif 1.** Déterminer si le changement microévolutif de l’AFR à l’île aux Coudres représente un phénomène général dans les populations préindustrielles en expansion, ou si la stase évolutive était plutôt la norme. Pour ce faire, j’ai utilisé comme répliques naturels trois populations canadienne-françaises régionales (Bois-Francis, Charlevoix et Côte-de-Beaupré). Pour les raisons évoquées en (a) dans la section précédente sur la question de recherche, mes hypothèses étaient que :

- L'AFR des femmes mariées dans chacune des populations étudiées présentait une composante génétique additive et donc héritable;
  - L'AFR était systématiquement sous sélection pour des valeurs plus précoces, car associées à des valeurs plus élevées de fitness;
  - Sous réserve que les deux hypothèses précédentes soient vérifiées, la sélection a causé un changement génétique dans l'AFR au cours de la période d'étude.
- **Objectif 2.** Évaluer et comparer la fiabilité des généalogies canadiennes-françaises pour les analyses de génétique quantitative, de même que l'impact du taux d'erreur généalogique et de divers attributs des pédigrées sur cette fiabilité à l'aide d'analyses de simulation. Mes hypothèses étaient que :
- Les erreurs généalogiques conduisent à une sous-estimation des paramètres génétiques, mais dont l'ampleur est modeste considérant l'ordre de grandeur du taux de ce type d'erreur dans la population canadienne-française;
  - La précision et la rectitude des estimations sont positivement corrélées avec la taille (nombre d'individus), la profondeur généalogique et la connectivité du pédigrée;
  - La précision et la rectitude des estimations sont impactés négativement lorsque l'architecture génétique du trait d'intérêt et la structure du modèle spécifié (paramètres de variance) sont plus complexes.
- **Objectif 3.** Étudier l'impact potentiel des fluctuations dans les conditions environnementales vécues en bas âge sur le rythme du changement microévolutif dans l'AFR qui avait été détecté dans la population préindustrielle de l'île aux Coudres. Mes hypothèses étaient que :
- L'AFR est modulé par des interactions génotype-environnement ( $G \times E$ ) qui se répercutent sur sa variance génétique additive et sa covariance avec le fitness;

- L'association génétique entre un AFR plus précoce et un fitness élevé à l'IAC était plus forte pour les femmes ayant été exposées à des conditions environnementales défavorables en début de la vie (le taux de mortalité infantile servant ici d'indicateur de la qualité des conditions environnementales).
- En raison de l'hypothèse précédente, les G×Es ont pu moduler la vitesse du changement microévolutif dans l'AFR (plus rapide sous des conditions environnementales défavorables).

### 1.12 Données et méthodes

Dans cette dernière section de l'introduction de cette thèse, je présente succinctement les ensembles de données que j'ai utilisés dans les différentes analyses, ainsi que les modèle et méthode de génétique quantitative retenus pour estimer les paramètres pertinents et tester mes hypothèses de recherche. Le lecteur ne devra pas s'étonner de retrouver certaines des explications ci-dessous répétées dans les chapitres suivants. En effet, il me semblait crucial de lui fournir ici les éléments nécessaires pour lui permettre de comprendre clairement, dès le début, la structure des jeux de données et du modèle général de génétique quantitative sur lesquels repose sur le travail de recherche décrit dans cette thèse.

*Les données.* Comme évoqué dans les sections précédentes, nous pouvons considérer les données utilisées dans ce travail de recherche de deux manières : 1- comme individuelle, chaque donnée représentant un événement de vie qui implique un ou plusieurs individus. C'est ainsi qu'on peut reconstruire les familles, les généalogies et les histoires de vie individuelles. Dans le tableau de données, chaque ligne correspond alors à l'enregistrement d'un individu, tandis que les colonnes contiennent les variables pertinentes (telles que l'année de naissance, le nombre total d'enfants de l'individu, etc.) ; 2- comme populationnelle, chaque individu étant connecté à d'autres par des liens généalogiques qui décrivent la structure de parenté. Dans ce cas, un tableau à trois colonnes (individu, mère et père) décrit le pédigrée. Les ancêtres doivent apparaître en premier dans ce tableau, suivis de leurs descendants (voir la Figure 1.2).

Ces ensembles de données ont été fournis par le biais des registres populationnels eux-mêmes basés sur les registres paroissiaux et civils contenant les actes de naissances,

mariages et décès. Pour la population de l'IAC, le jeu de données fut construit à l'origine par Pierre Philippe en 1967, avec Jacques Gomila, Jean Benoist et Guy Dubreuil (Université de Montréal). À partir de 1986, le registre fut informatisé et mis à jour par Francine M. Mayer, Mireille Boisvert, Yolande Lavoie et Pierre Philippe. Les jeux de données qui couvrent d'autres régions du Canada français, comme Charlevoix, proviennent du registre BALSAC, conservé à l'Université du Québec à Chicoutimi et qui a été lancé dans les années 1970 dans le but de reconstituer les familles à partir des registres paroissiaux. Les actes contenus dans ces registres ont été jumelés grâce aux informations nominatives, conduisant ainsi à la reconstruction des relations généalogiques et des structures de parenté de la population canadienne-française.

Il est important de noter que, bien que le registre de BALSAC soit continuellement mis à jour et corrigés, les informations contenues sont basées sur des événements socialement observés. Par conséquent, des erreurs et inexactitudes sont inévitables et nécessitent une attention particulière. Elles peuvent être le fait d'adoptions cachées, de paternités extraconjugales, et d'autres facteurs qui peuvent conduire à établir de mauvais liens généalogiques entre certains individus. Heureusement, selon les évaluations disponibles, le taux d'erreurs généalogiques dans BALSAC serait inférieur à 1% des liens. Néanmoins, la rigueur scientifique exige que nous puissions vérifier la fiabilité des données pour les analyses spécifiques à faire. C'est pourquoi un chapitre entier de cette thèse est consacré à l'évaluation de la fiabilité des données canadiennes-françaises pour l'analyse génétique quantitative.

*Le modèle.* Dans cette thèse, le principal cadre d'analyse est basé sur la génétique quantitative appliquée aux questions de biologie évolutive. Tel qu'expliqué dans la section 1.3 plus haut, la génétique quantitative utilise les connaissances sur les relations d'apparentement entre les individus d'une population et les données phénotypiques, pour faire des inférences sur l'héritabilité et le potentiel évolutif de traits d'intérêt. Une force de cette approche, nous l'avons vu, est que la connaissance explicite des loci génétiques impliqués dans l'architecture du trait étudié n'est pas requise.

J'ai utilisé l'inférence statistique pour estimer les paramètres génétiques d'intérêt au regard de mes questions de recherche sur l'évolution. Au début du XXI<sup>e</sup> siècle, la plupart des études de génétique quantitative reposaient sur des méthodes de maximum

de vraisemblance restreinte (REML ; Patterson & Thompson, 1971) combinées au modèle animal (Henderson, 1975) pour aborder les questions d'évolution dans les populations naturelles (Kruuk, 2004 ; Thompson, 2008). Cette approche s'inscrit dans le paradigme fréquentiste en statistiques. Le domaine de la génétique quantitative est historiquement lié à la théorie statistique et tous deux ont énormément bénéficié des avancées réalisées dans les deux domaines (Lynch et Walsh, 1998 et références qui y sont cités). Les méthodes bayésiennes d'inférence des paramètres génétiques quantitatifs furent quant à elles développées d'abord pour les études en reproduction animale, puis appliquées plus tard à l'étude des questions en biologie de l'évolution (Sorensen et Gianola, 2002).

Par rapport aux méthodes de vraisemblance, telles que REML, qui renvoient généralement des estimations ponctuelles des paramètres génétiques d'intérêt, ainsi qu'une mesure de leur incertitude, les méthodes bayésiennes, considérées comme une extension naturelle des méthodes de vraisemblance, génèrent une distribution postérieure de probabilité du paramètre, compte tenu des données observées, soit :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

$A$  et  $B$  étant deux événements distincts.  $P(A|B)$  est la probabilité postérieure, la probabilité que  $A$  soit vrai, étant donné que  $B$  est vrai.  $P(B|A)$  est la vraisemblance, c'est-à-dire la probabilité que  $B$  soit vrai, étant donné que  $A$  est vrai.  $P(A)$  est l'*a priori*, la probabilité que  $A$  soit vrai.  $P(B)$  est la marginalisation, la probabilité que  $B$  soit vrai. Morrissey et al. (2014) suggèrent qu'une distribution postérieure est plus interprétable en tant que représentation complète du support statistique pour un paramètre donné ayant une valeur donnée. En outre, un autre avantage est que nous pouvons utiliser des transformations des distributions postérieures de paramètres directement modélisés pour extraire une distribution postérieure valide d'un paramètre dérivé, la corrélation génétique, par exemple.

## CHAPITRE II

### LACK OF GENETIC RESPONSE TO SELECTION IN A HUMAN LIFE HISTORY TRAIT IN PREINDUSTRIAL POPULATIONS UNDERGOING EXPANSION

Le contenu de ce chapitre fera l'objet d'une soumission en anglais en tant que prépublication dans la revue *PCI Evol Bio* (Peer Community In) et comme soumission pour publication dans la revue *Evolution Letters* (Wiley)

#### 2.1 Contribution des auteurs

Walid Mawass a conçu et développé l'étude. Walid Mawass a effectué l'analyse des données et les analyses computationnelles avec la contribution de Emmanuel Milot. Walid Mawass a rédigé le manuscrit. Emmanuel Milot a contribué à la version finale du manuscrit. Emmanuel Milot a supervisé le projet.

#### 2.2 Résumé de l'article

L'existence de la microévolution dans la nature est largement acceptée, mais sa présence plus large en tant que processus adaptatif permanent dans les populations naturelles actuelles reste un sujet d'étude. Bien que l'on observe une sélection directionnelle généralisée sur les traits héréditaires, tels que les traits d'histoire de vie, le résultat prédominant apparaît souvent comme une stase évolutive, où une réponse microévolutive fait défaut. Malgré cela, la microévolution a été démontrée dans diverses populations et espèces, ce qui incite à se demander si des populations confrontées à des conditions écologiques et à des pressions sélectives comparables présentent des résultats microévolutifs similaires pour les mêmes traits. Dans cette étude, nous testons cette question en analysant l'évolution de l'âge des femmes mariées à la première reproduction (APR) dans trois populations humaines contemporaines. Outre leur pedigree, les trois populations semi-indépendantes sont considérées comme des "réplicas" car elles partagent un contexte génétique, démographique et historique. Comme prévu, nous constatons une sélection directionnelle vers un APR plus précoce au cours de la période étudiée dans les trois populations. D'après les modèles de génétique quantitative, l'APR présente une faible héritabilité dans les trois jeux de données ( $h^2$  moyen = 0,11), mais ce n'est que dans une seule population (Charlevoix) que nous avons détecté une corrélation



génétique négative entre l'APR et le fitness relatif. Dans tous les jeux de données, nous avons constaté que le fitness relatif était héritable. Dans la population de Charlevoix, le gradient de sélection génétique standardisé représente plus de 50 % du gradient de sélection phénotypique standardisé. Nous avons utilisé l'équation du sélectionneur et le second théorème de sélection de Robertson pour estimer la réponse quantitative attendue à la sélection dans l'APR. Seule la population de Charlevoix présente une réponse quantitative attendue de l'APR à la sélection. Il n'y a pas eu de tendance dans les valeurs phénotypiques et d'élevage moyennes de l'APR au cours de la période d'étude dans la population de Charlevoix. Ces résultats indiquent une stase évolutive dans ce trait. Notre étude suggère que la généralité des changements microévolutifs pour le même trait dans des populations répliquées n'est pas omniprésente, indépendamment d'une sélection directionnelle constante et du fait que le trait présente une variation génétique additive décente.

## 2.3 Lack of genetic response to selection in a human life history trait in preindustrial populations undergoing expansion

### Abstract

The existence of microevolution in the wild is widely accepted, yet its broader occurrence as an ongoing adaptive process in present-day natural populations continues to be a subject of inquiry. While widespread directional selection on heritable traits, such as life history traits, is observed, the predominant outcome often appears as evolutionary stasis, where a microevolutionary response is lacking. Despite this, microevolution has been demonstrated in various populations and species, prompting the inquiry into whether populations facing comparable ecological conditions and selective pressures exhibit similar microevolutionary outcomes for the same traits. In this study, we test this question by analyzing the evolution of married women's age at first reproduction (AFR) in three contemporary human populations. In addition to being pedigreed, the three semi-independent populations are considered 'replicates' given they share a genetic, demographic, and historical background. As expected, we find directional selection towards earlier AFR during the studied period across the three populations. Based on quantitative genetic models, AFR exhibited low heritability in all three datasets (average  $h^2 = 0.11$ ) but only in a single population (Charlevoix) did we detect a negative genetic correlation between AFR and relative fitness. Across all datasets, we found that relative fitness was heritable. In the Charlevoix population, the standardized genetic selection gradient accounts for more than 50% of the standardized phenotypic selection gradient. We used the breeder's equation and Robertson's second theorem of selection to estimate the expected quantitative response to selection in AFR. Only in the Charlevoix population, we find an expected quantitative response in AFR to selection. There was no trend in the average phenotypic and breeding values of AFR across the study period in the Charlevoix population. These results point to an evolutionary stasis in the AFR of married women. Our study suggests that the generality of microevolutionary change for the same trait in replicate populations is not ubiquitous regardless of constant directional selection and the trait exhibiting decent additive genetic variation.

**Keywords:** microevolution, selection, additive genetic variance, animal model, stasis

## Introduction

Evidence of microevolutionary change, defined as a change in the distribution of a phenotypic trait within a given population due to a genetic response to selection, has multiplied during the past few decades. The evidence comes from numerous experimental studies of local adaptation (e.g., Barrett *et al.* 2011; Stuart *et al.* 2014; Guo *et al.* 2015) and multiple studies documenting rapid evolutionary change in natural populations (e.g., Franks *et al.* 2007; Pigeon *et al.* 2016; Bonnet *et al.* 2019; see Franks *et al.* 2014 for a review). The expectation is that, if there are no genetic constraints, such as tradeoffs between the focal and other traits, natural selection acting on a trait exhibiting additive genetic variation (matching the direction of selection) should lead to an evolutionary change, i.e., change in the mean of the trait between generations (Teplitsky *et al.* 2014). The fact of microevolution in the wild is not seriously challenged anymore. However, its generality as an ongoing adaptive process in contemporary natural populations remains an open question.

A first broad question is how ubiquitous microevolution is. Although directional (and sometimes enduring) selection on heritable traits is widespread in nature (Byars *et al.* 2010; Kokko *et al.* 2017), the most frequent outcome appears to be ‘evolutionary stasis’, namely the lack of a microevolutionary response to it (e.g., O’Sullivan *et al.* 2019 for body size and Gienapp *et al.* 2006 for breeding time; see Merilä *et al.* 2001 and Pujol *et al.* 2018 for a review on this topic). Several explanations to this apparent paradox have been proposed, falling into two main categories (Kruuk *et al.* 2008; Merilä & Hendry 2014, Teplitsky *et al.* 2014; Pujol *et al.* 2018): i) statistical explanations that pertain to the ways that genetic parameters are estimated from empirical data, and ii) biological explanations that involve constraints on evolutionary potential (see below). However, it remains unclear why evolutionary stasis is apparently so common. Nevertheless, and as aforementioned, microevolution has now been documented for several populations and species, raising a second question pertaining to the generality of this process: do the same traits in populations exposed to similar ecological conditions and selective pressures show parallel microevolutionary responses?

Convergent phenotypic evolution has been observed in the wild (e.g., Colombo *et al.* 2013), so have cases of parallel evolution (e.g., Elmer *et al.* 2014). Parallel evolution

stands out in evolutionary biology as compelling evidence for adaptation, with independently evolved similar traits suggesting deterministic processes, often natural selection (Bolnick *et al.* 2018). Biologists utilize parallel evolution of genes, phenotypes, or ecotypes to infer analogous selective pressures, limited favored solutions, and indicative adaptations (Bailey *et al.* 2017). However, Bolnick *et al.* (2018) make a case that parallel evolution is best viewed as an extreme case in a continuum of evolutionary trajectories and might not represent replicated evolution generally. In this study, we examine this question by testing the generality of recent microevolution in semi-independent human population ‘replicates’ that have burgeoned after an initial colonization event during the preindustrial period, followed by demographic and geographic expansion.

Interestingly, with respect to investigating the generality of microevolution, humans may not be the first species to come to mind. Different arguments have been set forward to claim evolutionary stasis or even halt in human populations (Gould 2000), for instance that the transition to modernity has driven the decline in opportunity for selection, or that cultural evolution has replaced biological evolution (see Milot and Stearns 2020 for a discussion on this topic). Notwithstanding that it is likely shared only by a small minority nowadays, this view is soundly rejected by empirical evidence on different human populations (Byars *et al.* 2010; Moorad 2013; Bolund *et al.* 2015). Stearns *et al.* (2010) compiled a comprehensive list of estimates of selection (both directional and stabilizing) acting on a suite of heritable human traits (life history, physiological, and morphological traits) from across different populations and time periods. Evidence also comes from genomic data on several human complex traits (Tropf *et al.* 2015; Beauchamp 2016; Kong *et al.* 2017; Sanjak *et al.* 2018). Human populations indeed offer advantages for this type of study, given the availability of large-scale datasets, such as deep-rooted genealogies and ever-increasing genomic databases, and detailed information on historical demography, to tackle the question of the generality of microevolution.

For evolutionary change by selection to occur requires two key elements: 1) selection altering the phenotypic trait distribution in the parental generation (i.e., a covariance between trait and fitness), and 2) heritable genetic variation underlying the

phenotypic trait distribution, which allows the transmission of the selection-induced changes to the offspring generation. However, while these two elements are detectable in many wild populations, it is hard to robustly demonstrate a genetic response to selection, namely a change of the predicted amplitude in the mean breeding value (BV; also known as the individual's genetic merit) across generations (Kruuk & Hadfield 2007; Hadfield *et al.* 2010; Morrissey *et al.* 2012). There are three main explanations for this. First, a phenotypic correlation between a trait and that in fitness need not imply a causal relationship between them, even when the trait is heritable (Walsh & Lynch 2018). Hence, selection at the phenotypic level may be a poor predictor of microevolutionary change. This is especially true of predictions based on the Breeder's equation (BE; Lush 1937, see Morrissey *et al.* 2010 for an in-depth discussion on this topic). On the other hand, Robertson's secondary theorem of selection (STS) states that the expected evolutionary change from one generation to the next is equal to the additive genetic covariance of a trait with relative fitness (Robertson 1966, 1968). Both can be derived from Price's general theorem of selection (Price 1970, 1972a), which is an exact mathematical theorem of selection (BE and STS being approximations) given a completely general description of any evolutionary response (Queller 2017). Though the STS doesn't provide any information on causality, the main advantage of this model is that all the factors that contribute to trait-fitness covariation need not be known (Hadfield 2008). Therefore, Morrissey *et al.* (2012) suggest that using the STS to test for a response to selection might reveal that there is no expected response to selection, which explains the apparent evolutionary stasis found when using the BE.

Second, in part as a consequence of the former reason, phenotypic changes, even those occurring in the direction predicted by selection, may not be indicative of underlying genetic changes and thus cannot always be interpreted as strong evidence for a response to selection (Hadfield *et al.* 2010). Quantitative genetic (QG) methods based on mixed-effect linear models, such as the 'animal model', allow the estimation of individual BVs (Kruuk 2004). If a trend in BVs between generations (or cohorts) can be detected, then it can provide credible evidence for a genetic response of trait mean to natural selection (Wilson *et al.* 2010).

Third, when the focal trait is correlated to other (unmeasured) traits, any phenotypic change, or lack of, in the former, may reflect indirect selection induced by direct selection acting on correlated traits instead, or trade-offs among traits that impose limits on the evolution of the focal trait (Teplitsky *et al.* 2014; Stinchcombe *et al.* 2014). The assessment of multivariate selection on correlated traits is done through the estimation of the genetic variance-covariance matrix  $\mathbf{G}$  (Arnold *et al.* 2008). A central assumption of the BE, whether in its univariate or multivariate form, is that the covariance between trait(s) and fitness is causally explained by the trait(s) included in the model, such that no trait is missing that would partly explain this covariance (i.e., an unmeasured trait under selection that would be correlated to a measured trait; Queller 1992; Hadfield 2008; Morrissey *et al.* 2010). However, this assumption is relaxed in the case of the STS since trait evolution is predicted directly from the genetic component of the covariance between a trait and fitness. Therefore, the STS is well suited to quantify selection on genetically correlated traits and predict multivariate evolutionary change (Wilson & Poissant 2016).

Following the pioneering work of Lande & Arnold (1983) and Endler (1986) in the quantitative estimation of natural selection, evolutionary biologists are hard-pressed in trying to understand the major discrepancy between two general observations: i) there is evidence of frequent directional selection in nature on a variety of quantitative complex traits (Stearns *et al.* 2010; Courtiol *et al.* 2012; Hendry 2017), and ii) there is limited evidence of response to these selection pressures in many natural systems (Pujol *et al.* 2018). One challenge lies in the mismatch between what is expected and what is actually observed. The main mismatch we treat here is between the expectation of microevolutionary change in natural systems based on predictive models and the lack of any observed response to selection in these systems.

The French-Canadian (FC) population offers several advantages to investigate the generality of the microevolutionary process in contemporary populations exposed to similar situations. First, its history is characterized by sequential founder events that occurred in different geographical areas of what was then New France (today's Québec, Canada), through colonization waves followed by demographic expansions. Life history theory predicts that early reproductive onset will be selected for in a rapidly growing

iteroparous population (Cole 1954; Caswell 1989, 2001; Roff 1992; Stearns 1992). Milot *et al.* (2011) demonstrated empirically this relationship using LRS (defined as the number of children who survived to age 15 years old) and the age at first reproduction of married women in the preindustrial and endogamous FC population of île aux Coudres (IAC). They also report empirical evidence for rapid microevolutionary change in average AFR, towards younger ages, as a response to natural selection. In this study, we used three, much larger, preindustrial FC populations that share both a genetic, historical, and demographic background with the population of île aux Coudres.

All of these populations share a deep historical background and similar ecological and demographic conditions allowing us to consider them as “natural replicates”. The three populations studied here exhibit a similar demographic pattern of consistent expansion during the study period mirroring the expansion observe in the IAC population (Pelletier *et al.* 2017). During the study period, all of the populations can be considered preindustrial and exhibit a natural fertility regime (Gagnon 2015). The three populations share the same historical narrative of settler expansion into the regions around the St. Lawrence River following initial waves of French migrants (Moreau 2011). It follows then that these three populations also share a very close genetic background (Moreau *et al.* 2007). Moreover, the population of IAC was settled from a small number of families from the Charlevoix population (Martin 1957).

Given this setup, we aim to determine if the microevolutionary change in AFR detected in the population of IAC is a general phenomenon for populations under similar conditions or rather local (idiosyncratic) event exclusive to IAC. In this study, we applied a QG approach in each of the three FC populations to: i) investigate the genetic basis of the AFR of married women, ii) estimate the expected evolutionary change in the trait, and iii) determine if there was a response to selection. Our results revealed an evolutionary stasis in AFR, i.e., an absence of response to selection, which was shown to be heritable, even though there was consistent and unbiased selection for earlier values of AFR.

## **Materials and methods**

### **Study system**

The historical FC population is characterized by having detailed marriage, baptism and burial acts recorded in parish and civil registers that have been compiled and linked since the 1970s in two large databases: BALSAC (<http://balsac.uqac.ca/>; Bouchard 1989; Vézina & Bournival 2020) and the Ancient Québec Population Register (AQPR, Dillon *et al.* 2018). This extensive information provides deep-rooted pedigrees and makes it possible to reconstruct the life history of individuals and families descended from the French settlers that founded New France in 1608. We used this data to conduct QG analyses.

As mentioned above, the expansion of the FC population occurred through a series of founder effects followed by demographic expansion. We identified three preindustrial sub-(regional) populations for which detailed individual life histories are available and pedigrees of good quality to estimate genetic parameters with QG methods applied here (Mawass & Milot 2022). The three populations are those who were born and married in the following regions between foundation and 1833: (i) Charlevoix, a region located 100 km to the North-East of Québec City on the north bank of the St. Lawrence River; (ii) Côte-de-Beaupré, flanked by Québec City on the West and Charlevoix on the East; (iii) Bois-Francs (also known as Centre-du-Québec), located on the south shore of the St. Lawrence and lying roughly midway between Québec City and Montréal. Each of these populations received a number of the earliest French settlers who expanded in these regions from the main settlement of Québec City.

Given our focus on the age at first reproduction of women, we retained women who married only once with records of at least one reproductive event. An additional filter criterion of our datasets included selecting married women based on their age of last reproduction, setting an upper bound representing the end of natural fertility which usually coincides with menopause generally occurring at around 45 years old (Gold 2011). Our selection criteria are based on previous studies that used similar datasets (Milot *et al.* 2011; Pelletier *et al.* 2017)

Table 1 includes descriptive statistics of the married women's phenotypic traits in the three analyzed datasets. Though the three datasets cover very similar time periods, we offer additional information on how the study periods were selected. We selected a lower bound of the study period based on the earliest year of birth in the filtered



datasets. We selected an upper bound based on the last birth year with maximal life history information coverage. Only in the case of Charlevoix that experience a major immigration event towards the Saguenay-Lac-St-Jean area depleting the adult population, we selected the last birth year of women born and married in the region. With this in mind, we arrived at the following time periods: Bois-Francs [1688 – 1833]; Charlevoix [1683 - 1816]; Côte-de-Beaupré, [1661 – 1832]. This grouping was also required to calculate relative fitness, as explained below. Acts of marriage, baptism, and sepulture available for each of the datasets, including those prior to the study period, were used to reconstruct the pedigrees associated with each dataset.

### **Fitness measurements**

We measured relative fitness,  $w$ , for each married woman as her reproductive value at birth (Fisher 1930; Lenski & Service 1982), as calculated relative to vital rates in the studied population:

$$w_i = \sum_{x=0}^{\infty} \lambda^{-x} l_{x_i} m_{x_i},$$

where  $w_i$  is the fitness of woman  $i$ ,  $l_{x_i}$  her survival to age  $x$ ,  $m_{x_i}$  her age-specific fertility at age  $x$ , and the intrinsic rate of population growth  $\lambda$  (Moorad 2014). Given the size of the three populations, calculating  $\lambda$  separately for annual cohorts would be burdened by high imprecision (earlier years comprised less than <10 married individuals). Therefore, we calculated  $\lambda$  for 20-year cohorts.

### **Quantitative model fitting**

We performed all model fitting analyses under a Bayesian framework and avoided mixing Bayesian and frequentist approaches, which can lead to interpretation complications, if not incoherence when comparing parameter estimates provided by the two approaches. In a Bayesian framework, the posterior distribution is readily interpretable in terms of the probability of causes (i.e., probability that a parameter takes different values) as opposed to the frequentist's probability of effects (i.e., data) given the causes. Additionally, certain parameters of interest can be derived from the posterior distributions of fitted parameters, such as the genetic covariance, obtained from the distribution of traits' breeding values. All models were fitted to data using the *MCMCglmm* package (Hadfield 2010) for R (v4.2.1; R Development Core Team 2022).

### Phenotypic selection

We investigated the relationship between AFR and relative fitness using mixed-effect linear regression models. We standardized both our response variable (AFR) and covariates (i.e., mean = 0, s.d. = 1) to obtain standardized selection coefficients (Lande & Arnold, 1983; Postma 2006). We fitted both linear ( $\beta$ ) and quadratic effects ( $\gamma$ ) to estimate linear and non-linear selection gradients. The equation of the model is as follows:

$$\begin{aligned} \text{Fitness} \sim & \beta_0 + \beta_{AFR} + \gamma_{AFR} + \beta_{kinship} + \beta_{kinship^2} + \beta_{birthcohort} + \beta_{twin} \\ & + \beta_{infmort} + \beta_{parish} + m + \varepsilon \end{aligned}$$

The models included the following fixed effects: whether a married woman gave birth to at least one pair of twins, birth cohort, birth parish, and the number of her infants who died before age 1 yo. Demo-genetic studies had shown that the population of the Charlevoix region exhibited inbreeding ( $\sim 0.007$ ; Bouchard & De Braekeleer, 1991). In addition, at île aux Coudres, which is located in this region, inbreeding was non-linearly related to infant mortality rate (Boisvert & Mayer 1994). Thus, we included the quadratic effect of the parent's kinship coefficient in our final models. We also included a residual variance component and the variance due to the effect of the familial environment shared by siblings by fitting the identity of the mother as a random effect and recalling that couples were monogamous. All models were fitted with a Gaussian error structure. We used *MCMCglmm*'s default priors, equivalent to a Gaussian distribution for the fixed effects and an inverse-gamma prior distribution for the variance terms (i.e., shared familial environment).

### Bivariate quantitative genetic model

We used pedigree-based quantitative genetic analyses to assess the genetic basis of AFR and relative fitness, calculate trait heritability and selection based on the Robertson covariance, and detect potential genetic responses to selection in AFR. Specifically, we used the so-called “animal model” (Henderson 1975, 1986), a type of linear mixed-effects model that integrates the pairwise genetic relatedness matrix provided by the pedigree, using of all levels of relatives (e.g., siblings, grandparents, and

aunts/uncles as well as more distant relatives). Additionally, to a reasonable degree, the animal model allows the separation of genetic from nongenetic components of phenotypic change (Wilson & Poissant 2016). First, we fitted bivariate models with AFR and fitness as response variables to decompose their phenotypic variance into additive genetic, environmental, and residual components and to estimate genetic covariances among traits for each of our three analyzed datasets (Kruuk 2004; Wilson *et al.* 2010).

We fitted a bivariate model to quantify the amount of additive genetic variance,  $\sigma_A^2$  in AFR and fitness, as well as their genetic additive covariance. Similar to the phenotypic analyses, we fitted maternal identity as a random factor to control for the shared familial environment and to estimate the variance ( $\sigma_M^2$ ) explained by these effects. The structure of the model is as follows:

$$\mathbf{z} = \boldsymbol{\mu} + \beta_{kinship} + \beta_{kinship^2} + \beta_{birthcohort} + \beta_{twin} + \beta_{infmort} + \beta_{parish} \\ + \mathbf{a} + \mathbf{m} + \mathbf{e}$$

where  $\mathbf{z}$  is the response trait vector,  $\boldsymbol{\mu}$  is a vector of means,  $\mathbf{a}$  and  $\mathbf{m}$  represent the additive genetic effects and shared familial effects, respectively, while  $\mathbf{e}$  is the residual error matrix and  $\beta$  is a vector of partial regression coefficients for each fixed predictor. Additionally, we included as fixed effects whether a woman gave birth to at least one pair of twins and the number of her infants who died before age 1 (infant mortality count), as well as the focal individual's inbreeding coefficient as a covariate (quadratic effect). The woman's birth cohort and birth parish were entered as covariates to control for temporal trends or spatial effects that may be due to unmeasured factors.

Narrow-sense heritability ( $h^2$ ) was calculated by dividing  $\sigma_A^2$  by the sum of all variance components included in the model ( $\sigma_A^2 + \sigma_M^2 + \sigma_E^2$ ). However, narrow-sense heritability is not always considered an adequate parameter to compare the capacity of a trait to respond to selection across populations. Hence, we calculated the coefficient of

additive variance,  $CV_A = \frac{\sqrt{\sigma_A^2}}{\bar{X}}$ , with  $\bar{X}$  being the trait's mean, which is defined as a

measure of the evolvability of a trait or the magnitude of possible change due to selection relative to the trait's mean (Houle 1992).

Animal models were fitted using the *MCMCglmm* R package (Hadfield 2010). We ran the analyses in two ways: 1) with our response variables mean-standardized to get variance estimates at the observed scale as to calculate the coefficient of variation using the non-standardized mean, and 2) standardized with mean = 0 and s.d. = 1 to obtain standardized estimates of (co)variances. All of our response variables were fitted as Gaussian (see Figures S2 & S3 for distribution of AFR and fitness for the three analyzed datasets). To minimize autocorrelation in the MCMC chain samples, all models were run for 3,000,000 iterations after an initial burn-in of 500,000 iterations and sampled every 3,000 iterations to obtain the posterior distribution of parameters (i.e., 1,000 iterations kept from MCMC chains). We used an inverse Wishart distribution for the prior of the fitted variance-covariance matrices (Hadfield 2012; De Villemereuil *et al.* 2016).

### **Bias in phenotypic selection gradients**

To assess if there was bias in the phenotypic selection, we calculated the “environmental bias” metric which corresponds to the difference between the “environmental selection gradient”,  $\beta_E$ , the regression slope of environmental deviations for fitness on environmental deviation for AFR, and the “genetic selection gradient”,  $\beta_G$ , the regression slope of breeding values for fitness on breeding values for AFR (Rausher 1992, Hadfield 2008, Morrissey *et al.* 2012, Stinchcombe *et al.* 2014). This relates to the observation that phenotypic covariance can be generated between fitness and a focal trait through purely environmental causes (e.g., nutritional state) without the trait having any causal influence on fitness (Price, Kirkpatrick, & Arnold, 1988). Consequently, the phenotypic selection gradient would be biased and there should be no expectation of a response to selection, even if the trait is heritable, given that there is no actual selection on the underlying breeding values (Rausher 1992). We denote  $\Delta\beta = \beta_E - \beta_G$  as the environmental bias in selection, the posterior distribution of which was extracted from the Bayesian models fitted. We calculated  $\beta_G$  and  $\beta_E$  for AFR as  $\beta_G = \frac{\sigma_A(AFR, fitness)}{\sigma_A^2}$ ; for  $\beta_E$ , we divided the sum of all environmental covariance terms in the bivariate animal

model by the sum of the corresponding variance components for AFR,  $\beta_E = \frac{\sigma_M(\text{AFR}, \text{fitness}) + \sigma_E(\text{AFR}, \text{fitness})}{\sigma_M^e + \sigma_E^2}$ .

### Interpretation of posterior distributions

The 95% HPD interval provides a measure of the uncertainty about the true parameter value (they are not to be confused with confidence intervals). This means the larger the HPD interval is the wider the range of parameter values that are plausible (i.e., a lower precision). We also used two Bayesian indices to draw inferences from the posterior distribution about the existence and significance of a given genetic or environmental effect on a trait (see Makowski *et al.* 2019b). The Region of Practical Equivalence (or ROPE)-based index, specifically the *ROPE* (95%), corresponds to the proportion of the 95% HPD that lies within the ROPE (loosely speaking, this a Bayesian version of the frequentist’s concept of “power”). This type of index involves redefining the alternative hypothesis from the classic point-null (i.e., the value 0) to a range of values considered negligible or too small to be considered practically relevant to the biological question being studied (Kruschke 2014). According to Makowski *et al.* (2019b), the ROPE-based index offers information in favor of the null hypothesis and is sensitive to sample size (see SI for information on how the *ROPE* range was defined in our study). We also used the *Probability of Direction* (*pd*) which is the Bayesian equivalent of the frequentist *p*-value in that it reflects the existence of an effect. This index varies between 50% and 100%, a high value suggests that an effect exists while a low value indicates uncertainty around its existence. To calculate these indices in our study, we used the R package *bayestestR* (Makowski *et al.* 2019a).

### Microevolutionary response to selection

In this study, we used both the Breeder’s equation and Robertson’s secondary theorem of natural selection or Robertson’s covariance to predict the expected quantitative change in AFR between generations in terms of standard deviation. The main difference between BE and STS is that the former assumes environmental factors do not influence the covariance between fitness and the trait. If this assumption is violated, then BE will fail while STS can still hold, as the latter predict the response to

selection directly from the relationship between breeding values for trait and relative fitness (Morrissey *et al.* 2012)

BE states that the response to selection in one generation in a given trait is equal to the product of the narrow-sense heritability of that trait and the magnitude of selection acting on it:

$$\Delta\bar{z} = \sigma_P(z, w) \frac{\sigma_A^2}{\sigma_P^2} = sh^2 ,$$

where  $\Delta\bar{z}$  is the predicted change in trait mean phenotype after one generation,  $s$  represents the phenotypic selection differential and  $h^2$  is the narrow-sense heritability amounting to the proportion of phenotypic variance explained by the additive genetic variance. STS states that the evolutionary change is equal to the additive genetic covariance of a trait with relative fitness:

$$\Delta\bar{z} = \sigma_A(z, w)$$

We tested for a response to selection using a method accounting for random genetic drift, as the latter can produce temporal change in mean BVs not due to selection (Hadfield *et al.* 2010). The method consists of using the posterior estimate of  $\sigma_A^2$  from each of the 1,000 MCMC iteration sampled to randomly generate random breeding values (RBVs) along the pedigree, thereby mimicking a pure drift scenario, using the `rbv` function of the `MCMCglmm` package. Then, RBVs were regressed against birth cohort, and the slope ( $\beta_{RBV}$ ) was compared with that ( $\beta_{BV}$ ) for BVs estimated from real data for the same MCMC iteration. The proportion of times where the absolute value of  $\beta_{RBV}$  was as high or higher than the absolute value of  $\beta_{BV}$  was taken as the probability of obtaining a trend in breeding values as steep as the observed one as the result of drift only.

## Results

### Phenotypic selection

Table 1 provides summary statistics of the life history traits and fitness measure for each of the three populations used in this study. The results of the linear mixed models, which includes the standardized linear ( $\beta$ ) and quadratic ( $\gamma$ ) selection gradients for AFR, revealed a strong linear selection on AFR younger values across all three

populations (Table S1). Thus, a younger AFR is associated with higher values of fitness in all the analyzed populations. The standardized selection gradient on AFR was strongest for Charlevoix with  $\beta = -0.64$  [95% HPD:  $-0.68 - -0.60$ ], then for Côte-de-Beaupré  $\beta = -0.34$  [ $-0.37 - -0.30$ ], and for Bois-Francis  $\beta = -0.25$  [ $-0.29 - -0.22$ ] (Figure 1 and Table S1). Interestingly, we also detected a slight quadratic selection gradient in the Charlevoix population  $\gamma = 0.07$  [ $0.05 - 0.08$ ], but this parameter was negligible for the two other populations (Table S1).

### Quantitative genetic parameters

In this section, we detail the results of the bivariate animal model for each studied population. We report the results based on the standardized models to allow for comparison at the same scale. The bivariate model additive genetic variation and moderate heritability in both AFR and relative fitness (see Table 2.2 & 2.3). The coefficient of additive genetic variation was modest for AFR in Charlevoix ( $CV_A = 0.15$  [ $0.06 - 0.26$ ]) and very low in Bois-Francis ( $CV_A = 0.03$  [ $0.02 - 0.06$ ]) and Côte-de-Beaupré ( $CV_A = 0.04$  [ $0.02 - 0.06$ ]). Surprisingly, the coefficient of additive genetic variation was very low for relative fitness in Charlevoix ( $CV_A = 0.05$  [ $0.03 - 0.08$ ]), while it was slightly higher in Bois-Francis ( $CV_A = 0.24$  [ $0.15 - 0.35$ ]) and Côte-de-Beaupré ( $CV_A = 0.17$  [ $0.10 - 0.23$ ]).

The genetic correlation between AFR and relative fitness was completely different between the three populations: in Charlevoix, the genetic correlation negative was negative and strong in magnitude ( $r_g = -0.56$  [ $-0.77 - -0.30$ ]); in Bois-Francis, the correlation was negligible ( $r_g = -0.02$  [ $-0.36 - 0.25$ ]), the same result was found for Côte-de-Beaupré ( $r_g = 0.06$  [ $-0.35 - 0.30$ ]). The same pattern was found for the additive genetic covariance between AFR and relative fitness. Across all three populations, the shared familial (maternal) environment was found to be an important contributor of variance to the phenotypic variation in AFR and relative fitness. However, we note that this parameter might be underestimated based on sensitivity analyses previously done on these datasets (Mawass & Milot 2022).

We found that the additive genetic covariance between AFR and fitness, which also corresponds to the STS, differed between Charlevoix and the two other populations. For the Charlevoix population,  $\sigma_A = -0.07$  [ $-0.13 - -0.02$ ] with a *pd* index of 100%,

which means that is highly likely that it is negative in direction. The ROPE (25%) index reveals that there is a  $P = 0.007$  that this parameter falls in the  $[-0.03 - 0.03]$ . For the Bois-Francis population,  $\sigma_A = -0.002 [-0.034 - 0.025]$  with a  $pd$  index of 59.3%, which means that is equally likely that this parameter could be either negative or positive in direction. The ROPE (25%) index indicates that there is a  $P = 0.936$  that this parameter falls in the  $[-0.03 - 0.03]$ . The same applied to the Côte-de-Beaupré population,  $\sigma_A = 0.007 [-0.037 - 0.026]$  with a  $pd$  index of 50.8% with a  $P = 0.939$  that this parameter falls in the  $[-0.03 - 0.03]$ .

We used the approach of calculating the metric of bias in the standardized phenotypic selection gradient to reveal if there is any environmental bias. For the population of Charlevoix, we found  $\Delta\beta = -0.19 [-0.39 - 0.11]$  (Table 3). According to the ROPE (25%) index, there is a  $P = 0.084$  that  $\Delta\beta$  of Charlevoix would fall in the range  $[-0.03 - 0.03]$ . For the Bois-Francis population,  $\Delta\beta = -0.18 [-0.59 - 0.17]$  with a  $P = 0.077$  that it would fall in the range  $[-0.03 - 0.03]$ . For the Côte-de-Beaupré population,  $\Delta\beta = -0.45 [-0.69 - -0.04]$  with a  $P = 0.007$  that it would fall in the range  $[-0.03 - 0.03]$ .

### **Observed vs predicted microevolutionary change**

The results from the animal models clearly reveal that AFR is heritable, under unbiased selection and strongly correlated with our fitness measure. Using STS, given by the additive genetic covariance between AFR and fitness,  $\sigma_A$ , we found that the expected microevolutionary change from one generation to the next in AFR is an advancement of 0.07 phenotypic standard deviations in Charlevoix (or the equivalent of 25.5 days/per generation, or  $\frac{1}{2}$  a year for the whole study period), an advancement of 0.002 phenotypic standard deviations in Bois-Francis (or the equivalent of 0.73 days/per generation, or the equivalent of 4 days for the whole study period), and an increase of 0.007 phenotypic standard deviations in Côte-de-Beaupré (or the equivalent of 2.5 days/per generation, or the equivalent of 17.5 days for the whole study period). Using BE, given by multiplying the standardized phenotypic covariance between AFR and fitness by the trait's heritability, we found an expected advancement in AFR of 0.09 standard deviations in Charlevoix (or the equivalent 34.65 days/per generation, or the equivalent of  $\frac{2}{3}$  a year for the whole study period), an advancement of around 0.03



phenotypic standard deviations in both Bois-Francis and Côte-de-Beaupré (or the equivalent of around 10.5 days/per generation, or the equivalent of 63 and 73.5 days for the whole study period, respectively). The prediction based on the breeder's equation is ~50% higher than those made with the STS.

To establish if there was a trend in the average AFR, we ran a mixed effects linear regression model with birth cohort fitted as a covariate. Results from these models show that all three populations had completely different temporal trends in average AFR (Figure 3). The partial regression coefficient of AFR on birth cohort (i.e., slope of regression line) in Charlevoix,  $\beta = 0.02$  [-0.01 – 0.06], in Bois-Francis,  $\beta = -0.17$  [-0.20 – -0.13], and in Côte-de-Beaupré,  $\beta = 0.08$  [0.06 – 0.10]. We examined the trends in the average breeding values of AFR across the three populations to determine if there was a genetic response to selection. In all three populations, the regression coefficient of the average breeding values on birth cohort was negligible, in Charlevoix,  $\beta_{BV} = -0.01$  [-0.04 – 0.02], in Bois-Francis,  $\beta_{BV} = -0.01$  [-0.04 – 0.02], and in Côte-de-Beaupré,  $\beta_{BV} = -0.002$  [-0.01 – 0.006] (Figure 3). We found that for all three populations there is a posterior probability  $P \approx 0.5$  of observing trends of these magnitude or less under a pure drift scenario.

## Discussion

In this study, we tested the generality of microevolutionary change in response to selection in a fitness-related life history trait in three contemporary 'replicate' human preindustrial populations, part of the historic French-Canadian population. The AFR of women born and married in these three populations was found, using a demographically-informed measure of relative fitness, the reproductive value at birth, to be under sustained selection for earlier values, and the trait was found to be modestly heritable (average  $h^2 \sim 0.11$ ). Two different predictive models, the Breeder's equation and Robertson's secondary theorem of selection yielded an expected response to selection in average AFR only in the case of the Charlevoix population. However, this expectation did not match the observed temporal trends, at the phenotypic and additive genetic levels. Therefore, while selection gradients and, to a much lesser extent, heritability in AFR across the three populations were similar to those estimated for the île aux Coudres

subpopulation (Milot *et al.* 2011; Mawass *et al.* 2022), we did not detect a response to selection as was found in the latter study.

### **Microevolution by natural selection is not a general pattern across replicates**

In the île aux Coudres population, over a 140-year period, the average AFR of married women changed from 26 years to 22 years, while the average breeding values of AFR changed from +1 years to -2 years (Milot *et al.* 2011). We did not observe trends of these magnitudes in any of the populations studied here, especially in the Charlevoix population which encompasses the île aux Coudres population. Over their respective study period, all three populations were expanding demographically, so that the theoretical expectation is that earlier values of AFR should be favored. Indeed, AFR across all populations was under strong phenotypic selection for earlier values. This selection was consistent between the generations and is analogous to that reported in Milot *et al.* (2011) for île aux Coudres. More generally, other studies of modern human populations have demonstrated this directional selection for earlier values of AFR, using either quantitative genetic methods (e.g., Stearns *et al.* 2010) or genomic methods (e.g., Tropf *et al.* 2015).

Most of these studies point to the fact that there was either an observed phenotypic trend in the opposite direction of selection or simply no observed change in the average phenotypic trait value. We found similar results from our analyses, where in the Charlevoix population there was no change in the average value of AFR over time, and the population of Côte-de-Beaupré exhibited a trend towards later values. Only in the case of Bois-Francs, there was a trend towards earlier values of AFR. However, this phenotypic trend is not influenced by any underlying genetic change corresponding to a response to selection. In Bois-Francs, the genetic correlation between AFR and fitness was negligible suggesting no genetic association between the two variables in this population. The change in average AFR in this population is most likely driven by environmental factors that underlie the negative phenotypic association between AFR and fitness or due to a plastic response.

Based on quantitative genetic analyses, with both AFR and fitness as response variables, we arrived at multiple findings. Let's note beforehand that a strength of the animal model is its ability to leverage information from all relatives connected to each

other through the pedigree to infer and estimate quantitative genetic parameters. Additionally, we have shown elsewhere that the pedigree of Charlevoix provides enough power to detect a genetic effect (Mawass & Milot 2022). Using sensitivity analyses, all three pedigrees were able to detect a minimum value of additive genetic variance and genetic correlation of 0.1. While AFR exhibited a decent amount of additive genetic variance in the studied populations, its narrow-sense heritability is much lower than what was found by Milot *et al.* (2011) for the population of île aux Coudres ( $h^2 = 0.55$  [0.30 – 0.90]), but much closer to the magnitude of values presented in Stearns *et al.* (2010) for other human populations (average  $h^2 = 0.08 \pm 0.01$ ), which include measures from multiple modern populations, from different time periods.

### **Evolutionary stasis or evolutionary constraint**

The phenomenon of evolutionary stasis was first described with regard to the fossil record (Eldredge *et al.* 2005), where the record displays stasis in many species over long time periods. This observation of stasis was contrasted with rapid phenotypic evolution and genetic differentiation in extant populations (e.g., Pigeon *et al.* 2016). However, with advances in measurement techniques and increasing quality of datasets, there is evidence of evolutionary stasis occurring within extant populations in heritable traits under selection (e.g., Munoz *et al.* 2014; O'Sullivan *et al.* 2019). In the case of Charlevoix, both STS and BE predict an advancement in average AFR, albeit STS predicts a weaker response. However, there was no corresponding change in the phenotypic and breeding values of AFR in the direction of selection.

The predictive power and accuracy of BE for natural populations have been questioned. Though still strong and robust with controlled breeding designs, works such as Morrissey *et al.* (2010) point to the pitfalls of using BE to predict the expected response to directional selection in nature. Additionally, reported mismatches between the predicted response to selection in a trait and the observed phenotypic change were shown to be due to an overestimation of the predicted response when using BE (Morrissey *et al.* 2012). In this study, BE provided an expected response to selection in average AFR that is two times larger than that given by STS. The later model focuses on the additive genetic component of the covariance between the focal trait and fitness, possibly reducing the number of confounding factors (e.g., environmental factors) that

can otherwise create an association between fitness and a given trait (Morrissey *et al.* 2012; Queller 2017). Hence, estimates provided by STS should be more conservative than those provided by BE, which is what we find here. Nonetheless, both models predicted for Bois-Francs and Côte-de-Beaupré that no response to selection is expected which matches the trend in breeding values of AFR.

Burt (2001) proposed definitions of different evolutionary patterns in the context of comparative phylogenetic studies. His definitions of evolutionary stasis and evolutionary constraint also apply to studies of evolutionary change in extant populations. The AFR of women born and married in Bois-Francs and Côte-de-Beaupré was under strong and constant directional selection but there was no inferred selection on the breeding values of AFR, also confirmed by the predictions made by BE and STS. This pattern is consistent with evolutionary stasis. There is phenotypic selection acting on AFR in both of these populations, but the phenotypic change in the trait was in opposite directions. Based on the work of Rausher (1992), the observed phenotypic change in the absence of selection on breeding values can be due to selection on environmental deviations in the trait. Using our quantitative genetic analysis and the environmental bias metric (Rausher 1992; Stinchcombe *et al.* 2014), we were able to confirm that there is selection on environmental deviations of AFR in Bois-Francs and Côte-de-Beaupré.

In contrast, in the Charlevoix population, we did detect a genetic selection on breeding values and selection on environmental deviations which were in the same direction. Since there was no change in the breeding values of AFR, this incongruence is consistent with evolutionary constraint. The predictive models BE and STS provide an expectation of a response of selection which was absent in both in the inferred trend in the breeding values of AFR and its observed phenotypic trend. The measure of trait evolvability, the coefficient of additive genetic variation (Houle 1992), provide further evidence that AFR is amenable to microevolutionary change by natural selection in Charlevoix, which is not the case for Bois-Francs and Côte-de-Beaupré.

Pujol *et al.* (2018) offer two classes of explication of evolutionary stasis: statistical explanations and biological explanations. To mitigate statistical issues, we relied on two different but related predictive models, BE and STS (Walsh & Lynch 2018).

Additionally, we used the animal model across all of our analyses combined with previously assessed pedigrees for quantitative genetic study using sensitivity analyses. Our model accounted for non-genetic sources of variation that might have impacted the estimation of the additive genetic variance. However, one element that could impact our estimations is the choice of prior distribution in our models (Morrissey *et al.* 2014). Using more informative priors might increase estimation performance, but the posterior distributions would be more dependent on the prior than on the data.

In terms of biological explanations, multiple mechanisms could explain both the evolutionary stasis and constraint. The authors previously demonstrated in the population of île aux Coudres the role genotype-environment interactions can play in modulating the response to selection (Mawass *et al.* 2022). It is possible that these same interactions altered the magnitude and direction of selection acting on AFR (Hoffmann & Merilä 1999). AFR could be genetically correlated with other traits, either measured or unmeasured, that could limit the response to selection (Stearns 1989). A solution to this issue is to apply a multivariate approach and model relevant traits (e.g., age at last reproduction; Bürkli & Postma 2014). This approach yields a G-matrix (Arnold *et al.* 2008) which can be used to derive genetic correlations and how traits are genetically associated with fitness. Lastly, other more specific factors that might constrain response to selection might include: i) the high average inbreeding coefficient of the population of Charlevoix which can have an impact on the standing variation available for selection to act on since inbreeding ultimately leads to lower genetic diversity in a population (though see Buskirk & Willi 2006); ii) the fact that population of Charlevoix studied here was at its demographic beginning and small in size at the start which means there was possibly a larger role played by random drift in impacting trait changes.

It is clear that there is a very low expectation of a general microevolutionary pattern in response to selection across populations sharing similar ecological, genetic, and demographic conditions. Moreover, we reiterate the suggestion made by Burt (2001) that evolutionary stasis should be the null expectation when studying evolutionary pattern of a given heritable trait under selection. It follows then that evolutionary stasis is the expectation in nature and any change to this state highly depends on the idiosyncratic characteristic of the population in question. Though prudence towards

statistical issues impacting estimation and interpretation is required, there is a more important impetus in exploring and empirically studying the biological mechanisms underlying evolutionary constraints in natural populations.

## References

- Arnold, S. J., Bürger, R., Hohenlohe, P. A., Ajie, B. C., & Jones, A. G. (2008). Understanding the evolution and stability of the G-matrix. *Evolution: International Journal of Organic Evolution*, 62(10), 2451-2461.
- Bailey, S. F., Blanquart, F., Bataillon, T., & Kassen, R. (2017). What drives parallel evolution? How population size and mutational variation contribute to repeated evolution. *BioEssays*, 39(1), 1-9.
- Barrett, R. D., Paccard, A., Healy, T. M., Bergek, S., Schulte, P. M., Schluter, D., & Rogers, S. M. (2011). Rapid evolution of cold tolerance in stickleback. *Proceedings of the Royal Society B: Biological Sciences*, 278(1703), 233-238.
- Beauchamp, J. P. (2016). Genetic evidence for natural selection in humans in the contemporary United States. *Proceedings of the National Academy of Sciences*, 113(28), 7774-7779.
- Boisvert, M., & Mayer, F. M. (1994). Mortalité infantile et consanguinité dans une population endogame du Québec. *Population (French edition)*, 685-724.
- Bolnick, D. I., Barrett, R. D., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non) parallel evolution. *Annual Review of Ecology, Evolution, and Systematics*, 49, 303-330.
- Bolund, E., Hayward, A., Pettay, J. E., & Lummaa, V. (2015). Effects of the demographic transition on the genetic variances and covariances of human life-history traits. *Evolution*, 69(3), 747-755.
- Bonnet, T., Morrissey, M. B., Morris, A., Morris, S., Clutton-Brock, T. H., Pemberton, J. M., & Kruuk, L. E. (2019). The role of selection and evolution in changing parturition date in a red deer population. *PLoS Biology*, 17(11), e3000493.
- Bouchard, G., Roy, R., Casgrain, B., & Hubert, M. (1989). Fichier de population et structures de gestion de base de données: le fichier-réseau BALSAC et le système INGRES/INGRID. *Histoire & Mesure*, 39-57.

- Bouchard, G., & De Braekeleer, M. (1991). *Histoire d'un génome: Population et génétique dans l'est du Québec*. Sillery, Québec: Presses de l'Université du Québec.
- Bürkli, A., & Postma, E. (2014). Genetic constraints underlying human reproductive timing in a premodern Swiss village. *Evolution*, 68(2), 526-537.
- Burt, D. B. (2001). Evolutionary stasis, constraint and other terminology describing evolutionary patterns. *Biological Journal of the Linnean Society*, 72(4), 509-517.
- Buskirk, J. V., & Willi, Y. (2006). The change in quantitative genetic variation with inbreeding. *Evolution*, 60(12), 2428-2434.
- Byars, S. G., Ewbank, D., Govindaraju, D. R., & Stearns, S. C. (2010). Natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences*, 107(suppl 1), 1787-1792.
- Caswell, H. (1989). *Matrix Population Models*. Sinauer Associates.
- Caswell, H. (2001). *Matrix Population Models (2nd ed.)*. Sinauer Associates.
- Cole, L. C. (1954). The population consequences of life history phenomena. *The Quarterly Review of Biology*, 29(2), 103-137.
- Colombo, M., Diepeveen, E. T., Muschick, M., Santos, M. E., Indermaur, A., Boileau, N., ... & Salzburger, W. (2013). The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Molecular Ecology*, 22(3), 670-684.
- Courtiol, A., Pettay, J. E., Jokela, M., Rotkirch, A., & Lummaa, V. (2012). Natural and sexual selection in a monogamous historical human population. *Proceedings of the National Academy of Sciences*, 109(21), 8044-8049.
- Dillon, L., Amorevieta-Gentil, M., Caron, M., Lewis, C., Guay-Giroux, A., Desjardins, B., & Gagnon, A. (2018). The Programme de recherche en démographie historique: past, present and future developments in family reconstitution. *The History of the Family*, 23(1), 20-53.
- Eldredge, N., Thompson, J. N., Brakefield, P. M., Gavrilets, S., Jablonski, D., Jackson, J. B., ... & Miller, W. (2005). The dynamics of evolutionary stasis. *Paleobiology*, 31(S2), 133-145.

- Elmer, K. R., Fan, S., Kusche, H., Luise Spreitzer, M., Kautt, A. F., Franchini, P., & Meyer, A. (2014). Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature communications*, 5(1), 5168.
- Endler, J. A. (1986). Natural selection in the wild. *Monographs in Population Biology*, 21, 1-33.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press.
- Franks, S. J., Sim, S., & Weis, A. E. (2007). Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences*, 104(4), 1278-1282.
- Franks, S. J., Weber, J. J., & Aitken, S. N. (2014). Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary applications*, 7(1), 123-139.
- Gagnon, A. (2015). Natural fertility and longevity. *Fertility and Sterility*, 103(5), 1109-1116.
- Gienapp, P., Postma, E., & Visser, M. E. (2006). Why breeding time has not responded to selection for earlier breeding in a songbird population. *Evolution*, 60(11), 2381-2388.
- Gold, E. B. (2011). The timing of the age at which natural menopause occurs. *Obstetrics and Gynecology Clinics*, 38(3), 425-440.
- Gould, S. J. (2000). The spice of life. *Leader to Leader*, 15, 14-19.
- Guo, B., DeFaveri, J., Sotelo, G., Nair, A., & Merilä, J. (2015). Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC biology*, 13, 1-18.
- Hadfield, J. D. (2008). Estimating evolutionary parameters when viability selection is operating. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635), 723-734.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of statistical software*, 33, 1-22.
- Hadfield, J. (2012). The quantitative genetic theory of parental effects. *The evolution of parental care*, 267-284.



- Hadfield, J. D., Wilson, A. J., Garant, D., Sheldon, B. C., & Kruuk, L. E. (2010). The misuse of BLUP in ecology and evolution. *The American Naturalist*, 175(1), 116-125.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 423-447.
- Henderson, C. R. (1986). Estimation of variances in animal model and reduced animal model for single traits and single records. *Journal of Dairy Science*, 69(5), 1394-1402.
- Hendry, A. P. (2017). *Eco-evolutionary dynamics*. Princeton University Press.
- Hoffmann, A. A., & Merilä, J. (1999). Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution*, 14(3), 96-101.
- Houle, D. (1992). Comparing evolvability and variability of quantitative traits. *Genetics*, 130(1), 195-204.
- Kokko, H., Chaturvedi, A., Croll, D., Fischer, M. C., Guillaume, F., Karrenberg, S., ... & Stapley, J. (2017). Can evolution supply what ecology demands?. *Trends in Ecology & Evolution*, 32(3), 187-197.
- Kong, A., Frigge, M. L., Thorleifsson, G., Stefansson, H., Young, A. I., Zink, F., ... & Gudbjartsson, D. F. (2017). Selection against variants in the genome associated with educational attainment. *Proceedings of the National Academy of Sciences*, 114(5), E727-E732.
- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*.
- Kruuk, L. E. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1446), 873-890.
- Kruuk, L. E., & Hadfield, J. D. (2007). How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology*, 20(5), 1890-1903.
- Kruuk, L. E., Slate, J., & Wilson, A. J. (2008). New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annual Review of Ecology, Evolution, and Systematics*, 39, 525-548.

- Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, 1210-1226.
- Lenski, R. E., & Service, P. M. (1982). The statistical analysis of population growth rates calculated from schedules of survivorship and fecundity. *Ecology*, 63(3), 655-662.
- Lush, J. L. (1937). *Animal Breeding Plans*. Iowa State Press.
- Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, 4(40), 1541.
- Makowski, D., Ben-Shachar, M. S., Chen, S. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in psychology*, 10, 2767.
- Martin, Y. (1957). L'Île-aux-Coudres: Population et économie. *Cahiers de géographie du Québec*, 2(2), 167-195.
- Mawass, W., Mayer, F. M., & Milot, E. (2022). Genotype-by-environment interactions modulate the rate of microevolution in reproductive timing in humans. *Evolution*, 76(7), 1391-1405.
- Mawass, W., & Milot, E. (2022). Assessing the impact of pedigree quality on the validity of quantitative genetic parameter estimates. *bioRxiv*, 2022-11.
- Merilä, J., Sheldon, B. C., & Kruuk, L. E. B. (2001). Explaining stasis: microevolutionary studies in natural populations. *Genetica*, 112(1), 199-222.
- Merilä, J., & Hendry, A. P. (2014). Climate change, adaptation, and phenotypic plasticity: the problem and the evidence. *Evolutionary Applications*, 7(1), 1-14.
- Milot, E., Mayer, F. M., Nussey, D. H., Boisvert, M., Pelletier, F., & Réale, D. (2011). Evidence for evolution in response to natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences*, 108(41), 17040-17045.
- Milot, E., & Stearns, S. C. (2020). Selection on humans in cities. In *Urban Evolutionary Biology* (pp. 268-288). Oxford University Press.

- Moorad, J. A. (2013). A demographic transition altered the strength of selection for fitness and age-specific survival and fertility in a 19th century American population. *Evolution*, 67(6), 1622-1634.
- Moorad, J. A. (2014). Individual fitness and phenotypic selection in age-structured populations with constant growth rates. *Ecology*, 95(4), 1087-1095.
- Moreau, C., Vézina, H., & Labuda, D. (2007). Founder effects and genetic variability in Quebec. *Médecine Sciences: M/S*, 23(11), 1008-1013.
- Moreau, C., Bhérer, C., Vézina, H., Jomphe, M., Labuda, D., & Excoffier, L. (2011). Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science*, 334(6059), 1148-1150.
- Morrissey, M. B., Kruuk, L. E. B., & Wilson, A. J. (2010). The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology*, 23(11), 2277-2288.
- Morrissey, M. B., Parker, D. J., Korsten, P., Pemberton, J. M., Kruuk, L. E., & Wilson, A. J. (2012). The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution: International Journal of Organic Evolution*, 66(8), 2399-2410.
- Morrissey, M. B., de Villemereuil, P., Doligez, B., & Gimenez, O. (2014). Bayesian approaches to the quantitative genetic analysis of natural populations. In *Quantitative Genetics in the Wild* (pp. 228-253).
- Munoz, M. M., Stimola, M. A., Algar, A. C., Conover, A., Rodriguez, A. J., Landestoy, M. A., ... & Losos, J. B. (2014). Evolutionary stasis and lability in thermal physiology in a group of tropical lizards. *Proceedings of the Royal Society B: Biological Sciences*, 281(1778), 20132433.
- O'Sullivan, R. J., Aykanat, T., Johnston, S. E., Kane, A., Poole, R., Rogan, G., ... & Reed, T. E. (2019). Evolutionary stasis of a heritable morphological trait in a wild fish population despite apparent directional selection. *Ecology and Evolution*, 9(12), 7096-7111.
- Pelletier, F., Pigeon, G., Bergeron, P., Mayer, F. M., Boisvert, M., Réale, D., & Milot, E. (2017). Eco-evolutionary dynamics in a contemporary human population. *Nature Communications*, 8(1), 15947.
- Pigeon, G., Festa-Bianchet, M., Coltman, D. W., & Pelletier, F. (2016). Intense selective hunting leads to artificial evolution in horn size. *Evolutionary Applications*, 9(4), 521-530.

- Postma, E. (2006). Implications of the difference between true and predicted breeding values for the study of natural selection and micro-evolution. *Journal of evolutionary biology*, 19(2), 309-320.
- Price, G. R. (1970). Selection and covariance. *Nature*, 227, 520-521.
- Price, G. R. (1972). Extension of covariance selection mathematics. *Annals of human genetics*, 35(4), 485-490.
- Price, T., Kirkpatrick, M., & Arnold, S. J. (1988). Directional selection and the evolution of breeding date in birds. *Science*, 240(4853), 798-799.
- Pujol, B., Blanchet, S., Charmantier, A., Danchin, E., Facon, B., Marrot, P., ... & Winney, I. (2018). The missing response to selection in the wild. *Trends in ecology & evolution*, 33(5), 337-346.
- Queller, D. C. (1992). Quantitative genetics, inclusive fitness, and group selection. *The American Naturalist*, 139(3), 540-558.
- Queller, D. C. (2017). Fundamental theorems of evolution. *The American Naturalist*, 189(4), 345-353.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rausher, M. D. (1992). The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution*, 46(3), 616-626.
- Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Science*, 8(1), 95-108.
- Robertson, A. (1968). The spectrum of genetic variation. In R. C. Lewontin (Ed.), *Population Biology and Evolution* (pp. 5-7111). Syracuse, NY: Syracuse University Press.
- Roff, D. (Ed.). (1993). *Evolution of life histories: theory and analysis*. Springer Science & Business Media.
- Sanjak, J. S., Sidorenko, J., Robinson, M. R., Thornton, K. R., & Visscher, P. M. (2018). Evidence of directional and stabilizing selection in contemporary humans. *Proceedings of the National Academy of Sciences*, 115(1), 151-156.

- Stearns, S. C. (1989). Trade-offs in life-history evolution. *Functional Ecology*, 3(3), 259-268.
- Stearns, S. C. (1992). *The evolution of life histories*. Oxford University Press.
- Stearns, S. C., Byars, S. G., Govindaraju, D. R., & Ewbank, D. (2010). Measuring selection in contemporary human populations. *Nature Reviews Genetics*, 11(9), 611-622.
- Stinchcombe, J. R., Simonsen, A. K., & Blows, M. W. (2014). Estimating uncertainty in multivariate responses to selection. *Evolution*, 68(4), 1188-1196.
- Stuart, Y. E., Campbell, T. S., Hohenlohe, P. A., Reynolds, R. G., Revell, L. J., & Losos, J. B. (2014). Rapid evolution of a native species following invasion by a congener. *Science*, 346(6208), 463-466.
- Teplitsky, C., Robinson, M. R., & Merilä, J. (2014). Evolutionary potential and constraints in wild populations. In *Quantitative Genetics in the Wild* (pp. 190-208).
- Tropf, F. C., Stulp, G., Barban, N., Visscher, P. M., Yang, J., Snieder, H., & Mills, M. C. (2015). Human fertility, molecular genetics, and natural selection in modern societies. *PLoS One*, 10(6), e0126821.
- Vézina, H., & Bournival, J. S. (2020). An Overview of the BALSAC Population Database. Past Developments, Current State and Future Prospects. *Historical Life Course Studies*.
- De Villemereuil, P., Schielzeth, H., Nakagawa, S., & Morrissey, M. (2016). General methods for evolutionary quantitative genetic inference from generalized mixed models. *Genetics*, 204(3), 1281-1294.
- Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*. Oxford University Press.
- Wilson, A. J., Reale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., ... & Nussey, D. H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, 79(1), 13-26.
- Wilson, A. J., & Poissant, J. (2016). Quantitative genetics in natural populations. *Encyclopedia of evolutionary biology*, 361-371.

## Tables

**Table 2.1.** Average values of married women's life-history traits and standard deviations.

Trait	Charlevoix (N=2226)	Côte-de-Beaupré (N=3181)	Bois-Francs (N=4120)
Age at first birth (y)	23.4 ± 4.6	24.4 ± 5.0	23.9 ± 4.9
Age at last birth (y)	36.1 ± 6.3	35.2 ± 7.3	33.1 ± 7.5
Infant mortality (offspring death between 0-1 y)	0.9 ± 1.4	0.3 ± 0.7	0.2 ± 0.5
Fertility (completed family size)	7.2 ± 3.7	6.3 ± 3.9	5.03 ± 3.6
Reproductive value at birth	1.9 ± 1.0	3.8 ± 3.0	3.1 ± 3.12

**Table 2.2.** Results of the bivariate animal model fitting AFR and fitness as response variables for each dataset.  $\sigma^2_A$ : additive genetic variance;  $\sigma^2_M$ : shared familial environment variance;  $\sigma^2_R$ : residual variance;  $h^2$ : narrow-sense heritability;  $r_g$ : genetic correlation between AFR and relative fitness;  $CV_A$ : coefficient of additive variation; Values are posterior modes for each parameter, along with 95% highest posterior density (HPD) intervals in brackets. Values in italic represent model results using standardized variables.

	$\sigma_A^2$	$\sigma_M^2$	$\sigma_R^2$	$h^2$ *	$CV_A$ *	$r_g^*$
Charlevoix						
<b>AFR</b>	3.44 [1.29 – 6.05] <i>0.15 [0.09 – 0.26]</i>	1.60 [0.29 – 2.77] <i>0.10 [0.06 – 0.15]</i>	16.09 [14.41 – 18.27] <i>0.73 [0.65 – 0.81]</i>	0.17 [0.06 – 0.28] <i>0.16 [0.09 – 0.26]</i>	0.15 [0.06 – 0.26]	-0.70 [-0.83 – -0.35] <i>-0.56 [-0.77 – -0.30]</i>
<b>Relative Fitness</b>	0.10 [0.05 – 0.15]	0.06 [0.04 – 0.09]	0.65 [0.60 – 0.71]	0.13 [0.07 – 0.18]	0.05 [0.03 – 0.08]	-
Bois-Francis						
<b>AFR</b>	0.41 [0.14 – 1.71] <i>0.08 [0.05 – 0.13]</i>	2.74 [1.71 – 3.91] <i>0.12 [0.08 – 0.17]</i>	18.65 [16.98 – 19.57] <i>0.75 [0.70 – 0.81]</i>	0.02 [0.01 – 0.08] <i>0.09 [0.05 – 0.13]</i>	0.03 [0.02 – 0.06]	-0.01 [-0.50 – 0.78] <i>-0.02 [-0.36 – 0.25]</i>
<b>Relative Fitness</b>	0.57 [0.19 – 1.14] <i>0.10 [0.06 – 0.14]</i>	1.24 [0.86 – 1.67] <i>0.13 [0.09 – 0.16]</i>	5.51 [4.99 – 5.90] <i>0.54 [0.48 – 0.57]</i>	0.07 [0.03 – 0.16] <i>0.13 [0.09 – 0.19]</i>	0.24 [0.15 – 0.35]	-
Côte-de-Beaupré						
<b>AFR</b>	1.00 [0.19 – 2.36] <i>0.08 [0.06 – 0.16]</i>	5.43 [4.25 – 7.04] <i>0.21 [0.16 – 0.27]</i>	17.64 [16.15 – 19.06] <i>0.65 [0.60 – 0.72]</i>	0.03 [0.01 – 0.09] <i>0.09 [0.06 – 0.16]</i>	0.04 [0.02 – 0.06]	0.50 [-0.40 – 0.79] <i>0.06 [-0.35 – 0.30]</i>
<b>Relative Fitness</b>	0.29 [0.10 – 0.72] <i>0.08 [0.04 – 0.12]</i>	0.88 [0.59 – 1.37] <i>0.12 [0.08 – 0.16]</i>	5.55 [5.19 – 6.07] <i>0.61 [0.56 – 0.65]</i>	0.04 [0.01 – 0.10] <i>0.09 [0.06 – 0.15]</i>	0.17 [0.10 – 0.23]	-

\* derived parameter

**Table 2.3.** Continued results of the bivariate animal model fitting AFR and relative fitness as response variables for each dataset.  $\sigma_A$ : additive genetic covariance between AFR and relative fitness;  $\beta_G$ : genetic selection gradient;  $\beta_E$ : environmental selection gradient;  $\Delta\beta$ : bias metric in genetic selection coefficient; Values are posterior modes for each parameter, along with 95% highest posterior density (HPD) intervals in brackets. Values in italic represent model results using standardized variables.

$\sigma_A$	$\beta_G^*$	$\beta_E^*$	$\Delta\beta^*$
Charlevoix			
-0.30 [-0.64 – -0.06]	-0.09 [-0.16 – -0.05]	-0.12 [-0.13 – -0.11]	-0.03 [-0.08 – 0.04]
<i>-0.07 [-0.13 – -0.02]</i>	<i>-0.37 [-0.65 – -0.22]</i>	<i>-0.55 [-0.61 – -0.51]</i>	<i>-0.19 [-0.39 – 0.11]</i>
Bois-Francis			
-0.02 [-0.35 – 0.64]	-0.01 [-0.60 – 1.01]	-0.16 [-0.19 – -0.13]	-0.18 [-1.19 – 0.44]
<i>-0.002 [-0.034 – 0.025]</i>	<i>-0.03 [-0.38 – 0.32]</i>	<i>-0.24 [-0.29 – -0.20]</i>	<i>-0.18 [-0.59 – 0.17]</i>
Côte-de-Beaupré			
0.18 [-0.29 – 0.60]	0.13 [-0.25 – 0.75]	-0.21 [-0.25 – -0.19]	-0.42 [-0.90 – 0.06]
<i>0.007 [-0.037 – 0.026]</i>	<i>0.06 [-0.34 – 0.27]</i>	<i>-0.37 [-0.42 – -0.34]</i>	<i>-0.45 [-0.69 – -0.04]</i>

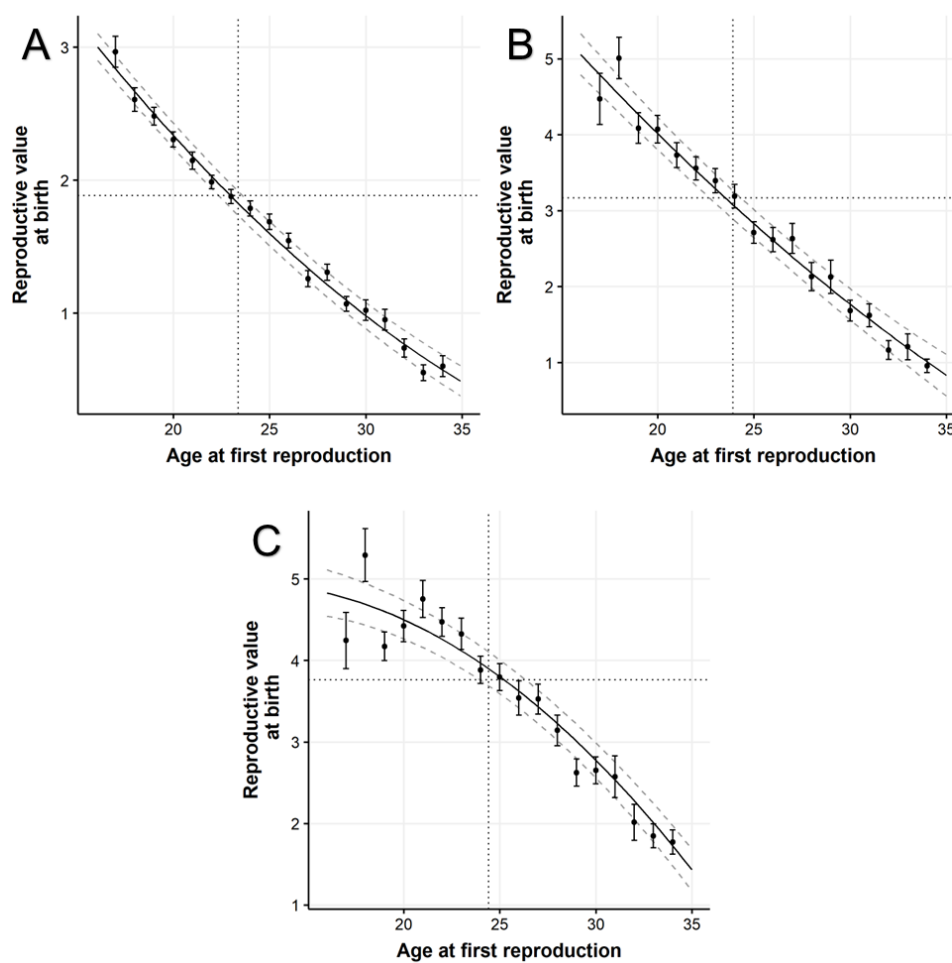
\* derived parameter



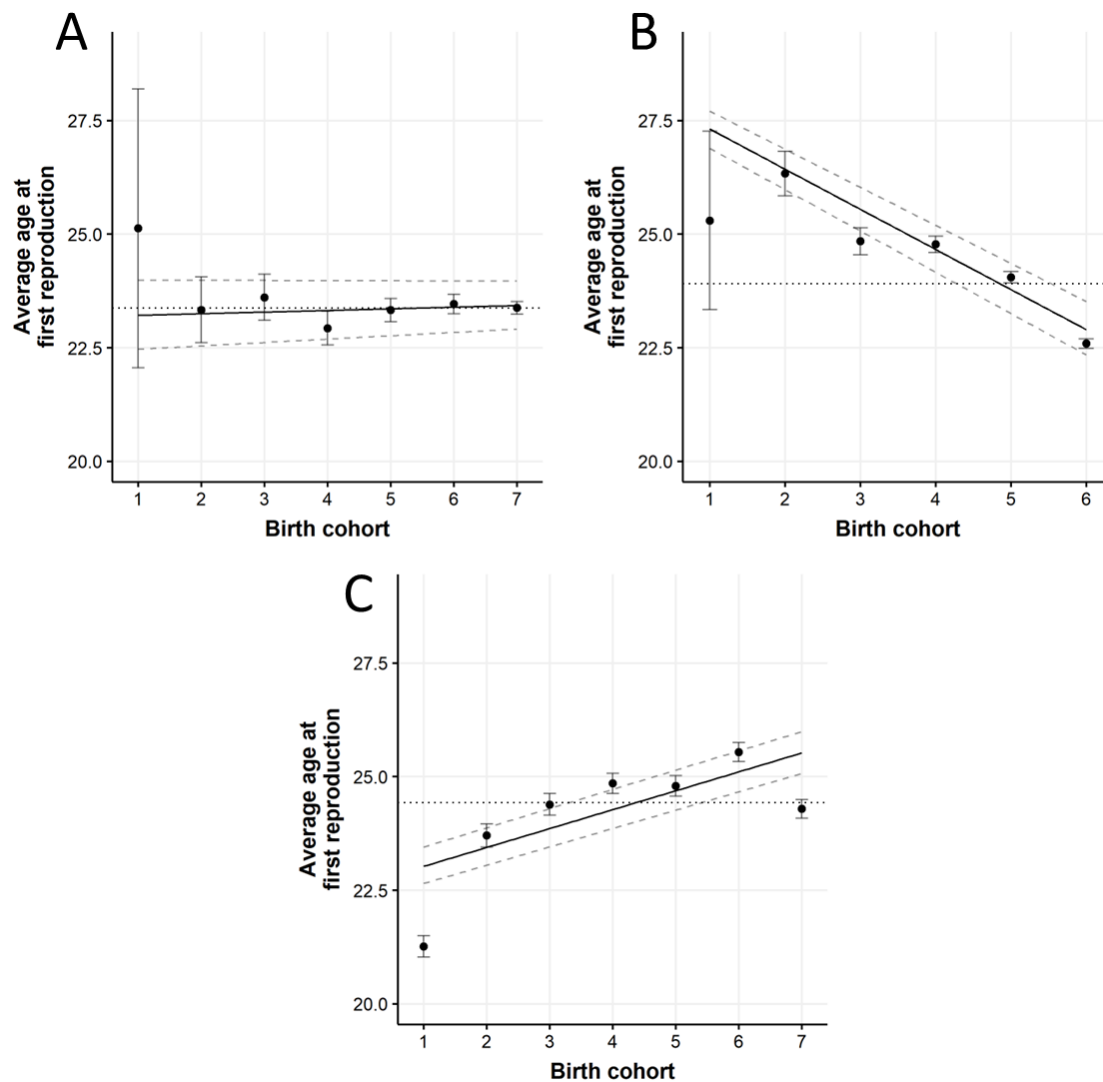
**Table 2.4.** Bayesian posterior indices for additive genetic covariance between AFR and relative fitness  $\sigma_A$ ; genetic selection gradient  $\beta_G$ ; environmental selection gradient  $\beta_E$ ; genetic selection gradient  $\Delta\beta$ . *pd* refers to the probability of direction, expressed here in percentage; *ROPE* refers to the proportion of probability that the 95% HPD interval that lies in the defined *Region of Practical Equivalence* based on two thresholds (10% - 25%; see Supplementary Material for details).

<b>Parameter</b>	<i>pd</i>	<i>ROPE</i> (25%)	<i>ROPE</i> (10%)
	Charlevoix		
$\sigma_A$	100%	0.007	0.000
$\beta_G$	100%	0.000	0.000
$\beta_E$	100%	0.000	0.000
$\Delta\beta$	87.80%	0.084	0.041
Bois-Francis			
-	59.30%	0.936	0.482
-	59.30%	0.101	0.040
-	100%	0.000	0.000
-	84.90%	0.077	0.030
Côte-de-Beaupré			
-	50.80%	0.939	0.515
-	50.80%	0.132	0.048
-	100%	0.000	0.000
-	97.80%	0.007	0.000

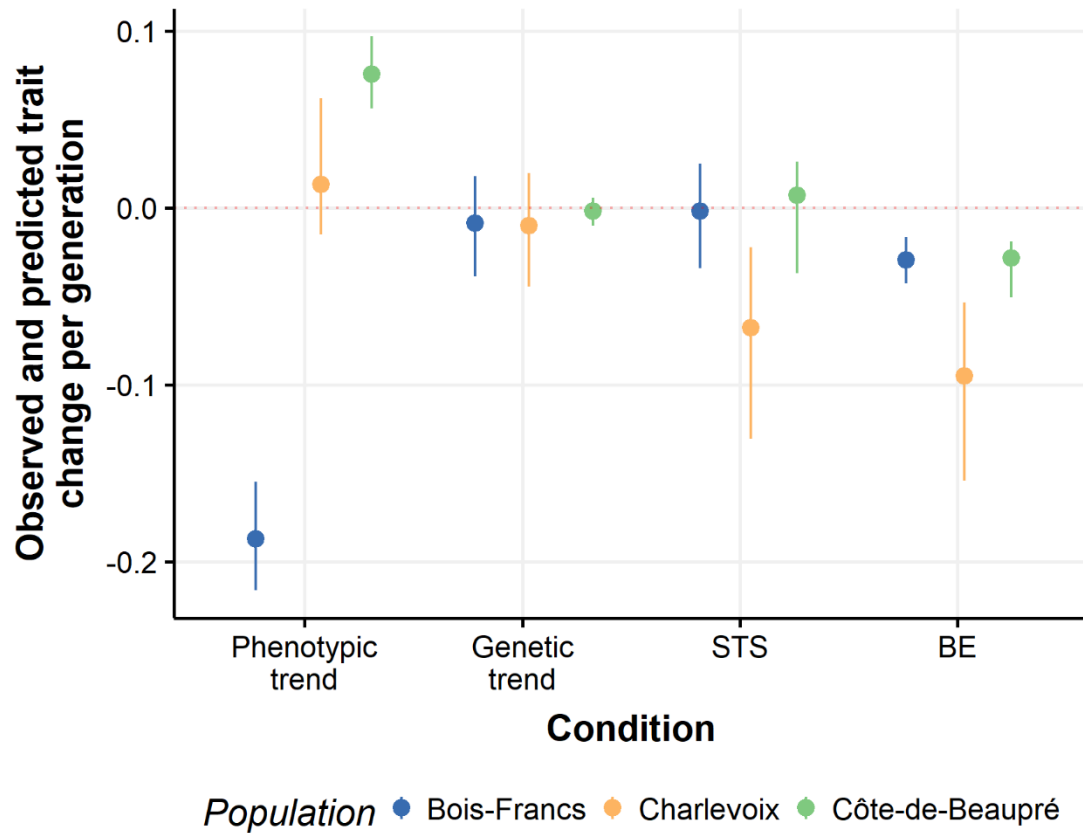
## Figures



**Figure 2.1.** The phenotypic relationship between married women's AFR and relative fitness. **A.** Charlevoix dataset; **B.** Bois-Francs dataset; **C.** Côte-de-Beaupré dataset. The regression line is based on predictions made using a linear mixed-effect model. The dots represent the observed average values. The dashed lines represent the 95% HPD intervals of the regression line. The vertical and horizontal dotted lines represent the population average of AFR and fitness, respectively.



**Figure 2.1.** Temporal trend in average AFR. **A.** Charlevoix dataset; **B.** Bois-Francs dataset; **C.** Côte-de-Beaupré dataset. The regression line is based on predictions made using a linear mixed-effect model. The dots represent the observed average values. The dashed lines represent the 95% HPD intervals of the regression line. The horizontal dotted line represents the population average of AFR.



**Figure 2.3.** The observed and predicted change in AFR per generation. The observed trend is the phenotypic trend based on the data, while the rest are predicted based on different models. The dots represent the posterior mode-estimates, along with the 95% HPD intervals. The dotted red horizontal line represents the zero value.

## CHAPITRE III

### ASSESSING THE IMPACT OF PEDIGREE QUALITY ON THE VALIDITY OF QUANTITATIVE GENETIC PARAMETER ESTIMATES

Le contenu de ce chapitre a fait l'objet d'une soumission en anglais dans la revue *Journal of Evolution Biology* (Wiley)

#### 3.1 Contribution des auteurs

Walid Mawass a conçu et développé l'étude. Walid Mawass a effectué l'analyse des données et les analyses computationnelles avec contribution de Emmanuel Milot. Walid Mawass a rédigé le manuscrit. Emmanuel Milot a contribué à la version finale du manuscrit. Emmanuel Milot a supervisé le projet.

#### 3.2 Résumé en français

L'étude de la dynamique évolutive de traits complexes dans la nature nécessite une évaluation précise de leur architecture génétique. À l'aide d'une approche de modélisation génétique quantitative, les informations sur la parenté provenant d'une généalogie combinée aux mesures phénotypiques peuvent être utilisées pour déduire la quantité de variance génétique additive dans les traits. Cependant, les informations généalogiques provenant de systèmes naturels ne sont pas parfaites et peuvent contenir des erreurs ou être de mauvaise qualité. Les analyses de sensibilité publiées ont révélé un impact limité des taux d'erreur attendus sur les estimations des paramètres. Cependant, les systèmes naturels seront différents à bien des égards, il serait donc inapproprié de généraliser les résultats d'un système à d'autres systèmes naturels. Les généalogies canadiennes-françaises sont étendues et profondément enracinées (jusqu'à 9 générations), ce qui les rend idéales pour étudier comment la qualité des pedigrees affecte les estimations génétiques quantitatives. Nous avons mené des analyses de simulation, dans un cadre bayésien, pour déduire la puissance, la précision et l'exactitude des pedigrees canadiennes-françaises en tant que facteur du taux d'erreur généalogique, de la variation de la qualité des pedigrees et de leurs propriétés.

### 3.3 Assessing the impact of pedigree quality on the validity of quantitative genetic parameter estimates

#### Abstract

Investigating the evolutionary dynamics of complex traits in nature requires the accurate assessment of their genetic basis. Using a quantitative genetic (QG) modeling approach (e.g., animal model), relatedness information from a pedigree combined with phenotypic measurements can be used to infer the amount of additive genetic variance in traits. However, pedigree information from natural systems is not perfect and might contain errors or be of low quality. Published sensitivity analyses revealed a limited impact of expected error rates on parameter estimates. Nonetheless, natural systems differ in many respects (e.g., mating system, data availability, etc.), thus patterns of missing data and pedigree attributes likely vary from one system to another and by consequence their impact on QG estimates. French-Canadian (FC) genealogies are extensive and deep-rooted (up to 9 generations in this study) making them ideal to study how pedigree quality and properties affect QG estimates. We conducted simulation analyses to infer the reliability of QG estimates using these human pedigrees and how it is impacted by genealogical errors and variation in pedigree attributes. Broadly, results show that pedigree size and depth are important determinants of precision but not of accuracy. Two pedigree properties, mean completeness (i.e., the ratio of known ancestors to expected ancestors) and mean genealogical depth (the amount of information contained in the pedigree), seem to be good indicators of accuracy. Including a source of similarity among relatives from a shared environment into the simulations leads to at least a 50% underestimation of the additive genetic variance. This has crucial implications for evolutionary studies aiming to estimate QG parameters given that many traits of interest, such as life history, exhibit important non-genetic sources of variation.

**Keywords:** quantitative genetics, additive (co)variance, sensitivity analysis, animal model, pedigree quality.

## Introduction

Advancing the predictive study of evolutionary dynamics in the wild rests on the proper estimation of genetic parameters underlying phenotypic variation, such as the additive genetic (co)variance. Over the last quarter of a century, the field of quantitative genetics (henceforth QG), drawing from the early work of animal breeders (Lush 1937; Henderson 1975) and pioneer work extending it to natural populations (Shaw 1987), popularized the application of pedigree-based methods to estimate genetic parameters in the wild (e.g., Lynch & Walsh 1998; Kruuk et al. 2000; Charmantier et al. 2014). These methods can exploit all known filial relationships between individuals in the pedigree to dissect phenotypic variation into its genetic and other components (e.g., environmental, maternal; Kruuk 2004; Wilson et al. 2010; Walsh & Lynch 2018). One popular method is to fit to the data the ‘animal model’, a special case of generalized linear mixed-effect models that can deal with unbalanced pedigrees typically associated with natural populations. Various implementations of the animal model are now available (Gilmour et al. 2002; Hadfield 2010; de Villemereuil et al. 2013, 2016; Rousset 2017; see van Benthem 2017 for a review) and have been used to study microevolutionary processes in nature (e.g., Foerster et al. 2007; Wilson et al. 2006, 2007; Fisher et al. 2019).

Pedigrees provide information on the expected pairwise relatedness between individuals, measured as twice the kinship coefficient  $\theta_{ij}$ , namely the probability that a gene copy randomly drawn from individual  $i$  is identical-by-descent to a copy randomly drawn from individual  $j$  for the same gene. This information is used in conjunction with phenotypic measurements to estimate QG parameters, based on the principle that the degree of additive genetic determination of a trait determines the phenotypic resemblance between relatives. How good parameter estimation is strongly depends on the information contained in the available pedigree, which can vary much among natural, hence uncontrolled, populations (Bonnet et al. 2022). In that regard, the validity of a method is its capacity to correctly measure what it should measure. In the context herein, i.e., a quantitative genetic modeling framework, we define validity as the capacity of the animal model to estimate parameters of interest (i.e., genetic) with acceptable accuracy (i.e., the proximity between the estimated value and the true value) and precision, and over the range of their potential values, given the pedigree. Based on

this operational definition, here we explore the impact pedigree quality and attributes have on the accuracy and precision of quantitative genetic estimates.

The sizable variation existing among species' and populations' characteristics results in a diversity of genealogical topologies in nature, some of which are more tractable than others by researchers. Herein *pedigree quality* refers to how close is the data-reconstructed pedigree to the true (but unobserved) pedigree. It is determined by attributes such as sample size, depth, genealogical error rate, and completeness (Pemberton 2008; see Table 1 for definition of some pedigree attributes). A pedigree may be of low quality because it lacks depth relative to the basal population or contains a high number of parental assignment errors, for example. However, among pedigree attributes that affect the precision and accuracy of parameter estimation, not all need be linked to empirical data quality. They can reflect intrinsic properties of the natural system under study. Consider, for instance, two pedigrees of identical size but differing by their degree of connectivity (e.g., average number of genealogical loops linking two individuals) due to population history or mating system. All else being equal, the one with the weaker connectivity should provide less precise parameter estimates (Charmantier & Réale 2005; Morrissey et al. 2007; Wilson et al. 2010; Firth et al. 2015). Precision of QG estimates is impacted by pedigree depth, sample size, and the complexity of the genetic basis underlying the focal trait (Morrissey et al. 2007; de Villemereuil et al. 2013). For example, using data sets from populations of two species with contrasting life histories, the great tit *Parus major* and the long-lived mute swan *Cygnus olor*, Quinn et al. (2006) found appreciable increase in the standard errors of heritability and genetic correlation estimates as the data depth and completeness decreased using restricted maximum likelihood analyses.

One assumption underlying QG is that, unless specified explicitly in the model structure, a pedigree provides errorless values of pairwise relationships between individuals. For it to be the case requires three conditions: i) genealogical links between parents and offspring are correct; ii) no links are missing between individuals who appear in the pedigree, such that all true genealogical paths connecting a pair of phenotyped individuals are known (relative to the basal population); iii) pedigree-based relatedness matches genetic relatedness. Regarding the first two of these conditions,



pedigree errors and incompleteness generally reduce the extent to which relatives appear to resemble each other, causing underestimation of genetic parameters. For example, in captive breeds of Hereford cattle, estimates of heritability in two traits, birth weight and weaning weight, decreased linearly as the rate of sire misidentification increased (Senneke et al. 2004). Now considering condition iii above, let us recall here that a pedigree provides the expected (mean) relatedness value for a given degree of relationship (e.g., cousins), which does not account for the variance introduced by Mendelian segregation and recombination. Consequently, even a perfectly reconstructed empirical pedigree – i.e., identical to the true one – cannot realistically provide exact relatedness values (Hill & Weir 2011), both for captive stocks in animal breeding programs (Visscher et al. 2002) and natural populations (Speed and Balding 2015). This means that condition iii is never fully met. Bérénos et al. (2014) used three pedigrees built from increasingly accurate estimates of relatedness (pedigree constructed from field observations; pedigree constructed from specific markers; pedigree constructed using whole-genome information) for an unmanaged Soay sheep population to show that field observation-based pedigree performed as well as SNP-based and whole-genome pedigree in terms of heritability estimates. By contrast, Perrier et al. (2018) found that using a socially-constructed pedigree of a blue tit natural population led to a minor underestimation of heritability, compared with estimates based on a genome-wide relatedness matrix, which was attributed to incorrect pedigree links due mostly to EPP and low information content.

An incorrect link can contribute to parameter underestimation in two ways that mirror each other: i) the link may wrongly connect an individual A to another individual B, while both are in reality little or unrelated to each other<sup>1</sup>. As a consequence, for a given value of heritability, the pair will exhibit no or limited phenotypic resemblance due to IBD genes, relative to what would be expected from their presumed but wrong genealogical relationship (this problem also extends to pairwise comparisons between A and other individuals who are related to B); ii) the wrong link will cause the analysis to ignore a true link (and resemblance) between the same individual A and others with

---

<sup>1</sup> Here ‘unrelated’ should be interpreted as referring to two individuals who share no common ancestors across the generations covered by the pedigree and with respect to the basal population (as biological evolution implies that ‘absolute unrelatedness’ cannot exist in a population).

whom he/she is more related than apparent in the erroneous pedigree. In certain species, it can be difficult to establish offspring-father links due to extra-pair paternities (EPPs) or because limited paternal care reduces the detectability of fathers. In other species, parentage is practically intractable in the field for both parents, and it may be necessary to perform genetic assignment to parental pairs (e.g., Perrier et al. 2013).

A pedigree is incomplete when the different lineages ascending from one or more focal (phenotyped) individuals are not equally covered by the data. This is the case, for example, when both paternal grand-parents but a single maternal grand-parent are known for an individual. Incompleteness can manifest in two ways: i) genealogical links between pairs of individuals (i.e., parent-offspring) are missing at random, that is, independently of other missing links in the pedigree. In this case, while incomplete, the reconstructed pedigree is nevertheless balanced with respect to the amount of missing ancestral links along maternal vs. paternal sides; ii) links are missing not at random, which can be due to various causes, such as individual heterogeneity in the probability of being included or not in the dataset resulting in an unbalanced pedigree or, in other words, one that exhibits greater lineage asymmetry than expected under the ‘links missing at random’ case (see Figure S1 for an illustration). For example, among-individual genetic differences in dispersal propensity could result in the underrepresentation of more dispersive individuals in the reconstructed pedigree.

Sensitivity and power analysis are two means of assessing the adequacy of a pedigree for QG parameter estimation (as per Morrissey et al. 2007). Sensitivity analysis consists in determining how parameter estimates change when genealogical errors are deliberately introduced in the pedigree. For example, Charmantier and Réale (2005) and Firth et al. (2015) conducted simulations on social pedigrees from blue tits (*Parus caeruleus*) and great tits (*P. major*; data from Firth et al. 2015) and showed that underestimation of heritability in body mass and tarsus length was fairly limited when extra-pair paternity (EPP) rates were up to 20% (note, however, that traits expressed in a single sex could react differently when genealogical errors involve mostly one parental sex, as in the case of EPPs). Pedigree errors and incompleteness generally reduce the extent to which relatives appear to resemble each other, causing genetic parameter underestimation. In captive breeds of Hereford cattle, estimates of heritability in two

traits, birth weight and weaning weight, decreased linearly as the rate of sire misidentification increased (Senneke et al. 2004).

The impact of a given reconstructed pedigree's quality (Table 1) on the validity of QG parameter estimates is rarely tested before using the pedigree for the intended study. Only a minority of empirical studies that fitted animal models to wild population data appear to quantify how good their pedigrees are for QG analyses. In a quick informal survey of studies published over the past 10 years ( $n=30$ ), we found only  $\sim 1/3$  of them performed a formal sensitivity analysis using simulations (e.g., Bourret et al. 2017; de Villemereuil et al. 2019; Bonnet et al. 2022). Among the remaining  $2/3$ , some studies limit their consideration of sensitivity issues to citing other studies, such as those aforementioned. This is of course of relevance, e.g., for result interpretation purposes, but does not amount to a quantitative assessment of the specific pedigree at hand. In addition, sensitivity analyses have mainly focused on genealogical link errors, such as those due to EPPs in social pedigrees, but less on missing links (i.e., pedigree completeness). Though both factors will generally decrease the apparent phenotypic resemblance between relatives, their quantitative impact on estimation can differ. Actually, Morrissey et al. (2007) showed that erroneous genealogical links cause a greater underestimation of QG parameters than missing links. Wolak and Reid (2017) found that parameters estimated with animal models will be biased when parentage links are missing in a systematic way with respect to genetic or phenotypic values of population founders. Moreover, pedigree errors or an incomplete pedigree can lead to a downward bias in the estimation of maternal effects, which are a relevant source of phenotypic similarity between relatives (Wilson et al. 2010), as well as of heritability, as aforementioned. All these issues relate to the problem of parameter identifiability in statistical models with complex variance structures (Bolker 2008). So far, it is still unclear how pedigree attributes impact the identifiability of variance components in animal models (but see Bourret et al. 2017).

In this study we used socially-constructed human pedigrees based on the genealogies of French-Canadian (henceforth FC) historical populations provided by the BALSAC database (<https://balsac.ugac.ca/>), which span all Québec regions since their colonization by European settlers. FC populations offer two advantages to explore the

relationship between pedigree attributes and genetic parameter estimation. The first one is the possibility to separate the huge Québec-wide pedigree into multiple sub-pedigrees with contrasting attributes, corresponding to several geographic subpopulations, and compare them in our analyses. The second is that we have good estimates of genealogical error rates from previous gene-dropping simulations (Doyon 2018.; further details below). Our objectives are: 1) to assess how pedigree attributes impact the validity of QG estimates; 2) evaluate the effect of different model parameterization and variance structures on the validity of QG estimates.

## **Material and Methods**

### **Study population**

For a number of human populations from the contemporary period, historical data on births, deaths, and marriages is available from civil or parish registers. This information has been used to reconstruct genealogies that were subsequently exploited in QG studies (e.g., Blomquist 2019; see Bolund et al. 2016 for a review and Stearns et al. 2010 for a list of medical- and clinical-based multigenerational datasets). The pedigrees used in this study were reconstructed from the history of Catholic families who lived in Québec, Canada, between the foundation of New France by French settlers, in 1608, and 1960. Marriage, baptism and burial acts recorded in parish and civil registers have been compiled and linked since the 1970s in two large databanks: BALSAC (Bouchard et al. 1989; Vézina & Bournival 2020) and the Ancient Québec Population Register (AQPR, Dillon et al. 2018). Altogether they contain all Catholic marriages celebrated during this period (first marriage in 1621), as well as births and deaths of all individuals for some regions and periods. For this study, we used data from BALSAC, which also contains part of AQPR data, and now comprises 4 million digitized records of events representing the lives of 5 million individuals. Those from the 1621-1900 period are duplicate copies of church registers, while those dating from 1900 onwards are from civil registers.

The Catholic colonization of Québec took place by multiple waves, in different regions and time periods. BALSAC data include the locality of records, i.e., parishes or municipalities, which are grouped into 23 regions that roughly correspond to the administrative regions of modern-day Québec (Figure S2) Some gene flow occurred

among regions, through sequential settlement and demographic expansion during the first few generations, and then migration (e.g., following marriages between spouses born in different regions).

### **Pedigree data**

We separated individual records by birth locations to reconstruct the regional pedigrees, among which we selected six for the present study: Bas-St-Laurent, Bois-Francs, Charlevoix, Côte-de-Beaupré, Gaspésie, and Laurentides. We used a number of criteria to choose them (detailed in SI). In brief, since pedigree size and depth are determining factors for quantitative genetic analyses (Quinn et al. 2006; de Villemereuil et al. 2013), we removed 10 regional datasets containing < 1,000 individuals and covering < 4 generations. Among the 13 pedigrees left, we then had to discard those with very large sample sizes, due to prohibitive computational run time required to reach convergence of the models fitted. For example, with > 400,000 individuals, the Saguenay-Lac-St-Jean pedigree took several weeks to run a single MCMC iteration of the hundreds of thousands necessary to fit a single model, disregarding any convergence issues due to the MCMC chain length. We thus ended up with six pedigrees for which QG analyses could be realistically done. Altogether, these regions total up ~250,000 individuals, i.e., 5% of all individuals in BALSAC. Table 2 provides attributes for each pedigree.

### **Missing parental links**

Missing parentages and the occurrence of some genealogical errors are hardly avoidable when pedigrees for entire human populations are reconstructed from historical civil and church registers (typically by linking acts such as baptisms, marriages, and burials). Incorrect genealogical links made between individuals (i.e., wrong biological parental assignments) can result from hidden child adoptions by social (non-biological) parents or false parentage declarations, for example. Missing parental assignments can be due to information that is lacking or unclear on historical acts. For more information on record linkage and data quality, we refer the reader to BALSAC documents available at <https://balsac.uqac.ca/bibliographie-selective/>, in addition to published works such as by Vézina & Bournival (2020) and Vézina et al. (2018).

### **Genealogical error rates**

The genealogical error rate in the BALSAC register had been previously estimated based on molecular data (mitochondrial DNA and Y chromosome haplotypes) from living individuals connected to the Québec genealogy. Specifically, Jomphe (2011) estimated the error rate to be respectively 0.55% and 0.82% for maternal and paternal links (using sequenced hypervariable regions, three genotyped positions of the mitochondrial DNA coding region and Y-chromosome markers). Doyon (2018) later recovered a very similar error rate from a larger sample of typed individuals. She found that 0.80% (s.e.=0.07%) of father-offspring links were erroneous, which includes 0.41% due to extra-pair paternities (EPP). As for mother-offspring links, the estimated error rate was 0.383% (s.e.=0.009%).

Since a pedigree necessarily contains equal numbers of mother- and father-offspring links, the global error rate is thus expected to be ~0.6%, hence one wrong link for every ~166. Note that a mismatch between parent's and offspring's DNA need not reflect a linkage error as it could be explained by a germline mutational event. Therefore, the per-marker mutation rate had also been accounted for in Doyon's (2018) estimations. Accordingly, we applied them in our analyses to check the sensitivity of QG results to genealogical error (i.e., 0.80% for patrilineal lines and 0.383% for matrilineal lines). We also reran the analyses by setting this time error rates to 1% for both parental lines, to assess how a slight increase impacts QG results (considering that 1% is nearly twice the 0.6% global rate reported by Doyon (2018)).

### **Pedigree simulations**

Starting from a given Québec regional pedigree and *a priori* information on genealogical error rates, we simulated a plausible complete pedigree (as described in the section *Simulation analyses* below). This pedigree then is completely based on the observed pedigree and the only difference between them is that the former is considered to have no missing links, to contain no errors (for the sake of the simulations) and is based on the breeding structure of the population under consideration (Morrissey & Wilson 2010). Then, phenotypic and breeding values were randomly generated along the simulated completed pedigree, based on pre-specified variance parameters for a Gaussian trait. The phenotypic values were then used as input data in animal model analyses. The simulated complete pedigree is used only when we generate individual

phenotypic values. While during the quantitative genetic analysis, we used the observed erroneous and incomplete pedigree as input data to the animal model. We compared genetic parameters estimated from the observed pedigree to the preset parameters used to generate breeding values using the simulated complete pedigree.

The specific analysis framework is the ‘reverse’ approach, as implemented in the R package *pedantics* (Morrissey & Wilson 2010). We simulated plausible ‘true’ pedigrees for each of the six regions using the *rpederr* function, which “probabilistically assigns “true” parents given an incomplete and potentially erroneous pedigree” (Morrissey 2018). To use this function, we re-arranged the pedigree dataset as follows. First, we filtered out maternal and paternal lineages with less than 10 individuals. Such small lineages result either from a true lineage end (i.e., no more descendants) or from the fact that subsequent descendants were born outside the focal region (i.e., emigration). Without this filtering of small lineages, pedigrees would be quite imbalanced in some cases. In the case of both maternal and paternal lineages, the average proportion of lineages < 10 individuals make up only 11% of total lineage sizes across the studied pedigrees. Each known social parent-offspring dyad remaining in the filtered dataset was assigned a probability of error corresponding to the genetically-estimated genealogical error rate (i.e., 0.0080 and 0.00383 for father- and mother-offspring links, respectively). If the identity of the mother or father was missing, the probability of erroneous parental assignment was set to 1. Importantly, this does not mean that the parental assignment is assumed to be erroneous with 100% certainty. Rather, setting this value to 1 indicates to the function that the parental assignment is missing or null (in the case of a founder) for these individuals, thus allowing their inclusion in the process of simulating plausible ‘true’ pedigrees. Founders or immigrants arriving later (i.e., post-foundation) are individuals with no record of birth year and parent IDs, hence were also assigned an erroneous parental assignment probability of 1 (we confirmed with BALSAC staff the founder/immigrant status of these individuals). Finally, the sex, birth year, and cohort of first and last offspring known for each individual were included as required input variables to identify the pool of potential parents for individuals with either missing or (simulated) erroneous parentage. The rearranged pedigree data was fed to the *rpederr* function, which provides as output a

plausible ‘true’ pedigree. The latter then served as input in the next step, namely the simulation of individual phenotypic and breeding values.

Since our aim was to assess how genealogical attributes affect the reliability of QG estimates, we calculated the following pedigree metrics for the six datasets: mean lineage completeness (Cazes & Cazes 1996), average inbreeding coefficient (Thompson 1986), total number of founders, and mean entropy (based on Kouladjian 1986, also see Cazes & Cazes 1996 and Milot et al. 2011). Genealogical entropy provides a quantitative assessment of the amount of information contained within a given genealogy. This measure calculates the expected value of the generation of all founders as the average of the genealogical depth. By considering the average number of generations, compared with other measures of genealogical depth, it makes comparison between genealogies of different populations (or subpopulations) possible. These metrics were obtained from the *GENLIB* R package (Gauvin et al. 2015; see SI for how metrics are calculated).

### **Animal model fitting**

The animal model is a mixed-effect linear regression which decomposes phenotypic trait variation into its genetic and non-genetic components (Kruuk 2004). In its simplest form, the phenotype  $z$  of an individual  $i$  can be written as:  $z_i = \mu + a_i + e_i$ , where  $\mu$  is the population mean,  $a_i$  is the additive genetic value for individual  $i$ , and  $e_i$  is the residual error (Wilson et al. 2010). Both the additive genetic effect and residual error are assumed to be normally distributed. The model also accommodates other sources of variation to be fitted as random effects, such as maternal or shared familial environment effects.

The pedigree’s individuals who are informative about the genetics of quantitative traits are either those with measured phenotypes and genetically related to one or more other phenotyped individuals, or ancestors along genealogical loops linking pairs of phenotyped individuals. Since computational algorithms to fit QG models are more efficient when uninformative individuals are removed from pedigrees, we pruned them to keep only informative individuals, using the `prunePed` function in the *MCMCglmm* R package (Hadfield 2010). With the appropriate genealogical information available, individual life history traits can be reconstructed. Even though it is possible that some



individuals could be phenotyped for other types of traits (e.g., morphological), it is highly likely that the pedigrees studied here are highly suitable to investigate the evolution and genetic basis of life history traits. For this reason, the phenotypic dataset supplied to the animal model was pruned to retain only individuals whose birth year was known and with known reproductive life. This is because only those individuals can be phenotyped by having their life history reconstructed. We used the Bayesian implementation of the animal model in the *MCMCglmm* R package employing Markov chain Monte Carlo methods (Hadfield 2010) in the R environment (v3.6.3; R Development Core Team 2019).

### Simulation analyses

We used the *phensim* function from *pedantics* to simulate phenotypic values for each individual in the simulated ‘true’ pedigree. These values were calculated as the sum of additive genetic effects (i.e., breeding values) and environmental effects simulated from user-specified (co)variance matrices. We ran scenarios with one and two response traits, in the latter case to assess the impact of pedigree quality on genetic covariance estimation. Additionally, in one scenario, we simulated maternal genetic effects to replicate phenotypic resemblance between siblings due to sharing the same influence of mother's genetic makeup on their traits, and in another scenario, we simulated maternal environmental effects to replicate phenotypic resemblance due to the sharing by siblings of the familial environment. In total, we ran four scenarios, as follows:

- **Scenario 1:** Univariate response with simple trait structure:  $z_i = \mu + a_i + e_i$ , where  $\mu$  is the population mean (with  $\mu = 0$  in our simulations),  $a_i$  is the additive genetic value for individual  $i$ , and  $e_i$  is the residual error. The phenotypic variance is  $V_P = V_A + V_E = 1$ , where  $V_A$  is the additive genetic variance and  $V_E$  is the residual (environmental) variance.  $V_A$  was set to 0.1 or 0.3, with  $V_E = 1 - V_A$ .
- **Scenario 2:** Univariate response as in Scenario 1 but adding the term  $mg_i$  for maternal genetic effects:  $z_i = \mu + a_i + mg_i + e_i$ . Here,  $V_P = V_A + V_{MG} + V_E = 1$ .  $V_A$  was set to 0.1 or 0.3,  $V_{MG}$  was set to 0.1, with  $V_E = 1 - (V_A + V_{MG})$ .
- **Scenario 3:** Univariate response as in Scenario 1 but adding the term  $m_i$  for shared familial (maternal) environment effects:  $z_i = \mu + a_i + m_i + e_i$ . Here,

$V_P = V_A + V_M + V_E = 1$ .  $V_A$  was set to 0.1 or 0.3,  $V_M$  was set to 0.1, with  $V_E = 1 - (V_A + V_M)$ .

- **Scenario 4:** Bivariate response with traits  $z_1$  and  $z_2$ :  $z_i = \mu + a_i + e_i$ . Again,  $V_P = V_A + V_E = 1$ .  $V_A$  was set to 0.1 or 0.3 for trait  $z_1$  and fixed to 0.1 for  $z_2$ . The genetic correlation  $r_g$  between  $z_1$  and  $z_2$  was set to 0.1 or 0.5, to simulate a weak and strong correlation, respectively.

Under the first two scenarios, we ran five simulations for each regional pedigree  $\times V_A$  value, for a total of 10 simulations. Scenario 2 was also repeated using an improper model parameterization, fitting an environmental maternal effect instead of a genetic maternal effect. Our aim here was to reproduce situations where researchers would enter maternal identity as a basic random effect in the model (i.e., with no explicit genetic component), whereas real maternal effects in the studied population would have a genetic basis. In the case of Scenario 4, time and computational resource constraints prevented us from fitting the bivariate animal model to more than one simulation per pedigree and parameter value, as it took several months to complete a single run with our largest pedigree.

We used uninformative priors for (co)variance terms in Bayesian animal model fitting. Following Hadfield (2010), we selected an inverse Gamma distribution with parameters [ $V = 1$ ,  $nu = 0.002$ ] as prior for variance terms in Scenarios 1, 2 and 3. In the case of Scenario 4, the prior was a  $2 \times 2$  matrix with diagonal terms (variances) equal to 1 and the off-diagonal terms (covariances) equal to 0. All models were run for 4,000,000 iterations after an initial burn-in of 500,000 iterations. We used a thinning interval of 4,000 iterations to obtain a total of 1,000 samples sampled from the Markov chain. In terms of diagnostics, first we report on the effective sample size, a metric that quantifies the amount of independent information by taking into account the correlation in the MCMC sample. Second, we checked for model convergence by directly checking the autocorrelation across different iteration intervals in the MCMC chain for every fitted term. R scripts used for conducting simulations and animal model analyses are accessible via a GitHub repository ([https://github.com/mawassw/pedigree\\_analysis](https://github.com/mawassw/pedigree_analysis)).

### **Reliability measures**

Precision and accuracy for  $V_A$  and  $V_M$  estimates were first checked visually. The accuracy of the estimates was assessed visually by plotting the posterior modes of the posterior distributions as point estimates and comparing its proximity to the true value. The precision was assessed from the width of the 95% highest posterior density (HPD) intervals of posterior distributions. HPD intervals provide a measure of the uncertainty about the true parameter value (and should not be confused with confidence intervals and never interpreted like them). The larger the HPD intervals and the wider the range of plausible parameter values, hence the lower the precision.

We also assessed precision and accuracy of parameter estimates with a quantitative metric. Specifically, we calculated the mean squared error as:  $MSE = E[(\hat{d} - d)^2]$ , where  $d$  is the true parameter value – i.e., the value specified to simulate phenotypes – and  $\hat{d}$  is the value estimated from the animal model fitted to simulated phenotypic data. The root mean squared error (RMSE) was then used to standardize the scale of estimates from the different models in order to compare them (de Villemereuil et al. 2013). We obtained the posterior distribution of the RMSE, for a given simulation and parameter, by calculating it for every MCMC iteration in the sample of 1,000 retained.

### **Interpretation of posterior distributions**

We applied two Bayesian indices to posterior distributions to draw further inferences about the existence and biological significance of a given genetic or environmental effect on a trait (Makowski et al. 2019b). The Region of Practical Equivalence (or ROPE)-based index, specifically the *ROPE* (95%), corresponds to the proportion of the 95% HPD that lies within the ROPE (loosely speaking, this a Bayesian version of the frequentist’s concept of “power”). This type of index involves redefining the alternative hypothesis from the classic point-null (i.e., the value 0) to a range of values considered negligible or too small to be considered relevant to the biological question being studied (Kruschke 2014; see SI for information on how the *ROPE* range was defined in our study). We also used the *Probability of Direction* ( $pd$ ), which has been interpreted as a Bayesian equivalent of the frequentist  $p$ -value in that it reflects the existence of an effect. This index varies between 50% and 100%, with a high value suggesting that an effect exists while a low value indicates uncertainty around its

existence. To calculate these indices in our study, we used the R package *bayestestR* (Makowski et al. 2019a).

## Results

### Pedigrees

There was a large variation between the regional pedigrees attributes (Table 2). We illustrate this by contrasting the pedigrees for the Gaspésie and Bois-Francs regions (Figures S3 and S4). On the one hand, the size and depth of the Gaspésie pedigree are, respectively, ~6 times and twice smaller than for the Bois-Franc pedigree. On the other hand, mean completeness and entropy are greater for Gaspésie (Table 2). Therefore, Gaspésie's pedigree is smaller but denser and better balanced than Bois-Francs' one. This reflects the more recent settlement in Gaspésie, its smaller population (nonetheless better covered by the data), and likely other demographic aspects. For example, it is interesting that the ratio of maternal over paternal grand-parents is 1.02 for Bois-Francs and 1.44 for Gaspésie (Table 2), an observation that might testify of more frequent men re-marriages in the latter region.

### Precision and accuracy

A visual inspection of posterior distributions of variance parameters clearly indicates that bigger and deeper pedigrees provide a higher precision in  $V_A$  estimates (i.e., narrower 95% HPD intervals; Figure 2). On the other hand, accuracy, represented by the proximity of the point estimate to the true value (dashed line) seemingly depends on trait structure and error rate. In terms of the impact of error rates on accuracy, increasing the error rates in our simulation to 1% led to a slight underestimation in  $V_A$  (Figure 3). This result suggests that there is a negative relationship between pedigree error rate and the accuracy of  $V_A$  independent of the trait structure.

The less biased is a parameter estimator and the greater the accuracy, i.e., the proximity between the expectation for point estimates and the true parameter value (dashed line in Figures 2 and 3). Our simulations show that increasing the genealogical error rate (from 0.6 to 1%) led to a slight underestimation in  $V_A$  (Figure 3), a result that is coherent with statements made by other authors (Charmantier & Réale 2005, Firth et al.

2015). Accuracy of QG estimates also varies with the genetic architecture of the trait. It dropped, especially with smaller and shallower pedigrees, for when the simulated trait had both additive and maternal genetic components (Scenario 2; Figure 4). Posterior modes and 95% HPD of RMSEs, across pedigrees and scenarios, are congruent with above observations about accuracy and precision (Table S1). Overall, pedigrees of larger size and depth showed little reduction in accuracy. Additionally, estimates of  $V_{MG}$  were also more accurate and precise with increasing sample size and pedigree depth.

When a shared familial environment component was added to the trait architecture (Scenario 3), precision and accuracy decreased (Figure 5; Table S1). On average,  $V_A$  was underestimated by ~50% and  $V_M$  was extremely underestimated. Even larger and deeper pedigrees produced inaccurate  $V_A$  and  $V_M$  estimates, to the extent that in some cases, the HPD intervals of the posterior distribution of  $V_A$ , averaged over five simulations, did not even include the expected value. RMSEs show that larger and deeper pedigrees were no longer performing better than the smaller ones when  $V_M$  was fitted. We redid the analysis with less restrictive phenotypic datasets, keeping individuals without a known year of birth, to increase the mean maternal sibship size to ~6 across pedigrees. Results from this analysis show the same estimation bias in  $V_A$  and  $V_M$  estimates (Figure S6).

Effective sample size of the MCMC chains increased with increasing pedigree sample size and depth. We found an interesting pattern whereby models of scenarios with a  $V_A = 0.1$  tended to have a smaller effective sample size, on average, compared to those when  $V_A = 0.3$ . We also performed diagnostics on our models by checking for autocorrelation in MCMC chains. Autocorrelation within the chain of  $V_A$  and between it and the chains of other modeled terms (e.g.,  $V_M$ ) within and between the chains was high ( $> 0.1$ ) only during the burn-in phase and decreased as the chain progressed. Given that we ran our models for a burn-in period of at least 500,000 iterations, the impact of autocorrelation should likely be minimal. We confirmed that after 50,000 iterations, autocorrelation was minimized and consistently below 0.05.

For Scenario 3, which involved a single simulation per pedigree and per parameter, a visual assessment was done by directly plotting the posterior distribution using the R package *bayesplot* (Gabry et al. 2019). Figures 6 and 7, in addition to RMSE measures, show a pattern consistent with the results for Scenario 1, with sample size and

pedigree depth being the most important determinants of precision, and, to a lesser degree, accuracy. Overall, the models performed better when the true value of the estimated parameters, especially in the case of  $Cov_A$ , were not close to zero. When that is the case, the model seems to find difficulty in converging based on MCMC diagnostics.

## Discussion

Assessing a population's evolutionary potential or change requires reliable estimates of genetic parameters. The sensitivity of parameter estimation to pedigree attributes has been explored mostly in (passerine) birds and (ungulate) mammals. This applies to the present study on a mammalian species. Therefore, extrapolation of system-specific conclusions should be done with caution, as other taxa may be characterized by on average different pedigree topologies or tractability (Pemberton 2008). It should not save one from performing sensitivity analysis for the empirical under investigation, when needed. Rather, aforementioned studies can help interpret QG results, design sampling schemes, or foresee the limits the empirical pedigree at hand may have to test genetic structure models of varying complexity. Additionally, we provide results on the extent to which variability in certain pedigree attributes, that are indicative of its quality, impact the reliability of QG estimates. Furthermore, we propose that certain pedigree attributes, rarely assessed (e.g., completeness, entropy, etc.) during sensitivity analyses, are likely useful metrics to determine pedigree performance for specific QG analysis designs. Overall, our simulation analyses show that all pedigrees tested herein convey enough information to detect and quantify small  $V_A$  values with modest uncertainty. Results for  $V_{MG}$ ,  $V_M$  and genetic covariances vary more. Below we discuss how pedigree attributes and errors affect the reliability of QG estimation in our population system, as well as the impact of adding an additional source of similarity between relatives.

### Validity and pedigree error rate

Using a genetically-based estimate of the genealogical error rate for the FC datasets (Doyon 2018), our study suggests that the reliability of QG estimates is little impacted by such errors. In fact, for a phenotype with simple structure (Scenario 1), we did not detect a downward bias in the additive genetic variance. However, increasing the

error rate to 1%, the upper bound of the expected genealogical error rate in the FC datasets, leads to a slight underestimation and reduction in accuracy across all pedigrees. Only the precision of the estimates is not affected by this increase, seemingly more impacted by the size of the pedigree instead. In a previous sensitivity study, Charmantier & Réale (2005) found that a paternity misassignment rate of 20% is the cutoff where the error rate begins to affect QG estimates. Generally, EPP rates in human populations are considered to be low (range: 0.4% – 11.8%; Andreson 2006; also see Larmuseau et al. 2016) with a median of 1.7%. Our results show only a slight increase in the genealogical error rate (i.e., from ~0.6% [average of maternal and paternal error rates] to 1%) is enough to negatively impact QG estimates. We observed this underestimation uniformly across all analyzed pedigrees which rules out the possibility that some of our pedigrees are more sensitive (in terms of QG analysis) to slight increases in the error rate. This result raises the issue of generalizing the results of the impact of error rates across different systems. Furthermore, it is more pertinent to investigate if certain particular attributes of human pedigrees explain this observed discrepancy when we increase the error rate and avoid generalizing results from studies using pedigrees from wild passerine and ungulate populations to social human pedigrees – and other newly-reconstructed pedigrees intended for QG analysis.

Currently, *pedantics* only allows for simulating traits that follow a Gaussian distribution which is a limitation of this analysis given that a number of traits are non-Gaussian (e.g., binomial, zero-inflated, etc.). There does exist a way to simulate non-Gaussian traits, however when we attempted to run our models with these traits as response variables, the time taken for a single simulation to conclude (without sacrificing the time the MCMC chain is ran for which maximizes its performance) using the smallest pedigree in our datasets was around several weeks. First, the execution time for building the relatedness matrix increases on  $O(N^2)$  (with N being the pedigree sample size). Additionally, the sampling algorithms used in the *MCMCglmm* package are sensitive to both dataset size and the distribution of the response variable. When the trait is non-Gaussian, estimating the posterior distribution of its underlying parameters is computationally intense and time-consuming. Given these reasons, it is practically impossible to run a simple model with a non-Gaussian trait with the rest of our datasets.

In *pedantics* phenotypic simulations, a trait value can be generated for each individual in the pedigree. However, in empirical studies, phenotypic data is often unavailable for a proportion of these individuals. In our analyses, to reflect this realistic expectation and the fact that these pedigrees will mostly likely be used to study the evolution and genetic basis of life history traits, each of the six phenotypic datasets were pruned to include only individuals with both a known year of birth and at least one reproductive event. Without this information it would be practically impossible to reconstruct individual life history, a common phenotypic trait of interest in natural populations.

Additionally, compared to a compilation of QG studies on wild populations by Postma (2014), the sample size (total number of records) of our pruned phenotypic datasets was considerably higher than the median number of records with life history trait data in these studies – which was 377 – and, in most cases, higher than the upper bound of the range (range 6–4992 records; from 39 studies on 19 species). Sample size and depth are two important factors that determine pedigree quality (Wilson et al. 2010). Accordingly, our results demonstrate that the larger and deeper the pedigree was the more precise the QG estimates were. In fact, both sample size and pedigree depth exhibited the strongest correlation with a pedigree's RMSE score (Figure S8), however it is difficult to separate their independent effect since they are strongly correlated with each other across datasets ( $r = 0.73$ ). Additionally, even with the pedigrees tested here being quite large and deep, they performed poorly in terms of accuracy under certain scenarios (for example, when additional source of similarity between relatives was included). In effect, it seems that accuracy is not influenced by pedigree sample size and depth.

We also found that the mean rate of missing links and mean expected genealogical depth exhibited a modest correlation with RMSE score. It is not completely clear in what manner these two pedigree attributes impact the precision and accuracy of QG estimates. Nonetheless, they should not be ignored in pedigree-based studies testing the validity of their QG estimates based on their respective pedigree. Depending on the availability of data, two reconstructed pedigrees can have the same degree of completeness, but the pedigree is unbalanced in terms of genealogical links. Therefore, it



would be interesting to investigate the impact of unbalanced completeness on the performance of pedigrees in specific QG model designs (see below for an example).

### **Impact of additional source of resemblance**

We observe biases in parameter estimates when an additional source of phenotypic similarity between relatives is included in the model (Scenario 2 and 3). In the case of a maternal genetic source, larger and deeper pedigrees still provided fairly precise and accurate estimates, while smaller and shallower pedigrees performed worse. Given this pattern, when applying this type of analysis to a reconstructed pedigree, it is highly recommended to run sensitivity analysis to ascertain that the pedigree can provide reliable estimates of both the additive genetic and maternal genetic variances. The major issue that could arise is an overestimation of the additive genetic variance if the model was not properly parametrized and the maternal genetic component was not specified in the model (Kruuk & Hadfield 2007; Wilson et al. 2010). In this case, incorrect qualitative interpretations are possible as the genetic basis of the studied trait would be overestimated and its evolutionary potential exaggerated (Morrissey et al. 2007).

On the other hand, in the case of a maternal environment source, there is systematic underestimation of the additive genetic variance across all pedigrees. This is a problem because a majority of quantitative traits are known to exhibit non-genetic sources of resemblance between relatives (Rice 2008; see Laugen et al. 2002, Pettay et al. 2005 for examples involving life history traits). In addition, when genetic parameters are estimated to test evolutionary hypotheses, accuracy is more important than precision because underestimation (or overestimation), even with low uncertainty, will lead to faulty interpretations. The proportion of individuals with less than two offspring was relatively high in pruned pedigrees, which could have negatively impacted the power to estimate shared familial environment effects (Wilson et al. 2010). However, redoing the analysis using less restrictive phenotypic datasets thus increasing the mean number of siblings per mother included in the pedigree did not eliminate the estimation bias, with the only salient effect being an increase in the precision of the estimates but not in their accuracy.

Another explanation is that, given our use of naïve priors in our models, there isn't enough information in the data to allow proper estimation of the posterior

distribution of the maternal environment variance. An obvious solution here would be to use a more informative prior based on available information on the proportion of phenotypic variation explained by this source of similarity between relatives. A speculative explanation that can explain why smaller and shallower pedigrees performed relatively as well as the larger and deeper ones is the ratio of available genealogical information across parental lineages. Gaspésie exhibited an imbalance between the number of maternal and paternal grandparents with a higher number of maternal grandparents reflecting the presence of remarriages. This specific imbalance in favor of the maternal lineage could provide more information in separating maternal genetic and maternal environmental effects which the former requires relatedness information, and the latter requires only the ID of the mother.

### **Validity of joint estimates when using bivariate animal models**

In evolutionary QG studies, reliable estimates of additive genetic covariances should be on par to that of additive genetic variances (Morrissey et al. 2012). Covariance between a focal trait and fitness (Robertson-Price identity) determines the genetic response to selection (Walsh & Lynch 2018), while covariance among multiple traits can set constraints on this response or reflect evolutionary trade-offs of interest (Teplitsky et al. 2014). Our Scenario 3 involved bivariate models mimicking such a situation where a life history trait is either strongly or weakly genetically correlated with fitness. Though it was difficult to perform multiple simulations per pedigree given prohibitive computation time, our results point out clear conclusions about the positive impact of pedigree sample size and depth on the reliability of additive genetic covariance estimates. Namely, bigger, and deeper pedigrees provided more precise estimates, but nonetheless inaccurate in some cases. An important factor affecting these estimates was the magnitude of the simulated covariance between the two traits: a smaller simulated covariance resulted in poorer estimates from subsequent animal model analyses. Though in all cases (but one) the most probable sign of the estimated covariance matched that of the simulated parameter. These results reveal that studies interested in genetic covariances and correlations, either in the context of studying genetic trade-offs or the genetic response to selection, should consider the possibility that, regardless of its attributes, certain pedigrees might be underpowered to detect weak genetic covariances.

## References

- Anderson, K., 2006. How well does paternity confidence match actual paternity? Evidence from worldwide nonpaternity rates. *Current anthropology*, 47(3), pp.513-520.
- Balloux, F., Amos, W. and Coulson, T., 2004. Does heterozygosity estimate inbreeding in real populations?. *Molecular ecology*, 13(10), pp.3021-3031.
- van Benthem, K.J., Bruijning, M., Bonnet, T., Jongejans, E., Postma, E. and Ozgul, A., 2017. Disentangling evolutionary, plastic and demographic processes underlying trait dynamics: a review of four frameworks. *Methods in Ecology and Evolution*, 8(1), pp.75-85.
- Béréanos, C., Ellis, P.A., Pilkington, J.G. and Pemberton, J.M., 2014. Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular ecology*, 23(14), pp.3434-3451.
- Blomquist, G.E., 2019. Unpacking the heritability of body mass index and other ratios. *American Journal of Human Biology*, 31(5), p.e23289.
- Bolker, B.M., 2008. *Ecological models and data in R*. Princeton University Press.
- Bolund, E., Hayward, A. and Lummaa, V., 2016. Life history evolution, human. *Encyclopedia of evolutionary biology*, pp.328-334.
- Bonnet, T., Morrissey, M.B., de Villemereuil, P., Alberts, S.C., Arcese, P., Bailey, L.D., Boutin, S., Brekke, P., Brent, L.J., Camenisch, G. and Charmantier, A., 2022. Genetic variance in fitness indicates rapid contemporary adaptive evolution in wild animals. *Science*, 376(6596), pp.1012-1016.
- Bouchard, G., Roy, R., Casgrain, B. and Hubert, M., 1989. Fichier de population et structures de gestion de base de données: le fichier-réseau BALSAC et le système INGRES/INGRID. *Histoire & Mesure*, pp.39-57.
- Bourret, A., Bélisle, M., Pelletier, F. and Garant, D., 2017. Evolutionary potential of morphological traits across different life-history stages. *Journal of Evolutionary Biology*, 30(3), pp.616-626.
- Bourret, A. and Garant, D., 2017. An assessment of the reliability of quantitative genetics estimates in study systems with high rate of extra-pair reproduction and low recruitment. *Heredity*, 118(3), pp.229-238.

- Cazes, M.H. and Cazes, P., 1996. Comment mesurer la profondeur généalogique d'une ascendance?. *Population (french edition)*, pp.117-140.
- Charmantier, A. and Réale, D., 2005. How do misassigned paternities affect the estimation of heritability in the wild?. *Molecular ecology*, 14(9), pp.2839-2850.
- Charmantier, A., Garant, D. and Kruuk, L.E. eds., 2014. *Quantitative genetics in the wild*. OUP Oxford.
- Dillon, L., Amorevieta-Gentil, M., Caron, M., Lewis, C., Guay-Giroux, A., Desjardins, B. and Gagnon, A., 2018. The Programme de recherche en démographie historique: past, present and future developments in family reconstitution. *The History of the Family*, 23(1), pp.20-53.
- Doyon, A (2018). *Dynamique des marqueurs génétiques liés au sexe dans la population canadienne-française pour l'interprétation des traces d'ADN en génétique forensique*. Mémoire. Trois-Rivières, Université du Québec à Trois-Rivières, 232 p.
- Firth, J.A., Hadfield, J.D., Santure, A.W., Slate, J. and Sheldon, B.C., 2015. The influence of nonrandom extra-pair paternity on heritability estimates derived from wild pedigrees. *Evolution*, 69(5), pp.1336-1344.
- Fisher, D.N., Wilson, A.J., Boutin, S., Dantzer, B., Lane, J.E., Coltman, D.W., Gorrell, J.C. and McAdam, A.G., 2019. Social effects of territorial neighbours on the timing of spring breeding in North American red squirrels. *Journal of evolutionary biology*, 32(6), pp.559-571.
- Foerster, K., Coulson, T., Sheldon, B.C., Pemberton, J.M., Clutton-Brock, T.H. and Kruuk, L.E., 2007. Sexually antagonistic genetic variation for fitness in red deer. *Nature*, 447(7148), pp.1107-1110.
- Gauvin, H., Lefebvre, J.F., Moreau, C., Lavoie, E.M., Labuda, D., Vézina, H. and Roy-Gagnon, M.H., 2015. GENLIB: an R package for the analysis of genealogical data. *BMC bioinformatics*, 16(1), pp.1-10.
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R., 2002. *ASReml user guide release 1.0*.
- Hadfield, J.D., 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of statistical software*, 33(1), pp.1-22.

- Henderson, C.R., 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, pp.423-447.
- Hill, W.G. and Weir, B.S., 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics research*, 93(1), pp.47-64.
- Jomphe, M., 2011. Validation des généalogies reconstituées à BALSAC à partir de données génétiques (Document IC-243). *Projet BALSAC, Chicoutimi*.
- Kouladjian, K., 1986. Une mesure d'entropie généalogique. *Chicoutimi, SOREP, document*, 3, p.C43.
- Kruschke, J., 2014. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.
- Kruuk, L.E., Clutton-Brock, T.H., Slate, J., Pemberton, J.M., Brotherstone, S. and Guinness, F.E., 2000. Heritability of fitness in a wild mammal population. *Proceedings of the National Academy of Sciences*, 97(2), pp.698-703.
- Kruuk, L.E., 2004. Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1446), pp.873-890.
- Kruuk, L.E.B. and Hadfield, J.D., 2007. How to separate genetic and environmental causes of similarity between relatives. *Journal of evolutionary biology*, 20(5), pp.1890-1903.
- Larmuseau, M.H., Matthijs, K. and Wenseleers, T., 2016. Cuckolded fathers rare in human populations. *Trends in ecology & evolution*, 31(5), pp.327-329.
- Laugen, A.T., Laurila, A. and Merilä, J., 2002. Maternal and genetic contributions to geographical variation in *Rana temporaria* larval life-history traits. *Biological Journal of the Linnean Society*, 76(1), pp.61-70.
- Lush, J.L., 1937. Animal breeding plans. Ames.
- Lynch, M. and Walsh, B., 1998. *Genetics and analysis of quantitative traits*.
- Makowski, D., Ben-Shachar, M.S. and Lüdtke, D., 2019. bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, 4(40), p.1541.

- Makowski, D., Ben-Shachar, M.S., Chen, S.H. and Lüdecke, D., 2019. Indices of effect existence and significance in the Bayesian framework. *Frontiers in psychology*, 10, p.2767.
- Milot, E., Mayer, F.M., Nussey, D.H., Boisvert, M., Pelletier, F. and Réale, D., 2011. Evidence for evolution in response to natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences*, 108(41), pp.17040-17045.
- Morrissey, M.B., Wilson, A.J., Pemberton, J.M. and Ferguson, M.M., 2007. A framework for power and sensitivity analyses for quantitative genetic studies of natural populations, and case studies in Soay sheep (*Ovis aries*). *Journal of evolutionary biology*, 20(6), pp.2309-2321.
- Morrissey, M.B. and Wilson, A.J., 2010. Pedantics: an R package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Molecular Ecology Resources*, 10(4), pp.711-719.
- Morrissey, M.B., Parker, D.J., Korsten, P., Pemberton, J.M., Kruuk, L.E. and Wilson, A.J., 2012. The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution: International Journal of Organic Evolution*, 66(8), pp.2399-2410.
- Morrissey, M.B., 2018. pedantics: Functions to Facilitate Power and Sensitivity Analyses for Genetic Studies of Natural Populations. R package version 1.7, <https://CRAN.R-project.org/package=pedantics>
- Pemberton, J.M., 2008. Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635), pp.613-621.
- Perrier, C., Baglinière, J. L., & Evanno, G., 2013. Understanding admixture patterns in supplemented populations: a case study combining molecular analyses and temporally explicit simulations in Atlantic salmon. *Evolutionary Applications*, 6(2), 218-230.
- Perrier, C., Delahaie, B., & Charmantier, A., 2018. Heritability estimates from genomewide relatedness matrices in wild populations: Application to a passerine, using a small sample size. *Molecular ecology resources*, 18(4), 838-853.
- Pettay, J.E., Kruuk, L.E., Jokela, J. and Lummaa, V., 2005. Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proceedings of the National Academy of Sciences*, 102(8), pp.2838-2843.

- Postma, E., 2014. Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data, better estimates. *Quantitative genetics in the wild*, 16, p.33.
- Quinn, J.L., Charmantier, A., Garant, D. and Sheldon, B.C., 2006. Data depth, data completeness, and their influence on quantitative genetic estimation in two contrasting bird populations. *Journal of evolutionary biology*, 19(3), pp.994-1002.
- R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rice, T.K., 2008. Familial resemblance and heritability. *Advances in genetics*, 60, pp.35-49.
- Rousset, F., 2017. An introduction to the spaMM package for mixed models.
- Senneke, S.L., MacNeil, M.D. and Van Vleck, L.D., 2004. Effects of sire misidentification on estimates of genetic parameters for birth and weaning weights in Hereford cattle. *Journal of animal science*, 82(8), pp.2307-2312.
- Shaw, R.G., 1987. Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution*, 41(4), pp.812-826.
- Speed, D. and Balding, D.J., 2015. Relatedness in the post-genomic era: is it still useful?. *Nature Reviews Genetics*, 16(1), pp.33-44.
- Stearns, S.C., Byars, S.G., Govindaraju, D.R. and Ewbank, D., 2010. Measuring selection in contemporary human populations. *Nature Reviews Genetics*, 11(9), pp.611-622.
- Teplitsky, C., Tarka, M., Møller, A.P., Nakagawa, S., Balbontin, J., Burke, T.A., Doutrelant, C., Gregoire, A., Hansson, B., Hasselquist, D. and Gustafsson, L., 2014. Assessing multivariate constraints to evolution across ten long-term avian studies. *PLoS One*, 9(3), p.e90444.
- Thompson, E.A., 1986. *Pedigree analysis in human genetics*. Johns Hopkins University Press. p25
- Vézina, H. and Bournival, J.S., 2020. An Overview of the BALSAC Population Database. Past Developments, Current State and Future Prospects. *Historical Life Course Studies*.

- Vézina, H., St-Hilaire, M., Bournival, J.S. and Bellavance, C., 2018. The linkage of microcensus data and vital records: An assessment of results on Quebec historical population data (1852–1911). *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 51(4), pp.230-245.
- Walsh, B., & Lynch, M., 2018. *Evolution and selection of quantitative traits*. Oxford University Press.
- de Villemereuil, P., Gimenez, O. and Doligez, B., 2013. Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. *Methods in Ecology and Evolution*, 4(3), pp.260-275.
- de Villemereuil, P., Schielzeth, H., Nakagawa, S. and Morrissey, M., 2016. General methods for evolutionary quantitative genetic inference from generalized mixed models. *Genetics*, 204(3), pp.1281-1294.
- de Villemereuil, P., Rutschmann, A., Lee, K.D., Ewen, J.G., Brekke, P. and Santure, A.W., 2019. Little adaptive potential in a threatened passerine bird. *Current Biology*, 29(5), pp.889-894.
- Wilson, A.J., Pemberton, J.M., Pilkington, J.G., Coltman, D.W., Mifsud, D.V., Clutton-Brock, T.H. and Kruuk, L.B., 2006. Environmental coupling of selection and heritability limits evolution. *PLoS biology*, 4(7), p.e216.
- Wilson, A.J., Nussey, D.H., Pemberton, J.M., Pilkington, J.G., Morris, A., Pelletier, F., Clutton-Brock, T.H. and Kruuk, L.E., 2007. Evidence for a genetic basis of aging in two wild vertebrate populations. *Current Biology*, 17(24), pp.2136-2142
- Wilson, A.J., Reale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A., Kruuk, L.E. and Nussey, D.H., 2010. An ecologist's guide to the animal model. *Journal of animal ecology*, 79(1), pp.13-26
- Wolak, M.E. and Reid, J.M., 2017. Accounting for genetic differences among unknown parents in microevolutionary studies: how to include genetic groups in quantitative genetic animal models. *Journal of Animal Ecology*, 86(1), pp.7-20.
- Visscher, P.M., Woolliams, J.A., Smith, D. and Williams, J.L., 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *Journal of dairy science*, 85(9), pp.2368-2375.



## Tables

**Table 3.1.** Table showing the key factors which affect pedigree quality and performance.

Attribute of	Attribute name	Description	Notes
True pedigree	Depth	Number of generations separating the current (most recent) cohort and population founders.	From a QG perspective, founders are those who represent the basal population.
	Pedigree structure	Differing mating systems and reproductive strategies lead to contrasting pedigree structures, a large pedigree with a short generation time vs smaller pedigree with a longer generation time	Different pedigree structures perform better for specific estimations. For example, pedigrees from polygynous species are good for estimating shared environmental effects, while pedigrees from long-lived species perform well in studies of genotype-by-environment interactions
Data-reconstructed (or empirical) pedigree	Quality	In a broad sense, quality means how close is the reconstructed pedigree to the true, but unobserved, pedigree.	Quality is determined by attributes such as sample size, completeness, depth, and the occurrence of genealogical errors.
	Sample size	Number of individuals who are known/used to build the pedigree.	Sometimes reported per generation.
	Completeness	Ratio of number of known ancestors to the number of expected ancestors of the sampled individuals.	A complete pedigree contains ancestors for a focal individual, where $g=0$ designates its own generation, $g=1$ that of its parents, and so forth. E.g., if $G=3$ , then the individual has 14 ancestors <sup>a</sup> . A pedigree lacking the identity of 4 of them would have a completeness value of 0.71.
	Depth	Number of generations covered by the reconstructed pedigree (e.g., average number of generations separating individuals at lineage tips from their ancestors at lineage bases)	Balloux <i>et al.</i> (2004) consider that when depth is $< 5$ generations, and depending on mating system and population structure, a large proportion of variance in parameters such as coancestry, relatedness, and inbreeding coefficient, won't be captured.
	Phenotypic coverage	The proportion of individuals in the pedigree for which phenotypic measurements are available.	Individuals with no measurements still contribute information to QG analyses when they lie on a genealogical path linking 2 individuals with measurements. However, genetic parameter accuracy and precision should increase with increasing phenotypic coverage, especially in cases where other sources of variation are present such as shared familial environments which require at least two or more sibs be phenotyped
	Genealogical error	An incorrect link between two related individuals	This type of error occurs due to over-linkage between individuals (i.e., wrong biological parental assignments). It can result from hidden adoptions by social (non-biological) parents or false parentage declarations, for example. Genealogical errors can occur also from under-linkage which refers to missing parental assignments due to information that is lacking

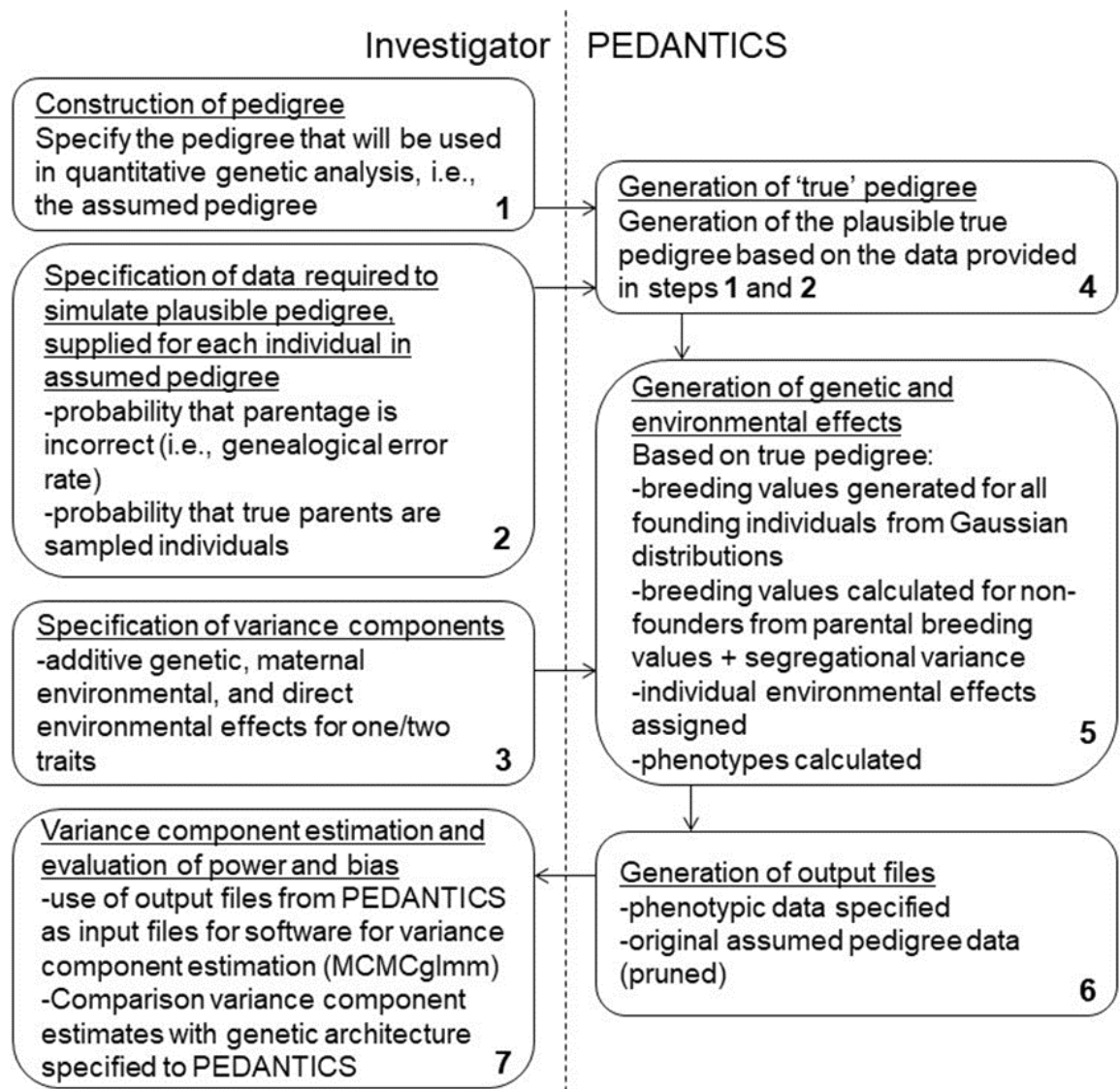
		or unclear on historical acts
Pedigree accuracy	The accuracy of pedigree links which declines with the presence of errors caused by extra-pair paternities and parental misassignment	Link errors are expected to lead to downward-biased estimations of heritability (or additive genetic variance), i.e., lower precision. Inaccurate pedigree links can cause environmental covariance to be misinterpreted as genetic covariance. This is most likely in cases of small sample size and highly heritable traits. However, according to Charmantier & Réale (2005), animal model heritability estimates are robust when EPP rates are lower than 20%
Entropy	This factor also relates to data availability. It is the proportion of missing links or missing parentage assignment to the total pedigree links. It provides a measurement of how sparse the information is on the ancestral generations	Lack of information on parental assignment and pedigree links differ from inaccuracy in links in that it might not lead to downward-biased estimations and imprecise measurements if sample size is large enough, but it is expected to negatively impact the accuracy of estimates

<sup>a</sup> When inbreeding occurs, these 14 ancestors need not be 14 different individuals. For example, one could be “twice” the great-grand-mother (i.e., through 2 genealogical paths) of an offspring born to a mating between cousins.

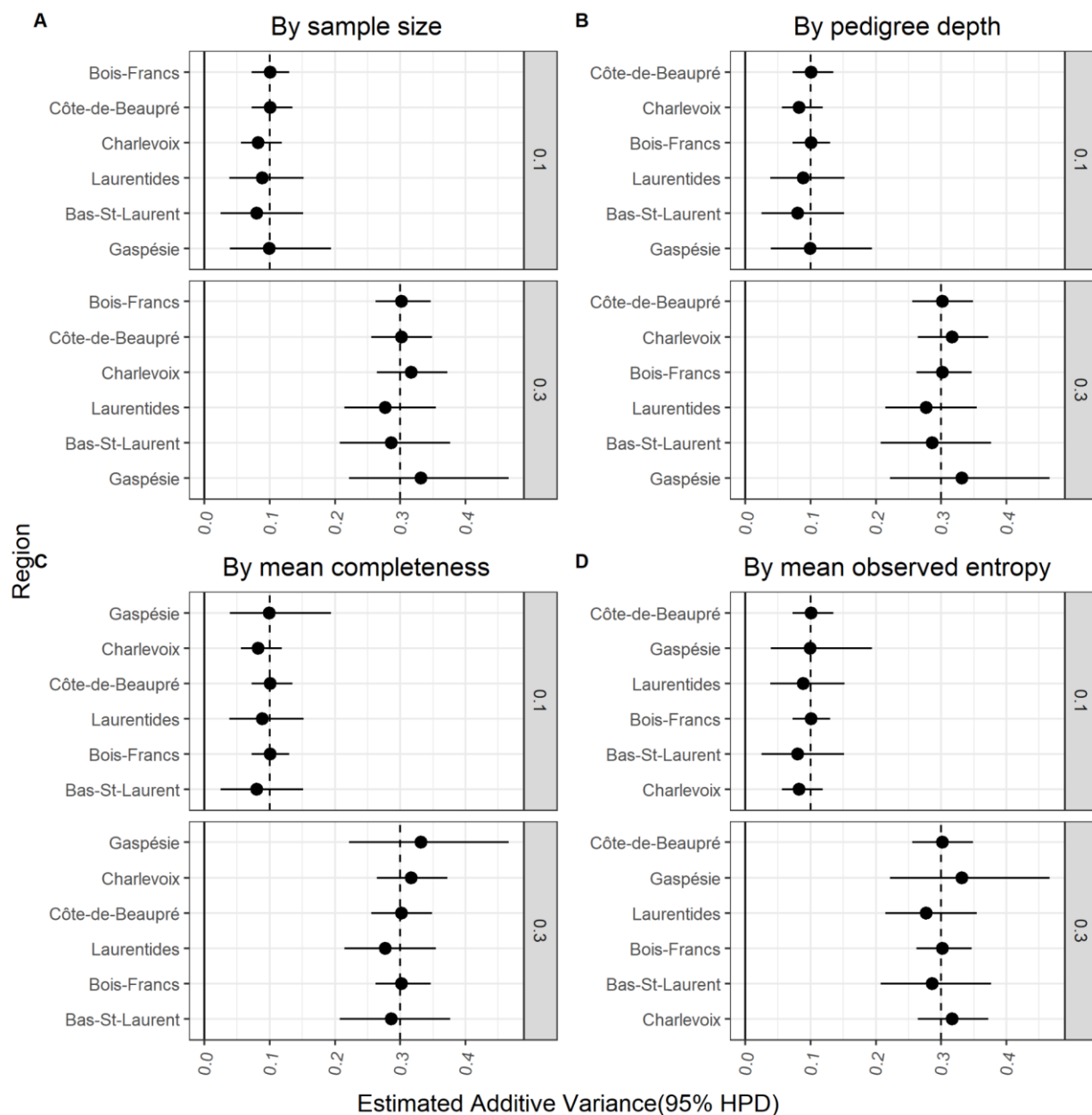
**Table 3.2.** Summary pedigree statistics and genealogical properties of the final six pedigrees pruned for quantitative genetic analyses.

	<b>Gaspeisie</b>	<b>Bas-St-Laurent</b>	<b>Laurentides</b>	<b>Charlevoix</b>	<b>Côte-de-Beaupré</b>	<b>Bois-Francis</b>	<b>Average</b>
<b>Records</b>	1,949	3,741	5,055	6,878	8,450	12,032	6350.83
<b>Maternities</b>	1,288	2,473	3,298	6,074	7,057	9,071	4876.83
<b>Paternities</b>	1,288	2,473	3,298	6,074	7,057	9,071	4876.83
<b>Full sibs</b>	1,755	4,054	4,213	12,145	9,543	14,281	7665.17
<b>Maternal sibs</b>	1,800	4,138	4,309	12,517	9,892	14,664	7886.67
<b>Paternal sibs</b>	1,839	4,279	4,559	13,512	10,307	15,330	8304.33
<b>Maternal grandmothers</b>	446	722	1,218	4,649	5,025	4,773	2805.50
<b>Maternal grandfathers</b>	446	722	1,218	4,649	5,025	4,773	2805.50
<b>Paternal grandmothers</b>	310	605	1,127	4,463	5,159	4,654	2719.67
<b>Paternal grandfathers</b>	310	605	1,127	4,463	5,159	4,654	2719.67
<b>Founders</b>	661	1,268	1,757	804	1,393	2,961	1474.00
<b>Max. pedigree depth</b>	4	6	6	8	9	7	6.67
<b>Mean maternal sibship size</b>	2.69	2.84	2.46	3.65	2.72	2.90	2.88
<b>Mean paternal sibship size</b>	2.71	2.9	2.57	3.83	2.80	3.01	2.97
<b>Mean inbreeding coefficient</b>	$5.89 \times 10^{-5}$	$5.42 \times 10^{-5}$	$6.44 \times 10^{-6}$	$2.48 \times 10^{-3}$	$9.36 \times 10^{-4}$	$8.68 \times 10^{-5}$	$6.04 \times 10^{-4}$
<b>Mean completeness</b>	58.88	39.03	41.45	48.10	43.61	40.65	45.29
<b>Mean expected entropy</b>	1.36	1.34	1.49	2.85	2.92	1.85	1.97
<b>Var. in expected entropy</b>	0.29	0.48	0.57	1.87	2.48	0.86	1.09
<b>Mean rate of missing links</b>	0.94	0.87	0.86	0.58	0.71	0.77	0.79
<b>Mean observed entropy</b>	1.26	1.05	1.08	1.00	1.31	1.07	1.13

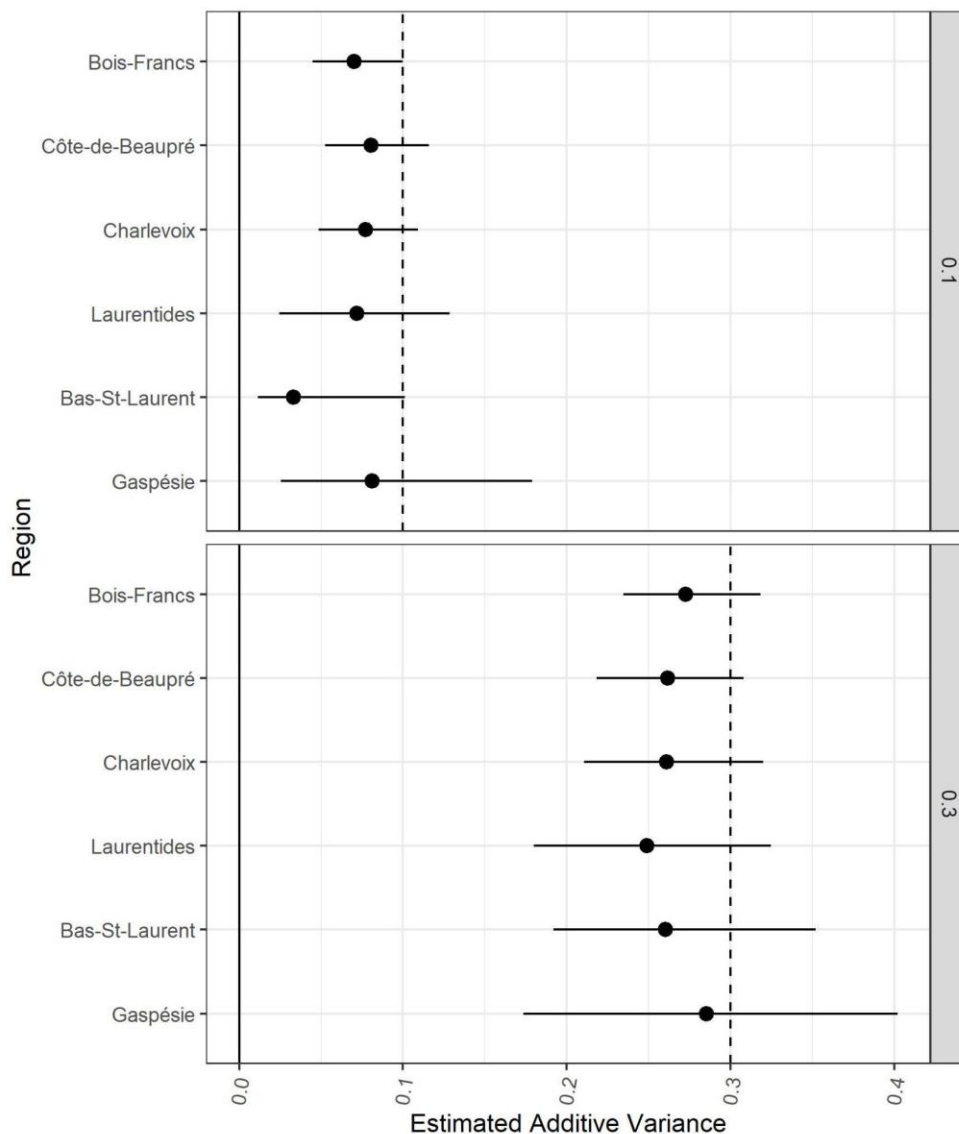
## Figures



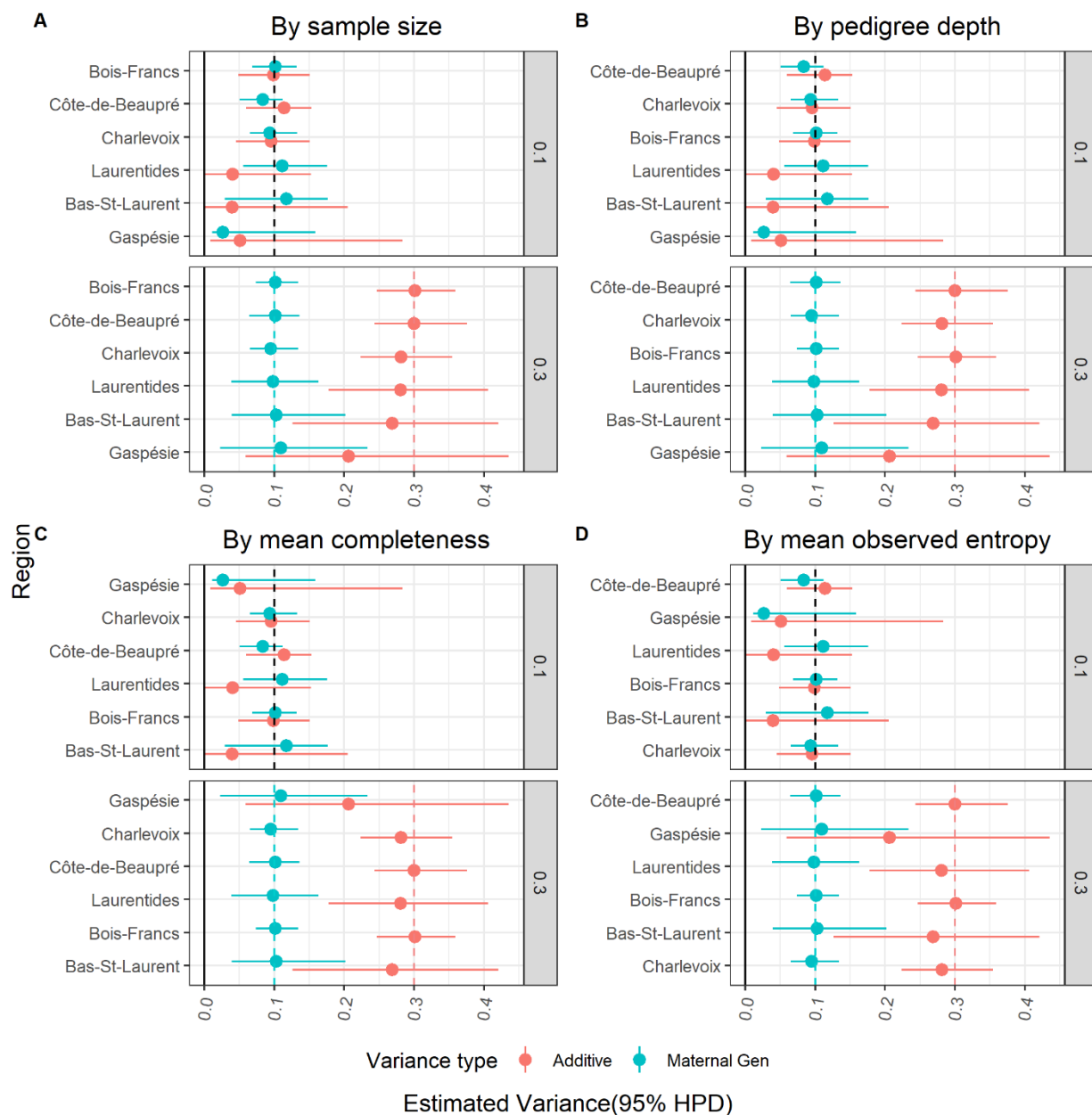
**Figure 3.1.** Schematic of the workflow design of this study. This design is based on Morrissey *et al.* (2007)



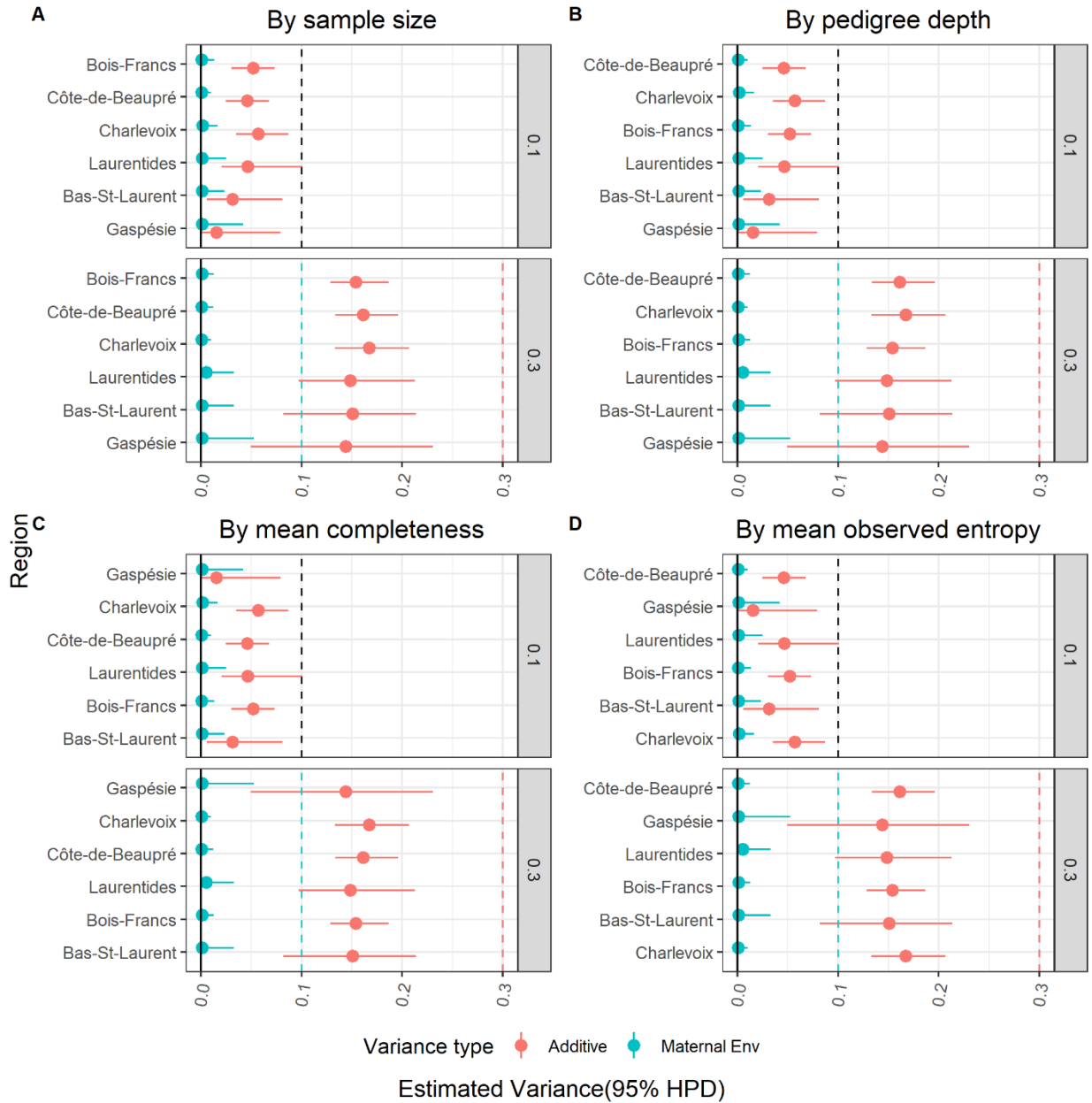
**Figure 3.2.** Estimates of additive genetic variance for each of the six pedigrees based on Scenario 1 (univariate simple phenotype;  $V_A = 0.1$  and  $0.3$  [represented by the values in the gray box]). In each plot, pedigrees are sorted, from bottom to top, by increasing: **A.** sample size, **B.** pedigree depth, **C.** mean completeness, and **D.** mean observed entropy. Each dot represents the average posterior mode-estimates, along with the average 95% HPD intervals based on the five simulations per pedigree per parameter value. Vertical dashed lines represent the expected value of the variance.



**Figure 3.3.** Estimates of additive genetic variance for each of the six pedigrees based on Scenario 1 (univariate simple phenotype;  $V_A = 0.1$  or  $0.3$ ) with a genealogical error rate increased to 1%. Pedigrees are sorted, from bottom to top, by increasing sample size. Each dot represents the average posterior mode-estimates, along with the average 95% credible intervals based on the five simulations per pedigree per parameter value. Vertical dashed lines represent the expected value of the variance.

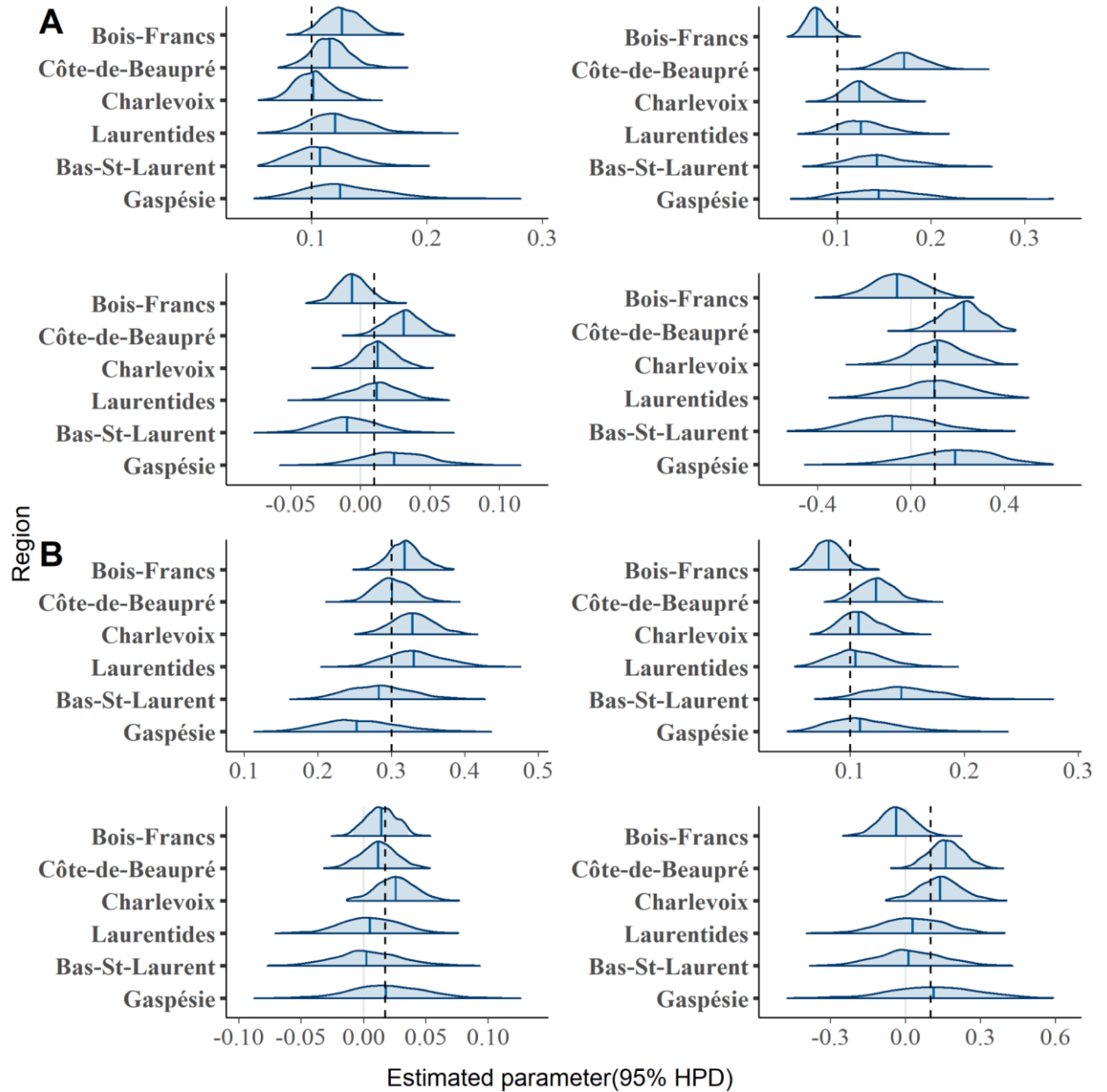


**Figure 3.4.** Estimates of additive genetic variance and maternal genetic variance for each of the six pedigrees based on Scenario 2 (univariate complex phenotype;  $V_A = 0.1$  and  $0.3$  [represented by the values in the gray box];  $V_{MG} = 0.1$ ). In each plot, pedigrees are sorted, from bottom to top, by increasing: **A.** sample size, **B.** pedigree depth, **C.** mean completeness, and **D.** mean observed entropy. Each dot represents the average posterior mode-estimates, along with the average 95% HPD intervals based on the five simulations per pedigree per parameter value. Vertical dashed lines represent the expected value for each variance component.

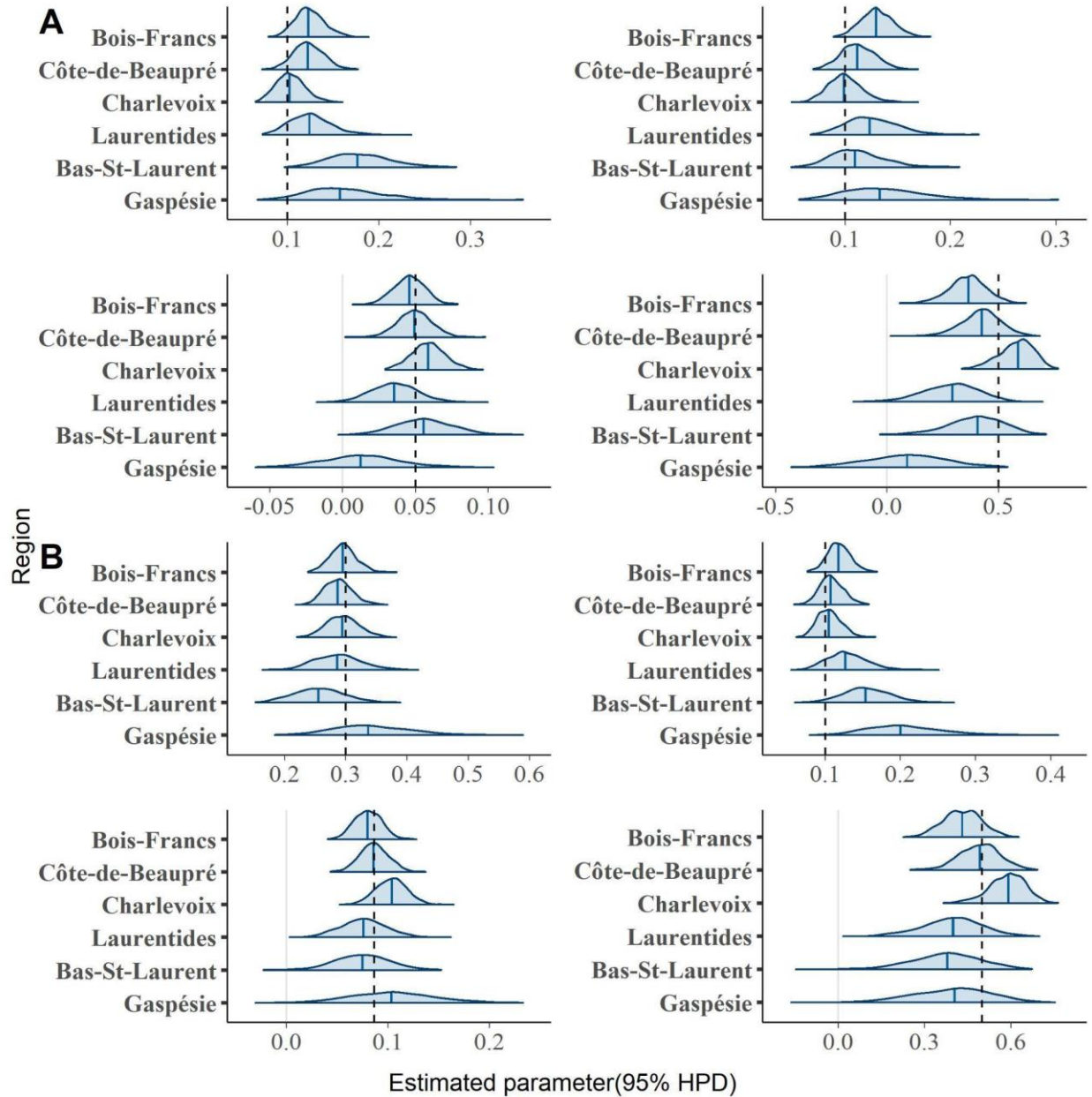


**Figure 3.5.** Estimates of additive genetic variance and maternal environmental variance for each of the six pedigrees based on Scenario 3 (univariate complex phenotype;  $V_A = 0.1$  and  $0.3$  [represented by the values in the gray box];  $V_M = 0.1$ ). In each plot, pedigrees are sorted, from bottom to top, by increasing: **A.** sample size, **B.** pedigree depth, **C.** mean completeness, and **D.** mean observed entropy. Each dot represents the average posterior mode-estimates, along with the average 95% HPD intervals based on the five simulations per pedigree per parameter value. Vertical dashed lines represent the expected value for each variance component.





**Figure 3.6.** Posterior distributions of estimated quantitative genetics parameters (with the 95% HPD interval) for each of the six pedigrees based on Scenario 4 (bivariate simple phenotype with genetic correlation  $r_g = 0.1$ ). Pedigrees are sorted, from bottom to top, by increasing sample size. Top left is additive variance of trait  $z_1$ , top right is additive variance of trait  $z_2$ , bottom left is additive covariance and bottom right is genetic correlation. **A)**  $V_{A,trait z_1} = 0.1$  and  $Cov_A = 0.01$ ; **B)**  $V_{A,trait z_1} = 0.3$  and  $Cov_A = 0.0173$ . In both cases,  $V_{A,trait z_2} = 0.1$ . Vertical dashed lines represent the true expected value of the variance.



**Figure 3.7.** Posterior distributions of estimated quantitative genetics parameters (with the 95% HPD interval) for each of the six pedigrees based on Scenario 4 (bivariate simple phenotype with genetic correlation  $r_g = 0.5$ ). Pedigrees are sorted, from bottom to top, by increasing sample size. Top left is additive variance of trait  $z_1$ , top right is additive variance of trait  $z_2$ , bottom left is additive covariance and bottom right is genetic correlation. **A)**  $V_{A,trait\ z_1} = 0.1$  and  $Cov_A = 0.05$ ; **B)**  $V_{A,trait\ z_1} = 0.3$  and  $Cov_A = 0.0866$ . In both cases,  $V_{A,trait\ z_2} = 0.1$ . Vertical dashed lines represent the true expected value of the variance

## CHAPITRE IV

### **GENOTYPE-BY-ENVIRONMENT INTERACTIONS MODULATE THE RATE OF MICROEVOLUTION IN REPRODUCTIVE TIMING IN HUMANS**

Le contenu de ce chapitre a fait l'objet d'une soumission acceptée avec modifications en anglais dans la revue *Evolution* (Wiley)

La référence de cet article est la suivante :

Mawass, W., Mayer, F. M., & Milot, E. (2022). Genotype-by-environment interactions modulate the rate of microevolution in reproductive timing in humans. *Evolution*.

#### **4.1 Contribution des auteurs**

Walid Mawass a conçu et développé l'étude. Walid Mawass a effectué l'analyse des données et les analyses computationnelles avec la contribution de Emmanuel Milot. Walid Mawass a rédigé le manuscrit. Emmanuel Milot et Francine M. Mayer ont contribué à la version finale du manuscrit. Emmanuel Milot a supervisé le projet.

#### **4.2 Résumé de l'article**

L'observation des populations naturelles montre que les changements des conditions environnementales peuvent entraîner des modifications rapides du potentiel évolutif des phénotypes, en partie par le biais d'interactions génotype-environnement (G×E). Par conséquent, le taux global de microévolution devrait dépendre des fluctuations des conditions environnementales, même lorsque la sélection directionnelle est maintenue sur plusieurs générations. Nous avons testé cette hypothèse dans une population humaine préindustrielle qui a connu un changement microévolutif de l'âge à la première reproduction (AFR) des mères, en utilisant le taux annuel de mortalité infantile (IMR) comme indicateur des conditions environnementales au début de leur vie.

### 4.3 Genotype-by-environment interactions modulate the rate of microevolution in reproductive timing in humans

#### Abstract

Evidence from natural populations shows that changes in environmental conditions can cause rapid modifications in the evolutionary potential of phenotypes, partly through genotype-by-environment interactions (G×E). Therefore, the overall rate of microevolution should depend on fluctuations in environmental conditions, even when directional selection is sustained over several generations. We tested this hypothesis in a preindustrial human population that experienced a microevolutionary change in age at first reproduction (AFR) of mothers, using the annual infant mortality rate (IMR) as an indicator of environmental conditions during their early life. Using quantitative genetics analyses, we found that G×Es explained a non-negligible fraction of the additive genetic variance in AFR and in relative fitness, as well as of the genetic covariance between AFR and fitness (i.e., the Robertson-Price covariance). The covariance was stronger for individuals exposed to unfavorable early-life environmental conditions. Our results unravel the presence of G×Es in an important life history trait and its impact on the rate of microevolution, which appears to have been sensitive to short-term fluctuations in local environmental conditions.

**Keywords:** microevolution, natural selection, environmental conditions, genetic (co)variance, G×E, fitness

#### Introduction

Phenotypic plasticity is often considered to be the main mechanism underlying short-term phenotypic changes in response to environmental changes. However, evolutionary dynamics can operate on ecological timescales and also facilitate adaptation, at the genetic level, to these rapid changes (Charmantier and Garant 2005, Gienapp *et al.* 2008, Engen and Sæther 2014; Ørsted *et al.* 2018). This evolutionary potential is expressed in the genetic covariance between a phenotypic trait and fitness, known as the Robertson-Price covariance, which quantifies the genetic change in mean trait value (microevolution) expected in the next generation, as a response to natural

selection (Price 1970; Robertson 1966, 1968; Morrissey *et al.* 2010). Several studies report rapid microevolution associated with trends in climate (e.g., global warming), resource availability, predation, or habitat deterioration (Réale *et al.* 2003; Franks *et al.* 2007; Barrett *et al.* 2011), or with nonlinear (e.g., yearly) environmental fluctuations (e.g., Grant & Grant 2002; Pigeon *et al.* 2016), such as genetic fluctuations in fish body size as a function of fishing pressures (Olsen *et al.* 2004; Swain *et al.* 2007). Microevolution in the wild can be detected as an intergenerational change in the mean breeding value for the focal trait. The breeding value (BV; also referred to as the ‘genetic merit’) quantifies the deviation of an individual’s phenotype from the population average that is explained by the additive effects of the genes carried by this individual. For quantitative traits, BVs are estimated from mixed-effects regression models incorporating kinship information obtained from pedigrees or dense genomic coverage.

A number of complications arise, though, in our efforts to understand microevolutionary responses (see Morrissey 2012 for an in-depth coverage of some major ones). One is environmental change/variation, either across generations or space, which can not only prompt microevolution, but also alter its pace in the course of the process. In particular, when the environment changes so does, in theory, the genetic (co)variation in traits – including the Robertson-Price covariance – if there exist genotype-by-environment interactions ( $G \times E$ s) involving those traits (Lynch and Walsh 1998; Wilson *et al.* 2006; Chevin *et al.* 2010). This prediction is also supported by studies on natural populations (reviewed in Hoffmann and Merilä 1999; Charmantier and Garant 2005; Wood and Brodie 2015; Rowiński and Rogell 2017). Moreover,  $G \times E$ s could partly explain a temporal trend in the breeding values for a phenotypic trait, simply as a result of an environmental trend, potentially confounding the measurement of responses to selection. In addition to these population-level environmental effects, the variation among individuals in the environment they experience, when the latter is heterogeneous, could influence the Robertson-Price covariance when  $G \times E$ s occur, as well as the variation in the expression of trade-offs between genetically correlated traits (Stearns 1989; Wilson *et al.* 2009).

A question that follows from above considerations is how environmental variability might constrain, accelerate, or change the pace of microevolution. Since the latter can determine how fast populations respond to perturbations such as habitat loss or climate change (Réale *et al.* 2003, Gienapp *et al.* 2014, Charmantier and Gienapp 2014), examining the role of short-term environmental variability in evolutionary processes is relevant both from a fundamental and applied (conservation) perspective. Here we address this question by deciphering a multigenerational genetic response to selection on a life-history trait, which occurred in a natural (human) population.

Variation in life-history traits like age at maturity, age at first reproduction (AFR), or fertility, comprises additive and non-additive genetic components that can be involved in G×Es (Hansen *et al.* 2011, Corbett *et al.* 2018, see Saltz *et al.* 2018 for review), and modulate the evolutionary potential of these traits (Roff 2002; Sgrò and Hoffmann 2004; Messer *et al.* 2016), including in humans (Bolund *et al.* 2015; Bolund and Lummaa 2017; Hayward *et al.* 2018). For instance, Bolund *et al.* (2015) compared the G-matrix for life-history traits in a Finnish population before and after the demographic transition, namely the decline of mortality and fertility in modernizing populations. The demographic transition was prompted by major changes – including cultural, nutritional, sanitary, and economic changes – that have had consequences on the expression of life-history traits and trade-offs (Mace 2000). They found that the additive genetic variance in several traits (age at first reproduction, age at last reproduction, and lifespan) was greater after than prior to the demographic transition, indicating an increase in the evolutionary potential of these traits.

Milot *et al.* (2011) showed that the age at first reproduction of married women in the preindustrial French-Canadian population of île aux Coudres (Québec, Canada) evolved over a period of 140 years. Quantitative genetic analyses support the hypothesis that the phenotypic change in average AFR, which went from 26 yo to 22 yo over the study period (~5-8 generations), was partially due to a genetic response to selection and that the change could not be accounted for by random genetic drift alone. Pelletier *et al.* (2017) further showed that the evolution of AFR at île aux Coudres had measurable consequences on the population growth rate. In addition, in the île aux Coudres population, infant mortality rates (IMR, 0–1-year-old), which captures most of local

juvenile (0-15 yo) mortality (up to 60%), fluctuated from 0 to nearly 30% among years. This is of high relevance here because juvenile mortality both conveys information about local environmental conditions and is predicted by theory to be involved in trade-offs with age at maturity and, by extension, with AFR (Stearns 1989).

Mainly two life-history models have linked juvenile survival to age at maturity/first reproduction. In brief, Stearns and Koella's (1986) optimization model predicts a plastic delay in the optimal age at maturity when juvenile mortality increases, and vice versa, which reflects the underlying relationship between organismal growth and mortality rates. In opposition, the "adaptive developmental plasticity model" posits that earlier maturation and reproductive onset are beneficial to individuals growing up in a stressful environment, i.e. associated with a higher mortality risk in early-life, by increasing their chances to leave some descendants (Bateson *et al.* 2004; Nettle *et al.* 2013; Nettle & Bateson 2015). Notwithstanding the discrepancy between these models, both converge on three points: First, environmental conditions affecting survival during the juvenile period will influence the phenotypic expression of life-history traits, including age at maturity/first reproduction. Second, this plasticity has a genetic basis, i.e. it has evolved under natural selection as an adaptive plastic response. Third, infant mortality at the population level is correlated to (and informative about) environmental quality (Habicht *et al.* 1973) and involved in trade-offs with AFR at the individual level (see also Coall and Chisholm 2010). In support of the latter point, the infant mortality rate is a well-known indicator of population health (Reidpath & Allotey 2003) and correlates with environmental conditions across human populations (Mattison 2010; Placek & Quinlan 2012). Demographic studies have also shown that a large proportion of juvenile mortality occurs within the first year of childhood, specifically in endogamous populations with high levels of inbreeding (Basaran *et al.* 1989; Bittles *et al.* 1991; for more details on IAC see Boisvert and Mayer 1994). Therefore, infant mortality is expected to capture a large part of the variation in juvenile survival that is important in the expression of trade-offs with age at maturity/first reproduction. Considering the above points and the large among-year variation in infant mortality at Île aux Coudres, the documented microevolution of AFR provides the opportunity to investigate how environmental conditions may have modulated the per-generation pace

of microevolution. Specifically, we examined how the Robertson-Price covariance changed as a function of infant mortality rate used as a proxy for environmental conditions.

## **Materials and methods**

### **Study system**

The dataset used in this study contains life-history traits of 572 married women born between 1772 and 1919 at île aux Coudres (IAC), a 34-km<sup>2</sup> island located to the northeast of Québec City along the St. Lawrence River (Canada; Figure S1). The IAC Population Register links the acts of baptism, burial, and marriage recorded in Catholic church parish books, civil registers and population censuses, providing information on all 8,378 individuals born on the island, as well as on all 1,393 marriages celebrated there, between its foundation in 1721 and 1973 (Philippe 1973; Boisvert and Mayer 1994). This largely endogamous population lived on subsistence farming and fishing, with a relative homogeneity among its people in socio-economic status, education and religion. After 1790, the saturation of the land prompted an increase in emigration, and population growth slowed down until a rebound after 1870, due to a diversification in economic activities, although the population still relied heavily on agriculture and fishing for a living until mid-20th century (Martin 1957; see Milot *et al.* 2011 for a recent summary of the socio-economic context).

### **Life-history data**

The dataset includes couples for which we had information from the birth of spouses to the end of their union (i.e. deceased spouse) or the end of natural fertility, coinciding with menopause, which occurs at around 45 yo at île aux Coudres, as indirectly estimated from historical birth records. According to Gold (2011), it is around this age that endocrine changes related to the onset of perimenopause (i.e. menopausal transition) begin signaling the end of the reproductive years. On the other hand, the median age at menopause (generally defined as the end of menstrual cycles) for white women from industrialized countries ranges between 50 and 52 years, however variation in that range exists with respect to different factors (e.g., race/ethnicity, BMI, and



physical and dietary activity; Gold 2011). Civil acts do not allow to distinguish emigrants, who spent their reproductive life partly or in totality outside of the island, from less or non-fecund couples (Boisvert and Mayer 1994). To tackle this problem, we applied the data filtering previously used in other studies on this population (Boisvert and Mayer, 1994; Milot *et al.* 2011; Pelletier *et al.* 2017). Briefly, analyses were performed by first including couples exhibiting long birth intervals, as if they truly had a lower birth rate (subfecundity hypothesis), then by excluding them, as if they emigrated from the island, such that some of their reproductions possibly went unrecorded in the IAC Register (migration hypothesis; Boisvert and Mayer, 1994). Since we detected no noticeable differences between the results obtained under these two hypotheses, we present only the results for the subfecundity hypothesis. The full dataset corresponds to the whole population of 572 mothers married between 1800 and 1939 with all offspring births (4,002). These women were attributed to eight 20-year birth cohorts, corresponding to the maximal number of generations covered by the data, as well as to cohorts used by Milot *et al.* (2011) to document AFR microevolution. Thus, women in the first cohort were born between 1772 and 1791 while those in the last (eighth) cohort were born between 1892 and 1901. This grouping was also required to calculate relative fitness, as explained below. All acts in the IAC Population Register, including those prior to the study period, were used to reconstruct the pedigree connecting these women.

### **Environmental conditions**

Our goal was to assess whether G×Es occurred and had potential to impact on the rate of AFR microevolution. Likely the best proxy of environmental conditions available for this purpose is the annual infant mortality rate. As aforementioned, IMR is, on the one hand, an important indicator of environmental conditions, and, on the other hand, predicted to engage in trade-offs with AFR. Consequently, IMR captures relevant environmental variation that has the potential to modulate AFR microevolution when G×E exist. Initially, we considered annual climatic indices as well (e.g., NAO, drought index, temperature) because climate and seasonal changes impacted resource availability and susceptibility to diseases in preindustrial Québec (Charboneau 1975; Gagnon 2012). Therefore, they would add more direct measurements of environmental factors that

impact on IMR and potentially on AFR (through IMR or directly). However, we faced the issue that historical climatic data at the small local scale of the island are sparse and unavailable for many years.

### **Environmental conditions window**

To assign a measure of early life environmental conditions to each mother in the data set, it was imperative to delineate the life period for which yearly variation in IMR was most strongly associated with our focal trait, i.e. AFR. To do so we applied the sliding window method described in Brommer *et al.* (2008). The method involves creating several time windows of different intervals around the year of birth and choosing the one for which the correlation coefficient is highest between IMR (averaged over the years included in the window) and AFR (see SI for more information). In this way, the IMR average for the chosen window can be assigned to every individual in our dataset based on their year of birth, as an indicator of environmental conditions. We tested a large number of windows covering the first 10 years of life. Each window (wd) is denoted by the years covered around birth, with the birth year represented by 1 (Table S1). For example, wd=123 includes the year of birth and the two subsequent years, while wd=-112 includes the birth year as well as the years preceding and following it (see SI for details). Thus, we tested for G×Es (genetic variance in plasticity) by first selecting a window to capture the most plasticity possible.

Two windows showed the highest Pearson's correlation between IMR and AFR ( $r \approx 0.13$ ): the five first years of life (wd=12345) and the fourth year of life (wd=4). Since both windows yielded very similar results, only those for wd=12345 are reported here (see Table S1 for the correlation coefficient associated with each tested window). Since the choice of these windows entirely relies on a statistical, data-driven criterion, its biological interpretation may not be obvious, especially for wd=4. To complement this purely statistically-based criterion, we also used two windows based on biological criteria: wd=12 represents the period of life during which juvenile mortality risk is by far the highest; wd=-112 includes one year pre- and two years post- birth, to also capture potential effects of environmental conditions during fetal growth.

### **Early-life environmental condition groups**

Life-history models predict either a later or earlier maturity with increasing harshness of the early-life environment, as measured here by IMR. Therefore, in quantitative genetic models we estimated the linear relationship between IMR and response traits (AFR, relative fitness) using a random regression animal model (RRAM; described later). However, IMR shows a complex and multimodal distribution, which could complicate estimation due to the uneven coverage of the range of values (min IMR=0, max IMR=0.29; Figure S2). Consequently, in a second model we fitted IMR as a 3-group categorical variable. The rationale is that we can at least crudely consider three types of early-life conditions to which individuals were exposed: average vs. “extremes”, i.e. harsher and milder conditions than those to which the bulk of the population was exposed. Therefore, women were split into three groups with cutoffs based on IMR quartiles (Figure S2): a low IMR group, representing mild conditions, composed of women in the 1st quartile (n= 141); a middle IMR group (average conditions) including women of the 2nd and 3rd quartiles (n= 288); a high IMR group (harsh conditions) of women in the 4th quartile (n= 143). We also ran analyses using different types of categorizations (e.g., terciles) arriving at similar results. Figure 1 shows the distributions of the life-history traits for women of each IMR group. As the population was growing, there was a modest temporal increase in average IMR, from ~0.08 (20-year cohort 1) to ~0.10 (cohort 8), with many fluctuations though (Figure S3A), especially among years (Figure S2). It is worth noting that for the 7th birth cohort there were no women who experienced mild early-life conditions (i.e. low IMR; Figure S4).

### **Fitness**

In a previous study conducted on this population, Milot *et al.* (2011) used the number of offspring surviving up to 15 yo, as the measure of fitness. Although still widely used in anthropological and behavioral ecology studies, this kind of metric is however inappropriate in the present context for two reasons: First, it conflates selection and inheritance and mixes the parents’ fitness with that of their offspring (Wolf and Wade 2001; Thomson and Hadfield 2017); Second, it assumes a constant population

size, which is incorrect for many human populations including IAC, which grew from 400 to ~1000 people over the study period (Martin 1957). Therefore, we measured relative fitness,  $w$ , for each woman as her reproductive value at birth (the  $v_0$  of Fisher 1958), which is not flawed by these problems:

$$w_i = \sum_{x=0}^{\infty} \lambda^{-x} l_{x_i} m_{x_i}$$

where  $w_i$  is the fitness of woman  $i$ ,  $l_{x_i}$  her survival to age  $x$ ,  $m_{x_i}$  her age-specific fertility at age  $x$ , and  $\lambda$  the intrinsic rate of population growth (see Moorad 2014 for a recent discussion of this fitness metric). Given the size of the IAC population, calculating  $\lambda$  separately for annual cohorts would be burdened by high imprecision (572 women / 140 years  $\approx$  4 women per year, on average). Therefore, we calculated  $\lambda$  for the eight 20-year cohorts mentioned earlier.

### Quantitative genetic analysis

We used pedigree-based quantitative genetic analyses to assess the influence of the early-life environment on the genetic architecture of AFR and relative fitness. Specifically, we used the so-called “animal model”, a type of generalized linear mixed-effects model (GLMM) that integrates the pairwise relatedness matrix provided by the pedigree. Thus, we fitted bivariate models with AFR and relative fitness as response variables to decompose their phenotypic variance into additive genetic, environmental, and residual components and to estimate genetic correlations ( $r_g$ ) among traits and environments (Kruuk 2004; Wilson *et al.* 2010).

### Random regression animal model

The first bivariate model was a random regression animal model (RRAM), which served to quantify how genes and the familial environment contributed to variance in the response of individuals to environmental conditions. Specifically, our model estimated three parameters for each effect (i.e., additive genetic and familial) and for each response variable (AFR, relative fitness): the variance in individual intercepts, the variance in individual slopes, and the covariance between intercepts and slopes. Thus,

IMR was standardized and then fitted as a continuous variable with interaction with random additive genetic and familial effects. We fitted maternal identity as a random factor to control for the familial environment shared by siblings (recalling that couples were monogamous) and to estimate the variance ( $\sigma_M^2$ ) explained by these effects. Each individual with no known mother ( $n=20$ ) was assigned a unique dummy mother ID. The structure of the model is as follows:

$$y_i = \mu + \beta_{IMR} \times IMR_i + \beta_{birthyear} \times birthyear_i + \beta_{infmort} \times infmort_i + \beta_{inbreeding} \times inbreeding_i + \beta_{inbreeding^2} \times inbreeding_i^2 + \beta_{twin} \times twin_i + a_i + b_i \times IMR_i + m_i + n_i \times IMR_i + e_i \quad (\text{eq. 1})$$

where  $y_i$  is the response trait (AFR or fitness) of woman  $i$ ,  $\mu$  the population mean trait, and  $\beta$ s are partial regression coefficients for fixed predictors. The latter include whether a woman gave birth to at least one pair of twins and the number of her infants who died before age 1 (infant mortality count or infmort), as well as the inbreeding coefficient as a covariate (quadratic effect; see Boisvert & Mayer 1994). The woman's birth year was entered as a covariate to control for temporal trends that may be due to unmeasured factors. The reaction norm of a woman's genotype is described by intercept  $a_i$ , which is woman  $i$ 's breeding value for the response trait when IMR is equal to the average value, and slope  $b_i$  that quantifies the change in breeding value as a function of the environmental gradient. Likewise,  $m_i$  and  $n_i$  are the intercept and slope of the reaction norm associated with maternal effects on the response trait. Lastly,  $e_i$  is the residual error. We first fitted the model to the complete data set encompassing the full IMR distribution. We then re-fitted the model without the observations for which IMR fell inside the 2.5% tails of the distribution. This second analysis aimed to verify that our conclusions were not driven by a few women with extreme IMR values.

### **Effect of environmental conditions on the rate of microevolution**

Under the hypothesis of an identical Robertson-Price covariance across environments, the pace of microevolution will not be influenced by fluctuations in IMR among cohorts. Thus, to quantify what would be the expected rate of microevolution under different environmental conditions, we quantified the Robertson-Price covariance

under the hypothetical scenario that all women would be born under identical conditions (i.e. had the same associated IMR value). The aim of this analysis was to assess the magnitude of the maximal impact that environmental conditions could have, in theory, on the response to selection, based on empirical estimates from the RRAM – all else being equal. To do so, we transformed the posterior **G**-matrix from the RRAM, which contains the parameters that describe the inter-individual variation in reaction norms, to a format that recovers the Robertson-Price covariance for one or more specific environments. This is known as the character state framework of G×Es assessment (Nussey *et al.* 2007; Roff & Wilson 2014). The transformation for the Robertson-Price covariance was done using the following equation:

$$\sigma_{A_x(AMR,w)} = \sigma_{a,a(AMR,w)} + x(\sigma_{a,b(AMR,w)} + \sigma_{b,a(AMR,w)}) + x^2\sigma_{b,b(AMR,w)} \quad (\text{eq. 2})$$

where  $\sigma_{A_x(AMR,w)}$  is the Robertson-Price covariance in environment  $x$  (here  $x$  being the IMR value) and the terms to the right of the equation are provided by the RRAM **G**-matrix:  $\sigma_{a,a(AMR,w)}$  and  $\sigma_{b,b(AMR,w)}$  are the additive genetic covariances between AMR and fitness for random intercepts ( $a$ ) or slopes ( $b$ ), respectively;  $\sigma_{a,b(AMR,w)}$  is the additive genetic covariance between AMR random intercepts and fitness random slopes, and vice versa for  $\sigma_{b,a(AMR,w)}$ . Given that we fitted a standardized form of IMR to the RRAM model, we used these standardized values in equation (2). We then transformed IMRs back to the observed scale only for plotting purposes.

### Categorical animal model

In this analysis, IMR was entered as the 3-group factor described above (i.e. low, middle, high IMR) instead of as a continuous variable as in the RRAM. The model included the interaction between IMR and genetic effects. The 6x6 unstructured **G**-matrix contained additive genetic variances ( $\sigma_A^2$ ) along the diagonal, i.e. one for each trait for each IMR group, as well as the additive genetic covariances ( $\sigma_{AA}$ ) off-diagonal. Hence, we did not assume that the genetic effects under different early-life conditions are independent but rather that they can covary (*per* Hadfield 2012). This allowed us to evaluate the cross-environment genetic correlations within each trait. The full **G**-matrix was:

$$\begin{bmatrix} \sigma_{A_1}^2(AFR) & \sigma_{AA_{1,2}}(AFR) & \sigma_{AA_{1,3}}(AFR) & \sigma_{AA_{1,1}}(AFR,w) & \sigma_{AA_{1,2}}(AFR,w) & \sigma_{AA_{1,3}}(AFR,w) \\ \sigma_{AA_{2,1}}(AFR) & \sigma_{A_2}^2(AFR) & \sigma_{AA_{2,3}}(AFR) & \sigma_{AA_{2,1}}(AFR,w) & \sigma_{AA_{2,2}}(AFR,w) & \sigma_{AA_{2,3}}(AFR,w) \\ \sigma_{AA_{3,1}}(AFR) & \sigma_{AA_{3,2}}(AFR) & \sigma_{A_3}^2(AFR) & \sigma_{AA_{3,1}}(AFR,w) & \sigma_{AA_{3,2}}(AFR,w) & \sigma_{AA_{3,3}}(AFR,w) \\ \sigma_{AA_1}(w,AFR) & \sigma_{AA_{1,2}}(w,AFR) & \sigma_{AA_{1,3}}(w,AFR) & \sigma_{A_1}^2(w) & \sigma_{AA_{1,2}}(w) & \sigma_{AA_{1,3}}(w) \\ \sigma_{AA_{2,1}}(w,AFR) & \sigma_{AA_{2,2}}(w,AFR) & \sigma_{AA_{2,3}}(w,AFR) & \sigma_{A_{2,1}}(w) & \sigma_{A_1}^2(w) & \sigma_{AA_{2,3}}(w) \\ \sigma_{AA_{3,1}}(w,AFR) & \sigma_{AA_{3,2}}(w,AFR) & \sigma_{AA_{3,3}}(w,AFR) & \sigma_{A_{3,1}}(w) & \sigma_{AA_{3,2}}(w) & \sigma_{A_1}^2(w) \end{bmatrix}$$

where a term in the format  $\sigma_{AA_{x,x}}(AFR,w)$  (or  $\sigma_{AA_{x,x}}(w,AFR)$ ) corresponds to the Robertson-Price covariance in environment  $x=1$  for low IMR, 2 for middle IMR, 3 for high IMR. A term in the format  $\sigma_{AA_{x,y}}(AFR,w)$  is the genetic covariance between AFR in environment  $x$  and fitness in environment  $y$ , where  $x \neq y$ . These latter terms provide further information about how the genetic architecture (e.g., genes, alleles, epistasis) underlying the Robertson-Price covariance differs among environments. This **G**-matrix provides evidence for G×Es across IMR categories when estimates for a parameter type (e.g., variance in a trait, covariance between traits, covariance within trait between environment, etc.) vary as a function of the specific environment(s). In addition, in the absence of G×Es cross-environment genetic correlations ( $\sigma_{AA_{x,y}}$  terms) are expected to be maximal (i.e. =1; *sensu* Falconer and Mackay 1996; Charmantier and Garant 2005).

Similar to additive genetic effects, we assumed that the familial effects could covary between different early-life conditions, hence we also fitted a 6×6 unstructured covariance matrix. We assumed that residual errors could covary between traits but were independent between environmental conditions represented by IMR groups. The IMR group was included as a fixed effect as well to capture differences in mean trait mean across environmental groups. We included the same fixed effects used for the RRAM model. The structure of the categorical model was as follows:

$$y_i = \mu + \beta_{IMR,2}x_{2,i} + \beta_{IMR,3}x_{3,i} + \beta_{infmort} \times infmort_i + \beta_{inbreeding} \times inbreeding_i + \beta_{inbreeding^2} \times inbreeding_i^2 + \beta_{twin} \times twin_i + \mathbf{a}_i \cdot \mathbf{x}_i + \mathbf{m}_i \cdot \mathbf{x}_i + e_i \cdot x_i \quad (\text{eq. 3})$$

where  $\beta_{IMR,2}$  and  $\beta_{IMR,3}$  are the partial regression coefficients associated respectively to the middle IMR and high IMR environments,  $x_{2,i}$  and  $x_{3,i}$  are indicator variables with value =1 for that corresponding to the early-life environment experienced by woman  $i$

and value =0 for the other. Thus,  $\mu$  is trait mean under the mild (low IMR) environment. Other fixed effects are as in equation (1). The  $\mathbf{a}_i \cdot \mathbf{x}_i$  term is the scalar product of two vectors:  $\mathbf{x}_i = \{x_{1,i}, x_{2,i}, x_{3,i}\}$  is the vector of indicator variables for woman  $i$ 's environment (e.g.,  $\mathbf{x}_i = \{1,0,0\}$  if she experienced a low IMR environment);  $\mathbf{a}_i = \{a_{1,i}, a_{2,i}, a_{3,i}\}$  is the vector of her breeding value predicted in each environment (note that this does not signify that 3 breeding values are estimated for a woman, since only one of these term remains as a result of the product  $\mathbf{a}_i \cdot \mathbf{x}_i$ ). The  $\mathbf{m}_i \cdot \mathbf{x}_i$  term fits per-environment maternal effects following the same model as for breeding values. To model independence in the residual errors between environments, the term  $e_{i \cdot \mathbf{x}_i}$  fits residual errors for each woman  $i$ 's environment independent of the residual errors of the two other environments. Environment-specific genetic correlation  $r_g$  between AFR and relative fitness were calculated from the posterior distributions of genetic parameters as

$$\frac{\sigma_A(\text{AFR}; \text{Fitness})}{\sqrt{\sigma_A^2(\text{AFR}) \times \sigma_A^2(\text{Fitness})}}$$

Animal models were fitted using the Bayesian Markov Chain Monte Carlo implementation in the MCMCglmm R package (Hadfield 2010). To facilitate MCMC sampling and make comparisons scale-invariant, we standardized all continuous variables (mean = 0, s.d. = 1), including response variables. Both response traits, AFR and relative fitness, were fitted using a Gaussian error distribution. To minimize autocorrelation in the MCMC samples, all models were run for 4,000,000 iterations after an initial burn-in of 500,000 iterations and sampled every 4,000 iterations (i.e. 1,000 samples kept) to obtain the posterior distribution of parameters. We used the least informative priors as possible based on the structure of the fitted variance-covariance matrices, as recommended by Hadfield (2012) and De Villemereuil *et al.* (2016). All models were run in R (v4.0.4; R Development Core Team 2019).

### **Interpretation of posterior distributions**

We extracted the posterior distributions of genetic parameters and applied three Bayesian metrics to compare them under different environmental conditions. While p-values are not typically associated with Bayesian reasoning, and indeed cannot be obtained from posterior distributions, they remain a statistical landmark to many. Thus,



for convenience, the first index we chose is the probability of direction (pd), which is, loosely speaking, a Bayesian version of the frequentist p-value through the relationship:  $p\text{-value}_{2\text{-tailed}} = 2(1\text{-pd})$  (Makowski *et al.* 2019a; see SI for interpretation details). Better known is the Bayesian highest posterior density (HPD) interval, providing the narrowest value range that totals up 95% of the posterior density distribution. It measures the uncertainty around the true parameter value, particularly in the case where the mode of the posterior distribution is taken as a point estimate (see SI): the larger the HPD interval the wider is the range of parameter values that are plausible (i.e., the smaller the precision). Finally, the Region of Practical Equivalence (ROPE) is the proportion of the HPD interval that is included within a specified range of parameter values that would be considered as not different enough from zero to be biologically meaningful (see Makowski *et al.* 2019b). As many readers might be unfamiliar with this metric, we explain it with more details in the Supplementary Information. While the ROPE cut-off is subjective, it has the advantage of reconnecting significance and effect size. The question is not anymore “is the estimate significantly different from zero” (a question coherent only under the frequentist approach) but rather “what is the probability that the parameter value is at least as large as the cut-off value”. We selected three cut-offs to calculate ROPEs, as explained in the Supplementary Information. Bayesian indices were computed with the bayestestR R package (Makowski *et al.* 2019a).

## Results

### Random regression animal model

*Variance parameters.* The bivariate animal model fitting random slopes for G×Es outperforms the model fitting random intercepts only (deviance information criterion equal to 2491.5 and 2763.4 for the RRAM and the model with random intercepts only, respectively). The RRAM provides non-zero additive genetic variances in random intercept and slope for both response traits (Table 1).  $\sigma_{A,b}^2$  is even slightly higher when the restricted data set, i.e., with the truncated IMR distribution, is used instead of the full data set (Table S2). *ROPE* indices indicate that the posterior probability is null for  $\sigma_{A,b}^2$

close to zero (Table 2). Interestingly, we also detected a similar amount of variation in random slopes associated with familial effects for both response variables (Table 1), which indicates that the magnitude of the change in trait expression, as a function of environmental conditions, also varied among families.

*Robertson-Price covariance.* We note that the Robertson-Price covariance between AFR and relative fitness was separated into two components: the covariance of random intercepts and that of random slopes (Figure 2A inset). The first is interpreted as the expected change in mean breeding value in the next generation due to selection when IMR is equal the average value of the distribution (at the observed scale, given that we used standardized IMR values for the character-state transformation, see Materials and Methods), i.e., in the most benign environment. This covariance was negative and, as expected, recovers the signal of selection for an advancement in AFR previously documented with the same data ( $\sigma_{a,a(AFR,w)} = -0.12 [-0.25 - 0]$ ,  $pd = 0.99$ ,  $ROPE = 0.004$  to  $0.021$ ). The genetic covariance of random slopes was also negative with  $\sigma_{b,b(AFR,w)} = -0.04 [-0.13 - 0]$  (or  $-0.29 [-0.55 - 0]$  when expressed as a genetic correlation). The  $pd$  index indicates that this covariance has a low probability (i.e., 0.03) to be  $\geq 0$ . Recalling that it represents a change in the mean breeding values due to genotype-by-environment interactions, the negative point estimate of  $\sigma_{b,b(AFR,w)}$  can be interpreted as a negative relationship between the fraction of AFR and relative fitness breeding values determined by the environment. Thus, the magnitude of the Robertson-Price covariance is influenced by G×Es. Specifically, it depends on environmental conditions to which a given birth cohort was exposed, the frequency of different genotypes in this cohort, and the reaction norm of these genotypes.

### **Categorical animal model**

*Variance parameters.* Results from the model with IMR fitted as a categorical factor are generally consistent with those from the RRAM, albeit less clear-cut. For all IMR groups, each variance component ( $\sigma_A^2$ ,  $\sigma_M^2$ ,  $\sigma_E^2$ ) accounts for roughly a third to a half of total variance ( $\sigma_{total}^2 = \sigma_A^2 + \sigma_M^2 + \sigma_E^2$ , i.e. the phenotypic variance after controlling for fixed effects) according to posterior-mode estimates, and heritabilities are large, ranging from 0.31 to 0.40 (Table 3). Bayesian indices indicate that the probability that any of the

genetic variance parameters has a value of (or close to) zero is essentially null (Table 4). Despite some non-negligible between-environment differences in genetic parameter estimates (e.g.,  $\sigma_A^2$  for relative fitness goes from 0.36 under low IMR to 0.55 under high IMR), Bayesian 95% HPD intervals are wide and largely overlapping for all parameters (Table 3). Additionally, between-environment differences in the additive variance provide no clear evidence that  $\sigma_A^2$  in a given environment is higher or lower than in another one. Specifically, the posterior distribution of  $\sigma_A^2$  differences ranges largely from negative to positive values:  $\sigma_{A_3}^2 - \sigma_{A_2}^2 = 0.16$  [-0.14 - 0.56];  $\sigma_{A_3}^2 - \sigma_{A_1}^2 = 0.03$  [-0.34 - 0.49];  $\sigma_{A_2}^2 - \sigma_{A_1}^2 = -0.03$  [-0.39 - 0.20].

*Robertson-Price covariance.* This covariance is negative for the three IMR groups (Table 5). The one for the high IMR group is lower (i.e. most negative) than for the two other groups ( $\sigma_{AA(AFR,w)} = -0.19$  [-0.51 - 0], -0.12 [-0.28 to -0.02], -0.11 [-0.30 - 0.04] for the high, middle, and low IMR group, respectively). However, there is a large overlap in their credible intervals (Table 5, Figure 2A). According to these intervals and other Bayesian indices, the probability is low for covariances to be  $\geq 0$  for the middle and high IMR groups (all  $ROPE \leq 0.015$  and  $pd \geq 0.98$ ; Table 4), and slightly higher for the low IMR group ( $ROPE \leq 0.067$ ,  $pd=0.95$ ). Transforming Robertson-Price covariances into genetic correlations ( $r_g$ ) between AFR and relative fitness,  $r_g$  was lower in magnitude (i.e. less negative) for the low IMR group ( $r_{g_1} = -0.27$  [-0.58 - 0.04]) relative to that of the other IMR groups ( $r_{g_2} = -0.42$  [-0.59 to -0.10] for the middle IMR group and  $r_{g_3} = -0.45$  [-0.68 to -0.06] for the high IMR group; Table 5). Like for  $\sigma_A^2$ , the posterior distribution does not reveal a systematic direction in between-environment differences in genetic correlations:  $r_{g_3} - r_{g_2} = -0.05$  [-0.43; 0.37];  $r_{g_3} - r_{g_1} = -0.15$  [-0.56; 0.33];  $r_{g_2} - r_{g_1} = 0.01$  [-0.49; 0.32]. The cross-environment genetic correlations in AFR, for every environmental combination, were lower than one ( $r_{g_{1,2}} = 0.09$  [-0.21 - 0.48],  $r_{g_{2,3}} = 0.07$  [-0.24 - 0.48], and  $r_{g_{1,3}} = 0.17$  [-0.24 - 0.48]; Figure 2B), which is an indication of G×Es (the same pattern applies to relative fitness). In this case, we are treating our focal trait within each environmental condition as a separate character (Via & Lande 1985).

### **G×Es and the pace of microevolution**

The Robertson-Price covariance was much more negative under the hypothetical scenario whereby all IAC women would be born under the worst early-life conditions (IMR=0.3; posterior mode  $\approx -0.79$ ) than average and mild conditions (posterior mode smaller than -0.15 for IMR=0.09 or 0; Figure 3). Therefore, all else being equal, exposure to unfavorable early-life conditions could lead to a microevolutionary change in AFR many times greater. However, we note that HPD intervals for the covariance are quite large when IMR approaches the upper bound of 0.3.

### **Discussion**

A previous study had shown that the AFR of married women in the population of île aux Coudres underwent a microevolutionary change in response to selection during the preindustrial period (18th-20th centuries; Milot *et al.* 2011). This had measurable consequences on demography, increasing the population growth rate relative to that expected had no evolutionary change occurred (Pelletier *et al.* 2017). In this study, we further show that short-term fluctuations in early-life environmental conditions likely impacted the overall rate of microevolutionary change. This is in line with studies on natural populations revealing that environmental changes modulate the genetic (co)variance of traits (reviewed in Hoffmann and Merilä 1999; Charmantier and Garant 2005; Charmantier *et al.* 2014). This is especially true for complex traits such as life-history traits, which are predicted to exhibit G×Es (Roff 2002) that can affect the rate of microevolution (Teplitsky *et al.* 2014).

### **Measurement of environmental conditions**

Due to the historical nature of IAC data, little information was available to associate environmental conditions with individual life histories. However, infant (0-1 yo) mortality, which is readily measured from descending genealogies when the timing of birth and death is known, is indicative of environmental factors (e.g., climate, nutrition, pathogens) that affect the physiological development of individuals surviving to maturity, who may respond by adjusting their fertility and reproductive timing

(Quinlan 2007; Nettle 2011; Giudice *et al.* 2015), in line with life-history theory (Stearns 1992; Ellis *et al.* 2009; Griskevicius *et al.* 2011). A recent genomic study using the UK Biobank and infant mortality rate as a proxy for environmental quality, found moderate and negative genetic correlations between regional, annual IMR and fertility and AFR (Wu *et al.* 2021). Other studies on the French-Canadian population further support the link between environmental conditions and survival. Bergeron *et al.* (2014) found a positive relationship between solar irradiance and infant survival at île aux Coudres for the same time period as the one covered in the current study. The variance in survival explained by solar irradiance was nevertheless small, possibly because this climatic index reflects a small fraction of the local variation in environmental conditions (due to the limits of climatic data explained in Methods). In addition, preliminary results for the preindustrial population of a much larger French-Canadian region show that infant mortality varies with climatic conditions (temperature, drought index, and NAO; Noël *et al.*, unpubl.). In this case, the scale of measurement of climatic variables better matches the region size.

### **Evidence for genotype-by-environment interactions**

The approach used in this study to detect G×E is based on selecting the environment (or a proxy of the environment) which captures the most plasticity in our focal trait and hence most likely to reveal G×E. However, we do not claim that this is the sole approach for predicting and detecting G×Es as there could be other evolutionary rationales that justify taking different approaches than the one used here. For example, we initially tried to the per-cohort per-environment genetic parameters, while accounting for the different per-cohort proportion of women in each IMR group. However, despite some interesting patterns, there were too many parameters to estimate given the sample (i.e. population) size and too much uncertainty around the estimates for any meaningful biological interpretation of these results.

According to the random regression model, the additive genetic variance in age at first reproduction and relative fitness reflects variation in both individual intercepts (i.e. unrelated to environmental conditions) and slopes as a function of the environmental gradient. This means that the plasticity in this trait has a genetic underpinning expressed

as a differential response of genotypes to the environmental gradient. Therefore, variation in the conditions of the early-life environment contributed, through interactions with the underlying genetic architecture of traits, to the additive variation in AFR. The same result is found for relative fitness.

Changes in genetic covariation between traits across environmental conditions are indicators of G×Es at play (see Aastveit & Aastveit 1993). Our results support the hypothesis that G×Es also modulated the Robertson-Price covariance at IAC. This translates to a stronger association between genes that influence age at first reproduction to occur earlier and genes that increase relative fitness under harsher conditions. In this population, a woman's relative fitness is heavily determined by her fertility (i.e. the total number of her children), and when she began reproduction early she tended to end it early as well. This raises the possibility that under unfavorable conditions, the tradeoff between early and late reproduction had a stronger genetic determinism. However, it remains difficult to map genetic correlations between life-history traits directly to their expressed trade-offs at the phenotypic level (Bolund 2020). In addition, based on our character state analysis, the cross-environment genetic correlations in AFR were estimated to be (much) less than one, which provides further evidence for G×Es (*sensu* Falconer and Mackay 1996; Charmantier and Garant 2005).

### **G×Es affect the rate of microevolution**

Regardless of early-life conditions, the direction of the genetic covariance between AFR and fitness (i.e. the Robertson-Price covariance) did not change at Île aux Coudres, meaning that selection consistently favored earliest AFR values during the study period. Indeed, environmental-driven changes are expected to affect the direction of selection more rarely than its strength (Siepielski *et al.* 2013). Nevertheless, both animal models fitted here (RRAM and categorical) point to stronger selection and expected response as the early-life environment gets harsher (albeit with greater uncertainty around covariance estimates with the categorical model). The possible role of G×Es in modulating the change in mean breeding values from one generation to the next is further highlighted by the fact that average IMR fluctuated much among cohorts (Figure S3). Rapid fluctuations in early-life environmental conditions coupled with

G×Es can thus play a role in modulating the expected rate of microevolutionary change even when selection is consistent overall (see Kolmodin *et al.* 2003).

### **Familial effects**

In both animal models a fair part of the variation in both AFR and fitness is explained by familial effects. The RRAM further suggests that among-environment variation partly arises from these familial effects (as attested by the non-zero  $\sigma_M^2$  due to individual slopes, of similar value to the  $\sigma_A^2$  in individual slopes). This is noticeable because Milot *et al.* (2011) estimated  $\sigma_M^2$  to be negligible or very small in their analysis of the same data that incorporated the same fixed and random effects, excepting IMR. One potential explanation is that variation in environmental conditions might have accentuated dissemblance in traits among full sibs that masked familial effects in the previous analysis. This could also reflect different effects of mothers on their daughters' phenotype given different early-life conditions to which their daughters are exposed.

### **Parameter uncertainty and interpretation**

G×Es can be hard to detect in the wild, especially when involving non-labile traits such as AFR, compared to in tailored experimental settings (Gienapp and Brommer 2014; Ingleby *et al.* 2010; Des Marais *et al.* 2013). We verified that our results were not driven by a few extreme IMR values or specific parts of the multimodal IMR distribution by running two additional analyses that revealed trends consistent with the full RRAM model. However, uncertainty around genetic parameter estimates from the categorical model, reflected in their posterior distributions, prevents us from drawing firm conclusions about differences among environmental conditions as categorized in this model. In terms of the magnitude of cross-environmental genetic correlations for AFR and relative fitness, we found them to be less than 1 across two different priors in the categorical model. This result increases our confidence that there are G×Es at play in both traits in the IAC population, which contrasts with a recent study that did not detect any G×Es in a non-human natural system (Hayward *et al.* 2018). However, since our categorical model implied the estimation of a large number of genetic parameters (about four times that of the RRAM), we suspect that the “sample” size was perhaps too small

to allow quantitative statements about  $G \times E$ s with some precision. Since our sample actually amounts to the whole population of women who reproduced at île aux Coudres during the study period, it would be hardly possible to increase the power. Therefore, the categorical model fitted here with its 21 genetic (co)variance parameters (and the same number of familial effects) is at the limit (or above) of the level of the complexity we may hope to analyze with Bayesian GLMM modeling applied to this population.

### **Conclusion: measuring microevolution when $G \times E$ s occur**

Recent studies have revealed that some natural populations can quickly adapt to rapid fluctuations in environmental conditions (e.g., climate change; Charmantier and Gienapp 2014) through microevolutionary change driven, in part, by a response to natural selection (though see Ozgul *et al.* 2010). On île aux Coudres, AFR underwent such an evolutionary change and impacted the demographic growth of the population (Milot *et al.* 2011; Pelletier *et al.* 2017). This study further shows that the pace of microevolution may be sensitive to short-term fluctuations in local environmental conditions. This has two implications for the empirical study of microevolutionary processes. First, modeling  $G \times E$ s whenever possible can help better understand how evolutionary potential fluctuates over short time periods due to environmental changes/fluctuations. Second, it may be important to separate (fractions of) temporal trends in breeding values that reflect real genetic response to selection from those resulting from the mere expression of  $G \times E$ s when the environment changes. In the case of île aux Coudres, both the additive variance in traits and the covariance between AFR and fitness varied with IMR. Nevertheless, the temporal trend in breeding values is linear and steady (Milot *et al.* 2011), while environmental conditions (IMR) fluctuated much with ups and downs but, in general, exhibited only a very slight linear change in time. Thus, this supports the hypothesis that the genetic advancement in AFR is largely due to a response to selection (with potential effects of  $G \times E$ s on this response). With a focus on estimating environment-specific additive (co)variance in nature, we can reach a better understanding of how evolutionary potential can fluctuate with time and space, and with environmental heterogeneity. Hopefully, future work will determine the importance in



other natural populations (including humans) that short-term environmental fluctuations may have on the microevolutionary dynamics of complex traits.

## References

- Aastveit, A. H., & Aastveit, K. 1993. Effects of genotype-environment interactions on genetic correlations. *Theor Appl Genet*, 86(8), 1007-1013.
- Barrett, R. D., Paccard, A., Healy, T. M., Bergeck, S., Schulte, P. M., Schluter, D., & Rogers, S. M. 2011. Rapid evolution of cold tolerance in stickleback. *Proc. R. Soc. B.*, 278(1703), 233-238.
- Basaran, N., Hassa, H., Basaran, A., Artan, S., Stevenson, J. D., and Sayli, B. S. 1989. The effect of consanguinity on the reproductive wastage in the Turkish population. *Clin. Genet.* 36(3):168-173.
- Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R.A., Gluckman, P., Godfrey, K., Kirkwood, T., Lahr, M.M. and McNamara, J. 2004. Developmental plasticity and human health. *Nature.* 430(6998):419.
- Bergeron, P., Milot, E., Mayer, F. M., Boisvert, M., Réale, D., & Pelletier, F. 2014. Solar Irradiance, Survival and Longevity in a Pre-industrial Human Population. *Hum. Ecol.*, 42(4), 645-650.
- Bittles, A. H., Mason, W. M., Greene, J., and Rao, N. A. 1991. Reproductive behavior and health in consanguineous marriages. *Science.* 252(5007):789-794.
- Boisvert, M., and Mayer, F. M. 1994. Infant mortality and consanguinity in an endogamous population in Québec. *Population.* 49(3):685-724.
- Bolund, E., Hayward, A., Pettay, J. E., and Lummaa, V. 2015. Effects of the demographic transition on the genetic variances and covariances of human life-history traits. *Evolution.* 69(3):747-755.
- Bolund, E., and Lummaa, V. 2017. The effects of resource availability and the demographic transition on the genetic correlation between number of children and grandchildren in humans. *Heredity.* 118(2):186–192.
- Bolund, E. 2020. The challenge of measuring trade-offs in human life history research. *Evol Hum Behav*, 41(6), 502-512.

- Brommer, J. E., Rattiste, K., and Wilson, A. J. 2008. Exploring plasticity in the wild: laying date–temperature reaction norms in the common gull *Larus canus*. *Proc. R. Soc. B.* 275(1635):687-693.
- Charbonneau H. 1975. *Vie et Mort de nos ancêtres*. Montréal: Les Presses de l'Université de Montréal.
- Charmantier, A., and Garant, D. 2005. Environmental quality and evolutionary potential: lessons from wild populations. *Proc. R. Soc. B.* 272(1571):1415-1425.
- Charmantier, A., and Gienapp, P. 2014. Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evol. Appl.* 7(1):15–28.
- Chevin, L. M., Lande, R., and Mace, G. M. 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLOS Biol.* 8(4):e1000357.
- Coall, D. A., and Chisholm, J. S. 2010. Reproductive development and parental investment during pregnancy: Moderating influence of mother's early environment. *Am. J. Hum. Biol.* 22(2):143–153.
- Corbett, S., Courtiol, A., Lummaa, V., Moorad, J. A., and Stearns, S. C. 2018. The transition to modernity and chronic disease: mismatch and natural selection. *Nat. Rev. Genet.* 19(7):419-430
- Des Marais, D. L., Hernandez, K. M., and Juenger, T. E. 2013. Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. *Annu. Rev. Ecol. Evol. Syst.* 44:5-29.
- De Villemereuil, P., Schielzeth, H., Nakagawa, S., and Morrissey, M. 2016. General methods for evolutionary quantitative genetic inference from generalized mixed models. *Genetics.* 204(3):1281-1294.
- Ellis, B. J., Figueredo, A. J., Brumbach, B. H., and Schlomer, G. L. 2009. Fundamental dimensions of environmental risk. *Hum. Nat.* 20(2):204-268.
- Engen, S., and Sæther, B. E. 2014. Evolution in fluctuating environments: Decomposing selection into additive components of the robertson-price equation. *Evolution.* 68(3):854–865.
- Falconer, D. S., and Mackay, T. F. C. 1996. *Introduction to quantitative genetics*. Harlow, UK.

- Fisher, R. A. 1958. *The genetic theory of natural selection*. Dover.
- Franks, S. J., Sim, S., & Weis, A. E. 2007. Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *PNAS*, 104(4), 1278-1282.
- Gagnon, A. 2012. Effect of birth season on longevity: thrifty and hopeful phenotypes in historical Quebec. *Am. J. Hum. Biol.* 24(5), 654-660.
- Gienapp, P., Teplitsky, C., Alho, J. S., Mills, J. A., and Merilä, J. 2008. Climate change and evolution: Disentangling environmental and genetic responses. *Mol. Ecol.* 17(1):167–178.
- Gienapp, P., and Brommer, J. E. 2014. Evolutionary dynamics in response to climate change. In *Quantitative Genetics in the Wild*. 254-273.
- Giudice, M. D., Gangestad, S. W., & Kaplan, H. S. 2015. Life history theory and evolutionary psychology. *The handbook of evolutionary psychology*, 1-27.
- Gold, E. B. 2011. The timing of the age at which natural menopause occurs. *Obstet. Gynecol. Clin.* 38(3):425-440.
- Grant, P. R., and Grant, B. R. 2002. Unpredictable evolution in a 30-year study of Darwin's finches. *Science*. 296(5568):707-711.
- Griskevicius, V., Delton, A. W., Robertson, T. E., and Tybur, J. M. 2011. Environmental contingency in life-history strategies: the influence of mortality and socioeconomic status on reproductive timing. *J. Pers. Soc. Psychol.* 100(2):241.
- Habicht, J.-P., Yarbrough, C., Lechtig, A. & Klein, R. Relationship of birthweight, maternal nutrition and infant mortality. *Nutrition Reports International* 7, 533-546 (1973)
- Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Softw.* 33(2):1-22.
- Hadfield, J. 2012. MCMCglmm course notes. Available at <http://cran.rproject.org/web/packages/MCMCglmm/vignettes/CourseNotes.pdf>.
- Hansen, T. F., Pélabon, C., and Houle, D. 2011. Heritability is not evolvability. *Evol. Biol.* 38(3):258.
- Hayward, A. D., Pemberton, J. M., Berenos, C., Wilson, A. J., Pilkington, J. G., and Kruuk, L. E. B. 2018. Evidence for selection-by-environment but not genotype-by-

- environment interactions for fitness-related traits in a wild mammal population. *Genetics*. 208(1):349–364.
- Hoffmann, A. A., and Merilä, J. 1999. Heritable variation and evolution under favourable and unfavourable conditions. *Trends Ecol. Evol.* 14(3):96-101.
- Ingleby, F. C., Hunt, J., and Hosken, D. J. 2010. The role of genotype-by-environment interactions in sexual selection. *J. of Evol. Biol.* 23(10):2031-2045.
- Kolmodin, R., Strandberg, E., Jorjani, H., & Danell, B. 2003. Selection in the presence of a genotype by environment interaction: response in environmental sensitivity. *Animal Science*, 76(3), 375-385.
- Kruuk, L. E. 2004. Estimating genetic parameters in natural populations using the ‘animal model’. *Philos. Trans. R. Soc. B.* 359(1446):873-890.
- Lynch, M., and Walsh, B. 1998. *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer.
- Mace, R. 2000. Evolutionary ecology of human life history. *Animal behaviour*, 59(1), 1-10.
- Makowski, D., Ben-Shachar, M. S., & Lüdtke, D. 2019a. bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, 4(40), 1541.
- Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdtke, D. 2019b. Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology* 2019;10:2767.
- Martin, Y. 1957. L’Île-aux-Coudres: Population et économie. *Cahiers de géographie du Québec*, 2(2) :167-195.
- Mattison, D. R. 2010. Environmental exposures and development. *Current opinion in pediatrics*, 22(2), 208.
- Messer, P. W., Ellner, S. P., and Hairston Jr, N. G. 2016. Can population genetics adapt to rapid evolution? *Trends Genet.* 32(7):408-418.
- Milot, E., Mayer, F. M., Nussey, D. H., Boisvert, M., Pelletier, F., and Réale, D. 2011. Evidence for evolution in response to natural selection in a contemporary human population. *Proc. Natl. Acad. Sci. U.S.A.* 108(41):17040-17045.

- Moorad, J. A. 2014. Individual fitness and phenotypic selection in age-structured populations with constant growth rates. *Ecology*. 95(4):1087-1095.
- Morrissey, M. B., Kruuk, L. E. B., & Wilson, A. J. 2010. The danger of applying the breeder's equation in observational studies of natural populations. *J. Evol Biol*, 23(11), 2277-2288.
- Morrissey, M.B., Parker, D.J., Korsten, P., Pemberton, J.M., Kruuk, L.E. and Wilson, A.J., 2012. The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution: International Journal of Organic Evolution*, 66(8), pp.2399-2410.
- Morrissey, M. B., de Villemereuil, P., Doligez, B., and Gimenez, O. 2014. Bayesian approaches to the quantitative genetic analysis of natural populations. In *Quantitative genetics in the wild*, 228-253.
- Nettle, D. 2011. Flexibility in reproductive timing in human females: integrating ultimate and proximate explanations. *Philos. Trans. R. Soc. B*. 366(1563):357-365.
- Nettle, D., Frankenhuis, W. E., and Rickard, I. J. 2013. The evolution of predictive adaptive responses in human life-history. *Proc. R. Soc. B*. 280(1766):20131343.
- Nettle, D., and Bateson, M. 2015. Adaptive developmental plasticity: what is it, how can we recognize it and when can it evolve? *Proc. R. Soc. B*. 282(1812):20151005.
- Nussey, D.H., Wilson, A.J. and Brommer, J., 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. *J. Evol. Bio.*, 20(3), pp.831-844.
- Ørsted, M., Hoffmann, A. A., Rohde, P. D., Sørensen, P., and Kristensen, T. N. 2018. Strong impact of thermal environment on the quantitative genetic basis of a key stress tolerance trait. *Heredity*. 122(3):315-325.
- Olsen, E. M., Heino, M., Lilly, G. R., Morgan, M. J., Brattey, J., Ernande, B., & Dieckmann, U. 2004. Maturation trends indicative of rapid evolution preceded the collapse of northern cod. *Nature*, 428(6986), 932-935.
- Ozgul, A., Childs, D.Z., Oli, M.K., Armitage, K.B., Blumstein, D.T., Olson, L.E., Tuljapurkar, S. and Coulson, T. 2010. Coupled dynamics of body mass and population growth in response to environmental change. *Nature*. 466(7305):482.

- Pelletier, F., Pigeon, G., Bergeron, P., Mayer, F. M., Boisvert, M., Réale, D., and Milot, E. 2017. Eco-evolutionary dynamics in a contemporary human population. *Nat. Commun.* 8:15947.
- Philippe, P. 1973. Fécondité, fécondabilité et consanguinité à l'Isle-aux-Coudres. *Rech. Sociogr.* 14(1):117-123.
- Pigeon, G., Festa-Bianchet, M., Coltman, D. W., & Pelletier, F. 2016. Intense selective hunting leads to artificial evolution in horn size. *Evol Appl*, 9(4), 521-530.
- Placek, C. D., and Quinlan, R. J. 2012. Adolescent fertility and risky environments: a population-level perspective across the lifespan. *Proc. R. Soc. B.* 279(1744):4003-4008.
- Price, G. R. 1970. Selection and covariance. *Nature.* 227(5257):520-521.
- Quinlan, R. J. 2007. Human parental effort and environmental risk. *Proc. R. Soc. B.* 274(1606):121–125.
- R Core Team 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Réale, D., McAdam, A. G., Boutin, S., and Berteaux, D. 2003. Genetic and plastic responses of a northern mammal to climate change. *Proc. R. Soc. B.* 270(1515):591–596.
- Reidpath, D. D., & Allotey, P. 2003. Infant mortality rate as an indicator of population health. *J. Epidemiol. Community Health*, 57(5), 344-346.
- Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Anim. Prod.* 8: 95–108.
- Robertson, A. 1968. Population Biology and Evolution. In *The spectrum of genetic variation*. Syracuse University Press, New York.
- Roff, D. A. 2002. Life-history evolution. Sinauer Associates
- Roff, D.A. and Wilson, A.J., 2014. Quantifying genotype-by-environment interactions in laboratory systems. *Genotype-by-environment interactions and sexual selection*, pp.100-136.

- Rowiński, P. K., and Rogell, B. 2017. Environmental stress correlates with increases in both genetic and residual variances: A meta-analysis of animal studies. *Evolution*. 71(5):1339–1351.
- Saltz, J. B., Bell, A. M., Flint, J., Gomulkiewicz, R., Hughes, K. A., and Keagy, J. 2018. Why does the magnitude of genotype-by-environment interaction vary? *Ecol. Evol.* 8(12):6342–6353.
- Sgrò, C. M., and Hoffmann, A. A. 2004. Genetic correlations, tradeoffs and environmental variation. *Heredity*. 93(3):241.
- Siepielski, A. M., Gotanda, K. M., Morrissey, M. B., Diamond, S. E., DiBattista, J. D., and Carlson, S. M. 2013. The spatial patterns of directional phenotypic selection. *Ecol. Lett.* 16(11):1382-1392.
- Stearns, S. C., and Koella, J. C. 1986. Evolution of phenotypic plasticity in life-history traits: Predictions of reaction norms for age and size at maturity. *Evolution*. 40(5):893–913.
- Stearns, S. C. 1989. Trade-offs in life-history evolution. *Funct. Ecol.* 3(3):259-268.
- Stearns, S. C. 1992. *The evolution of life histories*. Oxford University Press Oxford.
- Swain, D. P., Sinclair, A. F., & Mark Hanson, J. 2007. Evolutionary response to size-selective mortality in an exploited fish population. *Proc. R. Soc. B*, 274(1613), 1015-1022.
- Teplitsky, C., Robinson, M. R., & Merilä, J. 2014. Evolutionary potential and constraints in wild populations. *Quantitative genetics in the wild*, 190-208.
- Thomson, C.E. and Hadfield, J.D. 2017. Measuring selection when parents and offspring interact. *Methods Ecol. Evol.* 8:678–687.
- Via, S., & Lande, R. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, 39(3), 505-522.
- Wilson, A. J., Pemberton, J. M., Pilkington, J. G., Coltman, D. W., Mifsud, D. V., Clutton-Brock, T. H., and Kruuk, L. B. 2006. Environmental coupling of selection and heritability limits evolution. *PLOS Biol.* 4(7):e216.
- Wilson, A. J., Pemberton, J. M., Pilkington, J. G., Clutton-Brock, T. H., and Kruuk, L. E. B. 2009. Trading offspring size for number in a variable environment: selection on reproductive investment in female Soay sheep. *J. Anim. Ecol.* 78(2):354-364.

- Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A., Kruuk, L.E. and Nussey, D.H 2010. An ecologist's guide to the animal model. *J. Anim. Ecol.* 79(1):13-26.
- Wolf, J.B. and Wade, M.J. 2001. On the assignment of fitness to parents and offspring: whose fitness is it and when does it matter? *J. Evol. Biol.* 14:347-356.
- Wood, C. W., and Brodie, E. D. 2015. Environmental effects on the structure of the G-matrix. *Evolution.* 69(11):2927–2940.
- Wu, Y., Furuya, S., Wang, Z., Nobles, J. E., Fletcher, J. M., & Lu, Q. 2021. GWAS on Birth Year Infant Mortality Rates Provides New Evidence of Recent Natural Selection. *MedRxiv*, 2021.06.08.21258559.



## Tables

**Table 4.1** Results of the bivariate random regression animal model fitting IMR distribution as a first-order polynomial function.  $\sigma_A^2$ : additive genetic variance;  $\sigma_M^2$ : shared familial environment variance;  $\sigma_E^2$ : residual variance;  $h^2$ : narrow-sense heritability. Random intercept and slope represented by  $a$  and  $b$ , respectively. Values are posterior modes for each parameter, along with 95% highest posterior density (HPD) intervals in brackets.

Trait	Term	$\sigma_A^2$	$\sigma_M^2$	$\sigma_E^2$	$h^2$
AFR	$a$	0.32 [0.20; 0.54]	0.27 [0.14; 0.34]	0.33 [0.21; 0.45]	0.39 [0.25; 0.56]
	$b$	0.18 [0.13; 0.30]	0.18 [0.11; 0.26]	-	0.28 [0.19; 0.41]
$w$	$a$	0.26 [0.17; 0.45]	0.20 [0.12; 0.28]	0.36 [0.27; 0.50]	0.31 [0.21; 0.49]
	$b$	0.16 [0.12; 0.29]	0.18 [0.12; 0.25]	-	0.25 [0.16; 0.36]

**Table 4.1** Results of the bivariate random regression animal model fitting IMR distribution as a first-order polynomial function.  $pd$  refers to the probability of direction, expressed here in percentage;  $ROPE$  refers to the proportion of probability that the 95% HPD interval that lies in the defined *Region of Practical Equivalence* based on three thresholds (5% - 10% - 25%; see Supplementary Material for details).

Parameter	$pd$	$ROPE$ (5%)	$ROPE$ (10%)	$ROPE$ (25%)
$\sigma_{A,a}^2(AFR)$	100%	0	0	0
$\sigma_{A,b}^2(AFR)$	100%	0	0	0
$\sigma_{A,a}^2(w)$	100%	0	0	0
$\sigma_{A,b}^2(w)$	100%	0	0	0

**Table 4.3** Results of the bivariate animal model fitting unstructured covariance matrices with the IMR group as a random categorical interaction.  $\sigma_A^2$ : additive genetic variance;  $\sigma_M^2$ : shared familial environment variance;  $\sigma_E^2$ : residual variance;  $h^2$ : narrow-sense heritability. Values are posterior modes for each parameter, along with 95% highest posterior density (HPD) intervals in brackets. IMR groups are represented numerically: 1 for Low IMR group, 2 for Middle IMR group, and 3 for High IMR group.

Trait	IMR group	$\sigma_A^2$	$\sigma_M^2$	$\sigma_E^2$	$h^2$
AFR	1	0.48 [0.27; 0.75]	0.41 [0.24; 0.58]	0.39 [0.21; 0.63]	0.40 [0.22; 0.53]
	2	0.43 [0.25; 0.60]	0.31 [0.20; 0.46]	0.35 [0.21; 0.49]	0.36 [0.25; 0.52]
	3	0.54 [0.31; 0.94]	0.44 [0.29; 0.82]	0.51 [0.23; 0.78]	0.38 [0.23; 0.56]
w	1	0.38 [0.24; 0.63]	0.44 [0.23; 0.59]	0.28 [0.16; 0.47]	0.35 [0.23; 0.52]
	2	0.39 [0.23; 0.55]	0.30 [0.19; 0.42]	0.30 [0.19; 0.44]	0.37 [0.25; 0.52]
	3	0.54 [0.23; 0.88]	0.46 [0.26; 0.72]	0.57 [0.32; 0.88]	0.31 [0.19; 0.52]

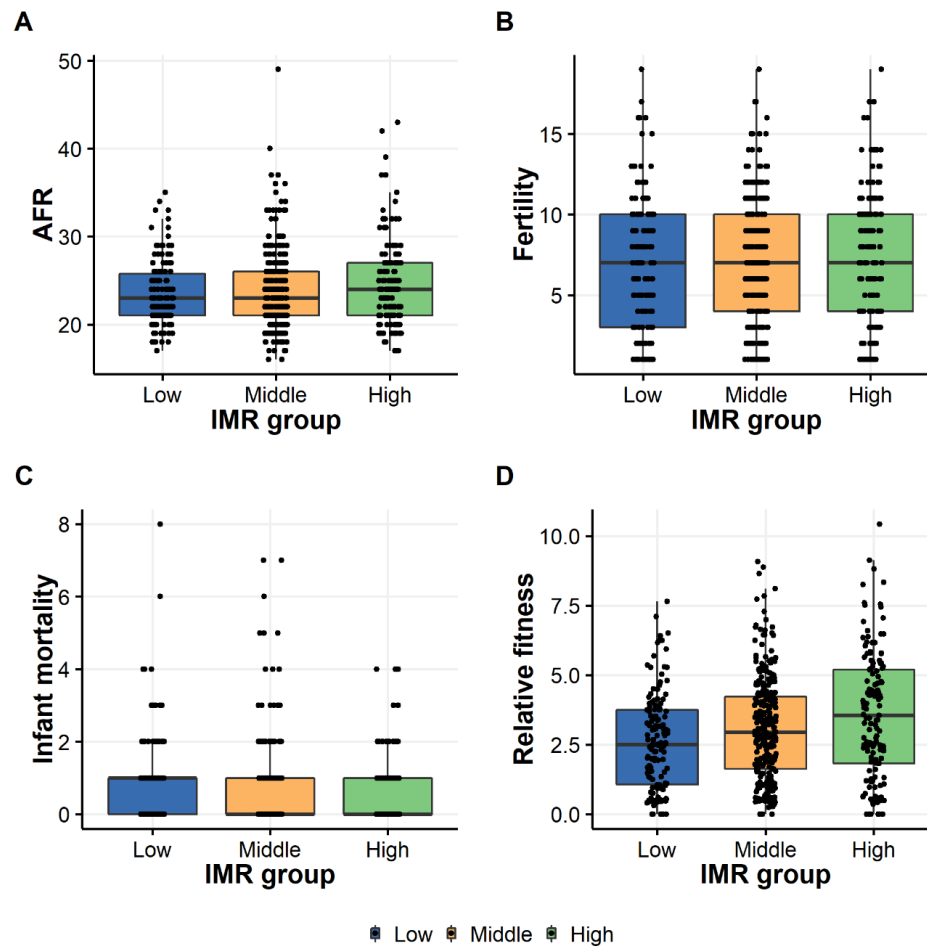
**Table 4.4** Bayesian posterior indices for environment-specific additive genetic variances  $\sigma_A^2$  and covariances  $\sigma_{AA}$ . *pd* refers to the probability of direction, expressed here in percentage; *ROPE* refers to the proportion of the 95% HPD interval that lies in the defined *Region of Practical Equivalence* based on three thresholds (5% - 10% - 25%; see Supplementary Material for details).

Parameter	<i>pd</i>	<i>ROPE</i> (5%)	<i>ROPE</i> (10%)	<i>ROPE</i> (25%)
$\sigma_{A_1(AFR)}^2$	100%	0	0	0
$\sigma_{A_2(AFR)}^2$	100%	0	0	0
$\sigma_{A_3(AFR)}^2$	100%	0	0	0
$\sigma_{A_1(w)}^2$	100%	0	0	0
$\sigma_{A_2(w)}^2$	100%	0	0	0
$\sigma_{A_3(w)}^2$	100%	0	0	0
$\sigma_{AA_1(AFR,w)}$	95%	0.013	0.025	0.067
$\sigma_{AA_2(AFR,w)}$	99%	0	0	0.004
$\sigma_{AA_3(AFR,w)}$	98%	0.004	0.005	0.015

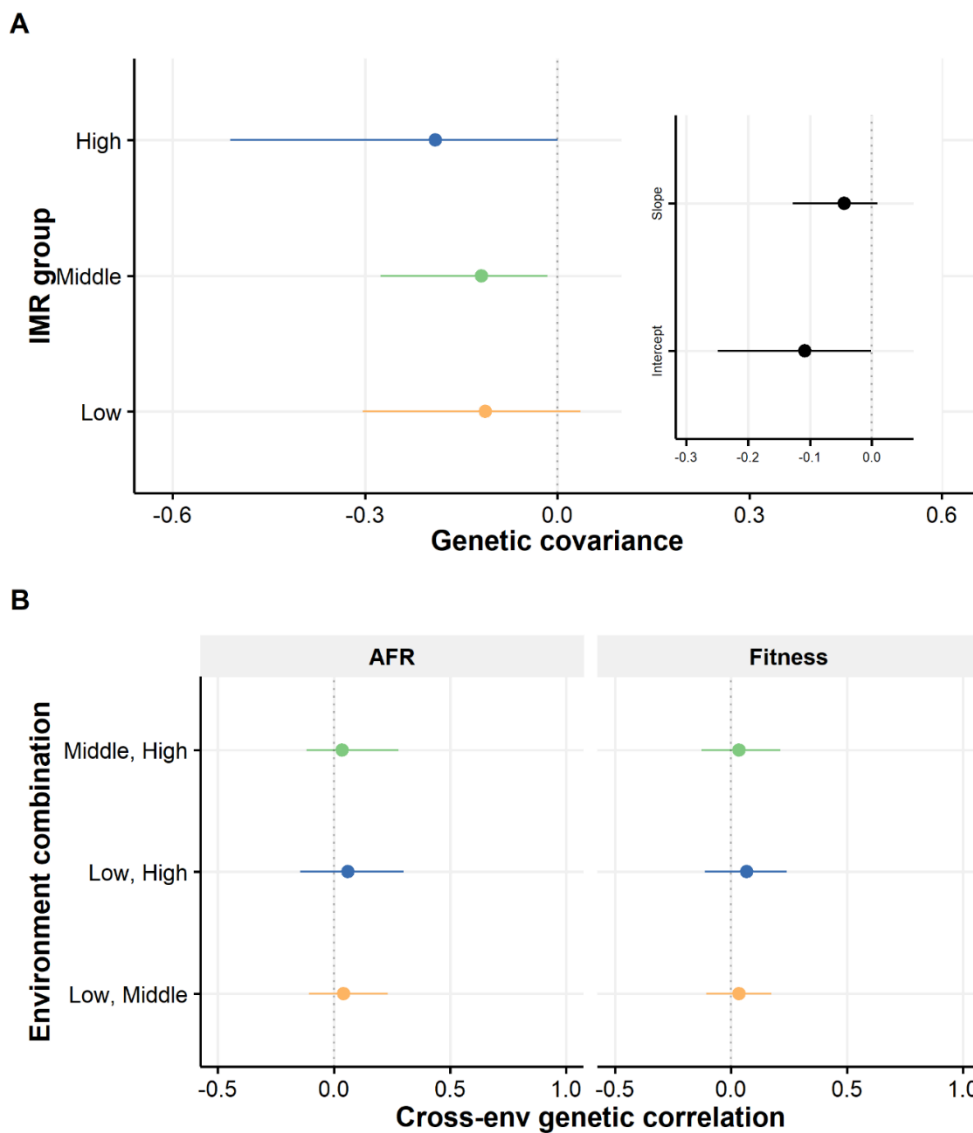
**Table 4.5** Covariance-correlation matrix (conditional on fixed effects, based on categorical bivariate model) for AFR and fitness. Variances are on shaded diagonals, covariances above and correlations below. Values are posterior modes for each parameter, along with 95% HPD intervals in brackets.

		<b>Trait</b>					
		<b>AFR</b>			<b><i>w</i></b>		
<b>Env.</b>		<b>Low IMR</b>	<b>Middle IMR</b>	<b>High IMR</b>	<b>Low IMR</b>	<b>Middle IMR</b>	<b>High IMR</b>
<b>Low IMR</b>		0.48 [0.27; 0.75]	0.04 [-0.11; 0.23]	0.06 [-0.15; 0.30]	-0.11 [-0.30; 0.04]	-0.03 [-0.21; 0.11]	-0.06 [-0.27; 0.17]
<b>Middle IMR</b>		0.15 [-0.21; 0.48]	0.43 [0.25; 0.60]	0.04 [-0.12; 0.28]	-0.02 [-0.18; 0.11]	-0.12 [-0.28; -0.02]	-0.03 [-0.24; 0.14]
<b>High IMR</b>		0.15 [-0.28; 0.45]	0.10 [-0.24; 0.48]	0.54 [0.31; 0.94]	-0.02 [-0.21; 0.17]	-0.03 [-0.23; 0.13]	-0.19 [-0.51; 0]
<b>Low IMR</b>		-0.27 [-0.58; 0.04]	-0.06 [-0.37; 0.28]	-0.04 [-0.39; 0.32]	0.38 [0.24; 0.63]	0.04 [-0.11; 0.17]	0.07 [-0.11; 0.24]
<b>Middle IMR</b>		-0.11 [-0.42; 0.28]	-0.42 [-0.59; -0.11]	-0.05 [-0.40; 0.28]	0.10 [-0.24; 0.40]	0.39 [0.23; 0.55]	0.03 [-0.13; 0.21]
<b>High IMR</b>		-0.14 [-0.47; 0.30]	-0.08 [-0.42; 0.30]	-0.46 [-0.68; -0.06]	0.19 [-0.24; 0.45]	0.12 [-0.25; 0.44]	0.54 [0.23; 0.88]

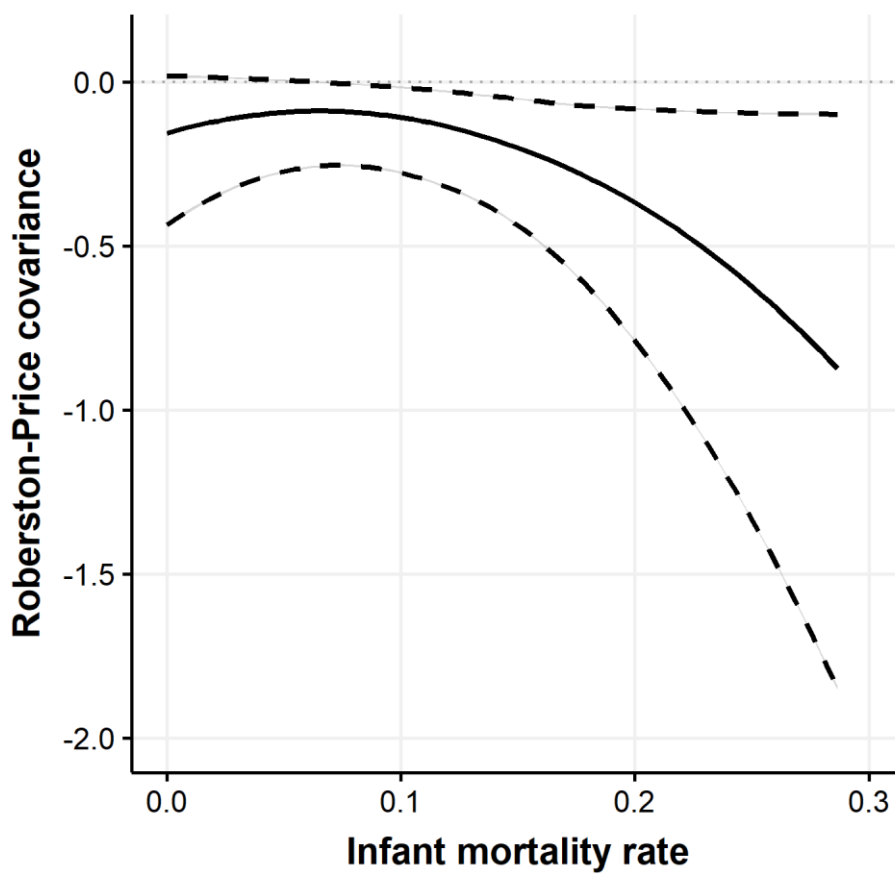
## Figures



**Figure 4.1** Boxplots of woman life-history traits within each of the three IMR groups (Low,  $n=141$ ; Middle,  $n=288$ ; High,  $n=143$ ). **A.** Age at first reproduction; **B.** Fertility (total number of children born); **C.** Infant mortality count (IMC: offspring died between 0-1 yo); **D.** Fitness. Each box shows the lower and upper quartiles of raw distributions and the horizontal line inside the box is the median. Black dots represent the raw data points. Colors associated with each IMR group are the same for all figures in this paper.



**Figure 4.2** **A.** Posterior estimates of environment-specific genetic additive covariance between AFR and relative fitness with an inset plot of the same additive covariance based on the RRAM model; **B.** Posterior estimates of cross-environmental genetic correlation for each of AFR and relative fitness. Each dot represents the posterior mode-estimates, along with the 95% HPD intervals.



**Figure 4.3** The expected Robertson-Price covariance if all individuals in the population were exposed to the same early-life environmental condition (i.e., same value of IMR). The solid line shows the posterior mode-estimate of the covariance, and the dashed lines are the 95% HPD intervals.

## CHAPITRE V

### CONCLUSION

La première édition de L'Origine des espèces de Charles Darwin a été publiée il y a 163 ans et, depuis, notre compréhension du processus d'évolution biologique et de ses mécanismes ne cesse de s'étendre. Étonnamment, Darwin n'a jamais mentionné le mot « évolution » dans cette première édition, préférant utiliser l'expression « descendance avec modification », qui reste une représentation valable de l'évolution biologique. Darwin et son contemporain Alfred Russel Wallace ont tous deux décrit le mécanisme de la sélection naturelle. Leur description, qui reste correcte à ce jour, a été faite à une époque où les notions de gènes et la connaissance de l'ADN/ARN étaient inexistantes. Ils ont tous deux insisté sur le fait que les caractères sont héréditaires, c'est-à-dire qu'ils sont transmis du parent à la progéniture, même s'ils ne comprenaient pas exactement le mécanisme biologique qui sous-tend cette hérédité. À cette époque, les découvertes de Gregor Mendel sur l'hérédité des caractères étaient peu connues et leur portée mal comprise et associée surtout à l'hybridation. En outre, Darwin et Wallace ont ajouté qu'il existe une variation entre les descendants d'un même parent et entre les individus d'une même population en général qui est héréditaire. Le dernier élément crucial de la sélection naturelle, inspiré des travaux de Thomas Malthus sur les populations humaines, est la notion de « lutte » pour la survie et la reproduction dans un contexte de ressources limitées. Ainsi, la sélection naturelle est un mécanisme qui peut conduire à l'évolution des phénotypes au sein des populations, à l'adaptation et, avec suffisamment de temps, à la spéciation. Selon les propres mots de Darwin :

*Comme il naît beaucoup plus d'individus de chaque espèce qu'il ne peut en survivre et que, par conséquent, la lutte pour l'existence se répète fréquemment, il s'ensuit que tout être, s'il varie, même légèrement, d'une manière qui lui soit profitable, dans les conditions complexes et parfois*

*variables de la vie, aura une meilleure chance de survivre et sera donc naturellement sélectionné. En vertu du puissant principe d'hérédité, toute variété sélectionnée aura tendance à propager sa nouvelle forme modifiée. (Darwin, De l'origine des espèces, 6<sup>e</sup> édition; traduite par Edmond Barrier 1876)*

## 5.1 Définir le changement évolutif

La définition la plus simple du changement évolutif est la suivante : « toute modification de la fréquence des allèles au sein d'une population d'une génération à l'autre ». Il est important de noter que cette définition concerne les changements microévolutifs et non l'émergence des espèces ou la macroévolution. D'aucuns ont eu tendance à restreindre le changement évolutif à ce dernier contexte, c'est-à-dire le changement lent et graduel des caractéristiques des populations menant à leur divergence et à l'émergence subséquente de nouvelles espèces. Cependant, des processus microévolutifs sous-tendent ces changements macroévolutifs qui conduisent à de nouvelles espèces. Bien que la définition ci-dessus semble lier l'évolution aux changements d'allèles, certains biologistes proposent des définitions plus larges de l'évolution. Prenons par exemple celle de John Endler :

*L'évolution peut être définie comme tout changement directionnel net ou tout changement cumulatif dans les caractéristiques des organismes ou des populations sur de nombreuses générations - en d'autres termes, la descendance avec modification... Elle inclut explicitement l'origine ainsi que la propagation des allèles, des variantes, des valeurs de trait ou des états de caractère. (Endler 1986 : 5)*

Par exemple, le sujet de l'évolution moléculaire traite les niveaux d'expression des ARNs et des protéines au niveau moléculaire comme des traits qui évoluent. D'autres présentent une vision non strictement génétique de l'évolution : par exemple, la définition de Charles Goodnight est centrée sur les changements du phénotype et le renouvellement continu des individus au sein des populations (Goodnight 2011).



Cependant, je note que chaque définition peut mener à des idées importantes dans un cadre analytique bien défini, et je ne favorise pas une définition plutôt qu'une autre.

## 5.2 La contribution des composantes non-additives

Les recherches présentées dans cette thèse s'inscrivent pleinement dans le domaine de la génétique évolutive, qui a énormément contribué à la compréhension de l'hérédité et des mécanismes impliqués dans l'évolution des traits quantitatifs et complexes. Notre étude de l'évolution des traits d'histoire de vie chez l'humain, en particulier l'âge à la première reproduction, considère principalement le phénotype comme l'objet d'intérêt basé sur les prémisses de la génétique quantitative. Le thème de l'unité de sélection est controversé et fait l'objet de discussions tant scientifiques que philosophiques (voir par exemple Franklin et Lewontin 1970 et Okasha 2006). Le débat sur l'unité de sélection ne date pas d'hier. Différents auteurs tendent à attribuer l'évolution de certains traits à l'action de la sélection naturelle à différents niveaux (Wilson, 1997). Par exemple, l'évolution de l'altruisme est-elle due à la sélection au niveau de l'individu ou de l'organisme ? G.C. Williams a été le premier à défendre une vision de l'évolution centrée sur le gène, arguant que le haut degré de permanence du gène en fait l'unité de sélection idéale. En revanche, d'éminents biologistes tels que Lewontin (1970) et Maynard-Smith et Szathmary (1997) ont défendu une perspective de sélection agissant à différents niveaux de la hiérarchie organisationnelle dans la nature, ce qu'on appelle la sélection multi-niveau.

L'étude de la contribution relative des différentes sources de variation à la variation phénotypique, y compris celles génétiques et non-génétiques, est au centre de l'approche analytique de la génétique quantitative. Notre enquête a utilisé l'analyse des diverses composantes de la variance qui constituent la variation phénotypique totale d'un trait, suivant le modèle général  $V_P = V_G + V_E$ , pouvant se décliner à différents niveaux de complexité. Ce cadre implique des simplifications qui supposent que les contributions de l'épistasie (ou interactions entre gènes) et de la dominance sont négligeables lorsque nous considérons les traits polygéniques. Ceci est vrai dans le contexte de la réponse évolutive à la sélection qui dépend de la variance génétique

additive. Ceci découle du fait que la composante additive contribue à la majeure partie de la variance génétique, même lorsque les interactions de dominance et épistatiques sont incluses. Mais l'importance de la dominance dans la réponse évolutive reste mal connue dû au manque d'études empiriques dans les populations naturelles.

Le principal problème est que l'absence de la variance de dominance dans les modèles QG peut gonfler l'estimation de la variance additive et l'héritabilité (Class et Brommer 2020). Wolak et Reid (2014) ont compilé des estimations de la variance de dominance basées sur des études sur l'élevage animal et végétal, sur des expériences de laboratoire et sur des données humaines, et ont conclu que ce type de variance non-additive contribue substantiellement à la variation génétique et phénotypique d'une variété de traits (dans les traits d'histoire de vie des animaux non-domestiqués, la fraction de la variance génétique expliquée par la dominance dépasse 60% en moyenne). De plus, les auteurs précisent que la contribution de la variance de dominance sera plus élevée dans les populations qui présentent une consanguinité élevée. Roff et Emerson (2006) ont également démontré, en utilisant des résultats publiés d'études expérimentales, que le rapport entre la composante de variance non additive et la composante additive est élevé pour les traits d'histoire de vie.

Dans le contexte de nos études, ces deux points sont extrêmement pertinents parce que i) nous étudions la réponse évolutive des traits d'histoire de vie, c'est-à-dire l'âge à la première reproduction ; ii) les populations que nous étudions présentent des niveaux de consanguinité élevés (Bouchard et De Braekeleer 1991 ; Boisvert et Mayer 1994). Dans les cas de l'île aux Coudres et de Charlevoix, nous ne pouvons pas exclure les contributions à la variance phénotypique de la dominance. L'emphase analytique mise sur l'estimation de la variance génétique additive ne doit pas se traduire par l'exclusion des autres sources de variation phénotypique. Au contraire, aucune estimation de la variance génétique additive ne peut être exacte et précise, quelle que soit la puissance statistique, si d'autres sources de variation génétique contribuent à la variance phénotypique et ne sont pas prises en compte par le modèle. Les seuls cas où ces sources de variation peuvent être parfois ignorées sont ceux des plans d'élevage méticuleux qui contrôlent l'environnement de tous les individus apparentés. Nos résultats du chapitre III constituent une démonstration de la manière dont l'absence de modélisation des

composantes connues de la variance phénotypique, comme l'environnement familial partagé, peut conduire à une surestimation de la variance génétique additive et par extension de l'héritabilité et de certaines corrélations génétiques.

Il est essentiel d'inclure ces facteurs non-génétiques confondants et de les séparer explicitement des effets génétiques. En outre, pour minimiser le biais dans les estimations génétiques pertinentes pour l'évolution, il est essentiel de déterminer dans quelle mesure les effets de dominance et d'épistasie impactent l'ampleur de la variance génétique additive dans les populations canadiennes-françaises. Avec les avancées méthodologiques des modèles de génétique quantitative et la disponibilité de pedigrees à grande échelle, l'estimation des sources génétiques non additives telles que la dominance est faisable pour mieux comprendre l'évolution des traits d'histoire de vie.

### **5.3 Sur la mesure de la sélection naturelle dans la nature**

L'estimation de la sélection naturelle dans cette thèse est basée sur une perspective phénotypique. Dans la génétique quantitative et la théorie d'histoire de vie, l'hypothèse est qu'il existe une mesure phénotypique de fitness maximisée par l'action de la sélection naturelle. Avec la formulation de la synthèse moderne par ses architectes (Julian Huxley en 1942, Ernst Mayr en 1959, G. Ledyard Stebbins en 1966 et Theodosius Dobzhansky en 1974), la sélection naturelle s'est assuré un rôle central dans le façonnement de l'évolution des populations dans la nature. Cependant, aucune tentative empirique de mesurer la sélection naturelle dans la nature n'a eu lieu par la suite, et son existence et son rôle ont toujours été déduits sur la base des caractéristiques de l'organisme étudié (voir Antonovics 1987 qui a été le premier à formuler cette critique à l'encontre des partisans de la synthèse moderne). Nous appliquons ici les méthodes de régression développées par Lande et Arnold (1983) et Arnold et Wade (1984a, 1984b) qui ont ouvert la voie à la mesure empirique de la sélection dans les populations naturelles.

Les travaux présentés dans cette thèse emploient les méthodes de régression pour démontrer que l'AFR dans plusieurs sous-populations canadiennes-françaises était soumis à une forte sélection pour des valeurs plus précoces. Cette mesure de sélection

est liée à la covariance phénotypique entre l'AFR et le fitness relatif. L'hypothèse essentielle ici est que cette covariance est en partie génétique. Néanmoins, l'environnement peut induire une covariance entre l'AFR et le fitness relatif, ce qui conduit à des interprétations erronées si la cause d'un changement phénotypique est attribuée à tort à une cause génétique. Cette confusion pourrait expliquer la stase évolutive de l'AFR détectée dans les populations étudiées dans le chapitre II. La sélection sur l'AFR pourrait être due dans une large mesure à une covariance environnementale entre le trait et le fitness relatif. Nos analyses démontrent que la covariance génétique additive entre AFR et le fitness dans les trois populations étudiées (c'est-à-dire l'identité Robertson-Price ; Robertson 1966, 1968 ; Price 1970) était de faible ampleur.

Un élément important dans la mesure de la sélection naturelle est le choix de la mesure du fitness individuel. En tant que concept, le fitness peut avoir différentes définitions, significations et façons de le mesurer. En général, le fitness d'un organisme comporte deux composantes principales : la survie et le succès reproductif. La sélection naturelle elle-même peut être séparée en sélection de viabilité (variabilité de la survie) et sélection de fécondité (variabilité de la reproduction). Selon le contexte et la conception d'une étude donnée, et la limitation des jeux de données, il est justifié de se concentrer sur l'une d'entre elles comme indicateur du fitness global, par exemple les taux de survie ou la fécondité. Chaque mesure de fitness a ses avantages et ses inconvénients. Le LRS (lifetime reproductive success) est une mesure du fitness individuel absolue fréquemment utilisée dans les études écologiques et anthropologiques, ainsi que dans les études de génétique quantitative (p. ex. Kruuk *et al.* 2000; Milot *et al.* 2011). Cependant, cette métrique présente des inconvénients et peut être inappropriée pour l'inférence de sélection dans des populations avec des dynamiques telles que les populations humaines. Dans le chapitre II et IV, nous avons plutôt opté pour la valeur reproductive à la naissance, basé sur Fisher (1958) et Lenski et Service (1982), qui est sensible au taux de croissance et intègre le moment de la reproduction. En utilisant cette métrique de fitness relative, dans le chapitre IV, nous avons confirmé les résultats publiés dans Milot *et al.* (2011), y compris l'ampleur de la sélection agissant sur l'AFR, une estimation plus

conservatrice de l'héritabilité de fitness, et enfin, la tendance génétique détectée en réponse à la sélection.

Le choix de la mesure de fitness a un impact sur la mesure de la sélection naturelle. Deux mesures différentes peuvent être congruentes dans la direction et même l'ampleur de la sélection linéaire, mais elles peuvent donner des résultats différents en termes de sélection non linéaire. Des résultats non publiés ici comparant le LRS et la valeur reproductive à la naissance ont montré que seule cette dernière était capable de détecter une composante non linéaire de la sélection sur AFR. Il s'agit d'une question pertinente à examiner à la fois dans le cadre de recherches futures et dans le cadre de la révision d'études antérieures qui auraient pu négliger une composante non linéaire de la sélection en raison du choix de la mesure de fitness. Avec les données appropriées, il existe d'autres façons possibles d'estimer le fitness individuel, et nous encourageons les travaux futurs à intégrer plus d'une mesure de fitness (par exemple, l'estimation de la valeur reproductive à l'aide de pedigrees proposée par Hunter *et al.* 2019).

Nous ne proposons pas cette suggestion dans le but de couvrir les éventuelles faiblesses de l'une ou l'autre mesure, mais dans le but d'obtenir des informations supplémentaires et de détecter des incongruités informatives pour l'analyse globale. Par exemple, des limitations et des complications apparaissent dans le cas de la valeur reproductive à la naissance que LRS ne possède pas : l'impact des événements démographiques extrêmes. Au chapitre II, la population de Charlevoix a connu un important événement migratoire au milieu du 19e siècle vers le nord, plus précisément vers la rivière Saguenay et le lac St-Jean. Cet événement migratoire a entraîné une contribution relativement disproportionnée aux générations suivantes des individus nés et mariés dans Charlevoix. Par conséquent, ces individus avaient des valeurs reproductives à la naissance gonflées, et à cause de cette inflation, l'analyse des périodes avant et après l'événement migratoire était pratiquement impossible en utilisant cette mesure de fitness. Le LRS, en revanche, n'avait pas cette limitation parce qu'il ne suit que le nombre d'enfants qui ont survécu jusqu'à la maturation. Par conséquent, le choix de la mesure de fitness dans les études de génétique quantitative ne doit pas se concentrer sur celle qui conduit à des résultats significatifs ou aux résultats souhaités. Il doit plutôt refléter la prudence et la transparence de l'investigateur qui tente de décrire la

relation possible entre le fitness et un caractère donné, tout en tenant compte des différents pièges que présente chaque mesure.

#### **5.4 Les contraintes sur les changements microévolutifs et la réponse à la sélection**

Une notion importante à retenir est que le potentiel évolutif ne se traduit pas toujours par un changement évolutif. Dans le cas des populations humaines, les preuves du potentiel évolutif de l'AFR ont été bien documentées (Stearns *et al.* 2010). La direction de la sélection est consistante, les âges plus précoces de la première reproduction sont associés à un meilleur fitness, et l'ampleur de la sélection peut varier entre les populations. Dans notre cas, nous avons présenté des preuves supplémentaires de cette tendance au chapitre II et nous avons également révélé que le potentiel évolutif de l'AFR est omniprésent dans les populations humaines de différentes périodes. Malgré cela, la stase évolutive reste le résultat observé le plus probable. Cette stase peut être attribuée et expliquée par de multiples possibilités, ce qui souligne l'importance d'étudier les facteurs qui limitent la réponse évolutive dans la nature. Les agences peuvent intégrer ces informations dans les stratégies impliquées dans la réponse des espèces au changement climatique, traiter les espèces invasives, et proposer de nouvelles façons de modéliser et de maintenir la conservation des populations naturelles. Dans notre analyse de la base génétique quantitative de l'AFR, nous avons détecté un biais environnemental dans la sélection phénotypique sur AFR. Cela signifie que la plus grande partie de la sélection sur AFR (c'est-à-dire la covariance avec fitness) est déterminée par des effets environnementaux et non génétiques. Alors, tout changement dans l'AFR moyen ne correspondrait pas à un changement génétique sous-jacent (ou à une réponse à la sélection), mais à un changement phénotypique associé à des facteurs environnementaux.

Les interactions entre le génotype et l'environnement,  $G \times E$ s, sont également un facteur qui a un impact sur le potentiel évolutif, et nous en avons apporté la preuve dans une population humaine pour un trait non-répété tel que l'AFR. Le fait que  $G \times E$ s puisse affecter le potentiel évolutif de l'AFR pour qu'il soit plus élevé dans des conditions difficiles peut expliquer la tendance actuelle d'un AFR plus tardif observé dans le

monde entier. L'amélioration des conditions environnementales à la fois au début et à la fin de la vie signifie que les G×Es pourraient avoir réduit la capacité de l'AFR à répondre à la sélection en affectant sa génétique sous-jacente et en découplant éventuellement la composante génétique dans sa relation avec le fitness. Cependant, il s'agit d'une spéculation et nous ne pouvons pas généraliser les preuves d'une seule petite population isolée à d'autres populations qui peuvent varier de nombreuses façons.

Une affirmation que nous pouvons généraliser sans risque est que les G×Es peuvent modifier le degré auquel les associations et les compromis sont génétiquement déterminés. Dans les conditions difficiles du début de la vie, la relation entre l'AFR précoce et un fitness relatif plus élevée était plus prononcée génétiquement. Dans les chapitres II et IV, l'association entre l'AFR et le fitness relatif était généralement négative en raison de facteurs environnementaux (ou culturels). Lorsque les conditions de développement étaient précaires, cette relation négative présentait un degré plus élevé de contrôle génétique. Cette observation donne plus de crédibilité à la possibilité que l'amélioration séculaire des conditions au début de la vie au XX<sup>e</sup> siècle ait limité la réponse à la sélection de l'AFR, étant donné que la relation génétique entre ce dernier et le fitness relatif est beaucoup plus faible. Avec l'augmentation de la disponibilité des études longitudinales et multigénérationnelles, l'étude du rôle des interactions entre le génotype et l'environnement représente la meilleure voie possible pour comprendre les limites de la réponse à la sélection dans les traits d'histoire de vie chez les populations humaines. Cela peut aller jusqu'à trouver des moyens d'intégrer ces facteurs dans la modélisation démographique et les prévisions des taux de fécondité qui jouent un rôle dans la santé publique et personnalisée.

## **5.5 Effets indirects sur l'évolution et la variation du rythme reproductif humain**

Les effets maternels représentent l'influence du trait phénotypique d'une mère sur l'expression phénotypique de sa progéniture à la génération suivante. Cet effet se produit par la « construction » d'un environnement maternel qui peut influencer directement l'expression phénotypique de la progéniture, alors que l'effet maternel est indirect. Nous avons inclus ce type d'influence environnementale dans nos analyses génétiques

quantitatives, étant donné qu'elle contribue à la variation globale des traits d'histoire de vie, en particulier chez les humains. Le raisonnement était le suivant : pour déduire et prédire la trajectoire évolutive d'un phénotype donné, nous devons quantifier toutes les sources de variation qui contribuent à la variation phénotypique globale. Cette nécessité implique de tenir compte des effets confondants tels que les effets maternels. Mais en plus de cet effet maternel, il peut y avoir des effets génétiques indirects, ou IGEs, qui impactent la variation et l'évolution des traits d'histoire de vie.

L'environnement social est une composante de l'environnement général d'un individu qui peut être composé des phénotypes d'autres individus et de leurs interactions (directes ou indirectes). Le phénotype d'un individu, qui est déterminé en partie par son génotype, peut affecter par des interactions sociales le phénotype des autres et influencer leur fitness. Cet effet est appelé un IGE (Moore *et al.* 1997 ; McAdam *et al.* 2014). En plus de contribuer à une variation héritable supplémentaire, les IGEs peuvent influencer l'évolution phénotypique dans la nature en fonction de la force de l'effet, de la sélection sur le trait focal et de la structure de parenté entre les individus en interaction (Wolf *et al.* 1998). Bien que des preuves récentes suggèrent que les IGEs peuvent être répandus (Bailey *et al.* 2018), leur étude est limitée par la nécessité de disposer d'ensembles de données à long terme avec des généalogies connues. L'ensemble de données historiques canadiennes-françaises présente une telle structure de données, avec des généalogies profondes et des informations sur l'histoire de vie de nombreux individus, ce qui permet d'étudier la contribution des IGEs à la variation génétique et à la dynamique évolutive.

En ce qui concerne l'étude de l'influence d'autres individus sur l'expression des traits d'histoire de vie, nous pouvons dépasser le cadre de l'interaction entre la mère et la progéniture et explorer les effets découlant des interactions avec d'autres membres sociaux dans chaque population donnée (Frank 2007). Au cours de sa vie, un individu interagit pendant de longues périodes avec un nombre limité de membres sociaux dans son voisinage. Il est probable que les gènes des autres membres puissent influencer, par leurs phénotypes, les traits d'histoire de vie d'un individu focal. La variation de l'AFR a une base génétique héritable ; néanmoins, sa détermination chez un individu focal dépend de l'AFR de son compagnon social et de la date du mariage. En d'autres termes,



la variation du moment de la première reproduction et, éventuellement, du fitness, pourrait être affectée par les gènes du partenaire social.

Il a été démontré que la proximité et la présence d'une grand-mère peuvent affecter le fitness et la fécondité de la fille (Engelhardt *et al.* 2019). Ce résultat indique que le phénotype des proches apparentés peut présenter des IGEs potentiels qui affectent le trait d'histoire de vie de l'individu focal. Nous pouvons utiliser des modèles génétiques quantitatifs basés sur une approche de partitionnement de la variance pour estimer les effets génétiques indirects des phénotypes des partenaires sociaux sur l'expression d'un trait focal (Bijma *et al.* 2007a, 2007b). De plus, l'ensemble de données canadiennes-françaises offre des informations sur la localisation spatiale des individus. Il est possible d'incorporer ces informations dans le modèle génétique quantitatif pour mettre à l'échelle les effets génétiques indirects des partenaires sociaux et des proches sur leur distance par rapport à l'individu focal (Fisher *et al.* 2019). Le chapitre III a montré que les ensembles de données canadiennes-françaises affichent une puissance élevée dans la détection d'un petit effet génétique sous-jacent à un trait donné, et nous supposons qu'il existe un potentiel statistique dans l'utilisation de ces ensembles de données pour détecter les IGEs et déterminer leur contribution relative à la variation génétique et à la dynamique évolutive des traits d'histoire de vie.

## **5.6 Le risque de mortalité prédit le début de la reproduction chez les humains**

Il est important de noter que les données et les résultats présentés dans Stearns *et al.* (2010) sur le potentiel évolutif de l'AFR sont biaisés en faveur des populations occidentales pour des raisons essentiellement liées à la disponibilité et à la qualité des données. Néanmoins, des études bioanthropologiques telles que Walker *et al.* (2006), qui ont inclus 20 sociétés à petite échelle (par exemple, Ache, Hiwi, Tsimane), présentant toutes un régime de fécondité naturelle, dans leur ensemble de données analysées, ont constaté que le début de la reproduction est étroitement lié au fitness, mais dépend fortement du contexte autour des conditions environnementales, telles que la disponibilité de l'énergie et la mortalité juvénile-adulte. Même les études sur les populations contemporaines (occidentales) ont révélé la prévalence d'une stratégie

d'histoire de vie « rapide », c'est-à-dire une ménarche et une reproduction précoce, dans des environnements caractérisés par une morbidité/mortalité élevée (par exemple, Chisholm *et al.* 2005 ; Nettle *et al.* 2011).

Cependant, ce schéma ne se généralise pas à d'autres populations humaines contemporaine ou historiques. Dans l'IAC, nous avons constaté que l'AFR moyen des femmes mariées avait tendance à se produire plus tard avec un taux de mortalité infantile accru (mesuré au niveau de la population) vécu au début de la vie (résultats non publiés). De plus, Placek et Quinlan (2012) suggèrent que les risques de mortalité précoces et actuels peuvent affecter la reproduction et son rythme. Par conséquent, les risques de mortalité à un stade avancé de la vie devraient être intégrés lorsqu'on essaie de comprendre comment les conditions au début de la vie ont un impact sur le rythme de la reproduction et si cet impact peut être modifié plus tard dans la vie (cela pourrait constituer un projet de recherche décent pour les étudiants en master avec les jeux de données étudié ici).

Il est intéressant de noter que Störmer et Lummaa (2014) montrent que le niveau auquel l'expérience du risque de mortalité se produit a de l'importance pour façonner la stratégie d'histoire de vie. Ils ont comparé les effets des expériences de mortalité au niveau individuel par rapport à celles au niveau familial sur le moment de reproduction et ont constaté que ce trait s'ajustait en fonction de l'influence de la mortalité au niveau familial, tant pour les femelles que pour les mâles. En d'autres termes, l'appartenance à une famille est un meilleur prédicteur de stratégies d'histoire de vie plus rapides que les expériences individuelles, du moins dans les trois sociétés historiques analysées dans leur étude (qui incluent la population historique canadienne-française). Ce résultat est directement lié aux conclusions de la génétique quantitative, ici et dans d'autres études, sur le rôle de l'environnement familial dans le façonnement de la variation du moment de reproduction.

Dans les analyses présentées dans cette thèse, pour capturer l'effet environnemental maternel, nous avons inclus l'identité de la mère dans les modèles sans connaissance préalable des voies médiatrices possibles de cet effet. Selon les résultats de Störmer et Lummaa (2014), l'une des voies pourrait être l'exposition à la mortalité en tant que membre de la même famille, qui peut façonner le rythme individuel de

reproduction. Néanmoins, sur la base de nos résultats au chapitre IV, nous avons déduit qu'une mère pouvait exercer des influences différentes sur le phénotype de ses filles (dans notre cas, le rythme de reproduction) en fonction de l'exposition différentielle à la mortalité au niveau individuel. Ce résultat pourrait être lié à celui de Störmer et Lummaa (2014) dans la mesure où il capture un effet au niveau familial, ou il pourrait indiquer un autre niveau de complexité dans la façon dont les mères influencent les phénotypes de leur progéniture, un niveau basé sur des signaux contextuelles provenant de l'environnement externe (par exemple, les risques de mortalité infantile).

La théorie d'histoire de vie prédit un lien fort entre le taux de mortalité et le moment de la reproduction, qu'il provienne d'une perspective d'optimalité ou de plasticité développementale. Bien que cette prédiction soit factuelle dans de multiples espèces et populations, chez les humains au moins, la direction de cette relation n'est pas omniprésente comme on le pensait auparavant (voir Griskevicius *et al.* 2011). Dans les ensembles de données analysés ici, nous avons détecté la relation négative entre le moment de reproduction et la mortalité infantile attendue par l'effet d'une fécondité plus élevée. Mais en termes d'effets moyens, nos résultats ne correspondent pas au résultat souvent généralisé d'une stratégie d'histoire de vie plus rapide en réponse à des conditions environnementales difficiles. En conséquence, nous recommandons qu'un cadre de catégorisation plus nuancé remplace le continuum "rapide-lent" des stratégies d'histoire de vie, au moins dans le cas des populations humaines. Ce cadre de catégorisation devrait : i) s'appuyer sur les résultats de nombreuses populations humaines de différentes périodes et origines, ii) prendre en compte l'expression différentielle des compromis et les facteurs qui l'influencent, et iii) reconnaître que différentes combinaisons contextuelles de taux de mortalité, d'état nutritionnel et de disponibilité énergétique peuvent conduire à l'expression de différentes stratégies d'histoire de vie (voir Tableau 5 dans Walker *et al.* 2006).

## 5.7 Pas seulement par les gènes

Prenons l'exemple suivant où on observe une augmentation de la taille moyenne dans les populations humaines, nous pouvons facilement conclure que les grandes tailles

sont sélectionnées positivement si les individus plus grands ont tendance à avoir plus d'enfants en moyenne et conclure qu'une réponse à la sélection a eu lieu. Cependant, le changement séculaire de la taille induit par les modifications du régime alimentaire et des conditions environnementales est à l'origine de cette relation. Cette observation introduit un autre processus, pour lequel les humains sont très doués, à savoir la construction d'une niche, définie comme un changement induit par l'organisme dans l'environnement immédiat - et éventuellement dans les environnements futurs (Odling-Smee *et al.* 2013). En général, la construction de niches peut créer des boucles de rétroaction évolutive et peut-être modifier les pressions sélectives. Les populations ou groupes humains induisent probablement des changements dans leur environnement qui conduisent à des associations imprévisibles entre les traits d'histoire de vie et à des changements possibles dans les compromis qui en découlent. La prise en compte de ce fait peut accroître la complexité de l'étude de l'évolution d'histoire de vie humaine, mais, en retour, nous obtenons une résolution accrue de la façon dont les différents processus évolutifs, y compris ceux qui ont été sous-estimés jusqu'à présent dans leur contribution à l'évolution humaine, interagissent et ont un impact sur l'histoire de vie humaine.

L'hérédité culturelle est un mécanisme qui a un impact discernable sur la diversité phénotypique et qui joue un rôle dans les populations humaines. Dans le chapitre IV, nous supposons que les effets culturels et socio-économiques sont minimes au sein de l'IAC parce que sa composition quant aux ces effets est homogène. Cependant, une complication des études d'analyse de variance sur les populations humaines est que les antécédents culturels et socio-économiques peuvent varier considérablement et doivent être mesurés et inclus dans l'analyse. Contrairement aux résultats présentés dans Milot *et al.* (2011), nous avons révélé dans notre analyse que les effets maternels ont un impact sur la variation de l'AFR dans la population de l'IAC. Le même résultat s'applique aux autres populations étudiées au chapitre II, ce qui signifie que les sœurs partageant une mère (ou un environnement maternel) ont très probablement un AFR similaire en raison de cet effet, en plus de la similarité phénotypique due au partage de parties de leurs génomes.

La question qui en découle est de savoir par quels mécanismes ou processus l'effet maternel influe sur le phénotype de leur progéniture. Une réponse probable est qu'il s'agit d'une composante culturelle héritée ou transmise liée à l'expression du moment attendu de la reproduction, qui pourrait être particulier à cette population à cette période. Le parent, ou la mère dans ce cas, utilise ce dispositif culturel pour "construire" et renforcer un environnement pour ses enfants au cours de leur développement, ce qui peut avoir un impact, dans une certaine mesure, sur l'expression de leur AFR à l'âge adulte. Nous pouvons étudier cette hypothèse en utilisant un cadre d'interaction gène-culture pour comprendre comment des éléments hérités de natures différentes peuvent interagir et avoir un impact sur l'expression et la variation d'un trait donné tel que l'AFR. Par exemple, dans le contexte de l'étude de l'hérédité comportemental, Danchin *et al.* (2013) proposent un cadre basé sur la génétique quantitative et ses modèles pour démêler les contributions des sources génétiques et non génétiques (comme la culture) de l'hérédité. Il est vrai qu'il est difficile de réaliser des plans expérimentaux (par exemple, l'adoption croisée) sur des populations humaines, mais avec le bon type de données et des suppositions raisonnables, nous pouvons appliquer ces méthodes aux pedigrees humains comme les généalogiques canadiennes-françaises.

## **5.8 L'évolution de l'histoire de vie humaine et la culture**

L'étude de l'histoire de vie et sa théorie impliquent de multiples disciplines : écologie, endocrinologie, physiologie, évolution, génétique et psychologie. La principale différence entre ces domaines réside dans leur champ d'investigation général, qui implique la recherche d'explications immédiates ou ultimes pour les modèles d'histoire de vie dans la nature (ce qui se traduit grossièrement par la différence entre répondre aux questions « comment » et « pourquoi »). L'évolution de l'histoire de vie implique l'étude du rôle de la sélection naturelle dans les patrons observés de variation de l'histoire de vie entre et au sein des espèces, et la raison pour laquelle ces résultats évolutifs sont adaptatifs. Cette étude ne se limite toutefois pas à la sélection naturelle. Le rôle d'autres forces évolutives (par exemple, la dérive aléatoire, la migration, l'accouplement non assorti) doit être examiné et élucidé pour comprendre les patrons

observés, étant donné que les patrons d'histoire de vie peuvent être neutres sur le plan adaptatif ou même mal-adaptés<sup>1</sup>.

L'une de ces forces est la culture, qui peut influencer l'expression de l'histoire de vie en ayant un impact sur les résultats comportementaux et les effets cognitifs. Cet élément est pertinent dans le cas des populations humaines où la culture joue un rôle dans nos interactions sociales, en termes de coopération et d'adversité, et cela a des implications importantes pour l'évolution de l'histoire de vie humaine, à la fois dans les temps anciens et contemporaines, peut-être par le biais des effets de la sélection de groupe culturelle (Muehlenbein et Flinn 2011). Dans les populations humaines, la culture joue un rôle au cours du développement précoce, lorsqu'elle peut avoir un impact maximal sur l'expression des traits de l'histoire de vie à l'âge adulte (Stearns *et al.* 2010). Le point essentiel est que la culture peut moduler les pressions sélectives. Ce fait peut conduire au piège de conclure qu'un patron d'histoire de vie observé est dû à la sélection naturelle et qu'il est adaptatif alors qu'en réalité, il est dû à des forces culturelles et historiques dont l'influence n'est pas nécessairement adaptative. Par conséquent, nous devons souligner que les patrons de variation de l'histoire de vie humaine, en particulier à l'époque contemporaine, sont tout aussi susceptibles d'être causés par des changements dans la culture et les traditions que par la génétique. La question qui se pose est de savoir quel impact cela a sur notre compréhension de la nature humaine<sup>2</sup>.

Par exemple, chez l'IAC, le changement phénotypique de 4 ans de l'AFR vers des âges plus précoces n'est que partiellement expliqué par le changement génétique en réponse à la sélection. Cette observation signifie que des forces principalement

---

<sup>1</sup> Bien que la théorie de l'histoire de vie trouve ses racines dans Cole (1954), Stearns (1989) et Roff (2002) sont crédités de l'exposé formel des prémisses et cadres fondamentaux pour l'étude de l'évolution de l'histoire de vie, et nous recommandons aux lecteurs ces deux volumes pour la bonne compréhension de l'évolution de l'histoire de vie et de ses fondements théoriques et méthodologiques.

<sup>2</sup> Cela concerne également la critique du programme adaptationniste et le déluge d'histoires « justes » pour expliquer l'évolution des propriétés et des traits des organismes. Pour une solide critique de cette question, nous recommandons l'article fondateur de Gould et Lewontin (1979) et, pour un examen plus récent, voir Pigliucci et Kaplan (2000).

environnementales sont à l'origine, et peut-être abordées, de la tendance à la reproduction précoce dans cette population. Les chercheurs doivent en tenir compte dans toute enquête et investigation sur l'histoire de vie humaine, que ce soit entre ou au sein des populations (Stearns et Rodriguez 2020). C'est le cas non seulement parce que les études sur l'évolution de l'histoire de vie humaine peuvent avoir des implications éthiques, mais aussi parce que cela peut conduire à des circonvolutions sur les traits qui sont dus à l'action de la sélection naturelle alors qu'ils sont dus à d'autres causes (comme la culture ou la dérive aléatoire), ou à des interactions entre ces dernières et la sélection naturelle - c'est la cause la plus probable. Pour cela, j'invite ceux qui s'intéressent à l'évolution de l'histoire de la vie humaine (et du comportement humain, en général) à adopter le cadre des "hypothèses de travail multiples", proposé par Chamberlin (1890).

Dans plusieurs des populations canadiennes-françaises préindustrielles étudiées, l'intervalle entre le mariage et la première reproduction est un déterminant fort de l'AFR. Les deux variables sont corrélées positivement dans les deux cas. Ce résultat signifie que dans les couples qui avaient tendance à commencer à concevoir directement après le mariage, la femme était plus susceptible d'avoir un AFR précoce et, par conséquent, d'avoir une meilleure fitness grâce à une fécondité élevée. Les populations d'IAC et Charlevoix avaient des périodes d'expansion importante, les terres étaient divisées et vendues ou offertes aux familles de colons arrivant dans la région. Les parents aidaient leurs fils à acquérir rapidement des terres en se mariant tôt pour fonder leur propre famille. Ce schéma socioculturel pourrait expliquer les courts intervalles entre le mariage et la première reproduction observés dans ces populations. D'autres éléments sociétaux et culturels qui offrent une explication similaire sont liés à l'aspect patriarcal de l'Église catholique romaine, où le rôle principal des femmes dans la société était de concevoir et d'élever les enfants dans la foi (et dans la population croissante des colons). Cette forme religieuse de planification de la famille et de la fécondité, à l'époque, est évidente si l'on considère que la fécondité moyenne d'une femme, qui ne s'est mariée qu'une seule fois, était de sept enfants (avec une fourchette allant de 1 à 19 enfants). Cependant, nous insistons sur le fait qu'il ne s'agit pas d'une inférence causale mais plutôt d'une reconnaissance du rôle que ces éléments sociétaux, considérés comme

faisant partie de l'environnement écologique humain, peuvent avoir sur l'expression et la détermination de l'histoire de vie.

## 5.9 La popularité croissante des approches bayésiennes

L'approche bayésienne est devenue populaire dans les études de génétique quantitative des populations naturelles. Cette popularité signifie que les lecteurs intéressés doivent reconsidérer certaines notions d'interprétation de l'estimation des paramètres héritées de l'approche fréquentiste. Par exemple, la distribution postérieure extraite de la variance génétique additive d'un caractère donné dépend de l'a priori fourni (c'est-à-dire de la connaissance préalable, ou *prior*) ; une forte dépendance à l'égard de cet a priori indique qu'il est possible qu'il n'y ait pas de signal génétique dans les données, tandis que si la distribution postérieure n'est pas affectée par différents a priori, alors les données sont très informatives. Par conséquent, et comme nous l'avons fait dans toutes nos analyses, nous avons testé plusieurs distributions a priori non informatives pour vérifier toute dépendance de la distribution postérieure à l'égard des a priori. Ce test est une bonne pratique qui devrait être mentionnée explicitement dans toute étude de génétique quantitative utilisant des méthodes bayésiennes.

La distribution a priori n'est pas un élément du théorème de Bayes qu'il faut définir puis ignorer rapidement ; au contraire, il incombe au chercheur d'établir une relation adéquate entre le type d'a priori utilisé et le test de différentes a priori. Par exemple, dans le chapitre IV, nous avons dérivé de nos modèles les distributions postérieures des corrélations génétiques inter-environnementales dans l'AFR afin de tester les G×Es. Nous avons utilisé deux a priori différentes et obtenu le même résultat, ce qui nous a permis de conclure que les distributions postérieures de ces corrélations génétiques ne dépendent pas des connaissances préalables et qu'il existe des preuves crédibles de G×Es.

Dans l'inférence statistique bayésienne, on peut utiliser des a priori informatifs lorsqu'on dispose d'informations pertinentes ou fiables sur les paramètres du modèle statistique avant d'observer les données (Lemoine 2019), par exemple, connaissances préalables sur la proportion de variance phénotypique expliqué par les composantes



généétiques et non-généétiques. Les a priori informatifs permettent d'incorporer cette connaissance existante dans l'analyse et peuvent conduire à une estimation des paramètres plus efficace et plus précise (voir Morris *et al.* 2015). Toutefois, la décision d'utiliser des a priori informatifs doit être fondée sur un examen minutieux du contexte et des objectifs spécifiques de l'analyse (Banner *et al.* 2020). Voici quelques situations dans lesquelles l'utilisation d'a priori informatif dans des analyses génétiques quantitatives peut s'avérer utile : Dans les cas où on dispose de données limitées, les a priori informatifs peuvent fournir une régularisation et aider à stabiliser les estimations. De plus, les a priori informatifs peuvent aider à éviter les estimations de paramètres extrêmes lorsque les données sont rares (Lenk et Orme 2009). Dans un autre cas, si certaines valeurs de paramètres ne sont pas plausibles ou sont limitées par la théorie ou les lois naturelles, des a priori informatives peuvent contribuer à renforcer ces contraintes et à orienter l'analyse vers des résultats plus réalistes. Il convient de noter que les a priori informatifs peuvent également être variables en termes de degré d'information préalable qu'ils fournissent. Certains sont fortement informatifs, d'autres le sont faiblement.

Alors que les statistiques fréquentistes basent les probabilités de effets sur de multiples répétitions (réelles ou théoriques) de l'expérience, les méthodes bayésiennes fournissent directement les probabilités (a posteriori) des causes, en l'occurrence des paramètres de variance estimés (Morrissey *et al.* 2014 ; Walsh et Lynch 2018). Dans nos analyses, nous avons suivi la convention d'utiliser un estimateur ponctuel, le mode, pour décrire la distribution postérieure, en combinaison avec l'utilisation d'une représentation statistique des intervalles de crédibilité, la densité postérieure la plus élevée à 95 %. Pour accroître la confiance dans nos déclarations, nous avons utilisé des indices postérieurs nouvellement suggérés (c'est-à-dire la probabilité de direction, *pd*, et le *ROPE*) qui permettent une interprétation plus directe de la distribution postérieure complète (Makowski *et al.* 2019b). À notre connaissance, il s'agit de la première utilisation de ces indices dans une étude de génétique quantitative, et nous encourageons d'autres chercheurs à y avoir recours.

Bien entendu, tout cela ne signifie pas que les statistiques bayésiennes soient meilleures que les statistiques fréquentistes. Par exemple, le fait d'augmenter la taille de

l'échantillon dans toute analyse peut accroître considérablement les performances des méthodes fréquentistes, mais cela ne s'applique pas aux méthodes bayésiennes. Le débat entre les paradigmes fréquentiste et bayésien est également philosophique et c'est une conversation dans laquelle nous nous abstiendrons d'entrer. Notre objectif est de nous concentrer sur l'application scientifique de ces méthodes et sur l'équilibre entre leurs coûts et leurs avantages. Néanmoins, comme dans le cas du *p-hacking* et des interprétations erronées des seuils de falsification (Head *et al.* 2015), il est possible d'abuser de ces indices pour atteindre un seuil prédéfini de « significativité », dans l'espoir de fournir des preuves de son hypothèse. Non seulement cette pratique n'est pas scientifique, mais elle est aussi pratiquement impossible dans un cadre bayésien. Dans le paradigme bayésien, la vérification des hypothèses n'est pas la même que dans le paradigme fréquentiste classique ; dans ce cas, l'hypothèse alternative est l'a priori. En outre, il n'y a pas de seuils prédéfinis de significativité dans le paradigme bayésien puisque le concept lui-même est purement subjectif. Cette notion est liée à l'une des forces de l'inférence bayésienne, les déclarations probabilistes (qui incluent implicitement une mesure de l'incertitude) au lieu, dans le paradigme fréquentiste, des déclarations de significativité autour d'une estimation ponctuelle nécessitant le calcul d'une mesure distincte de l'incertitude. Ainsi, les indices postérieurs ne remplacent pas et ne devraient pas remplacer la puissance statistique de l'interprétation de la distribution postérieure elle-même pour faire des inférences sur le paramètre inconnu.

### **5.10 La valeur des études de génétique évolutive humaine**

Cette thèse est centrée sur l'investigation et la réponse aux questions d'évolution en utilisant un système modèle humain, la population historique canadienne-française. Ce cadre soulève de multiples questions qui convergent autour de la généralité des résultats basés sur des études humaines. Le principal argument contre la généralisation des résultats des études humaines à d'autres taxons est lié à l'exceptionnalisme humain, un point de vue rarement défendu dans les sciences de la vie mais que l'on retrouve souvent dans les sciences sociales (p. ex., Ingold 2007). L'idée d'exceptionnalisme humain accepte le fait que les humains soient le produit d'une descendance avec

modification (c'est-à-dire d'un changement évolutif par sélection naturelle) tout en considérant notre espèce comme un résultat si différent des autres espèces que les généralisations et les comparaisons entre espèces peuvent être peu informatives et trompeuses. Par exemple, comme cela a été mentionné plus haut, la transmission culturelle et la sélection de groupe jouent un rôle dans l'évolution des hominiens par rapport aux autres lignées (voir Boyd 2017).

De nos jours, l'argument de l'exceptionnalisme est invoqué autour de la pertinence des études humaines sur des phénomènes très généraux, tels que la cognition, l'apprentissage social et le comportement, et moins autour de phénomènes évolutifs, tels que les contraintes sur le changement évolutif et l'impact de la sélection naturelle sur la variation des traits au sein des populations et entre elles (Sterelny 2017). Cependant, un autre aspect de l'exceptionnalisme humain, moins courant, mais toujours présent dans certains milieux académiques, est l'argument selon lequel la sélection naturelle n'opère plus ou est très limitée dans son impact sur le changement phénotypique dans les populations humaines contemporaines. Les résultats de cette thèse remettent directement en question ce point de vue tout en reconnaissant d'autres aspects qui jouent un rôle dans l'évolution de l'humain moderne. De plus, les conclusions tirées des travaux de recherche présentés ici augmentent notre connaissance des processus évolutifs généraux tels que les interactions génotype-environnement, la généralité de la réponse à la sélection, et le potentiel évolutif.

Dans leur enquête sur les opinions des chercheurs en biologie de l'évolution et dans des domaines connexes, Briga *et al.* (2017) ont constaté que la plupart de leurs répondants ( $n = 442$ ) sont d'accord pour dire que les humains évoluent toujours, mais ils sont moins nombreux à penser que la sélection naturelle est le principal mécanisme de cette évolution. Un résultat plus intéressant est que la plupart sont d'accord pour dire que les études humaines ont contribué à l'avancement de la biologie évolutive, mais que les personnes moins exposées à ce domaine ont des opinions négatives à ce sujet. Ce patron est troublant et exige davantage d'efforts de la part des biologistes de l'évolution pour accroître la communication d'exemples de résultats qui démontrent l'impact et la contribution des études humaines à la compréhension fondamentale des processus évolutifs. L'enquête présente une liste décente d'exemples de contributions de différents

types des études humaines à la biologie de l'évolution et aux domaines de recherche connexes (p. ex., développements méthodologiques, big data, nouvelles hypothèses). Les études humaines ont énormément contribué à la génétique quantitative et évolutive. Ces contributions ne sont pas spécifiques à l'homme, et les résultats présentés dans cette thèse nous permettent de mieux comprendre les processus et mécanismes de l'évolution en fournissant des preuves empiriques qui confirment ou réfutent les hypothèses théoriques.

Le fait de fournir de telles informations empiriques constitue-t-il un argument solide en faveur de la valeur des études génétiques évolutives humaines pour le grand public ? C'est une chose que de faire changer d'avis les universitaires et les chercheurs sur la pertinence générale des études sur l'évolution humaine, mais c'en est une autre de convaincre le public que ces études méritent d'être financées et qu'elles peuvent contribuer à l'amélioration de la condition humaine. Une façon d'y parvenir est d'affirmer sans cesse que les processus évolutifs et l'adaptation peuvent se produire en peu de temps, ce qui augmente la probabilité d'interactions avec les processus écologiques et d'effets sur la dynamique démographique. L'impact de l'évolution de l'AFR dans la population IAC est un exemple de la manière dont l'étude et la mise en évidence de tout processus évolutif peuvent aider à mieux comprendre les différents contributeurs aux changements démographiques qu'une population humaine présente dans son histoire.

D'autre part, montrer qu'il existe une stase évolutive constitue une démonstration que l'attente d'un changement évolutif n'est pas toujours satisfaite dans la nature, ce qui accroît la nécessité d'examiner plus avant les sources possibles de cette contrainte dans les populations humaines. Des fonds sont pratiquement toujours disponibles pour des études sur la fécondité humaine et la santé publique. Étant donné que les traits reproductifs humains présentent un potentiel évolutif et qu'il existe de grandes quantités de données sur les populations humaines contemporaines, les études en génétique évolutive humaine devraient être financées afin de fournir des informations sur les processus évolutifs pouvant affectant les patrons de fécondité actuels (au lieu de supposer que l'action de la sélection naturelle est absente). Il n'est pas difficile de contextualiser l'importance des processus évolutifs par rapport aux problèmes de la vie

réelle puisque nous, en tant qu'espèce, sommes un produit de ces processus et soumis aux mécanismes qui affectent les changements de fréquence des allèles et les modifications des distributions phénotypiques. Puisque la recherche de bonnes politiques publiques pour améliorer la santé des gens ou prévenir des événements désastreux tels que l'effondrement d'une population, est un objectif général des sociétés humaines, il est impératif d'intégrer le rôle de la sélection naturelle dans ces considérations. Cela ne peut que fournir davantage d'informations et de connaissances sur notre constitution naturelle et notre dynamique aux niveaux individuel, populationnel et de l'espèce.

### **5.11 Dernières remarques**

Cette thèse traite des questions évolutives soulevées par les preuves récentes de l'évolution contemporaine par la sélection naturelle. Bien que les réponses apportées à ces questions ne soient pas définitives, l'étendue de notre confiance dans ces explications est conditionnée par la précision des outils statistiques et la qualité des ensembles de données analysés. En gardant cela à l'esprit, les hypothèses évolutives, et leurs prédictions correspondantes, que nous avons cherché à tester en utilisant l'ensemble de données canadiennes-françaises ont des implications directes sur les questions liées à l'impact de l'hétérogénéité environnementale sur la composante génétique de l'évolution phénotypique et la généralité des changements évolutifs des traits d'histoire de vie par la sélection naturelle dans les populations en expansion. Ces implications sont également associées à une mesure formelle et explicite de la performance des généalogies canadiennes-françaises lorsqu'elles sont analysées avec les dernières méthodes d'estimation génétique quantitative. À ce stade, je peux affirmer avec confiance que cette thèse apporte des réponses solides à des problèmes biologiques intéressants, présente une vérification convaincante de la qualité de nos données et, surtout, génère de nouvelles questions et de nouvelles avenues à explorer scientifiquement. Il se peut que le lecteur ne soit pas convaincu par une telle affirmation ou qu'il ou elle ait des réserves quant aux explications que nous proposons. Dans ce cas, je suis à l'aise avec une telle position car elle constituerait une invitation à plus

d'examen, de coopération et de discussions pour parvenir à des explications plus nuancées mais plus claires.

En termes de perspectives, les travaux de cette thèse présentent des possibilités de marier les approches de la génétique quantitative aux avancées de la génomique. Ces opportunités de recherche peuvent impliquer : i) des modèles théoriques de génétique des populations qui tentent de reconstruire les données génomiques passées en utilisant les données de séquençage des individus canadiens-français vivants reliés aux généalogies (ex., Nelson *et al.* 2018 ; Baumdicker *et al.* 2022), et ii) une évaluation empirique du rôle et de l'impact des processus évolutifs dans les générations plus récentes étant donné que les généalogies canadiennes-françaises sont mises à jour continuellement avec de nouvelles informations. Il est important de se rappeler que la variation génétique que nous avons examinée dans ce travail est une variabilité perdue. Par conséquent, notre recherche est rétrospective et ne dit pas grand-chose sur l'état de la population actuelle. Pour mieux comprendre les populations actuelles, nous devons étudier la variation génétique courante des traits de l'histoire de vie avec le flux croissant de données de séquençage génomique (Vézina et Bournival 2020). D'autre part, l'ADN ancien représente un puit profond d'informations sur les variations génétiques passées qui ont très probablement façonné les caractéristiques des populations passées vivant dans la région (e.g., Harding *et al.* 2020). Un exemple de programme de recherche perspicace en génétique des populations peut, à mon avis, faire le lien entre le passé et le présent, non seulement en termes de variation génétique mais aussi d'impact des processus qui façonnent cette variation, c'est-à-dire l'évolution, l'écologie et la démographie.

Dans une perspective plus large, les traits examinés dans cette thèse, les traits d'histoire de vie, ont une forte relation avec les caractéristiques d'un organisme tels que le développement, la reproduction et la survie. Cette relation dans les populations humaines se traduit par un lien substantiel avec la dynamique démographique, la santé publique et les interventions cliniques. Le fait que ces traits présentent une composante génétique, un potentiel évolutif et sont sensibles à l'hétérogénéité environnementale signifie que l'examen de ces traits dans tout autre contexte, social, démographique ou médical, par exemple, doit intégrer ces aspects (Milot 2021). Les connaissances acquises

en incluant les informations évolutives et génétiques peuvent mieux expliquer les résultats observés et possiblement adaptatifs de processus tels que le vieillissement, le déclin de la fertilité et la prévalence des maladies chroniques au niveau des individus et des populations. Ce type d'intégration est une caractéristique de la médecine évolutive qui vise à apporter de nouvelles connaissances dans la pratique clinique et les politiques de santé publique (Stearns 2012). Au-delà des traits de l'histoire de vie, le rôle de la sélection naturelle dans les populations humaines contemporaines devient plus pertinent lorsque nous considérons que, dans nos environnements urbains, nous sommes soumis à des pressions sélectives nouvelles et peut-être encore inconnues (Milot et Stearns 2020). Dans ce contexte, nous renversons l'affirmation selon laquelle les humains ont échappé à l'action de la sélection naturelle en soulignant que le rôle de ce processus évolutif d'adaptation est toujours aussi important aujourd'hui qu'au cours du paléolithique. Et je suis d'avis que l'action de la sélection naturelle est encore substantielle, même si elle est cryptique, étant donné qu'elle peut avoir un impact sur les populations en un petit nombre de générations et que sa direction et action peut correspondre à ceux d'autres processus évolutifs pertinents tels que la sélection de groupes culturels, la sélection sociale et la sélection sexuelle.

## REFERENCES BIBLIOGRAPHIQUES

- Anderson-Trocmé, L., Nelson, D., Zabad, S., Diaz-Papkovich, A., Kryukov, I., Baya, N., ... & Gravel, S. (2023). On the genes, genealogies, and geographies of Quebec. *Science*, 380(6647), 849-855.
- Antonovics, J. (1987). The evolutionary dys-synthesis: which bottles for which wine?. *The American Naturalist*, 129(3), 321-331.
- Ariew, A., & Lewontin, R. C. (2004). The confusions of fitness. *British Journal for the Philosophy of Science*, 55(2).
- Arnold, S. J., & Wade, M. J. (1984). On the measurement of natural and sexual selection: theory. *Evolution*, 709-719.
- Arnold, S. J., Bürger, R., Hohenlohe, P. A., Ajie, B. C., & Jones, A. G. (2008). Understanding the evolution and stability of the G-matrix. *Evolution*, 62(10), 2451-2461.
- Bailey, N. W., Marie-Orleach, L., & Moore, A. J. (2018). Indirect genetic effects in behavioral ecology: does behavior play a special role in evolution?. *Behavioral Ecology*, 29(1), 1-11.
- Banner, K. M., Irvine, K. M., & Rodhouse, T. J. (2020). The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in Ecology and Evolution*, 11(8), 882-889.
- Barton, N. H., Briggs, D., Eisen, J., Goldstein, D., & Patel, N. (2007). *Evolution*.
- Barton, N. H., Etheridge, A. M., & Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical population biology*, 118, 50-73.
- Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., ... & Kelleher, J. (2022). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220(3), iyab229.
- Belsky, J., Steinberg, L., & Draper, P. (1991). Childhood experience, interpersonal development, and reproductive strategy: An evolutionary theory of socialization. *Child development*, 62(4), 647-670.



- Béréanos, C., Ellis, P. A., Pilkington, J. G., & Pemberton, J. M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular ecology*, 23(14), 3434-3451.
- Bijma, P., Muir, W. M., & Van Arendonk, J. A. (2007). Multilevel selection 1: quantitative genetics of inheritance and response to selection. *Genetics*, 175(1), 277-288.
- Bijma, P., Muir, W. M., Ellen, E. D., Wolf, J. B., & Van Arendonk, J. A. (2007). Multilevel selection 2: estimating the genetic parameters determining inheritance and response to selection. *Genetics*, 175(1), 289-299.
- Bird, J. B. (1972). *The Natural Landscapes of Canada: A Study in Regional Earth Science*. Canada: Wiley Publishers of Canada.
- Boag, P. T., & Grant, P. R. (1978). Heritability of external morphology in Darwin's finches. *Nature*, 274(5673), 793-794.
- Boag, P. T., & Grant, P. R. (1981). Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galapagos. *Science*, 214(4516), 82-85.
- Boisvert, M., & Mayer, F. M. (1994). Mortalité infantile et consanguinité dans une population endogame du Québec. *Population (french edition)*, 685-724.
- Bolnick, D. I., Barrett, R. D., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non) parallel evolution. *Annual Review of Ecology, Evolution, and Systematics*, 49, 303-330.
- Bolund, E., Hayward, A., Pettay, J. E., & Lummaa, V. (2015). Effects of the demographic transition on the genetic variances and covariances of human life-history traits. *Evolution*, 69(3), 747-755.
- Bolund, E., Hayward, A., & Lummaa, V. (2016). *Life-history evolution, Human*.
- Bolund, E., & Lummaa, V. (2017). The effects of resource availability and the demographic transition on the genetic correlation between number of children and grandchildren in humans. *Heredity*, 118(2), 186-192.
- Bouchard, G. (1989). Population studies and genetic epidemiology in northeast Quebec. *Canadian Studies in Population [ARCHIVES]*, 61-86.

- Bouchard, G., & De Braekeleer, M. (1991). *Histoire d'un genome: Population et genetique dans l'est du Quebec*. Sillery, Québec: Presses de l'Université du Québec.
- Boyd, R. (2017). *A different kind of animal*. Princeton University Press.
- Braendle, C., Heyland, A., & Flatt, T. (2011). Integrating mechanistic and evolutionary analysis of life history variation. *Mechanisms of life history evolution. The genetics and physiology of life history traits and trade-offs*, 3-10.
- Briga, M., Griffin, R. M., Berger, V., Pettay, J. E., & Lummaa, V. (2017). What have humans done for evolutionary biology? Contributions from genes to populations. *Proceedings of the Royal Society B: Biological Sciences*, 284(1866), 20171164.
- Brommer, J. E. (2000). The evolution of fitness in life-history theory. *Biological Reviews*, 75(3), 377-404.
- Brommer, J. E., Rattiste, K., & Wilson, A. J. (2008). Exploring plasticity in the wild: laying date-temperature reaction norms in the common gull *Larus canus*. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635), 687-693.
- Brommer, J. E. (2013). Phenotypic plasticity of labile traits in the wild. *Current Zoology*, 59(4), 485-505.
- Bulmer, M. (1971). The effect of selection on genetic variability. *The American Naturalist*, 105(943), 201-211.
- Bulmer, M. G. (1973). Inbreeding in the great tit. *Heredity*, 30(3), 313-325.
- Bürkli, A., & Postma, E. (2014). Genetic constraints underlying human reproductive timing in a premodern Swiss village. *Evolution*, 68(2), 526-537.
- Byars, S. G., Ewbank, D., Govindaraju, D. R., & Stearns, S. C. (2010). Natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences*, 107(suppl\_1), 1787-1792
- Caswell, H. (1989). Analysis of life table response experiments I. Decomposition of effects on population growth rate. *Ecological Modelling*, 46(3-4), 221-237.

- Caudell, M. A., & Quinlan, R. J. (2012). Resource availability, mortality, and fertility: A path analytic approach to global life-history variation. *Human Biology*, 84(2), 101-125.
- Charmantier, A., & Garant, D. (2005). Environmental quality and evolutionary potential: lessons from wild populations. *Proceedings of the Royal Society B: Biological Sciences*, 272(1571), 1415-1425.
- Charmantier, A., Brommer, J. E., & Nussey, D. H. (2014). The quantitative genetics of senescence in wild animals. *Quantitative genetics in the wild*, 68-83.
- Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science*, (366), 92-96.
- Charnov, E. L., & Schaffer, W. M. (1973). Life-history consequences of natural selection: Cole's result revisited. *The American Naturalist*, 107(958), 791-793.
- Chisholm, J. S., Ellison, P. T., Evans, J., Lee, P. C., Lieberman, L. S., Pavlik, Z., ... & Worthman, C. M. (1993). Death, hope, and sex: Life-history theory and the development of reproductive strategies [and comments and reply]. *Current anthropology*, 34(1), 1-24.
- Chisholm, J. S., Quinlivan, J. A., Petersen, R. W., & Coall, D. A. (2005). Early stress predicts age at menarche and first birth, adult attachment, and expected lifespan. *Human nature*, 16(3), 233-265.
- Class, B., & Brommer, J. E. (2020). Can dominance genetic variance be ignored in evolutionary quantitative genetic analyses of wild populations?. *Evolution*, 74(7), 1540-1550.
- Colautti, R. I., & Barrett, S. C. (2013). Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science*, 342(6156), 364-366.
- Cole, L. C. (1954). The population consequences of life history phenomena. *The Quarterly review of biology*, 29(2), 103-137.
- Coltman, D. W., O'Donoghue, P., Jorgenson, J. T., Hogg, J. T., Strobeck, C., & Festa-Bianchet, M. (2003). Undesirable evolutionary consequences of trophy hunting. *Nature*, 426(6967), 655-658.
- Corbett, S., Courtiol, A., Lummaa, V., Moorad, J., & Stearns, S. (2018). The transition to modernity and chronic disease: mismatch and natural selection. *Nature Reviews Genetics*, 19(7), 419-430.

- Cowperthwaite, M. C., & Meyers, L. A. (2007). How mutational networks shape evolution: lessons from RNA models. *Annual review of Ecology and Systematics*, 38, 203-230.
- Cui, J., Hopper, J. L., & Harrap, S. B. (2002). Genes and family environment explain correlations between blood pressure and body mass index. *Hypertension*, 40(1), 7-12.
- Dammhahn, M., Dingemanse, N. J., Niemelä, P. T., & Réale, D. (2018). Pace-of-life syndromes: a framework for the adaptive integration of behaviour, physiology and life history. *Behavioral Ecology and Sociobiology*, 72, 1-8.
- Danchin, E., Pujol, B., & Wagner, R. H. (2013). The double pedigree: a method for studying culturally and genetically inherited behavior in tandem. *PLoS one*, 8(5), e61254.
- Darwin, C. (1859). *On the origin of species*.
- De Jong, G. (1994). The fitness of fitness concepts and the description of natural selection. *The Quarterly Review of Biology*, 69(1), 3-29.
- DeFaveri, J., & Merilä, J. (2014). Local adaptation to salinity in the three-spined stickleback?. *Journal of Evolutionary Biology*, 27(2), 290-302.
- Descamps, S., Boutin, S., Berteaux, D., McAdam, A. G., & Gaillard, J. M. (2008). Cohort effects in red squirrels: the influence of density, food abundance and temperature on future survival and reproductive success. *Journal of Animal Ecology*, 305-314.
- Doyon, A (2018). *Dynamique des marqueurs génétiques liés au sexe dans la population canadienne-française pour l'interprétation des traces d'ADN en génétique forensique*. Mémoire. Trois-Rivières, Université du Québec à Trois-Rivières, 232 p.
- Ellis, B. J. (2004). Timing of pubertal maturation in girls: an integrated life history approach. *Psychological bulletin*, 130(6), 920.
- Elmer, K. R., Fan, S., Kusche, H., Luise Spreitzer, M., Kautt, A. F., Franchini, P., & Meyer, A. (2014). Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature communications*, 5(1), 5168.
- Endler, J. A. (1986). Natural selection in the wild. *Monogr. Popul. Biol.*, 21, 1-33.

- Engelhardt, S. C., Bergeron, P., Gagnon, A., Dillon, L., & Pelletier, F. (2019). Using geographic distance as a potential proxy for help in the assessment of the grandmother hypothesis. *Current Biology*, 29(4), 651-656.
- Evans, S. R., Waldvogel, D., Vasiljevic, N., & Postma, E. (2018). Heritable spouse effects increase evolutionary potential of human reproductive timing. *Proceedings of the Royal Society B: Biological Sciences*, 285(1876), 20172763.
- Falconer, D. S., & Mackay, T. F. C. (1996). Introduction to quantitative genetics. Essex, UK: Longman Group.
- Finch, C. E., & Tanzi, R. E. (1997). Genetics of aging. *Science*, 278(5337), 407-411.
- Fischer, D. H. (1970). *Historians' fallacies: Toward a logic of historical thought*.
- Fisher, D. N., Haines, J. A., Boutin, S., Dantzer, B., Lane, J. E., Coltman, D. W., & McAdam, A. G. (2019). Indirect effects on fitness between individuals that have never met via an extended phenotype. *Ecology Letters*, 22(4), 697-706.
- Fisher, R. A. (1918). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2), 399-433.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press.
- Fisher, R. A. (1958). *The genetical theory of natural selection*. Oxford Univ. Press, Oxford, U. K.
- Flatt, T., & Heyland, A. (Eds.). (2011). *Mechanisms of life history evolution: the genetics and physiology of life history traits and trade-offs*. Oxford university press.
- Frank, S. A. (2007). All of life is social. *Current Biology*, 17(16), R648-R650.
- Franklin, I., & Lewontin, R. C. (1970). Is the gene the unit of selection?. *Genetics*, 65(4), 707.
- Futuyma, D. J. (2009). Natural selection and adaptation. *Evolution*, 279-301.
- Futuyma, D. J. (2013). *Evolution*. Third edition. Sunderland, Massachusetts U.S.A, Sinauer Associates, Inc. Publishers.

- Gagnon, A., Tremblay, M., Vézina, H., & Seabrook, J. A. (2011). Once were farmers: Occupation, social mobility, and mortality during industrialization in Saguenay-Lac-Saint-Jean, Quebec 1840–1971. *Explorations in Economic History*, 48(3), 429-440.
- Gardner, A. (2009). Adaptation as organism design. *Biology letters*, 5(6), 861-864.
- Gauvreau, D., & Jette, R. (1992, January). Histoire démographique et génétique humaine dans une région du Québec avant 1850. In *Annales de démographie historique* (pp. 245-267). Société de Démographie Historique.
- Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional ecology*, 21(3), 394-407.
- Ghalambor, C. K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., & Hughes, K. A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, 525(7569), 372-375.
- Gienapp, P., Reed, T. E., & Visser, M. E. (2014). Why climate change will invariably alter selection pressures on phenology. *Proceedings of the Royal Society B: Biological Sciences*, 281(1793), 20141611.
- Goodnight, C. J. (2011). Evolution in metacommunities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1569), 1401-1409.
- Gould, S. J., & Lewontin, R. C. (1979). *The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme* (pp. 204-221). Routledge.
- Gould, S. J. (1994). Tempo and mode in the macroevolutionary reconstruction of Darwinism. *Proceedings of the National Academy of Sciences*, 91(15), 6764-6771.
- Grant, B. R., & Grant, P. R. (1993). Evolution of Darwin's finches caused by a rare climatic event. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 251(1331), 111-117.
- Grant, P. R., & Grant, B. R. (1995). Predicting microevolutionary responses to directional selection on heritable variation. *Evolution*, 49(2), 241-251.
- Griskevicius, V., Delton, A. W., Robertson, T. E., & Tybur, J. M. (2011). Environmental contingency in life history strategies: the influence of mortality and socioeconomic

- status on reproductive timing. *Journal of personality and social psychology*, 100(2), 241.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of statistical software*, 33, 1-22.
- Harding, T., Milot, E., Moreau, C., Lefebvre, J. F., Bournival, J. S., Vézina, H., ... & Labuda, D. (2020). Historical human remains identification through maternal and paternal genetic signatures in a founder population with extensive genealogical record. *American journal of physical anthropology*, 171(4), 645-658.
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS biology*, 13(3), e1002106.
- Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, 9(2), 226-252.
- Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 423-447.
- Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, 69-83.
- Henderson, C. R. (1977). Best linear unbiased prediction of breeding values not in the model for records. *Journal of Dairy Science*, 60(5), 783-787.
- Henderson, C. R. (1984). Estimation of variances and covariances under multiple trait models. *Journal of Dairy Science*, 67(7), 1581-1589.
- Hendry, A. P., & Kinnison, M. T. (1999). Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution*, 53(6), 1637-1653.
- Heyer, E. (2009). One founder/one gene hypothesis in a new expanding population: Saguenay (Quebec, Canada). *Human biology*, 81(5/6), 645-655.
- Hill, K., & Kaplan, H. (1999). Life history traits in humans: Theory and empirical studies. *Annual review of anthropology*, 28(1), 397-430.
- Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537), 73-85.

- Hoffmann, A. A., & Merilä, J. (1999). Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution*, 14(3), 96-101.
- Huang, Y., Tran, I., & Agrawal, A. F. (2016). Does genetic variation maintained by environmental heterogeneity facilitate adaptation to novel selection?. *The American Naturalist*, 188(1), 27-37.
- Hunter, D. C., Pemberton, J. M., Pilkington, J. G., & Morrissey, M. B. (2019). Pedigree-based estimation of reproductive value. *Journal of Heredity*, 110(4), 433-444.
- Ingold, T. (2007). The trouble with 'evolutionary biology'. *Anthropology today*, 23(2), 13-17.
- Jensen, H., Szulkin, M., & Slate, J. (2014). Molecular quantitative genetics. *Quantitative genetics in the wild*, 209-227.
- Johnson, M. T., & Munshi-South, J. (2017). Evolution of life in urban environments. *Science*, 358(6363), eaam8327.
- Kang, M. S. (2002). Genotype-environment interaction: progress and prospects. *Quantitative genetics, genomics and plant breeding*, 221-243.
- Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. M. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 9(4), 156-185.
- Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., ... & Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*, 157(3), 245-261.
- Kinnison, M. T., & Hendry, A. P. (2001). The pace of modern life II: from rates of contemporary microevolution to pattern and process. In *Microevolution rate, pattern, process* (pp. 145-164). Springer, Dordrecht.
- Kokko, H., Chaturvedi, A., Croll, D., Fischer, M. C., Guillaume, F., Karrenberg, S., ... & Stapley, J. (2017). Can evolution supply what ecology demands?. *Trends in Ecology & Evolution*, 32(3), 187-197.
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*, 9(1), 1-9.



- Koskela, E., Jonsson, P., Hartikainen, T., & Mappes, T. (1998). Limitation of reproductive success by food availability and litter size in the bank vole, *Clethrionomys glareolus*. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1401), 1129-1134.
- Kruuk, L. E., Clutton-Brock, T. H., Slate, J., Pemberton, J. M., Brotherstone, S., & Guinness, F. E. (2000). Heritability of fitness in a wild mammal population. *Proceedings of the National Academy of Sciences*, 97(2), 698-703.
- Kruuk, L. E. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1446), 873-890.
- Kruuk, L. E., Slate, J., & Wilson, A. J. (2008). New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annual Review of Ecology, Evolution, and Systematics*, 39, 525-548.
- Kruuk, L. E. B., & Hill, W. G. (2008). Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635), 593-596.
- Kruuk, L. E., Charmantier, A., & Garant, D. (2014). The study of quantitative genetics in wild populations. *Quantitative genetics in the wild*, 1-15.
- Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, 1210-1226.
- Larmuseau, M. H., Matthijs, K., & Wenseleers, T. (2016). Cuckolded fathers rare in human populations. *Trends in ecology & evolution*, 31(5), 327-329.
- Lavoie, E. M., Tremblay, M., Houde, L., & Vézina, H. (2005). Demogenetic study of three populations within a region with strong founder effects. *Community genetics*, 8(3), 152-160.
- Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, 128(7), 912-928.
- Lenk, P., & Orme, B. (2009). The value of informative priors in Bayesian inference with sparse data. *Journal of Marketing Research*, 46(6), 832-845.
- Lenski, R. E., & Service, P. M. (1982). The statistical analysis of population growth rates calculated from schedules of survivorship and fecundity. *Ecology*, 63(3), 655-662.

- Leroy, G. (2011). Genetic diversity, inbreeding and breeding practices in dogs: results from pedigree analyses. *The Veterinary Journal*, 189(2), 177-182.
- Lewontin, R. C. (1970). The units of selection. *Annual review of ecology and systematics*, 1(1), 1-18.
- Lummaa, V. (2007). Life history theory, reproduction and longevity in humans. *The Oxford handbook of evolutionary psychology*, 397-414.
- Lush J (1937) *Animal breeding*. Plans. Iowa State College Press, Ames
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*.
- Makowski, D., Ben-Shachar, M. S., Chen, S. H., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in psychology*, 2767.
- Marlowe, F. (2000). Paternal investment and the human mating system. *Behavioural processes*, 51(1-3), 45-61.
- Martin, Y. (1957). L'Île-aux-Coudres: Population et économie. *Cahiers de géographie du Québec*, 2(2), 167-195.
- Martin, O. C., & Wagner, A. (2007). New structural variation in evolutionary searches of RNA neutral networks. *Biosystems*, 90(2), 475-485.
- Martin, T. E. (2015). Age-related mortality explains life history strategies of tropical and temperate songbirds. *Science*, 349(6251), 966-970.
- Mayer, F. M., Lavoie, Y., Létourneau, E., & Lavoie, J. (1988). Un programme de recherches sur la dynamique bioculturelle. *Cahiers québécois de démographie*, 17(2), 289-298.
- Maynard-Smith, J., & Szathmary, E. (1997). *The major transitions in evolution*. OUP Oxford.
- Mayr, E. (2001). *What evolution is*. Basic books.
- McAdam, A. G., Garant, D., & Wilson, A. J. (2014). The effects of others' genes: maternal and other indirect genetic effects. *Quantitative genetics in the wild*, 84-103.

- McDonald, J. F., & Ayala, F. J. (1974). Genetic response to environmental heterogeneity. *Nature*, 250(5467), 572-574.
- Meyer, K. (1991). Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genetics Selection Evolution*, 23(1), 67-83.
- Meyer, K. (1992). Variance components due to direct and maternal effects for growth traits of Australian beef cattle. *Livestock Production Science*, 31(3-4), 179-204.
- Milot, E., Mayer, F. M., Nussey, D. H., Boisvert, M., Pelletier, F., & Réale, D. (2011). Evidence for evolution in response to natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences*, 108(41), 17040-17045.
- Milot, E., & Stearns, S. C. (2020). Selection on humans in cities. *Urban Evolutionary Biology*, 268.
- Milot, E. (2021). How natural selection shapes our later years. *Nature Ecology & Evolution*, 5(3), 271-272.
- Monaghan, P., Uttley, J. D., Burns, M. D., Thaine, C., & Blackwood, J. (1989). The relationship between food supply, reproductive effort and breeding success in Arctic Terns *Sterna paradisaea*. *The Journal of Animal Ecology*, 261-274.
- Monaghan, P., Charmantier, A., Nussey, D. H., & Ricklefs, R. E. (2008). The evolutionary ecology of senescence. *Functional Ecology*, 371-378.
- Montiglio, P. O., Dammhahn, M., Dubuc Messier, G., & Réale, D. (2018). The pace-of-life syndrome revisited: the role of ecological conditions and natural history on the slow-fast continuum. *Behavioral Ecology and Sociobiology*, 72, 1-9.
- Moogk, P. N. (2000). *La Nouvelle France: The Making of French Canada-A Cultural History*. MSU Press.
- Moreau, C., Vézina, H., & Labuda, D. (2007). Founder effects and genetic variability in Quebec. *Medecine Sciences: M/S*, 23(11), 1008-1013.
- Moreau, C., Bhérier, C., Vézina, H., Jomphe, M., Labuda, D., & Excoffier, L. (2011). Deep human genealogies reveal a selective advantage to be on an expanding wave front. *Science*, 334(6059), 1148-1150.
- Morris, W. K., Vesk, P. A., McCarthy, M. A., Bunyavejchewin, S., & Baker, P. J. (2015). The neglected tool in the Bayesian ecologist's shed: a case study testing

informative priors' effect on model accuracy. *Ecology and Evolution*, 5(1), 102-108.

- Morrissey, M. B., Wilson, A. J., Pemberton, J. M., & Ferguson, M. M. (2007). A framework for power and sensitivity analyses for quantitative genetic studies of natural populations, and case studies in Soay sheep (*Ovis aries*). *Journal of Evolutionary Biology*, 20(6), 2309-2321.
- Morrissey, M. B., Kruuk, L. E. B., & Wilson, A. J. (2010). The danger of applying the breeder's equation in observational studies of natural populations. *Journal of evolutionary biology*, 23(11), 2277-2288.
- Morrissey, M. B., & Hadfield, J. D. (2012). Directional selection in temporally replicated studies is remarkably consistent. *Evolution*, 66(2), 435-442.
- Morrissey, M. B., de Villemereuil, P., Doligez, B., & Gimenez, O. (2014). Bayesian approaches to the quantitative genetic analysis of natural populations. *Quantitative genetics in the wild*, 228-253.
- Moorad, J. A., & Promislow, D. E. (2009). What can genetic variation tell us about the evolution of senescence?. *Proceedings of the Royal Society B: Biological Sciences*, 276(1665), 2271-2278.
- Moorad, J. A. (2013). A demographic transition altered the strength of selection for fitness and age-specific survival and fertility in a 19th century American population. *Evolution*, 67(6), 1622-1634.
- Moore, A. J., Brodie III, E. D., & Wolf, J. B. (1997). Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution*, 51(5), 1352-1362.
- Muehlenbein, M. P., & Flinn, M. V. (2011). Patterns and processes of human life history evolution. *Mechanisms of life history evolution: The genetics and physiology of life history traits and trade-offs*, 153-68.
- Muthukrishna, M., Doebeli, M., Chudek, M., & Henrich, J. (2018). The Cultural Brain Hypothesis: How culture drives brain expansion, sociality, and life history. *PLoS computational biology*, 14(11), e1006504.
- Nelson, D., Moreau, C., de Vriendt, M., Zeng, Y., Preuss, C., Vézina, H., ... & Gravel, S. (2018). Inferring transmission histories of rare alleles in population-scale genealogies. *The American Journal of Human Genetics*, 103(6), 893-906.

- Nettle, D. (2010). Dying young and living fast: Variation in life history across English neighborhoods. *Behavioral ecology*, 21(2), 387-395.
- Nettle, D., Coall, D. A., & Dickins, T. E. (2011). Early-life conditions and age at first pregnancy in British women. *Proceedings of the Royal Society B: Biological Sciences*, 278(1712), 1721-1727.
- Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, 280(1766), 20131343.
- O'Brien, E. L., & Dawson, R. D. (2011). Plumage color and food availability affect male reproductive success in a socially monogamous bird. *Behavioral Ecology*, 22(1), 66-72.
- Odling-Smee, J., Erwin, D. H., Palkovacs, E. P., Feldman, M. W., & Laland, K. N. (2013). Niche construction theory: a practical guide for ecologists. *The Quarterly review of biology*, 88(1), 3-28.
- Okasha, S. (2006). *Evolution and the levels of selection*. Clarendon Press.
- Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3), 545-554.
- Pebley, A. R., Casterline, J. B., & Trussell, J. (1982). Age at first birth in 19 countries. *International Family Planning Perspectives*, 2-7.
- Peischl, S., Dupanloup, I., Foucal, A., Jomphe, M., Bruat, V., Grenier, J. C., ... & Excoffier, L. (2018). Relaxed selection during a recent human expansion. *Genetics*, 208(2), 763-777.
- Pelletier, F., Pigeon, G., Bergeron, P., Mayer, F. M., Boisvert, M., Réale, D., & Milot, E. (2017). Eco-evolutionary dynamics in a contemporary human population. *Nature Communications*, 8(1), 15947.
- Pemberton, J. M. (2008). Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences*, 275(1635), 613-621.
- Pettay, J. E., Helle, S., Jokela, J., & Lummaa, V. (2007). Natural selection on female life-history traits in relation to socio-economic class in pre-industrial human populations. *Plos one*, 2(7), e606.

- Philippe, P., & Gomila, J. (1971). Structure de population et mariages consanguins à l'Isle-aux-Coudres (Québec). *Population (french edition)*, 707-716.
- Philippe, P. (1973). Analyse statistique des intervalles protogénésiques et intergénéésiques à l'Isle-aux-Coudres. Etude de démographie historique. *Population (french edition)*, 81-93.
- Pigeon, G., Festa-Bianchet, M., Coltman, D. W., & Pelletier, F. (2016). Intense selective hunting leads to artificial evolution in horn size. *Evolutionary Applications*, 9(4), 521-530.
- Pigliucci, M., & Kaplan, J. (2000). The fall and rise of Dr Pangloss: adaptationism and the Spandrels paper 20 years later. *Trends in ecology & evolution*, 15(2), 66-70.
- Pigliucci, M. (2010). Genotype–phenotype mapping and the end of the ‘genes as blueprint’ metaphor. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1540), 557-566.
- Placek, C. D., & Quinlan, R. J. (2012). Adolescent fertility and risky environments: a population-level perspective across the lifespan. *Proceedings of the Royal Society B: Biological Sciences*, 279(1744), 4003-4008.
- Plante, M., Claveau, S., Lepage, P., Lavoie, È. M., Brunet, S., Roquis, D., ... & Laprise, C. (2008). Mucopolidosis II: a single causal mutation in the N-acetylglucosamine-1-phosphotransferase gene (GNPTAB) in a French Canadian founder population. *Clinical genetics*, 73(3), 236-244.
- Postma, E. (2014). Four decades of estimating heritabilities in wild vertebrate populations: improved methods, more data, better estimates. *Quantitative genetics in the wild*, 16, 33.
- Price, G. R. (1970). Selection and covariance. *Nature*, 227, 520-521.
- Promislow, D. E., & Harvey, P. H. (1990). Living fast and dying young: A comparative analysis of life-history variation among mammals. *Journal of Zoology*, 220(3), 417-437.
- Queller, D. C. (2017). Fundamental theorems of evolution. *The American Naturalist*, 189(4), 345-353.
- Quinlan, R. J. (2010). Extrinsic mortality effects on reproductive strategies in a Caribbean community. *Human Nature*, 21, 124-139.

- Raveneau, J. (1977). Analyse morphologique, classification et protection des paysages: le cas de Charlevoix. *Cahiers de géographie du Québec*, 21(53-54), 135-186.
- Richerson, P. J., & Boyd, R. (2020). The human life history is adapted to exploit the adaptive advantages of culture. *Philosophical Transactions of the Royal Society B*, 375(1803), 20190498.
- Rickard, I. J., Holopainen, J., Helama, S., Helle, S., Russell, A. F., & Lummaa, V. (2010). Food availability at birth limited reproductive success in historical humans. *Ecology*, 91(12), 3515-3525.
- Ricklefs, R. E., & Wikelski, M. (2002). The physiology/life-history nexus. *Trends in ecology & evolution*, 17(10), 462-468.
- Robinson, M. R., Wilson, A. J., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Kruuk, L. E. (2009). The impact of environmental heterogeneity on genetic architecture in a wild population of Soay sheep. *Genetics*, 181(4), 1639-1648.
- Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Science*, 8(1), 95-108.
- Robertson, A. (1968). The spectrum of genetic variation. *Population biology and evolution*, 5-16.
- Roff, D. A. (1993). *Evolution of life histories: theory and analysis*. Springer Science & Business Media.
- Roff, D. A. (2002). *Life history evolution*. United Kingdom: Sinauer.
- Roff, D. A., & Emerson, K. (2006). Epistasis and dominance: evidence for differential effects in life-history versus morphological traits. *Evolution*, 60(10), 1981-1990.
- Roff, D. A. (2012). *Evolutionary quantitative genetics*. Springer Science & Business Media.
- Romagosa, I., & Fox, P. N. (1993). Genotype× environment interaction and adaptation. In *Plant breeding: principles and prospects* (pp. 373-390). Dordrecht: Springer Netherlands.
- Rosenberg, A. (1983). Fitness. *The Journal of Philosophy*, 80(8), 457-473.

- Roy-Gagnon, M. H., Moreau, C., Bherer, C., St-Onge, P., Sinnett, D., Laprise, C., ... & Labuda, D. (2011). Genomic and genealogical investigation of the French Canadian founder population structure. *Human genetics*, 129, 521-531.
- Sabourin, R. J. E. (1962). Notes sur la géomorphologie d'une partie de la Côte de Beaupré. *Cahiers de géographie du Québec*, 7(13), 27-36.
- Sanjak, J. S., Sidorenko, J., Robinson, M. R., Thornton, K. R., & Visscher, P. M. (2018). Evidence of directional and stabilizing selection in contemporary humans. *Proceedings of the National Academy of Sciences*, 115(1), 151-156.
- Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livestock Production Science*, 86(1-3), 35-45.
- Schluter, D., Clifford, E. A., Nemethy, M., & McKinnon, J. S. (2004). Parallel evolution and inheritance of quantitative traits. *The American Naturalist*, 163(6), 809-822.
- Scriver, C. R. (2001). Human genetics: lessons from Quebec populations. *Annual review of genomics and human genetics*, 2(1), 69-101.
- Sella, G., & Barton, N. H. (2019). Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annual review of genomics and human genetics*, 20.
- Seress, G., Sándor, K., Evans, K. L., & Liker, A. (2020). Food availability limits avian reproduction in the city: An experimental study on great tits *Parus major*. *Journal of Animal Ecology*, 89(7), 1570-1580.
- Siikamäki, P. (1998). Limitation of reproductive success by food availability and breeding time in pied flycatchers. *Ecology*, 79(5), 1789-1796.
- Sorensen, D., & Gianola, D. (2002). *Likelihood, Bayesian and MCMC methods in quantitative genetics*.
- Stearns, S. C. (1976). Life-history tactics: a review of the ideas. *The Quarterly review of biology*, 51(1), 3-47.
- Stearns, S. C. (1983). The influence of size and phylogeny on patterns of covariation among life-history traits in the mammals. *Oikos*, 173-187.
- Stearns, S. C. (1989). Trade-offs in life-history evolution. *Functional ecology*, 3(3), 259-268.



- Stearns, S. C. (1992). *The evolution of life histories*. Oxford: Oxford University Press.
- Stearns, S. C., Byars, S. G., Govindaraju, D. R., & Ewbank, D. (2010). Measuring selection in contemporary human populations. *Nature Reviews Genetics*, *11*(9), 611-622.
- Stearns, S. C. (2012). Evolutionary medicine: its scope, interest and potential. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1746), 4305-4321.
- Stearns, S. C., & Rodrigues, A. M. (2020). On the use of “life history theory” in evolutionary psychology. *Evolution and Human Behavior*, *41*(6), 474-485.
- Steppan, S. J., Phillips, P. C., & Houle, D. (2002). Comparative quantitative genetics: evolution of the G matrix. *Trends in ecology & evolution*, *17*(7), 320-327.
- Sterelny, K. (2017). Humans as model organisms. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1869), 20172115.
- Störmer, C., & Lummaa, V. (2014). Increased mortality exposure within the family rather than individual mortality experiences triggers faster life-history strategies in historic human populations. *PloS one*, *9*(1), e83633.
- Stuart, Y. E., Campbell, T. S., Hohenlohe, P. A., Reynolds, R. G., Revell, L. J., & Losos, J. B. (2014). Rapid evolution of a native species following invasion by a congener. *Science*, *346*(6208), 463-466.
- Teplitsky, C., Robinson, M. R., & Merilä, J. (2014). Evolutionary potential and constraints in wild populations. *Quantitative genetics in the wild*, 190-208.
- Thompson, R. (2008). Estimation of quantitative genetic parameters. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1635), 679-686.
- Vézina, H., & Bournival, J. S. (2020). An Overview of the BALSAC Population Database. Past Developments, Current State and Future Prospects. *Historical Life Course Studies*.
- Via, S., & Lande, R. (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, *39*(3), 505-522.
- Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*(1), 7-24.

- Walker, R., Gurven, M., Hill, K., Migliano, A., Chagnon, N., De Souza, R., ... & Yamauchi, T. (2006). Growth rates and life histories in twenty-two small-scale societies. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, 18(3), 295-311.
- Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*. Oxford University Press.
- West-Eberhard, M. J. (1989). Phenotypic plasticity and the origins of diversity. *Annual review of Ecology and Systematics*, 20(1), 249-278.
- West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.
- Westcott, B. (1986). Some methods of analysing genotype—environment interaction. *Heredity*, 56(2), 243-253.
- Williams, G. C. (2001). Pleiotropy, Natural Selection, and the Evolution of Senescence: Evolution 11, 398-411 (1957). *Science of Aging Knowledge Environment*, 2001(1), cp13-cp13.
- Wilson, A. J., Reale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., ... & Nussey, D. H. (2010). An ecologist's guide to the animal model. *Journal of animal ecology*, 79(1), 13-26.
- Wilson, D. S. (1997). Introduction: Multilevel selection theory comes of age. *The American Naturalist*, 150(S1), s1-s21.
- Wolak, M. E. (2012). nadv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. *Methods in Ecology and Evolution*.
- Wolak, M. E., & Keller, L. F. (2014). Dominance genetic variance and inbreeding in natural populations. *Quantitative genetics in the wild*, 104-127.
- Wolf, J. B., Brodie III, E. D., Cheverud, J. M., Moore, A. J., & Wade, M. J. (1998). Evolutionary consequences of indirect genetic effects. *Trends in ecology & evolution*, 13(2), 64-69.
- Wong, M. Y., Day, N. E., Luan, J. A., Chan, K. P., & Wareham, N. J. (2003). The detection of gene—environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement?. *International journal of epidemiology*, 32(1), 51-57.

Wray, N. & Visscher, P. (2008) Estimating trait heritability. *Nature Education* 1(1):29

Wright, S. (1932). *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R (Vol. 574, p. 574). New York: springer.

## SUPPLEMENTARY INFORMATION (CHAPITRE 3)

Walid Mawass and Emmanuel Milot (submitted) Assessing the impact of pedigree quality on the validity of quantitative genetic parameter estimates

### Pedigree filtering criteria

The BALSAC project includes over 4 million records on more than 5 million individuals over the span of almost four centuries and covering 23 different subregions of French-Canada (Québec, see Figure S2). We separated individual records based on the region of birth and reconstructed the pedigree of each subregion using this information. There was immense variation between the different subregions in terms of sample size, number of generations, number of founders. The major determining factors in quantitative genetic analyses are the size and depth of the pedigree. Hence, our first filtering criteria involved removing the datasets which had a very small sample size  $< 1,000$  individuals and pedigree depth  $< 4$  generations.

Following this, we were left with 13 pedigrees. Our second filtering criteria concerned the long processing and run time for a single model to converge in the case of pedigrees with very large sample sizes. For example, the Saguenay-Lac-St-Jean pedigree with a sample size  $> 4e+5$  required several weeks to run a single iteration of a quantitative genetic model. For this reason, we cut out at least half of the remaining pedigrees exhibiting this issue, which left us with the final 6 pedigrees used in this study. We note that even after these two filtration processes, the variation between the final 6 pedigrees was still large enough to allow for comparison and assessing the impact of differences in pedigree properties on validity (to better visualize this, we provided the visual representation of the pedigree of *Gaspésie*, a small and shallow pedigree, and the pedigree of *Bois-Francs*, a large and deep pedigree, in Figure S3 and S4 respectively).

### Calculation of genealogical properties

The pedigrees analysed in this study are based on reconstructed genealogies. To assess the impact of variability in the properties and characteristics of these genealogies, we calculated for each of our six pedigrees the following genealogical metrics:

1. *Mean completeness* (Cazes & Cazes 1996): primarily based on the completeness index of a pedigree table  $C_x$ , i.e., the ratio of the number of known ancestors to the number of expected ancestors, at each generation  $x$ . Thus, it is calculated using the following formula:  $C_x = \frac{\text{number of known ancestors}}{\text{number of expected ancestors}}$ , where the number of expected ancestors at a generation  $x$  is equal to  $2^x$ , the parents' generation being the first one. Finally, we compute the mean value for each pedigree over all the generations.
2. *Average inbreeding coefficient* (Thompson 1986): the inbreeding coefficient of an individual  $i$ ,  $f_i$ , is the probability that the homologous genes in the two gametes uniting to form individual  $i$  are identical-by-descent. Hence, the inbreeding coefficient of an individual is equal to the coefficient of kinship  $\psi$  of the individual's parents, i.e.,  $f_i = \psi(M_i, F_i)$ , where  $M_i$  and  $F_i$  are individual  $i$ 's parents. Finally, we compute the average inbreeding coefficient over the whole pedigree.
3. *Total number of founders*: in this study, we defined a founder (or a recent immigrant) as an individual missing information on both their parents and their year of birth. We also used the summary statistics from the R package *pedantics* which provides the

total number of founders which matched the raw number calculated from the data. We also had BALSAC run a confirmation analysis on the founders we identified to confirm their founder status in their database.

4. *Expected genealogical depth* (Cazes & Cazes 1996, based on Kouladjian 1986): this metric is based on an initial metric known as the genealogical entropy introduced by Kouladjian (1986). The three metrics above do not take into account the average number of generations in a pedigree table which makes comparing them between populations (or subpopulations) dubious. To allow comparison, a measure of genealogical depth of a reconstructed pedigree is needed. The measure of entropy in a given genealogy, designating the amount of information contained within it, is calculated by first transforming it into a binary tree where individuals who appear more than once as ancestors are considered as distinct individuals, and then  $p_i$  is defined as the probability of the origin of the gene coming from the founder  $i$  in this tree. Thus, the entropy parameter  $S_B$  is calculated as:

$$S_B = -\sum_i p_i \log^2 p_i = \sum_i N_i \frac{1}{2^{N_i}},$$

where  $N_i$  is the generation of founder  $i$  and  $p_i = \frac{1}{2^{N_i}}$ . This parameter gives the expected value of the founders' generation and can serve as a measure of the degree of ancestry. We also calculated the variance in this parameter which is defined as:

$$V = \sum_i p_i \log^2 p_i - S_B^2 = \sum_i N_i^2 \frac{1}{2^{N_i}} - S_B^2,$$

The calculation of the expected genealogical entropy assumes that there are no known missing links. To take into account the impact of missing links, we recalculated the expected genealogical entropy while incorporating an empirical estimate of the rate of missing links:

$$S_{B(\text{missing links})} = \sum_{g=1}^G N_g \frac{\delta^g}{2^g}$$

where  $G$  is the total number of generations over which we are calculating the entropy ( $g = 1$  corresponds to the parent generation,  $g = 2$  the grand-parent generation, etc.).  $N_g$  represents the number of ancestors available at generation  $g$ .  $\delta$  represents the rate of missing links, which is the ratio of present genealogical links to the expected genealogical links.

### Assigning the Region of Practical Equivalence (ROPE)

Under the frequentist paradigm it is common practice to test the null hypothesis  $\theta=0$ , where  $\theta$  is a population parameter of interest. However, this approach to some extent disconnects “statistical significance” and “effect size” (i.e., biological significance). In this study, we used the *ROPE* Bayesian index to re-interpret ‘animal model’ results by considering, instead of  $\theta=0$ , a range of  $\theta$  values that would be considered as not different enough from zero to be biologically meaningful (see Makowski *et al.* 2019b). The *ROPE* gives the proportion of a posterior distribution (within 95%-HPD range) that lies below the threshold (i.e., in the range of values considered as not biologically meaningful). While this is based on a subjective choice, this choice is not devoid of scientific criteria, which we make transparent here.

We selected the *ROPE* range on the basis of the parameter values used when simulating phenotypic values, i.e.,  $V_A$  was either 0.1 or 0.3. Specifically, having estimates for additive genetic variance ( $V_A$ ) and additive genetic covariance ( $Cov_A$ ) for different datasets, we asked how large the estimates of these parameters should be to be considered biologically meaningful given the initial simulated parameter values. Since what should be considered as “biologically meaningful” is partly subjective, we set different *ROPE*. First, we decided to

apply three different ROPE instead of a single one to allow both more conservative (higher ROPE) and more liberal (lower ROPE) interpretations, though in the main text we only report the moderate interpretation. The first one is 5% of the magnitude of the parameter estimate. For example, if simulated  $V_A$  was 0.3, then any estimate  $\geq 0.015$  would be considered as non-negligible, which might be viewed as liberal. We also applied a conservative 25% threshold, meaning that a substantial fraction ( $\geq 1/4$ ) of a parameter estimate should be within the range to be considered as meaningful. Finally, we also applied a moderate 10% threshold which is the measure reported in the main text.

## Tables

**Table S1.** RMSE values calculated for each pedigree for each scenario and simulated parameter value. Values are posterior modes of the derived posterior distributions of RMSE, along with 95% highest posterior density (HPD) intervals in brackets. The asterisk corresponds to estimates based on an average of 5 simulations.

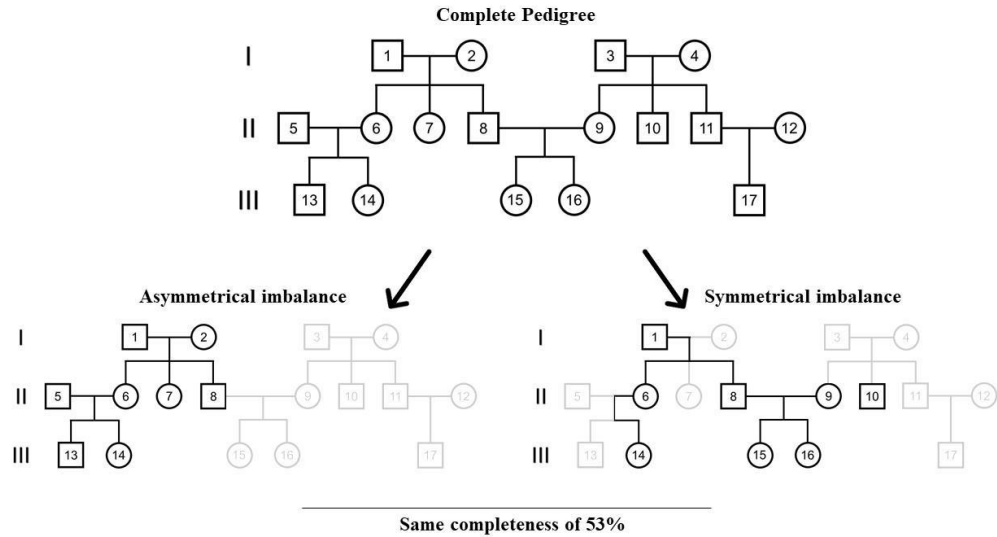
Scenario	Gaspésie	Bas-St-Laurent	Laurentides	Charlevoix	Côte-de-Beaupré	Bois-Francs
Scenario 1 ( $V_A=0.1$ )	0.043 [0.009 - 0.118]*	0.026 [0.002 - 0.076]*	0.015 [0 - 0.066]*	0.016 [0 - 0.042]*	0.005 [0 - 0.035]*	0.011 [0 - 0.037]*
Scenario 1 ( $V_A=0.3$ )	0.024 [0 - 0.157]*	0.039 [0 - 0.115]*	0.008 [0 - 0.080]*	0.023 [0 - 0.078]*	0.015 [0 - 0.058]*	0.012 [0 - 0.059]*
Scenario 2 ( $V_A=0.1$ )	0.107 [0.002 - 0.189]*	0.085 [0.002 - 0.120]*	0.062 [0.010 - 0.099]*	0.017 [0 - 0.067]*	0.007 [0 - 0.050]*	0.019 [0 - 0.069]*
Scenario 2 ( $V_A=0.3$ )	0.089 [0.003 - 0.261]*	0.082 [0.013 - 0.200]*	0.008 [0 - 0.136]*	0.029 [0.002 - 0.088]*	0.015 [0 - 0.077]*	0.028 [0.007 - 0.088]*
Scenario 3 ( $V_A=0.1$ )	0.080 [0.033 - 0.098]*	0.069 [0.025 - 0.090]*	0.054 [0.014 - 0.075]*	0.043 [0.014 - 0.065]*	0.054 [0.032 - 0.075]*	0.048 [0.027 - 0.070]*
Scenario 3 ( $V_A=0.3$ )	0.156 [0.071 - 0.250]*	0.149 [0.087 - 0.218]*	0.148 [0.088 - 0.203]*	0.133 [0.093 - 0.167]*	0.139 [0.104 - 0.167]*	0.146 [0.113 - 0.172]*
Scenario 3 ( $V_A=0.1$ with $r_g=0.1$ )	0.013 [0 - 0.088]	0.009 [0 - 0.045]	0.013 [0 - 0.064]	0.004 [0 - 0.042]	0.004 [0 - 0.039]	0.003 [0 - 0.028]
Scenario 3 ( $V_A=0.3$ with $r_g=0.1$ )	0.015 [0 - 0.151]	0.019 [0 - 0.091]	0.007 [0 - 0.078]	0.011 [0 - 0.059]	0.007 [0 - 0.049]	0.008 [0 - 0.046]
Scenario 3 ( $V_A=0.1$ with $r_g=0.5$ )	0.102 [0 - 0.215]	0.062 [0 - 0.127]	0.003 [0 - 0.045]	0.007 [0 - 0.045]	0.004 [0 - 0.033]	0.011 [0 - 0.036]
Scenario 3 ( $V_A=0.3$ with $r_g=0.5$ )	0.011 [0 - 0.122]	0.035 [0 - 0.140]	0.076 [0 - 0.149]	0.032 [0 - 0.097]	0.005 [0 - 0.051]	0.054 [0 - 0.090]
Scenario 3 ( $Cov_A=0.01$ with $r_g=0.1$ )	0.008 [0 - 0.062]	0.005 [0 - 0.037]	0.003 [0 - 0.044]	0.003 [0 - 0.025]	0.003 [0 - 0.027]	0.003 [0 - 0.021]

Scenario 3 ( $Cov_A=0.0173$ with $r_g=0.1$ )	0.072 [0 - 0.133]	0.005 [0 - 0.069]	0.005 [0 - 0.052]	0.024 [0 - 0.046]	0.004 [0 - 0.030]	0.005 [0 - 0.036]
Scenario 3 ( $Cov_A=0.05$ with $r_g=0.5$ )	0.009 [0 - 0.074]	0.016 [0 - 0.067]	0.014 [0 - 0.043]	0.005 [0 - 0.027]	0.009 [0 - 0.034]	0.005 [0 - 0.029]
Scenario 3 ( $Cov_A=0.0866$ with $r_g=0.5$ )	0.051 [0 - 0.094]	0.006 [0 - 0.066]	0.005 [0 - 0.061]	0.010 [0 - 0.048]	0.012 [0 - 0.043]	0.031 [0 - 0.052]

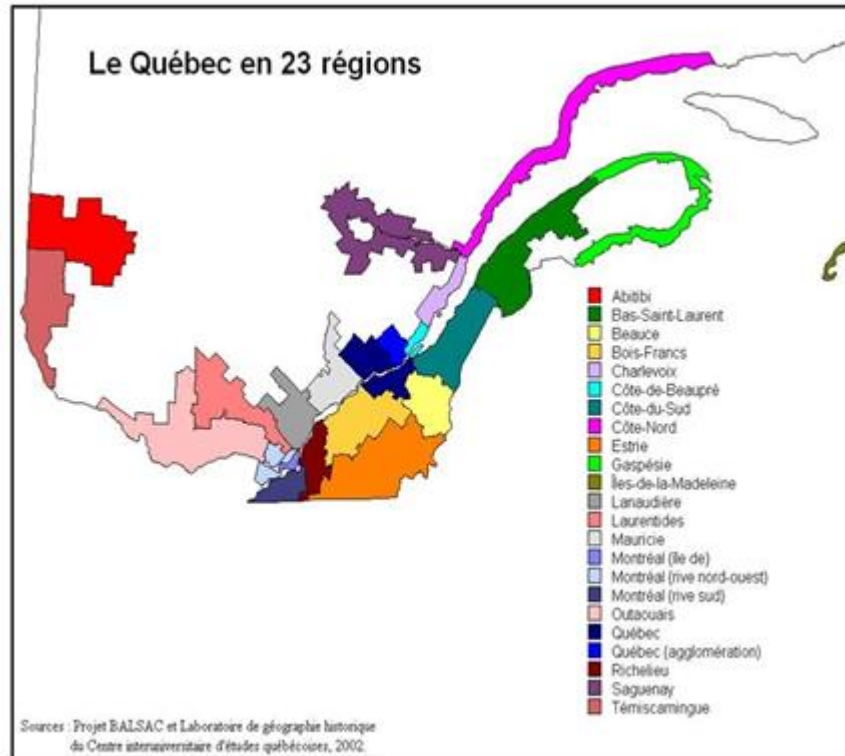
---



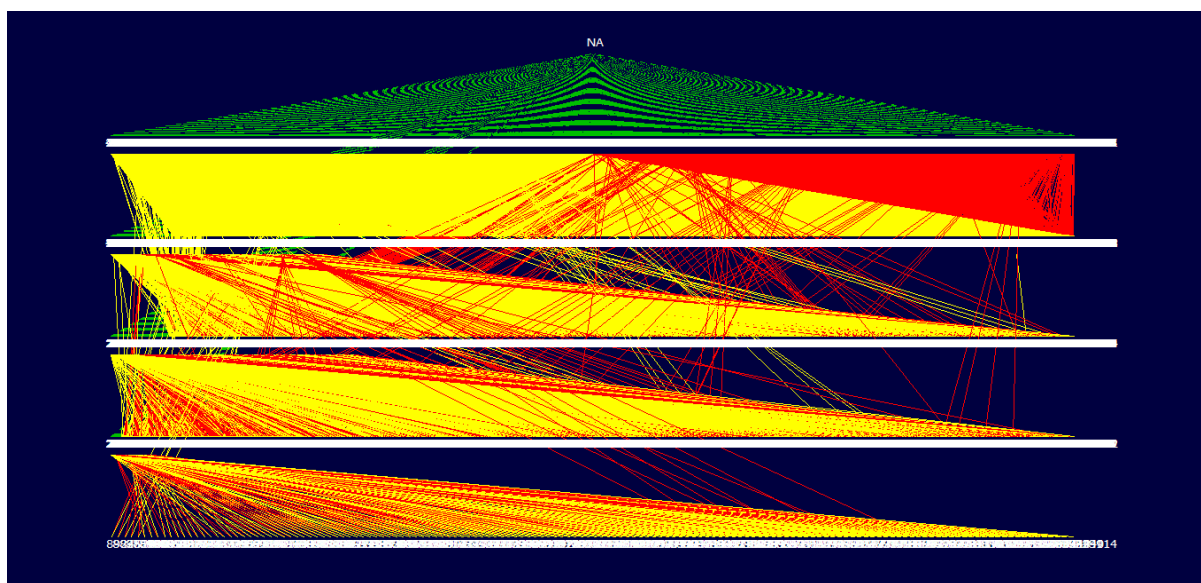
## FIGURES



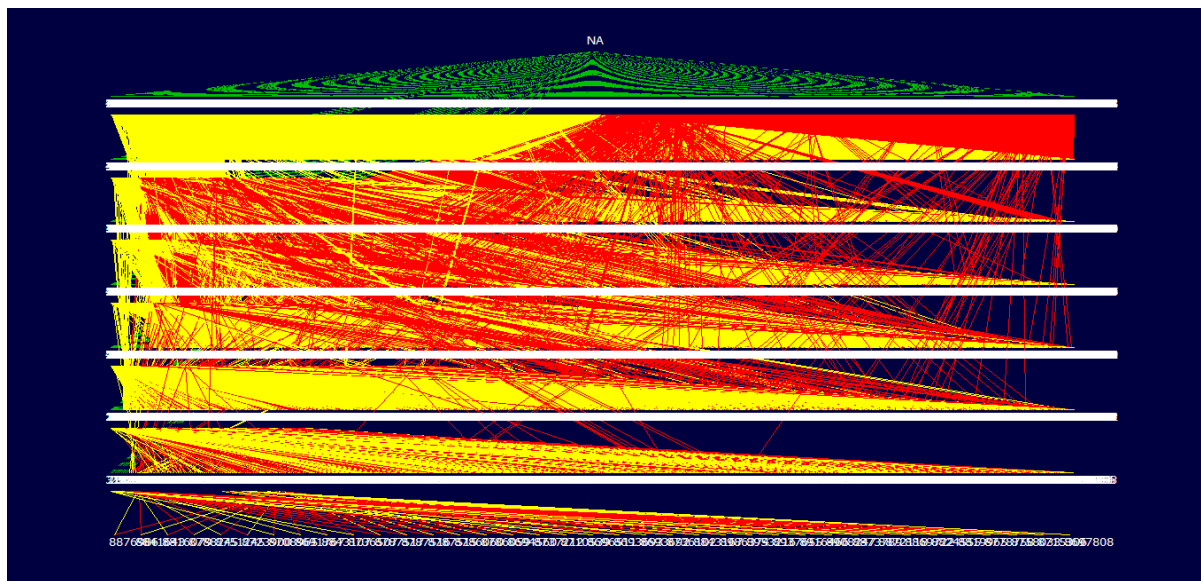
**Figure S1.** Illustration of the two types of pedigrees resulting, for the same degree of completeness, in either an asymmetrical or symmetrical incomplete pedigree.



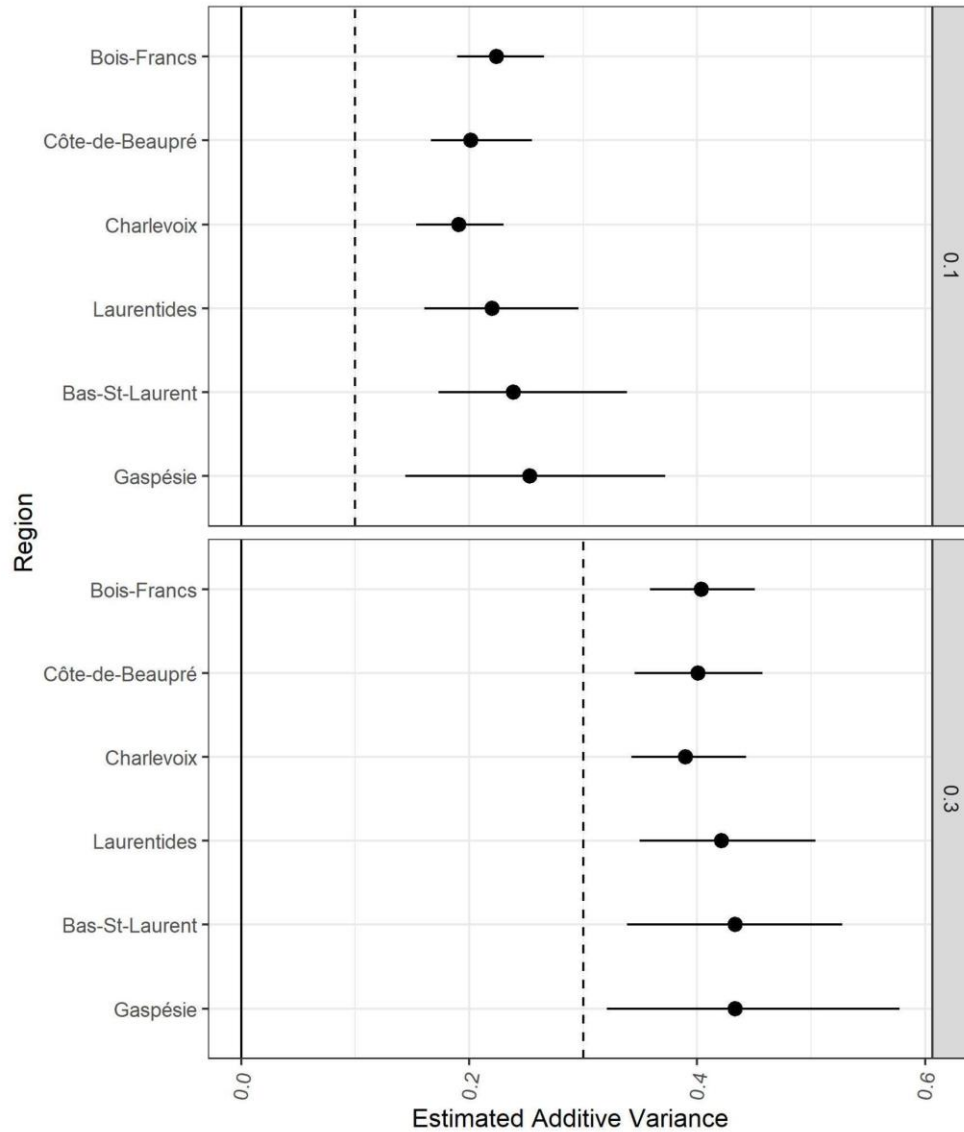
**Figure S2.** Map representing the 23 regions of French-Canada (Québec) for which the BALSAC project contains genealogical records. *Source: Project BALSAC and the laboratory of historical geography at the Inter-University Center for Québec Studies.*



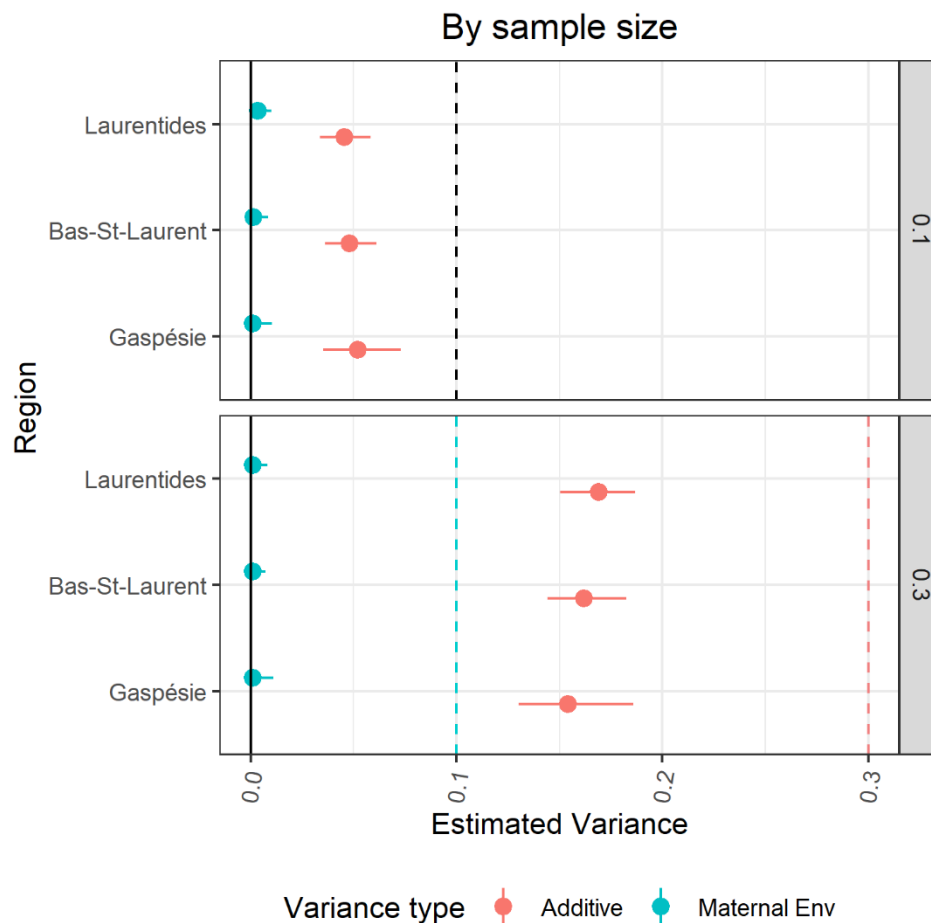
**Figure S3.** Visual representation of the pedigree of the subregion of *Gaspésie*. Red lines correspond to paternal links and yellow lines correspond to maternal links. Green lines correspond to individuals with no known parentage. Drawn using the program *Pedigree Viewer*



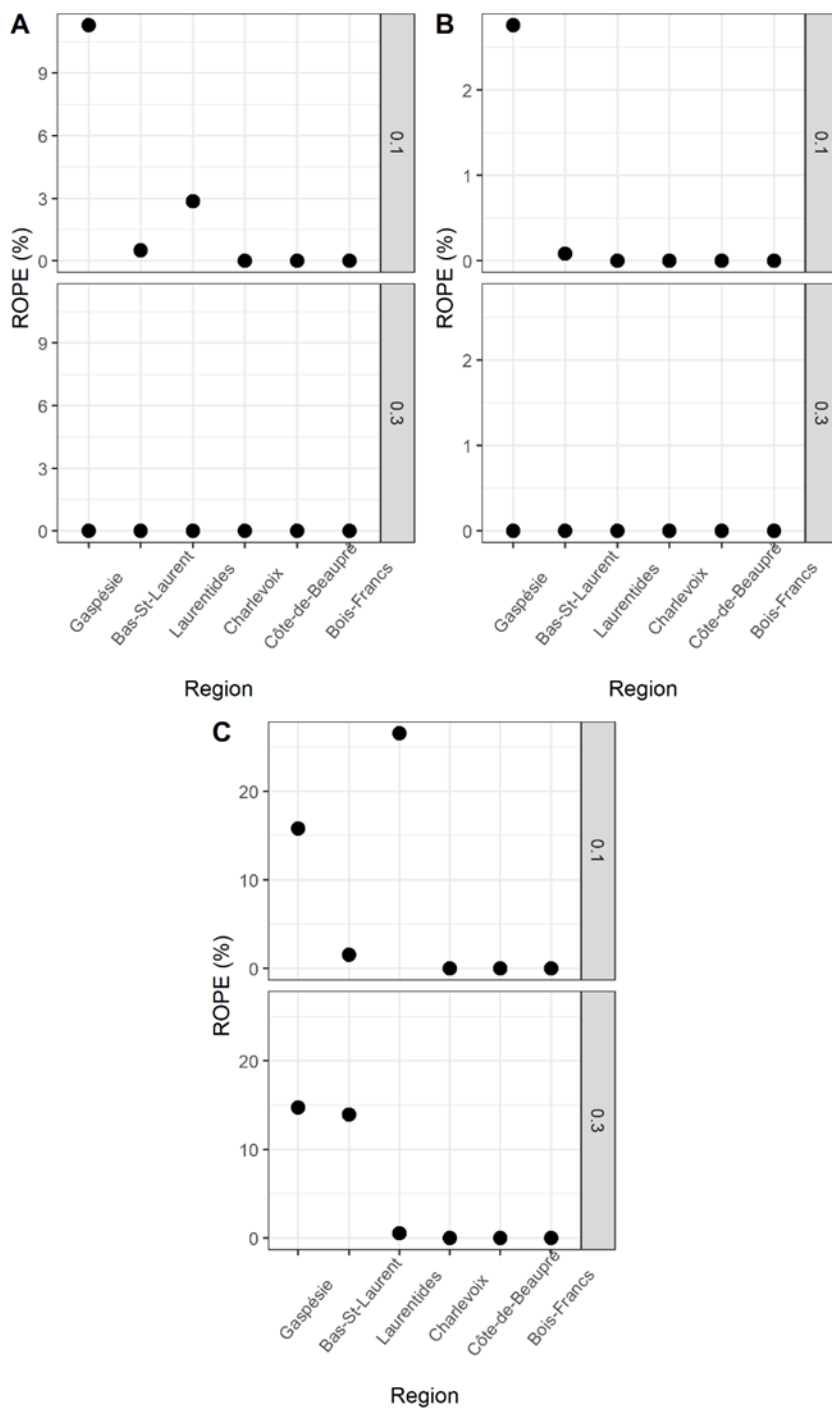
**Figure S4.** Visual representation of the pedigree of the subregion of *Bois-Francs*. Red lines correspond to paternal links and yellow lines correspond to maternal links. Green lines correspond to individuals with no known parentage. Drawn using the program *Pedigree Viewer*.



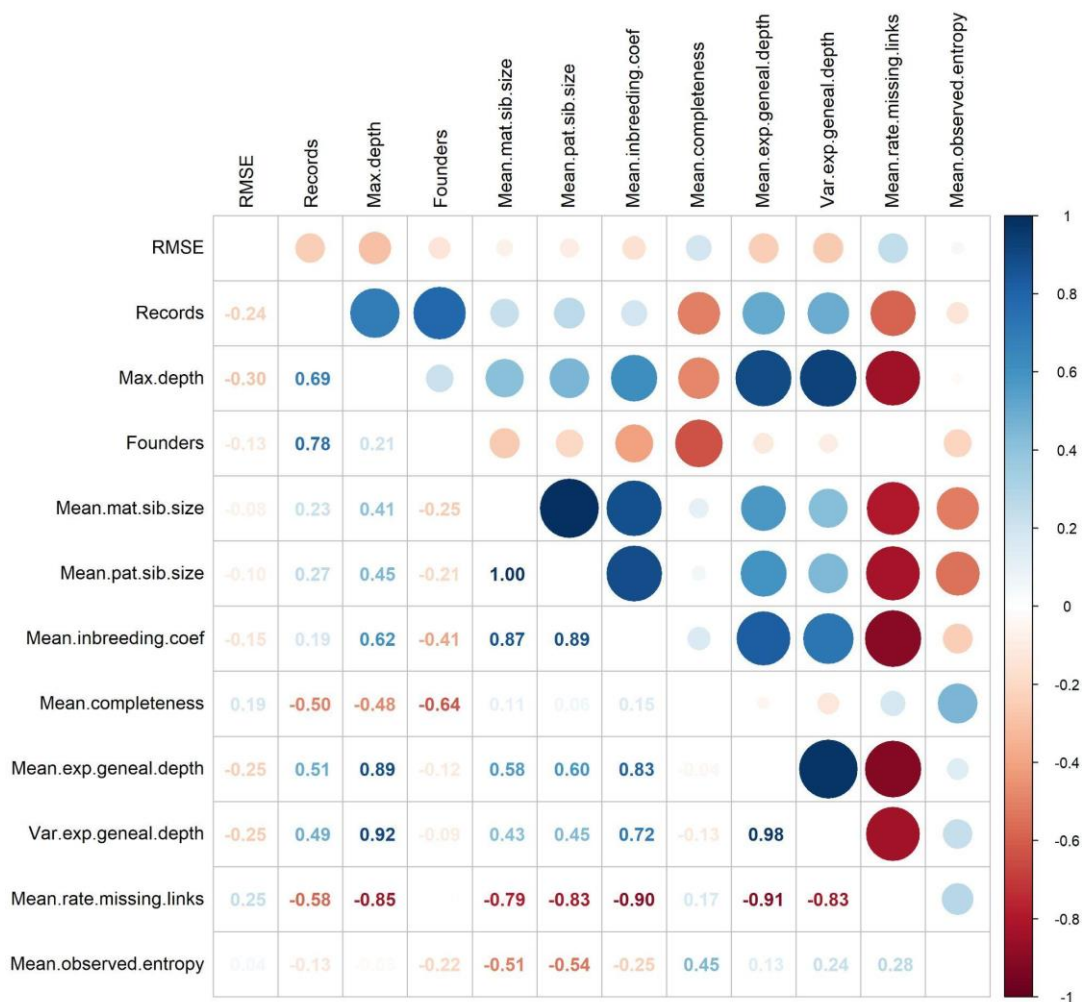
**Figure S5.** Estimates of additive genetic variance for each of the six pedigrees based on Scenario 2 (univariate complex phenotype;  $V_A = 0.1$  and  $0.3$  [represented by the values in the gray box];  $V_{MG} = 0.1$ ) when  $V_{MG}$  is improperly not estimated by the model. Each dot represents the average posterior mode estimates, along with the average 95% credible intervals based on the five simulations per pedigree per parameter value. Vertical dashed lines represent the expected value of the variance



**Figure S6.** Estimates of additive genetic variance and maternal variance for three of the pedigrees (due to long and prohibitive computation time of using larger pedigrees on the order of several weeks) based on Scenario 2 (univariate complex phenotype;  $V_A = 0.1$  or  $0.3$ ;  $V_M = 0.1$ ) using less restrictive datasets which increase the average maternal sib size. Each dot represents the average posterior mode-estimates, along with the average 95% credible intervals based on the five simulations per pedigree per parameter value. Vertical dashed lines represent the expected value for each variance component.



**Figure S7.** Bayesian index, ROPE, for each of the six pedigrees averaged over five simulations. **A.** additive genetic variance, Scenario 1 (univariate simple phenotype,  $V_A = 0.1$  and  $0.3$  [represented by the values in the gray box]), **B.** additive genetic variance, Scenario 2 (univariate complex phenotype,  $V_A = 0.1$  and  $0.3$  [represented by the values in the gray box],  $V_M = 0.1$ ). **C.** maternal variance, Scenario 2.



**Figure S8.** Results of pairwise correlation between different pedigree and genealogical properties across all analyzed datasets and scenarios. The main focus is the correlation of these variables with RMSE, the quantitative measure of precision and accuracy calculated in this study. The list of variables included in this plot is the following (by descending order): *Records* - pedigree sample size; *Max. depth* - maximum pedigree depth; *Founders* - number of founders in the pedigree; *Mean.mat.sib.size* - average number of maternal siblings; *Mean.pat.sib.size* - average number of paternal siblings; *Mean.inbreeding.coef* - mean inbreeding coefficient across the pedigree; *Mean.completeness* - mean completeness of the pedigree; *Mean.exp.geneal.depth* - average expected genealogical depth (or entropy); *Var.exp.geneal.depth* - variance in expected genealogical depth; *Mean.rate.missing.links* - mean rate of missing links or the ratio of present genealogical links to the expected genealogical links (higher value means low number of missing links); *Mean.observed.entropy* - average observed genealogical depth (incorporating mean rate of missing links)

## SUPPLEMENTARY INFORMATION (CHAPITRE 4)

### Subfecundity hypothesis vs migration hypothesis

Briefly, some couples were excluded after confirmation of their temporary or permanent emigration from the island based on multiple nominative censuses (including a 1967 census conducted by Philippe (1973) as part of a project to establish the IAC Population Register, as well as governmental lists conducted for the second half of the 19<sup>th</sup> century). The remaining couples were classified as “sub-fecund” if at least one of their birth intervals was unusually long, i.e. protogenesic (>36 months between marriage and first birth), intergenesic (>48 months between two births), or pausigenesic (>84 months between the last recorded birth and the end of reproductive lifespan; Henry 1958).

### Sliding window method

We varied the time windows sequentially in two ways: 1- by changing the starting point of the window relative to the birth year or BY (e.g. 0 means window starts at BY; 5 means window starts 5 years post BY), 2- by changing the length of the window (e.g. +0 means window includes only starting year; +4 means window includes starting year plus the 4 subsequent years). The correlation between IMR and AFR was calculated separately for each year from birth to the 9<sup>th</sup> year of life. The strongest relationship was captured both by the BY and the third year following birth (i.e. window 3+0; Table S2). To complement this purely statistically-based criterion, we also used two windows based on biological criteria: window 0+1 represents the period of life during which infant mortality risk is the highest (rate of infant deaths occurring before 1 yo at IAC was around 10%); window -1+2 includes as well the year before birth, to capture potential effects of environmental conditions during fetal growth. Results obtained with the biological-based windows were qualitatively similar to those obtained with statistically-based windows.

### Tables

**Table S1.** Estimated Pearson’s correlation coefficients between the average AFR and average IMR for different time windows. These windows were generated sequentially

by changing the starting year relative to birth year and the length of the window, ranging from 0 to 5 years. Two windows showed the highest and similar coefficients (highlighted in bold). *BY 0+4* was chosen as an indicator of environment conditions during the early life of women in this study. The *BY -1+2* windows was selected based on a biological criterion and encompasses a year pre- and post-birth.

<b>Window</b> (start+length)	<b>Correlation</b> <b>coefficient</b>
Birth Year -1+2	0.081
Birth Year 0	0.121
Birth Year 0+1	0.085
Birth Year 0+2	0.093
Birth Year 0+3	0.129
<b>Birth Year 0+4</b>	<b>0.131</b>
Birth Year 1+0	0.016
Birth Year 1+1	0.050
Birth Year 1+2	0.101
Birth Year 1+3	0.106
Birth Year 1+4	0.072
Birth Year 2+0	0.063
Birth Year 2+1	0.124
Birth Year 2+2	0.121
Birth Year 2+3	0.077
Birth Year 2+4	0.071
<b>Birth Year 3+0</b>	<b>0.133</b>
Birth Year 3+1	0.121
Birth Year 3+2	0.064
Birth Year 3+3	0.058
Birth Year 3+4	0.091
Birth Year 4+0	0.060
Birth Year 4+1	0.001
Birth Year 4+2	0.008
Birth Year 4+3	0.054
Birth Year 4+4	0.051
Birth Year 5+0	-0.059
Birth Year 5+1	-0.028
Birth Year 5+2	0.037
Birth Year 5+3	0.036
Birth Year 5+4	0.033
Max. corr. coef.	0.133
Min. corr. coef.	-0.059



**Table S2.** Results of bivariate animal models fitting unstructured covariance matrices with IMR group as a random categorical interaction based on a biological-based window ( $BY -1+2$ ).  $Va$ : additive genetic variance;  $Vm$ : shared familial environment variance;  $Vr$ : residual variance;  $h^2$ : narrow-sense heritability;  $r_g$ : genetic correlation. Values are posterior modes for each parameter, along with 95% highest posterior density (HPD) intervals in parentheses.

Trait	IMR group	$Va$	$Vm$	$Vr$	$h^2$	Genetic correlation $r$
AFR	Low	0.31 (0.10 - 0.56)	0.14 (0.07 - 0.34)	0.43 (0.20 - 0.66)	0.27 (0.11 - 0.56)	-0.47 (-0.76 to 0.12)
	Middle	0.41 (0.15 - 0.61)	0.18 (0.09 - 0.37)	0.47 (0.26 - 0.63)	0.32 (0.14 - 0.55)	-0.31 (-0.68 to 0.22)
	High	0.47 (0.17 - 1.05)	0.34 (0.15 - 0.72)	0.55 (0.20 - 0.95)	0.24 (0.11 - 0.64)	-0.57 (-0.86 to 0.05)
Fitness	Low	0.19 (0.08 - 0.44)	0.23 (0.11 - 0.44)	0.31 (0.13 - 0.45)	0.24 (0.12 - 0.46)	-
	Middle	0.23 (0.08 - 0.43)	0.17 (0.09 - 0.31)	0.43 (0.30 - 0.63)	0.24 (0.11 - 0.46)	-
	High	0.32 (0.11 - 0.83)	0.24 (0.11 - 0.61)	0.58 (0.29 - 0.95)	0.34 (0.10 - 0.55)	-

**Table S3.** Summaries of regression models arranged in the following order: Linear regression model with AFR as the response variable and IMR as a covariate; Linear regression model with AFR as the response variable and including the interaction between IMR and woman's birth cohort; Linear regression model with fitness as the response variable and including the interaction between IMR and woman's AFR; Generalized linear model with IMC as the response variable and including the interaction between IMR and woman's AFR.

Response variable	Predictor variable	Regression coefficient	SE	$P$ -value
AFR	IMR - Middle	0.68	0.45	0.1310
	IMR - High	1.13	0.51	<b>0.0280**</b>
	Birth cohort	-0.35	0.09	<b>&lt;0.001**</b>
	Twin (0 no, 1 yes)	-1.06	0.69	0.4475
	Intercept	24.85	0.52	<b>&lt;0.001**</b>
	Adjusted $R^2$	0.03		<b>&lt;0.001**</b>

Response variable	Predictor variable	Regression coefficient	SE	P-value
AFR	IMR - Middle	-0.49	1.14	0.6661
	IMR - High	-0.14	1.27	0.9103
	Birth cohort	-0.58	0.21	<b>0.0055**</b>
	Twin (0 no, 1 yes)	-1.08	0.69	0.1180
	Birth cohort : IMR-Middle	0.28	0.25	0.2495
	Birth cohort : IMR-High	0.31	0.27	0.2601
	Intercept	25.77	0.90	<b>&lt;0.001**</b>
	Adjusted R <sup>2</sup>	0.03		<b>0.002**</b>

Response variable	Predictor variable	Regression coefficient	SE	P-value
Fitness	AFR	-0.15	0.03	<b>&lt;0.001**</b>
	IMR - Middle	1.00	1.00	0.3151
	IMR - High	2.68	1.08	<b>0.0130**</b>
	Inbreeding coef. (linear)	30.3	13.92	<b>0.0299**</b>
	Inbreeding coef. (quadratic)	-635.6	253.9	<b>0.0126**</b>
	Birth cohort	0.07	0.04	<b>0.0977*</b>
	Twin (0 no, 1 yes)	1.40	0.27	<b>&lt;0.001**</b>
	AFR : IMR-Middle	-0.04	0.04	0.3599
	AFR : IMR-High	-0.09	0.04	<b>0.0386**</b>
	Intercept	5.94	0.84	<b>&lt;0.001**</b>
	Adjusted R <sup>2</sup>	0.26		<b>&lt;0.001**</b>

**Table S3.** (continued)

Response variable	Predictor variable	Regression coefficient	SE	P-value
IMC	AFR (linear)	4.47	3.36	0.1834
	AFR (quadratic)	8.74	4.89	<b>0.0744*</b>
	AFR (cubic)	6.95	5.70	0.2236

IMR - Middle	0.09	0.16	0.5707
IMR - High	0.09	0.17	0.6061
Birth cohort	-0.01	0.03	0.8745
Twin (0 no, 1 yes)	0.52	0.15	<b>0.0005**</b>
AFR (linear) : IMR - Middle	-2.86	4.29	0.5055
AFR (quadratic) : IMR - Middle	-9.72	5.75	<b>0.0912*</b>
AFR (cubic) : IMR - Middle	-6.24	6.24	0.3183
AFR (linear) : IMR - High	-2.23	4.34	0.6083
AFR (quadratic) : IMR - High	-7.00	5.65	0.2514
AFR (cubic) : IMR - High	-7.23	6.75	0.2840
Intercept	-2.33	0.18	<b>&lt;0.001**</b>

**\*\* $P < 0.05$ ;**

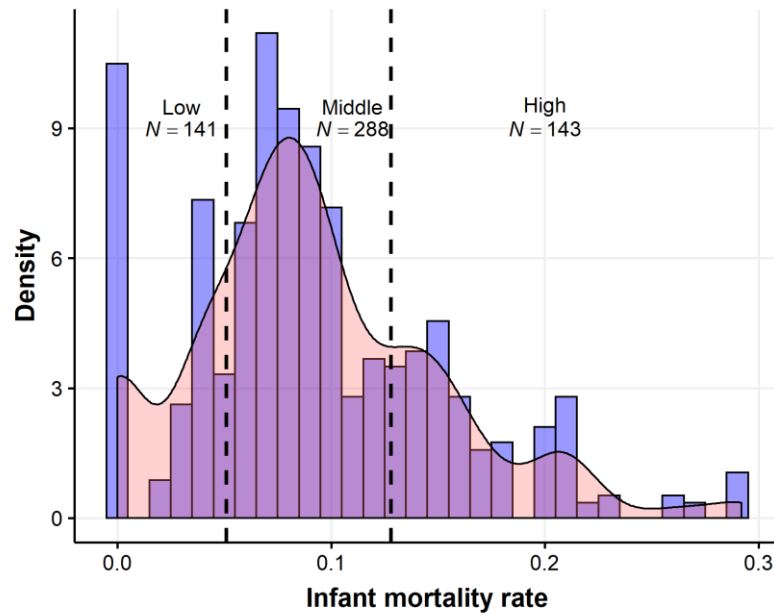
**\*  $P < 0.1$**

***Dispersion parameter = 1.369***

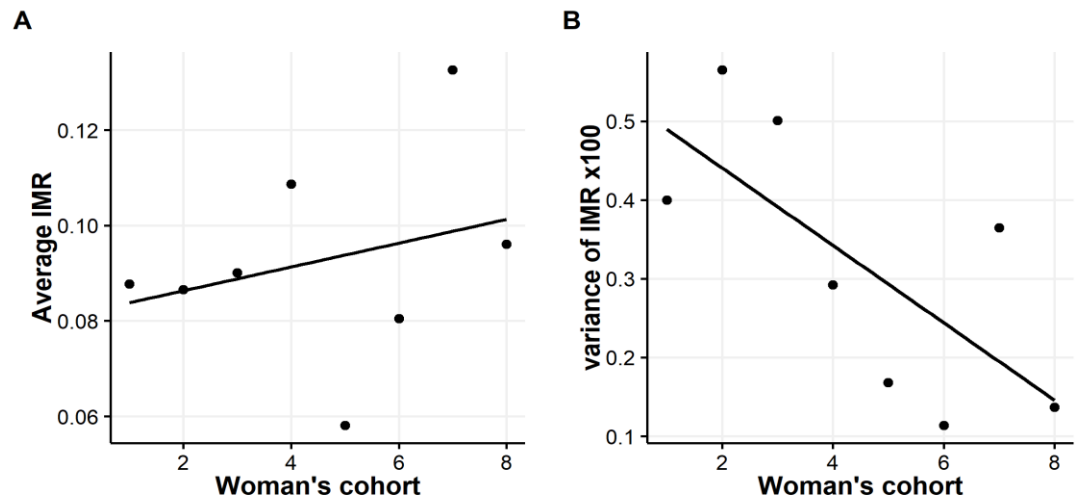
## Figures



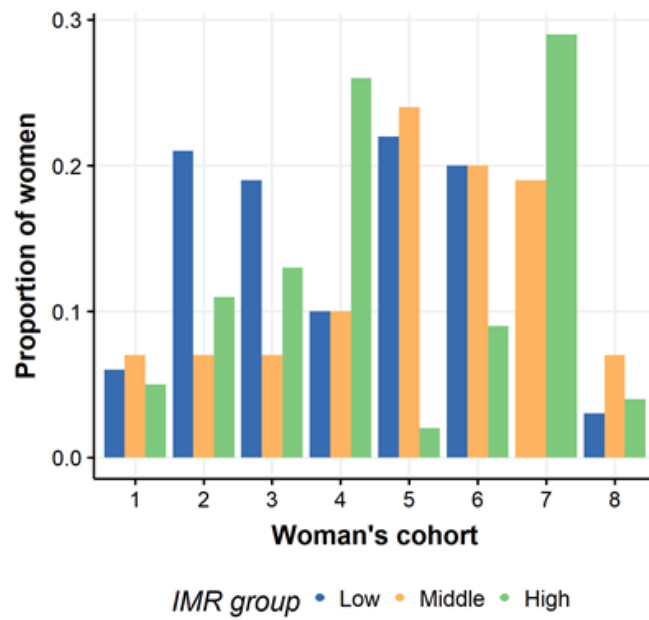
**Figure S1.** Satellite image of Ile aux Coudres located on the St. Lawrence river (coordinates: 47.4 N 70.4 W); *Google Maps*, 2018.



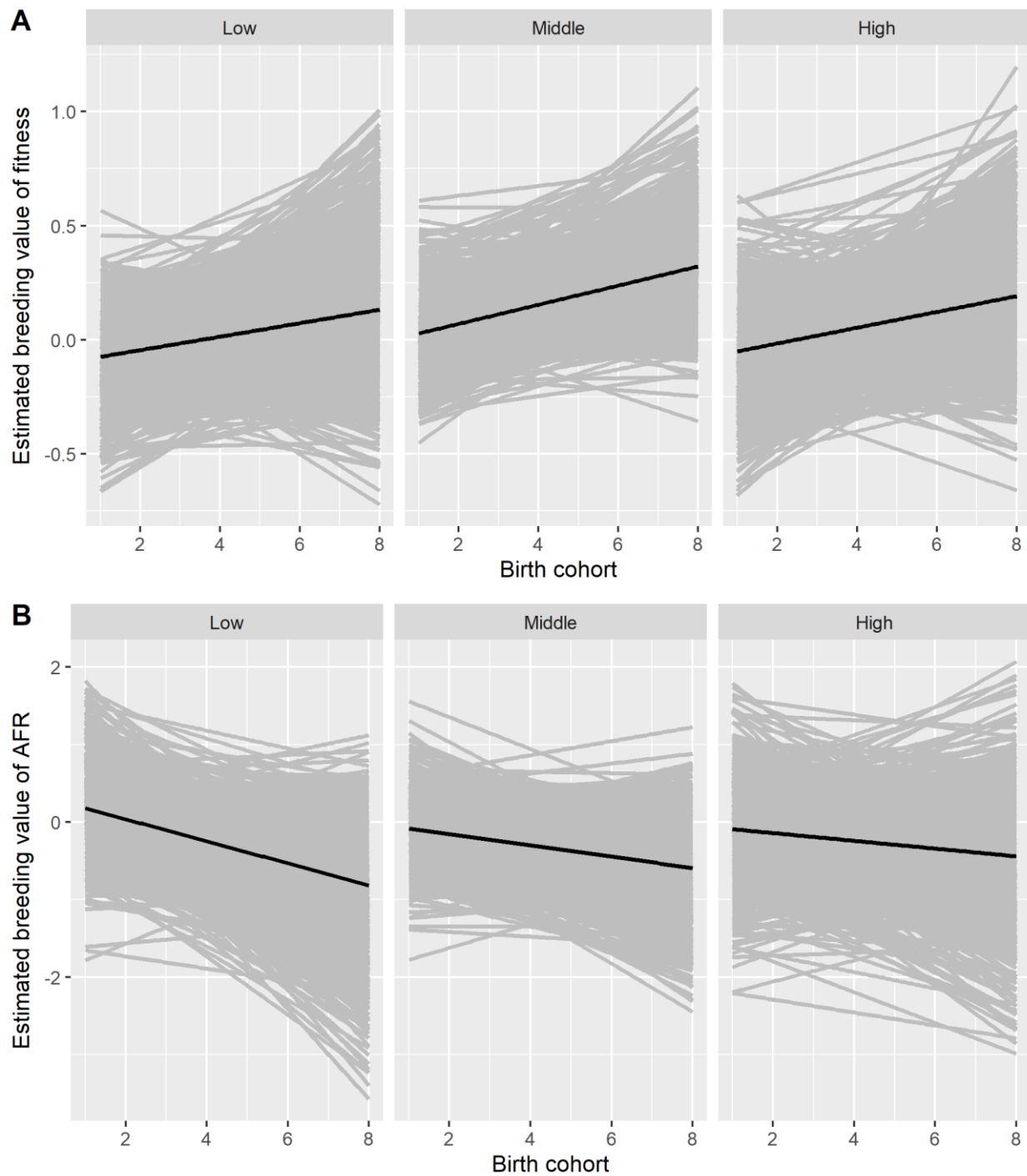
**Figure S2.** Histogram and density distribution of infant mortality rates at Île aux Coudres between 1772 and 1919 (i.e. the range of birth years of the women included in this study). The vertical dashed lines represent the specific quantile-cuts dividing the distribution into three early-life environment groups.



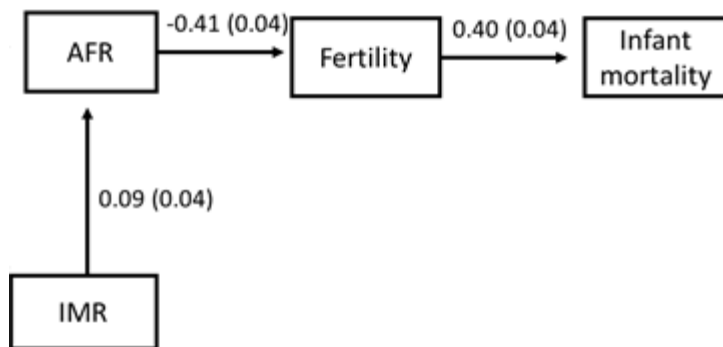
**Figure S3.** Population-level temporal trends in annual infant mortality rate (IMR) for the mean (A) and variance (B).



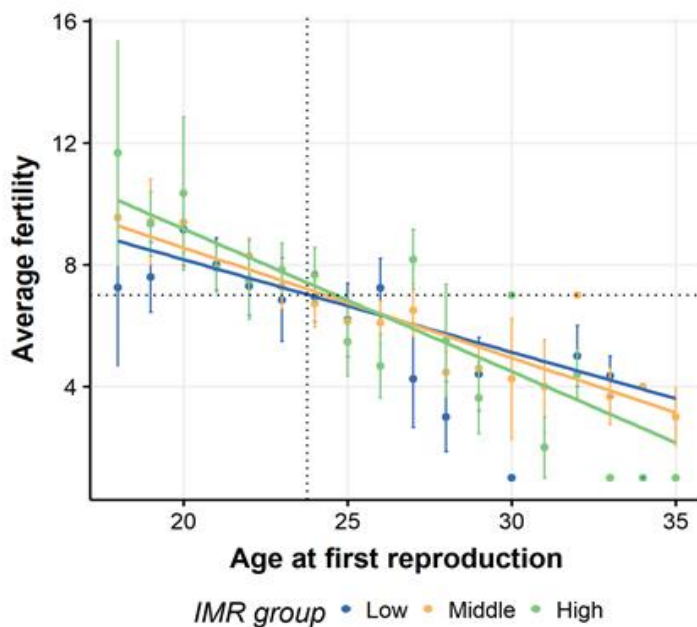
**Figure S4.** Per-cohort proportion of women in each of the three IMR groups.



**Figure S5.** Temporal trends in estimated breeding values (EBVs) for each IMR group, based on each of the 1,000 iterations retained in the MCMC sample from the bivariate animal model. **A.** Estimated breeding values of fitness; **B.** Estimated breeding values of AFR. Lines for each iteration are in grey and the black line is the average slope.



**Figure S6.** Path analysis showing the mediating effect of fertility between AFR and IMC. Values ( $\pm$ SE) next to solid arrows are standardized regression coefficients.



**Figure S7.** Interaction effect of IMR group and AFR on woman's fertility (total number of children). The regression lines are based on the best-fit generalized linear model (fertility was modeled as following a Poisson distribution). The dotted horizontal and vertical line represents the population average of fertility and AFR respectively. Vertical bars are standard errors of the mean.