Title

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

Running title

Pedigree quality impact on quantitative genetic parameter estimates

Authors

Walid Mawass [a*†] and Emmanuel Milot [b]

Institutional affiliation

[a] Department of Biomedical Sciences, University of Québec at Trois-Rivières, Québec, Canada; Email: Walid.Mawass@uqtr.ca

[b] Department of Chemistry, Biochemistry and Physics, University of Québec at Trois-Rivières, Québec, Canada; Email: Emmanuel.Milot@uqtr.ca

*Corresponding author

[†]Currently at: Department of Ecology & Evolutionary Biology, University of Arizona, Arizona, USA

Author contributions

W.M. conceived and developed the study. W.M. performed data and computational analysis with input from E.M.. W.M drafted the manuscript. E.M. contributed to the final version of the manuscript. E.M. supervised the project.

Conflict of interest

Author and co-author declare no conflict of interest.

Abstract

Investigating the evolutionary dynamics of complex traits in nature requires the accurate assessment of their genetic architecture. 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. However, natural systems will differ in many respects (e.g., mating system, data availability, pedigree structure), thus it can be inappropriate to generalize outcomes from one system to another. French-Canadian (FC) genealogies are extensive and deep-rooted (up to 9 generations in this study) making them ideal to study how the quality and properties (e.g., errors, completeness) of pedigrees affect QG estimates. We conducted simulation analyses to infer the reliability of QG estimates using FC pedigrees and how it is impacted by genealogical errors and variation in pedigree structure. Broadly, results show that pedigree size and depth are important determinants of precision but not of accuracy. While the mean genealogical entropy (based on missing links) seems to be a good indicator of accuracy. Including a shared familial component into the simulations led to on average a 46% overestimation 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: additive (co)variance, sensitivity analysis, animal model, natural populations, 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), including in contemporary human populations (reviewed in Stearns et al. 2010 and Bolund et al. 2016).

Pedigrees provide information on the expected pairwise relatedness between individuals, as measured by the kinship coefficient Θ_{ij}, namely the probability that an allele randomly drawn from individual *i* is identical-by-descent to an allele 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. Pedigree-based QG analyses strongly

depend on the provided pedigrees, which can vary much for natural, hence uncontrolled, populations (Bonnet et al. 2022). The validity of a method is its capacity to correctly measure what it should measure. In the context herein, we define it as the capacity of a QG method to estimate parameters of interest (i.e., genetic) with acceptable accuracy and precision, and over the range of their potential values, given the pedigree. In other words, is it valid to fit animal models to this particular pedigree? This operational definition is tailored to our goal, which is to study how good pedigrees (with given attributes) can be for the estimation of genetic parameters (as opposed to how good the animal model is in general, i.e., across pedigrees). While what is considered as "acceptable accuracy and precision" may vary among people or contexts, for the purpose of this study we will consider them as acceptable when they allow the reaching of correct qualitative conclusions.

Pedigree attributes influencing the validity of a QG analysis can be roughly divided into two categories: those related to data, such as completeness and errors, which reflect our state of knowledge about the true pedigree; those related to topology, which result from the biology of the population itself (see Table 1 for a summary of concepts). Regarding the first category, 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 perfectly matches genetic relatedness, recalling that a pedigree provides expected (mean) relatedness values, not accounting for the variance introduced by Mendelian segregation and recombination. Thus, obtaining errorless relatedness values is

unrealistic (Hill & Weir 2011), both for captive stocks in animal breeding programmes (Visscher et al 2002) and natural populations (Speed and Balding 2015). Pedigree errors and incompleteness generally reduce the extent to which relatives appear to resemble each other, causing parameter underestimation. For example, in captive breeds of Hereford cattle, Senneke et al. (2004) showed that estimates of heritability in two traits, birth weight and weaning weight, decreased linearly as the rate of sire misidentification increased.

Pedigree topology reflects the properties of the natural population *per se*, such as its mating system, life history, and vital rates. For example, a polygynous population will exhibit a greater variance in male reproductive success than a monogamous population (all else being equal) which will lead to different kinship structure in their pedigree and, consequently, different amount of information available to infer parameters. In addition, it will not be possible to separate sire and dam effects in a monogamous species (unless monogamy is sequential, in which case information to measure these effects may still remain quite limited).

Sensitivity and power analysis (as per Morrissey et al. 2007) are two means by which one can assess the adequacy of a pedigree for QG. Sensitivity consists in determining how parameter estimates change when genealogical errors are deliberately introduced in the pedigree. Power analysis measures the capacity to recover from QG analyses preset parameter values (V_a, h^2, etc.) used to simulate breeding values along the empirical pedigree. For example, Charmantier & Réale (2005) and Firth et al. (2015) conducted simulations on social pedigrees from blue tits (*Parus caeruleus*; Charmantier & Réale 2005) and great tits (*P. major*; 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).

Bérénos et al. (2014) used three pedigrees with increasingly accurate and complete estimates of relatedness (by using SNP and whole-genome information) of an unmanaged population of Soay sheep to show that the observation-based pedigree performed as well as marker-based pedigree in terms of heritability estimates.

The impact of pedigree quality and topology on the validity of QG parameter estimates has received limited attention in evolutionary genetics. A minority of empirical studies that fitted animal models to wild population data appear to quantify how good their pedigrees are when fed to QG analyses. In a quick informal survey of such studies published over the past 10 years (n=30), we found only ~⅓ of them performed an in-depth sensitivity analysis using simulations. Others cite the aforementioned sensitivity studies, which 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 errors in genealogical links, such as those due to EPP, and 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 may differ. Morrissey et al. (2007) showed that erroneous genealogical links cause a greater underestimation of QG parameters than missing links. While Wolak & Reid (2017) examined the impact of non-random missing parentage (with respect to a genetic or phenotypic value) of non-founder individuals on the prediction of breeding values and the estimation of additive genetic variance in wild populations and found that animal model parameter estimates will be biased in cases where there are differences between the mean additive genetic values of founders and those of other individuals with unknown parents.

In this study, we used socially-constructed human pedigrees based on the genealogies of the French-Canadian (henceforth FC) historical population provided by the BALSAC database

(https://balsac.uqac.ca/), which span all Québec regions from their foundation. This study system offers two advantages: i) the possibility to separate the huge full pedigree into multiple sub-pedigrees with contrasting attributes, corresponding to 23 geographic subpopulations in order to compare them in our analyses, and ii) we have good estimates of genealogical error rates from previous gene-dropping simulations (Doyon et al. *in prep.*). Our objectives are: 1) to assess how pedigree attributes (i.e., sample size, depth, etc.) impact the validity of QG estimates; 2) evaluate the impact of different model parameterization and variance structures on the validity of QG estimates.

Materials and Methods

Pedigree quality

The sizable variation in the properties of different species and populations results in a diversity of genealogical topologies in nature, some of which being 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; Table 1). 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. An incorrect link can contribute to parameter underestimation in two ways that mirror each other: i) it may wrongly connect an individual A to another individual B, while both are in reality little related to each other. 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) it will cause the analysis to

ignore a true link (and resemblance) between the same individual A and others with whom he/she is more related than apparent in the erroneous pedigree. If not accounted for, pedigree errors can also impact QTL analyses through biased estimation of genetic map distances (Slate 2005). In certain systems, it can be difficult to establish offspring-father links because of extra-pair paternity (EPP) or limited paternal care, hence detectability. In others, parentage is practically intractable in the field for both parents, and it may be necessary to perform genetic assignment to parental pairs (see Perrier et al. 2013 for an example in a salmonid population).

Precision of QG estimates is impacted by pedigree depth, sample size, and the complexity of the genetic architecture underlying the focal trait (Morrissey et al. 2007; de Villemereuil et al. 2013). Two pedigrees of identical size will differ in the precision of their estimates if they differ in the degree of connectivity between individuals, the pedigree with a weaker connectivity providing less precise estimates (Wilson et al. 2010). Accuracy is the proximity between the estimated value and the true value. All pedigrees, even those from controlled animal breeding designs, are likely to contain link errors affecting the accuracy of estimation (Charmantier & Réale 2005; Morrissey et al. 2007; Firth et al. 2015).

A pedigree's completeness varies when different lineages ascending from a focal individual are not uniformly covered (for example, when both paternal grand-parents are known for an individual but one only maternal grand-parent). This is not a property of the population itself, but instead results from the incompleteness of available data on which the pedigree is based. This incompleteness can manifest in two ways: i) when genealogical links between pairs of individuals (i.e. parent-offspring) are missing at random, i.e. independently of other missing links in the pedigree, which lead to a symmetrical incomplete pedigree, or ii) when links are not missing at random given that there could be individual heterogeneity in being included in the dataset or not

(e.g. due to among-individual genetic differences in dispersal propensity, individuals that tend to disperse far are less likely to be included in the pedigree), giving an asymmetrical incomplete pedigree (see Figure S1 for an illustration). Missing links also occur as a result of incomplete genealogical coverage (i.e., not due to their replacement by erroneous links). A pedigree said to be complete over *G* generations is one that contains all ancestors of the focal (i.e., phenotyped) individuals back to the first generation covered, as well as no missing genealogical links. For example, certain individuals have no known parentage because either i) they are founders or immigrants or ii) they were never successfully linked to possible parents in the reconstruction of the genealogy.

For a number of modern human populations, historical information on births, deaths, and marriages, available from civil or parish registers, was used to reconstruct genealogies that were subsequently exploited in QG studies (Bolund et al. 2016; see Stearns et al. 2010 for a list of medical- and clinical-based multigenerational datasets). However, since parental assignments are entirely reconstructed from social records, a pedigree of a human population can contain a non-trivial number of errors resulting from EPP, hidden adoptions, false parental declarations, record matching errors, and so forth.

Study system

The pedigrees used in this study were reconstructed from the history of Catholic individuals and families who lived in Québec, 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 progressively compiled and linked since the 1970s in two large databanks: BALSAC (http://balsac.uqac.ca/; Bouchard 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 event records representing the lives of 5 million individuals. Those dating from 1621 are duplicate copies of church registers, while those dating from 1900 onwards are from civil registers.

It is important to note that this dataset does not represent a homogenous population, rather it encapsulates the colonization history of all of Québec which occurred in multiple waves and expanded into different regions in different time periods. BALSAC contains information about the locality or parish origin of records, which are grouped into 23 regions (see http://balsac.uqac.ca/ for more details and Figure S2). The construction of pedigrees was based on the region of birth of individuals. We ended up with the reconstructed pedigrees for 23 subpopulations each with its own particular structure and properties. In this way, we were able to compare how differing pedigree qualities and genealogical properties impacted validity in QG analysis. We defined criteria to select six regions having suitable pedigrees for the analyses conducted here (see SI for details on the criteria): Bas-St-Laurent, Bois-Francs, Charlevoix, Côte-de-Beaupré, Gaspésie, Laurentides. Altogether, these regions comprise 5% of individuals in BALSAC, or around ~250,000 individuals in total (Table 2 provides quantitative descriptors of each pedigree).

We calculated the degree of missing parentage for each pedigree analyzed in this study as the number of individuals with no known parentage over the total number of individual records. The range of the degree of missing parentage extends from 5% to 31%. It is worth noting that the pedigree with the highest degree of missing parentage is the one with the highest number of founders. If we exclude those individuals which have been identified as founders, the degree of

missing parentage drops dramatically in most of the pedigrees. This means that most non-founder individuals in our datasets were successfully assigned parents (with a chance that this assignment might be erroneous). Additionally, based on summary statistics, pedigree records of paternities and maternities (including grandparentage) are equal within each pedigree, meaning there is no imbalance in terms of missing parentage assignments (Table 2).

Historical civil and church registers have provided information to reconstruct pedigrees for entire human populations. This is done by linking acts such as baptisms, marriages, and death. For such pedigrees, over-linkage refers to incorrect genealogical links made between individuals (i.e., wrong biological parental assignments). They can result from hidden adoptions by social (non-biological) parents or false parentage declarations, for example. Under-linkage refers to missing parental assignments due to information that is lacking or unclear on historical acts. For more information on linkage and data quality, we refer the reader to internal BALSAC documents available at https://balsac.uqac.ca/bibliographie-selective/, in addition to published works such as Vézina & Bournival (2020) and Vézina et al. (2018).

Using haplogroup genetic data, the genealogical error rate related to overlinking for BALSAC data was estimated to be 0.55% in maternal genealogical lineages and 0.82% in paternal genealogical lineages (Jomphe 2011). This was confirmed recently by Doyon et al. (*in prep.*), who analyzed mitochondrial DNA sequences (matrilineal inheritance) and Y-chromosome markers (patrilineal inheritance) from living individuals connected to the Québec genealogy. Conducting gene dropping simulations, they estimated that 0.80% (s.e.=0.07%) of father-offspring links are erroneous, which includes 0.41% of extra-pair paternities (EPP). As for mother-offspring links, the error rate was estimated to be 0.383% (s.e.=0.009%). Since a pedigree necessarily contains equal numbers of mother- and father- offspring links, the total error rate is thus expected to be

~0.6%. Note that a mismatch between parent's and offspring's DNA need not reflect a linkage error but could be due to a mutation. Therefore, the per-marker mutation rate was also accounted for in the estimation of the above genealogical error rates. We used the latter in the analyses of pedigrees performed here. Nonetheless, to check the sensitivity of the results to these parameter values, we also ran the same analyses assuming a 1% error rate for both parental lines, i.e., nearly twice the overall rate estimated by Doyon et al. (see SI for these results).

Pedigree simulation

The pedigrees analyzed in this study are considered 'observed' pedigrees, potentially containing a non-negligible number of errors. To assess the impact of pedigree properties, we have to compare parameter estimates given by the observed pedigree of phenotypic data simulated using a hypothetical 'true' pedigree. We use *a priori* information on the types and rates of genealogical errors to simulate plausible 'true' pedigrees. Note that the term 'true' refers to the true pedigree structure relative to the phenotypic data that are used in power and sensitivity analysis and does not necessarily reflect a pedigree that actually exists in nature (Morrissey et al. 2007). The 'true' pedigree then is completely based on the observed pedigree and the only difference between them is that the former is considered to contain no errors and is based on the true breeding structure of the population under consideration (Morrissey & Wilson 2010). We refer readers to the framework outlined in Morrissey et al. (2007) which we based our study on.

The specific analysis framework is called the 'reverse' approach which we implemented here using the R package *pedantics* (Morrissey & Wilson 2010). This package was designed for power and sensitivity analysis such as the one conducted here. We start with an observed pedigree containing errors from which we simulate a plausible 'true' pedigree, based on information about

rates and types of errors. Phenotypic data is simulated across the simulated ‘true’ pedigree. We simulated plausible ‘true’ pedigrees for each of the final six regions using the `rpederr` function. The function works as follows according to the *pedantics* manual: “Given estimates of individual life histories and rates and patterns of errors in pedigree links, `rpederr` probabilistically assigns "true" parents given an incomplete and potentially erroneous pedigree”. We input to this function the original pedigree dataset integrated with specifically required variables assigned for each individual included in the dataset which we outline below.

First, each known social parent-offspring dyad 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. It is important to note here that the assignment of a value of 1 does not mean that the parental assignment is 100% erroneous, rather setting this value helps us achieve two things: i) provide an indication to the simulating function that there is missing or null parental assignment, and ii) integrate sampled individuals with missing parental assignment in the process of simulating plausible ‘true’ pedigrees. In the case of founders (or recent immigrants), which were defined as any individual with no known year of birth and no parental assignment, we also assigned them an ‘erroneous’ parental assignment probability of 1. Additionally, to increase confidence in our criteria, we confirmed with BALSAC the founder/immigration status of individuals to ensure that precise information was provided to the function. Finally, the sex, year of birth, and cohort of first and last offspring were included as required variables to assign probable parentages based on this information. The final output of this function was a plausible ‘true’ pedigree which served as the dataset input for the next step of simulating individual phenotypic trait values.

We also note that, before simulating 'true' pedigrees, the datasets used in our analyses were subjected to a censoring filter based on the size of the maternal and paternal lineages (lineages including < 10 individuals were excluded). Such small lineages can be due either to a reproductive end of the lineage or that no new individual from the same lineage was born in the same designated region, i.e., possible emigration. This filtering allowed us to remove the small lineages which would have increased imbalance within the pedigree. Additionally, since we were interested in how genealogical properties could affect the reliability of QG estimates, we calculated for each of the six datasets, using the R package *GENLIB* (Gauvin et al. 2015; see SI for details on how each metric was calculated), the following metrics: i) the mean completeness of the genealogical data (Cazes & Cazes 1996), ii) the average inbreeding coefficient (Thompson 1986), iii) the total number of founders, and iv) the mean genealogical entropy given the observed rate of missing links (based on Kouladjian 1986, also see Cazes & Cazes 1996 and Milot et al. 2011).

The animal model

The animal model is a mixed-effects linear regression model which deconstructs phenotypic trait variation into its genetic and non-genetic variance components. In this model, the individual's breeding value is modeled as the deviation of the phenotypic value from the population average explained by the additive effects of the individual's genes. In its simplest form, the phenotype *y* of an individual *i* can be written as: $y_i = \mu + a_i + e_i$, where μ is the population mean, a_i is the additive genetic value for individual *i*, and e_i is a random residual error (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, e.g., the shared familial environment. In statistical terms, the pedigree provides the relatedness $n \times n$ matrix **A** of the relationship between every pair made of the *n* individuals in the population. The animal model uses the inverse ($\mathbf{A}^{-1}$) of

this relationship matrix in combination with individual phenotypic data to estimate QG parameters. This demonstrates why the estimates of QG parameters partly depend on the properties of the pedigrees provided to the model.

Parametrization of the animal model has a considerable effect on the accuracy of QG estimates (Kruuk & Hadfield 2007; Wilson 2008; Wolak et al. 2015). For example, certain phenotypic traits might not follow a normal (i.e., Gaussian) distribution, or they might be skewed towards a given value (e.g., zero-inflated distribution), properly parameterizing the model then is crucial (de Villemereuil 2018). For example, according to Bonnet et al. (2019), wrongly parameterizing fitness as a Gaussian distribution in animal models can cause inaccuracies in the estimates of the additive genetic variance. Given that most variation in phenotypic traits are explained by more than one source, it is not only important to include all the relevant sources but design the variance structure of the model in a way to properly reflect the contribution of each element. Not including additional sources of variation in the model will affect the estimation of other variance terms leading to precise but inaccurate results (Kruuk and Hadfield 2007). Additionally, pedigree errors 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). All these issues also relate to the concept of 'identifiability' in certain statistical models with complex variance structures (Bolker 2008). So far, it is still unclear how pedigree topology and properties impact the identifiability of variance components in animal models (Bourret et al. 2017).

Pedigrees also reflect important aspects of individual life histories, providing measures of individual fitness components, namely on fertility and, when based on records with accurate dates, on the timing of reproduction (Hunter et al. 2019). Only individuals whose phenotypes are measurable, or who or are related by ancestry or descent to two or more phenotyped individuals,

are informative with respect to the genetics of the given trait. To take this into account in our analysis, we pruned the supplied pedigree to the model to include only the informative individuals. This is done using the `prunePed` function in the *MCMCglmm* R package (Hadfield 2010). Additionally, the phenotypic dataset supplied to the model was pruned to retain only individuals for which the year of birth is known and had at least one successful reproductive event. The reasoning behind this is that, at the least, only individuals with a known year of birth and reproductive event can be phenotyped by having their life history reconstructed. Furthermore, this decision allowed us to be conservative by checking for the reliability of estimates based on a restricted phenotypic dataset reflective of the case of QG analysis in natural populations.

We used the Bayesian implementation of the animal model in the *MCMCglmm* R package employing Markov chain Monte Carlo methods (Hadfield 2010). Previous pedigree power and precision analyses similar to this study had been performed using a frequentist approach (e.g., Charmantier & Réale 2005, Morrissey et al. 2007, Bérénos et al. 2015, Bourret & Garant 2017). Though the frequentist approach offers the benefit of a shorter computation time especially with very large datasets, we opted for the Bayesian approach for its intuitive and straightforward interpretation of estimates. This point relates to the fact that Bayesian methods deal with posterior density distributions of estimated variables. Morrisey et al. (2014) suggest that a posterior distribution is more interpretable as a complete representation of the statistical support for a given parameter having a given value. In the context of our analyses, this provided us with a probabilistic way of inferring the precision and accuracy of QG estimates. One additional benefit is that we can employ transformations of the posterior distributions of parameters that are directly modeled to extract the posterior distribution of derived parameters (e.g., derive the posterior distribution of

genetic correlations based on that of genetic covariances). All models were run in R (v3.6.3; R Core Team 2019).

Simulation analyses

We used the `phensim` function from the *pedantics* R package to simulate phenotypic values for each individual in the simulated 'true' pedigree. Phenotypic values were calculated as the sum of additive genetic effects (or breeding values) and environmental effects simulated from user-specified (co)variance matrices. We ran scenarios with one and two response traits, as we were also interested in the impact of pedigree quality on the validity of genetic covariance estimates. Additionally, in one scenario, we simulated maternal environmental effects to replicate phenotypic resemblance due to the sharing by siblings of the familial environment, an important source of pairwise resemblance to control for in QG studies. In total, we ran three scenarios, as follows:

- **Scenario 1**: Univariate (one response trait, z) with simple trait architecture: $z_i = \mu + a_i + e_i$, where μ is the population mean (with μ = 0), a_i is the additive genetic value for individual i, and e_i is a random 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 environmental (or residual) variance. V_A was set to 0.1 or 0.3, with V_E = 1-V_A.
- **Scenario 2**: Univariate with complex architecture: $z_i = \mu + a_i + m_i + e_i$, similar architecture as Scenario 1 with the added term m_i which represents the shared familial environment effect for individual 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 3**: Bivariate (two response traits, z_1 and z_2) with a similar simple architecture for both traits: $z_i = \mu + a_i + e_i$. Here, $V_P = V_A + V_E = 1$. V_A of trait z_1 was set to 0.1 or 0.3

and that of trait z_2 was fixed to 0.1, with V_E = 1-V_A. The genetic correlation r_g between z_1 and z_2 was set to 0.1 or 0.5, replicating a weak or strong correlation, respectively.

In the first two scenarios, we ran five simulations for each regional pedigree × V_A value. Due to prohibitive computational time to run bivariate models multiple times on large datasets, we only ran one simulation per pedigree and parameter value in the case of Scenario 3.

Reliability measures

For Scenarios 1 and 2, precision and accuracy for V_A and V_M estimates were checked visually using dot and bar plots, with the average – over five simulations – posterior mode (represented by the dot) for accuracy and the average 95% Highest Posterior Density interval, or 95% HPD, (represented by the bars) of the posterior distribution for precision. The Bayesian 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). For Scenario 3, which involved only one simulation per pedigree and per parameter, the visual check was done by directly plotting the posterior distribution using the R package *bayesplot* (Gabry et al. 2019). The mean squared error or MSE is defined as $E[(\hat{d} - d)^2]$, where d is the true value (for example, the simulated parameter value of $V_A$or r_g) and $\hat{d}$ is the estimated value. A small MSE indicates high precision and accuracy (Bolker 2008). Here, to quantify precision and accuracy, we used the root mean squared error (RMSE) which allows for comparisons at the scale of the estimates (de Villemereuil et al., 2013). We calculated the posterior distribution of the RMSE for every parameter (averaging over the five simulations for Scenario 1 and Scenario 2) by applying the same calculation over every saved

MCMC iteration (1,000 iterations in total) and then extracting the corresponding posterior mode, which would represent the accuracy, and the 95% HPD, which would represent the precision.

Interpretation of posterior distributions

We interpret posterior distributions based on two Bayesian indices which help to infer information on the existence and significance of an effect based on the data (i.e., genetic in this case) - with prior knowledge (see Makowski et al. 2019b); in the case of this study, we used uninformative priors. We aimed to assess the power of a particular dataset to detect a genetic effect of a given size. Within the frequentist paradigm, the power of a particular dataset represents the probability of detecting an effect using the dataset given that this effect is true, and it is influenced by the same factors as precision, mainly sample size and pedigree depth (Quinn et al. 2006, Morrissey et al. 2007). Conversely, this definition of power does not exist in the Bayesian paradigm. However, we can use the Region of Practical Equivalence (or ROPE)-based index, specifically the *ROPE* (95%), which refers to the percentage of the 95% HPD of the posterior distribution that lies within the ROPE, as a Bayesian equivalent to 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).

Results

Pedigrees

Following pedigree reconstruction and filtering, we ended up with six datasets corresponding to six Québec regions. There was a large amount of variation between the datasets in terms of structure and properties (Table 2 and 3). This variation is evident when we contrast two very different datasets: Gaspésie and Bois-Francs. The sample size of the pedigree of Bois-Francs is five times larger than that of Gaspésie; pedigree depth (i.e., total number of generations) is 4 for Gaspésie and 7 for Bois-Francs; even the number of founders of Bois-Francs is almost 5 times that of Gaspésie. However, these two pedigrees did have a few similar properties, such as the mean inbreeding coefficient and expected genealogical depth (see Figure S3 and S4 for visual representation of the two pedigrees). This highlights that between our six datasets there were both important structural differences and similar genealogical properties.

Power

We note, first, that the simulated values of V_A or r_g are highly correlated ($r > 0.99$) with the realized value within each simulated population. Overall, across all scenarios, our results revealed that all of our datasets detected the existence of a genetic effect underlying the modeled trait. Whether the simulated additive genetic variance was low, 0.1, or moderate, 0.3, all our datasets, to the exception of Gaspésie, produced no evidence for the alternate hypothesis of no genetic effect according to the *ROPE* index (Figure 6). To further support this result, we ran a frequentist power analysis employing a parent-offspring regression to estimate the additive genetic variance (in this analysis the simulated V_A ranged from 0 to 0.5 by 0.01 increments) for each dataset. Five of our datasets reached a power of 100% at $V_A = 0.05$, this value represents the lower bound of additive genetic

variance estimates in most human traits (Figure S4). Only Gaspésie, which was our smallest and shallowest pedigree, reached a 100% power at $V_A > 0.1$.

Precision and accuracy

Visual inspections of the posterior distributions (Figures 1-5) reveal clear information on the precision and accuracy of V_A and Cov_A estimates: larger and deeper pedigrees provided the highest estimate precision but estimate accuracy depended heavily on trait architecture (i.e., simple vs complex architecture) and model parameterization. This information was congruent with comparisons of RMSE values between pedigrees across simulation scenarios (Table S1). In Scenario 1, the univariate model with simple trait architecture, the larger and deeper pedigrees such as Bois-Francs and Côte-de-Beaupré had the lowest average posterior-mode RMSE values between all of the six datasets when the simulated V_A was equal to 0.1 (0.011 [0 - 0.037] and 0.005 [0 - 0.035], respectively); When simulated V_A was equal to 0.3, the pedigree of Laurentides performed slightly better compared to the much larger pedigree of Bois-Francs (0.008 [0 - 0.080] vs 0.012 [0 - 0.059], respectively), however with larger 95% HPD intervals, which translates to a lower estimate precision (Table S1).

When another source of resemblance between related individuals was added in Scenario 2 (i.e., maternal effects), precision and accuracy of estimates were reduced according to the average RMSE values which was also confirmed visually (see Figure 2). We detected an average overestimation of V_A by 46% and an average underestimation of V_M by 42%. Though the larger and deeper pedigrees produced greater precise estimates, these estimates were very inaccurate based on visual inspections. So much so that when the simulated V_A was equal to 0.1, the estimated 95% HPD intervals did not even contain the expected value. Comparing RMSE values in this case

revealed that the larger and deeper pedigrees were no longer performing better than the smaller ones in terms of precision and accuracy of their estimates. To confirm this pattern, we ran the same scenario by setting the simulated V_A to 0.1 and then increasing V_M to 0.3 and 0.5, respectively. Results from this analysis also revealed the issue of overestimation of the additive variance and underestimation of the maternal variance.

Given that the animal model attempts to separate different variance components partly based on the phenotypic similarity between relatives, it is expected that the additive genetic variance may include the maternal variance (leading to an upward bias). This is because the model cannot distinguish if the similarity between siblings in the dataset is due to sharing genetic material or sharing the same mother. We confirmed this by running Scenario 2 with improper parametrization (i.e., without fitting a maternal variance component) which clearly revealed that the totality of the expected maternal variance (i.e., V_M = 0.1) was incorporated into the estimated additive variance (Figure 3). We attempted the same analysis using a less restrictive phenotypic datasets wherein the mean maternal family sibship size was ~6 across pedigrees. Results from this analysis, which are not shown here, detected the same bias in additive genetic and maternal variance estimates.

Lastly, we looked at the identifiability of our models in Scenario 2. Given that important variance components such as the additive and maternal variance can be confounded as our results showed, it is vital that the results being used to infer about biological processes come from models with adequate identifiability. This was done by calculating the Spearman's correlation coefficient between the posterior distribution of V_A and V_M for each simulation. Results from this analysis revealed that identifiability was low for most datasets as correlations were high between the two

variance components (also confirmed through MCMC autocorrelation checks). The identifiability of the model only increased slightly for datasets with larger depth and sample size.

Discussion

Assessing a population's evolutionary potential or change requires precise and accurate estimates of genetic parameters. However, when datasets for natural populations are exploited for this purpose, the validity of QG estimates using these reconstructed pedigrees is seldom investigated and quantified (but see Bourret & Garant 2017), whether in the case of a social (observation-based) or a genetic (marker-based) pedigree. The sensitivity of parameter estimation to pedigree attributes has been explored mostly in (passerine) birds and (ungulate) mammals. Therefore, extrapolation or generalization should be done with caution to other taxa, which may be characterized by on average different pedigree topologies or tractability (Pemberton 2008). Evidently this also applies to this study since we are specifically using a mammalian study system (pedigrees from the historical French-Canadian population) to assess the impact of pedigree topology and genealogical errors on the validity of QG parameter estimation.

Overall, our simulation analyses confirmed that all the pedigrees tested in our study have enough 'power' to detect a small V_A, with sample size, pedigree depth, and mean genealogical entropy being important determinants of the validity of QG estimates, while inclusion of other sources of similarity (ex. maternal variance) leads both to an upward estimation bias of V_A and an identifiability issue. We also reveal that in the case of a trait with a complex architecture (i.e., inclusion of V_M), a pedigree with a higher mean completeness performed slightly better in terms of accuracy of estimates. Additionally, the validity of genetic covariance estimates was primarily affected by the magnitude of the simulated parameter. We discuss in the following sections (1) the

impact of pedigree errors and other properties on validity, (2) the issues that arise with model parameterization and identifiability, and (3) the interpretation of the bivariate analysis results.

Validity unaffected by reasonable pedigree error rate

Using a reliable, genetically-based estimate of the genealogical error rate for the FC datasets, our study revealed that the impact of this rate on the validity of QG estimates was very limited. In fact, in the case of a phenotype with a simple architecture (Scenario 1), we did not find any sort of downward bias in the estimates of the additive genetic variance (increasing the error rate to 1%, the upper bound of the expected genealogical error rate in the FC datasets, lead to a slight underestimation and reduction in accuracy; see Figure S5). This is in line with previous sensitivity studies which found that with plausible error rates of pedigree errors bias would generally be small (Charmantier and Réale, 2005; Bérénos et al., 2014; Firth et al., 2015). Additionally, EPP rates in human populations are considered to be low (range: 0.4–11.8; Anderson 2006; also see Larmuseau et al. 2016) with the median of 1.7%, being much lower than the cutoff rate of 20% where misassigned paternity begins to affect QG estimates as reported by Charmantier & Réale (2005). Though it is unwise to generalize results from studies using pedigrees from wild passerine and ungulate populations to social human pedigrees, our study serves as clear confirmation that the previously reported pattern does indeed apply to human datasets as well.

In our analyses, the final datasets supplied to the animal model reflected the realistic expectation that not all individuals included in the pedigree can be phenotyped. Each of the six 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 done on wild populations by Postma (2014), the sample size (total number of records) of our pruned pedigrees was considerably higher than the median number of records for life history traits – which was 377 – and in most cases, higher than the upper bound of the range in records as well (range 6–4992; from 39 studies, covering 19 species). Our results confirmed our expectation that our datasets are powerful enough to detect a genetic signal in a phenotypic trait, using both a parent-offspring regression framework and a Bayesian implementation of the animal model. However, sample size alone is not the only factor that determines pedigree quality (Wilson et al. 2010), pedigree depth is another important indicator of pedigree quality. Our results demonstrated that the deeper the pedigree the more precise its QG estimates. In fact, both sample size and pedigree depth exhibited the strongest correlation with a pedigree's RMSE score, followed by mean completeness and genealogical entropy (see Figure S6). These last two factors are rarely considered when it comes to the validity of QG pedigree-based methods. We suggest that pedigree-based studies calculate and incorporate these factors when testing the validity of their QG estimates based on their respective pedigree.

An estimation bias was detected in our estimates in Scenario 2 following inclusion of an additional source of phenotypic similarity between relatives. Though larger and deeper pedigrees continued to provide strongly precise estimates, only estimates from pedigrees with a high mean completeness exhibited relatively adequate accuracy. This is problematic in terms of validity for two reasons: 1) most life history traits are expected to exhibit non-genetic sources of resemblance between relatives and 2) when it comes to reporting genetic parameters which will then be used to test evolutionary hypotheses, accuracy is more important than precision because overestimation (or underestimation), even with low uncertainty, will lead to faulty interpretations. In our opinion, the detected overestimation of the additive genetic variance could be due to the following reasons:

(i) the animal model we ran was explicitly "bare", meaning there were no predictors or fixed effects included in the model, (ii) the models were run with uninformative priors, in other words, no *a priori* knowledge on the relative contribution of the different variance components to the total phenotypic variation, (iii) the rate of individuals having less than two offspring in the pruned dataset was relatively high which negatively impacts estimation of shared maternal effects. In terms of the latter reason, further analysis using less restrictive phenotypic datasets revealed that increasing the mean number of siblings per mother did not eliminate the estimation bias (Figure S7). This places an additional burden on investigators to make sure that their datasets are adequately informative to the specific modeled variance components they are attempting to estimate, in our example, it is the shared familial environment.

Proper parameterization and low identifiability

Another important factor affecting precision of estimates, and validity in general, is proper model parameterization (Morrissey et al. 2007). This involves the proper specification of the model to account for different sources of variation in the focal trait, both genetic and non-genetic. A specific model provided with a dataset with high power (i.e., power of detecting an effect given the presence of this effect) could be missing a critical variance component which will lead to precise but inaccurate results (Kruuk & Hadfield 2007). We tested this explicitly by misspecifying our animal model in Scenario 2 by dropping the maternal variance, and results revealed that all the expected maternal variance was included into the additive genetic variance estimates. Hence, without the proper variance structure in a model explicitly testing for a genetic effect in a given trait, there is a high risk of overestimation of the genetic component leading to inflated conclusions. Additionally, the issue of identifiability (i.e., discrimination between two variance terms) must be taken into account since it can affect the validity of model estimates. Our results revealed low

identifiability between the genetic and shared familial environment components (though it slightly increased with sample size). This is similar to the results found by Bourret & Garant (2017) using a frequentist framework, and here we confirm the same pattern using a Bayesian framework. Though it is important to note that all our models were run with uninformative priors which will impact identifiability negatively.

Validity of joint estimates when using bivariate animal models

Scenario 3 involved bivariate models which we structured in a way to mimic the situation where a life history trait (either with relatively small or moderate additive genetic variance) is either strongly (i.e., high genetic correlation) or weakly (i.e., low genetic correlation) genetically associated with fitness (represented by a secondary trait with little additive genetic variation). Recently in the field of quantitative genetics, researchers have focused on estimating additive covariances either between multiple traits (ex. genetic constraints on response to selection or trade-offs, see Bolund 2020) or between a focal trait and a measure of fitness (ex. Robertson-Price identity, see Morrissey et al. 2012). Hence, emphasis on providing valid estimates of the additive genetic covariance should be on par to that of the additive genetic variance. We also note that our study explicitly tested the validity of the estimates of this parameter instead of its derivation, the genetic correlation (though the latter is more easily interpretable than the former).

Though it was difficult to perform multiple simulations per pedigree given the exorbitant amount of computation time required, we were able to demonstrate the same pattern of reliability for the additive covariance as the additive variance. Bigger and deeper pedigrees provided more precise estimates but inaccurate in a few cases. An important factor affecting validity of estimates was the magnitude of the simulated covariance parameter between the two traits where inaccuracy

was more common when the magnitude simulated was small. Though in all the cases (except one) the most probable sign of the estimated covariance matched that of the simulated parameter. Careful attention must be paid to ensuring that estimates of additive covariance are truly valid before, for example, inferring the expected quantitative change in a trait in response to selection.

Conclusion

High density genetic markers have become more available for certain study systems allowing a more accurate estimate of the genome-wide average of the realized relatedness. This will definitely improve the estimation of QG parameters, in addition to opening the possibility of studying the QG of a natural system where it is difficult to obtain a multi-generational pedigree. In their review of empirical and simulation studies of QG in wild populations using marker-based estimates of relatedness, Gay et al. (2013) determined that the method of pedigree reconstruction using genetic markers performs poorly in parameter estimation relative to the method of relatedness matrix reconstruction using high density genetic markers (i.e., GRM, or genomic relatedness matrix). The results of our study contribute additional support to previous results affirming that socially-constructed pedigrees still perform remarkably well in terms of QG parameter estimation (ex. Bérénos et al. 2014). The combination of inexpensive sequencing of genomes and increased coverage of genetic information along the pedigree will no doubt better the validity of key genetic parameter estimates.

Though REML (Restricted Maximum Likelihood) is considered a better estimator of additive variance for Gaussian traits, we recommend using a Bayesian approach for simulation studies aiming to assess the validity of estimates for any type of trait distribution for the following reasons: (i) its intuitive interpretation of estimates, (ii) possibility of assigning different degrees of

priors, and (iii) checking and controlling for autocorrelation between variance components (but see Houle & Meyer 2015). Though large and deep pedigrees can be *a priori* considered adequate for QG analyses, this comes with the caveats of considering the impact of trait architecture, pedigree quality, and model parameterization. We also highlight in this study the value of investigating the impact of pedigree properties such as completeness and genealogical entropy. With routine sensitivity analyses becoming easier to implement with accompanying packages in the statistical environment R making the programming easier, there is high expectation of producing validity measures alongside reporting QG estimates for either a new pedigree or when testing a more complex model using a previously-studied pedigree.

Data availability

The data is part of the BALSAC Project demographic database which contains vital human information. The data sets generated and/or analyzed during the current study are available upon request and review to the BALSAC Project.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery grant. We thank the BALSAC Project located at the Université du Québec à Chicoutimi for its role in the ongoing collection, maintenance, and support of its demographic database. This study would not have been possible without the original work of Gérard Bouchard who started the BASLAC project in 1976 and built its population registry. Since 2010, Hélène Vézina, professor in the Department of Human and Social sciences at Université du Québec à

Chicoutimi, has been managing the BALSAC Project. None of these organizations had any role in the design, analysis, interpretation, or report of the study.

Acknowledgments

We thank R. Landry and L. Lavergne for their comments and ideas during the realization of this study. We also extend our thanks to J.-S. Bournival with the generation and calculation of life history data, as well as providing important insights into the BALSAC database.

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.

Bérénos, 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.

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.

Bolund, E., 2020. The challenge of measuring trade-offs in human life history research. *Evolution and Human Behavior*, *41*(6), pp.502-512.

Bonnet, T., Morrissey, M.B. and Kruuk, L.E., 2019. Estimation of genetic variance in fitness, and inference of adaptation, when fitness follows a log-normal distribution. *Journal of Heredity*, 110(4), pp.383-395.

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., 1989. Population studies and genetic epidemiology in northeast Quebec. *Canadian Studies in Population* [ARCHIVES], pp.61-86..

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.

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.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. and Gelman, A., 2019. Visualization in Bayesian workflow. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*(2), pp.389-402.

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.

Gay, L., Siol, M. and Ronfort, J., 2013. Pedigree-free estimates of heritability in the wild: promising prospects for selfing populations. PloS one, 8(6), p.e66983.

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.

Houle, D. and Meyer, K., 2015. Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. *Journal of Evolutionary Biology*, *28*(8), pp.1542-1549.

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.

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 Lüdecke, 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.

Pemberton, J.M., 2008. Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1635), pp.613-621.

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/.

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. *J. Anim. Sci.*, 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.

Slate, J., 2005. Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Molecular ecology*, 14(2), pp.363-379.

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.

Thompson, E.A., 1986. *Pedigree analysis in human genetics*. Johns Hopkins University Press. p25

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.

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*.

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. *J. Dairy Sci.*, 85(9), pp.2368-2375.

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., 2018. Quantitative genetic methods depending on the nature of the phenotypic trait. *Annals of the New York Academy of Sciences*, 1422(1), pp.29-47.

Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*. Oxford University Press.

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., 2008. Why h2 does not always equal VA/VP?. *Journal of evolutionary biology*, *21*(3), pp.647-650..

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.s

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.

Tables

Table 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[a]. 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 et al. (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 to be phenotyped
	Genealogical error	An erroneous link between a pair of individuals (e.g., parent-offspring) due to misassignment in the reconstruction process	In pedigrees reconstructed for natural populations, this error is mainly due to extra-pair paternities. Though, depending on the organism, the parental misassignment can also be attributed to hidden adoptions or false parentage declarations.
	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%
	Rate of missing links	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

[a] 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 2. Summary pedigree statistics and genealogical properties of the final six pedigrees pruned for quantitative genetic analyses.

	Gaspesie	Bas-St-Laurent	Laurentides	Charlevoix	Côte-de-Beaupré	Bois-Francs
Records	1,949	3,741	5,055	6,878	8,450	12,032
Maternities	1,288	2,473	3,298	6,074	7,057	9,071
Paternities	1,288	2,473	3,298	6,074	7,057	9,071
Full sibs	1,755	4,054	4,213	12,145	9,543	14,281
Maternal sibs	1,800	4,138	4,309	12,517	9,892	14,664
Paternal sibs	1,839	4,279	4,559	13,512	10,307	15,330
Maternal grandmothers	446	722	1,218	4,649	5,025	4,773
Maternal grandfathers	446	722	1,218	4,649	5,025	4,773
Paternal grandmothers	310	605	1,127	4,463	5,159	4,654
Paternal grandfathers	310	605	1,127	4,463	5,159	4,654
Founders	661	1,268	1,757	804	1,393	2,961
Max. pedigree depth	4	6	6	8	9	7
Mean maternal sibsip size	2.69	2.84	2.46	3.65	2.72	2.90
Mean paternal sibsip size	2.71	2.9	2.57	3.83	2.80	3.01
Mean inbreeding coefficient	0	$5.42*10^{-5}$	$6.44*10^{-6}$	$2.48*10^{-3}$	$9.36*10^{-4}$	$8.68*10^{-5}$
Mean completeness	58.88	39.03	41.45	48.10	43.61	40.65
Mean expected entropy	1.36	1.34	1.49	2.85	2.92	1.85
Var. in expected entropy	0.29	0.48	0.57	1.87	2.48	0.86
Mean rate of missing links	0.94	0.87	0.86	0.58	0.71	0.77
Mean observed entropy	1.26	1.05	1.08	1.00	1.31	1.07

Figures

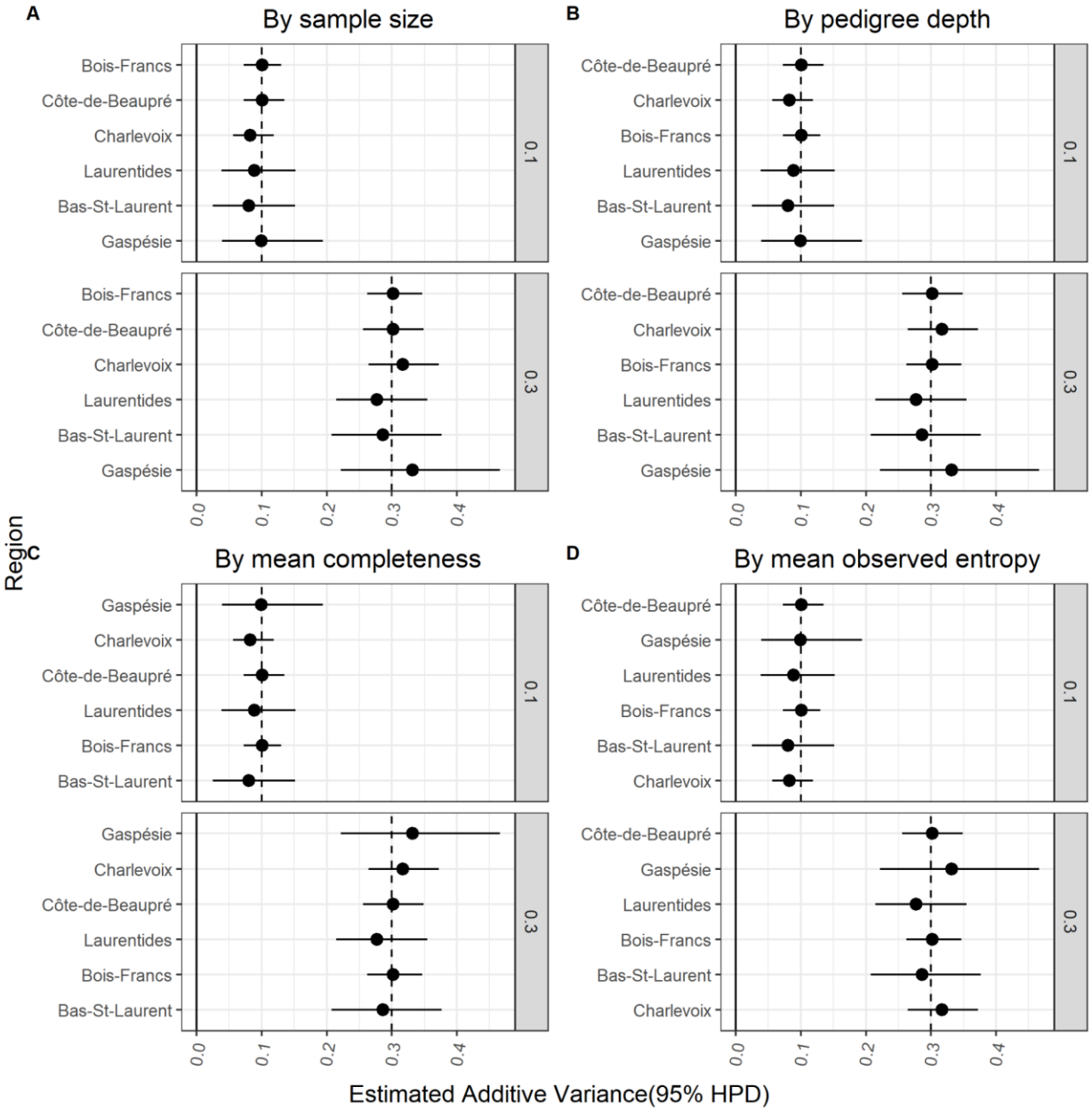

Figure 1. Estimates of additive genetic variance for each of the six pedigrees based on Scenario 1 (univariate simple phenotype; $V_A = 0.1$ *and* 0.3). In each plot, pedigrees are sorted, from bottom to top, by increasing: **A.** sample size, **B.** pedigree depth, **C.** mean completeness, and **D.** 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.

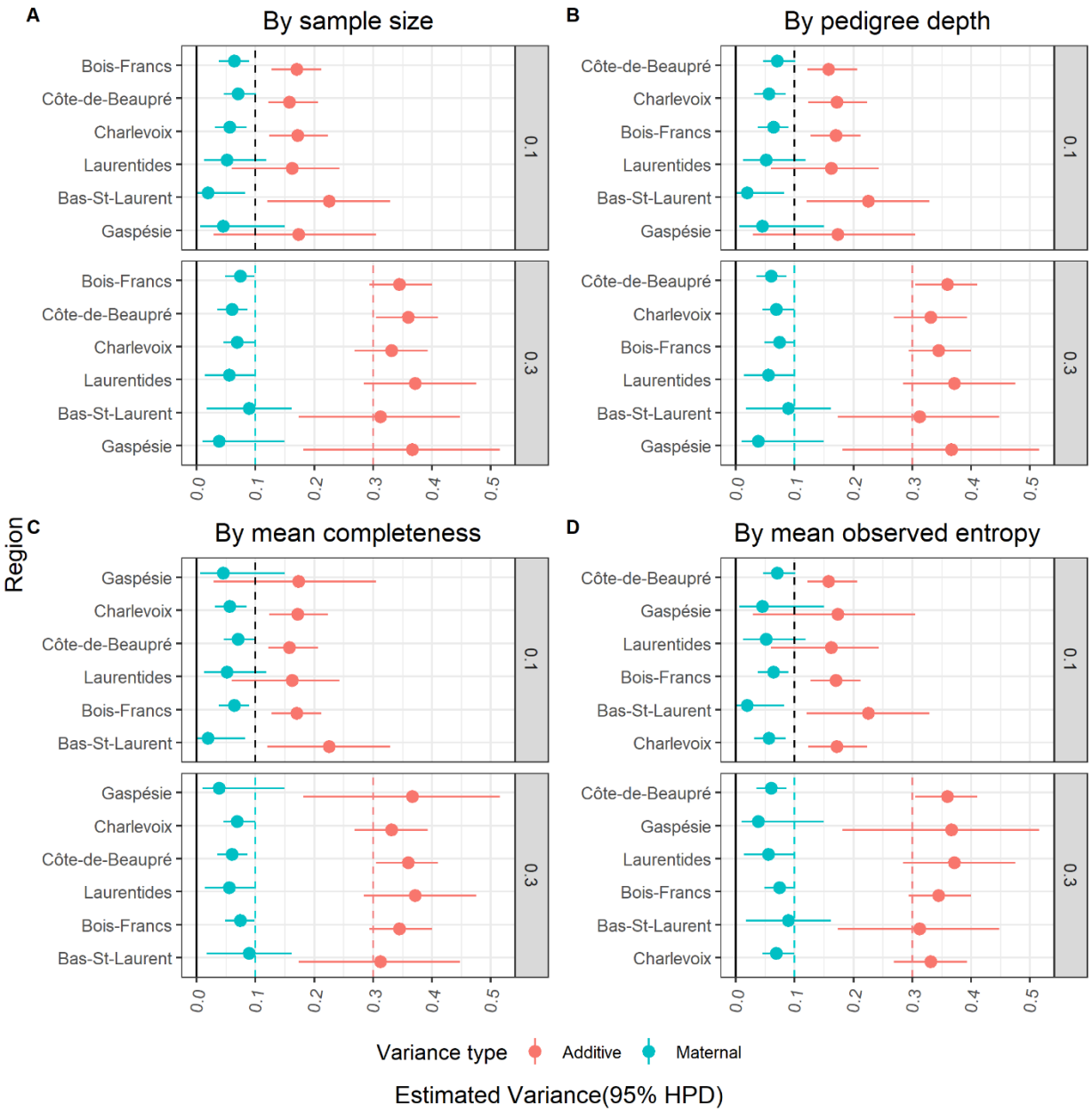

Figure 2. Estimates of additive genetic variance and maternal variance for each of the six pedigrees based on Scenario 2 (univariate complex phenotype; $V_A = 0.1$ *and* 0.3;$V_M = 0.1$). In each plot, pedigrees are sorted, from bottom to top, by increasing: **A.** sample size, **B.** pedigree depth, **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.

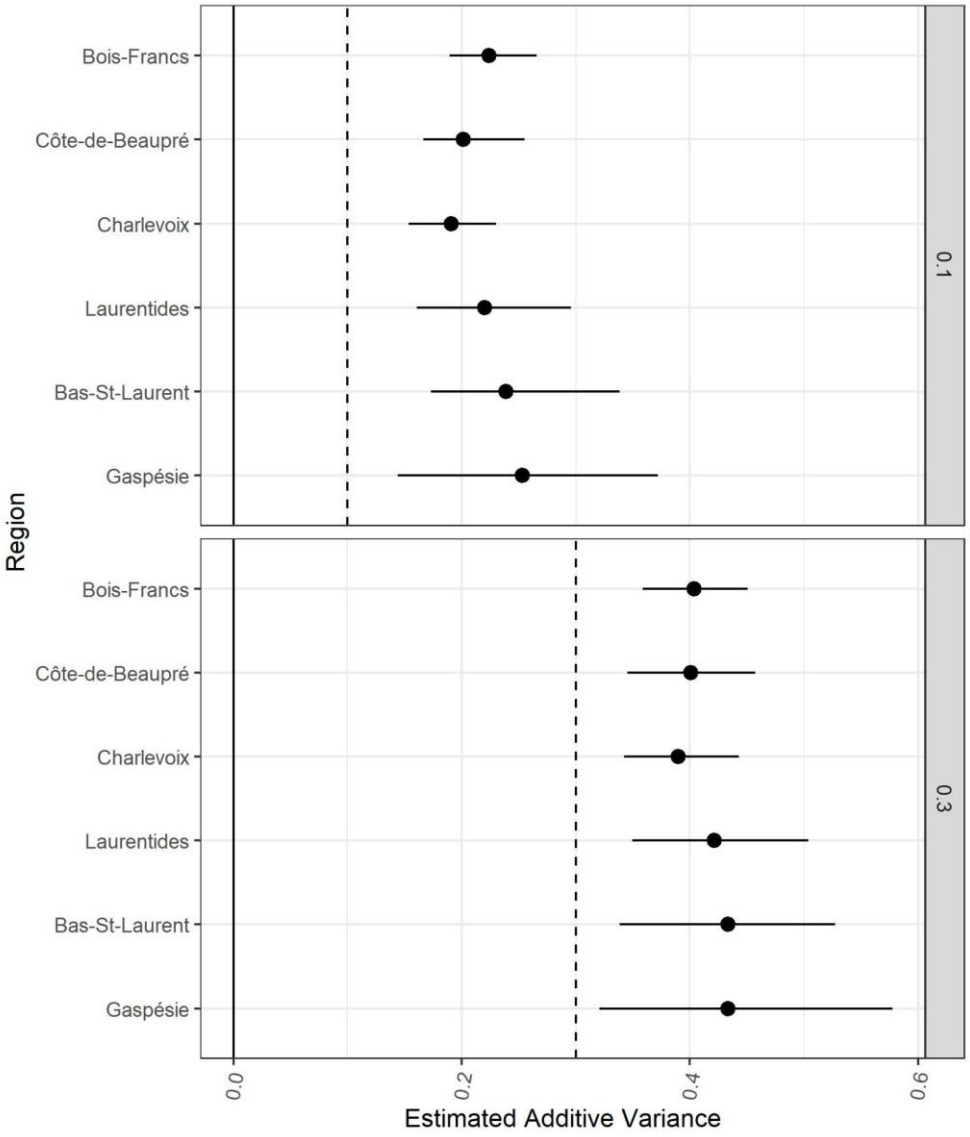

Figure 3. Estimates of additive genetic variance for each of the six pedigrees based on Scenario 2 with improper parameterization (univariate complex phenotype; $V_A = 0.1$ *and* 0.3;$V_M = 0.1$). Each dot represents the average posterior mode-estimates, along with the average 95% credible intervals based on the five simulations per pedigree per parameter value. Vertical dashed lines represent the expected value of the variance

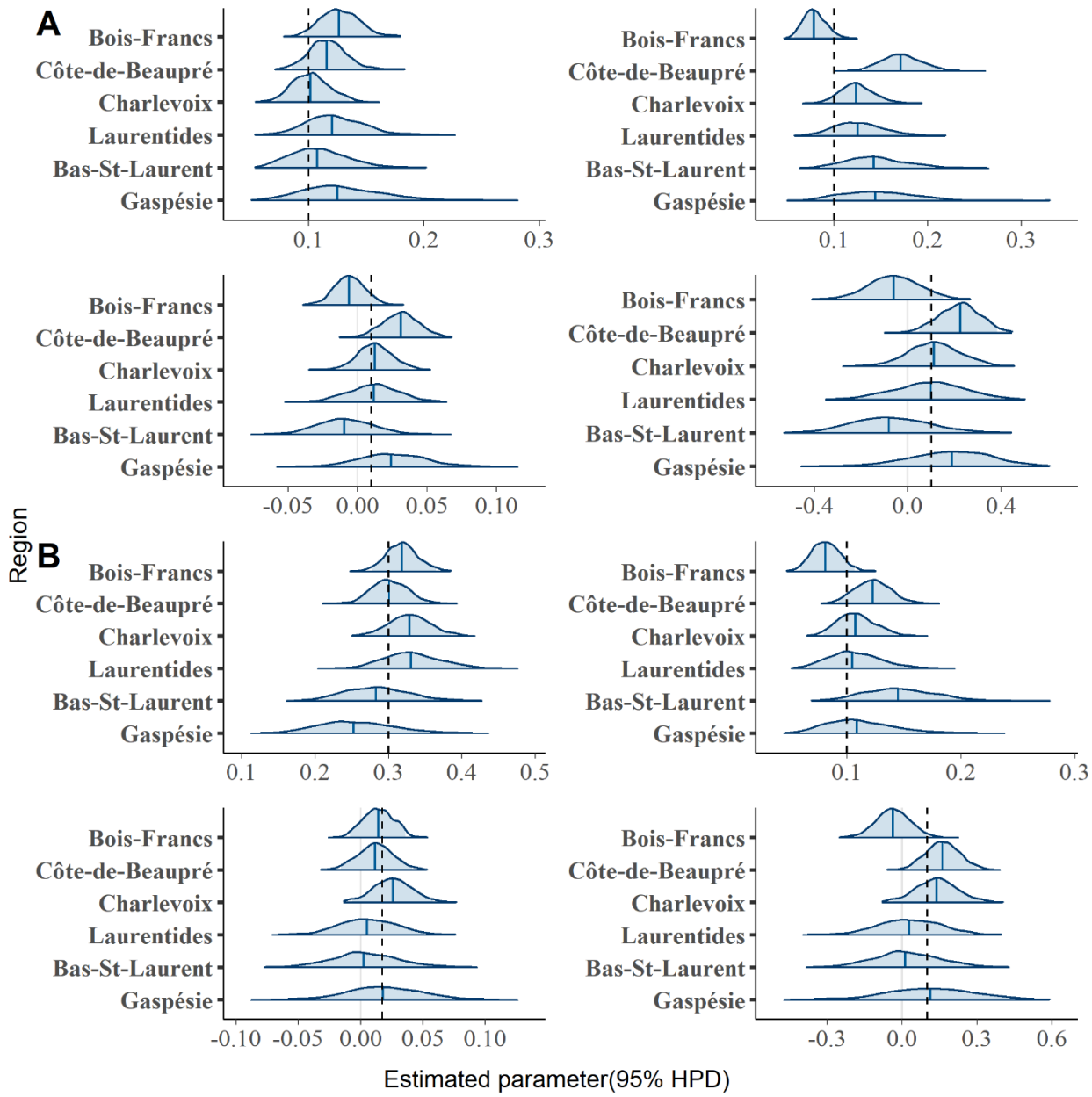

Figure 4. Posterior distributions of estimated quantitative genetics parameters (with the 95% HPD interval) for each of the six pedigrees based on Scenario 3 (bivariate simple phenotype with genetic correlation r_g= 0.1). Top left is additive variance of trait z_1, top right is additive variance of trait 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.

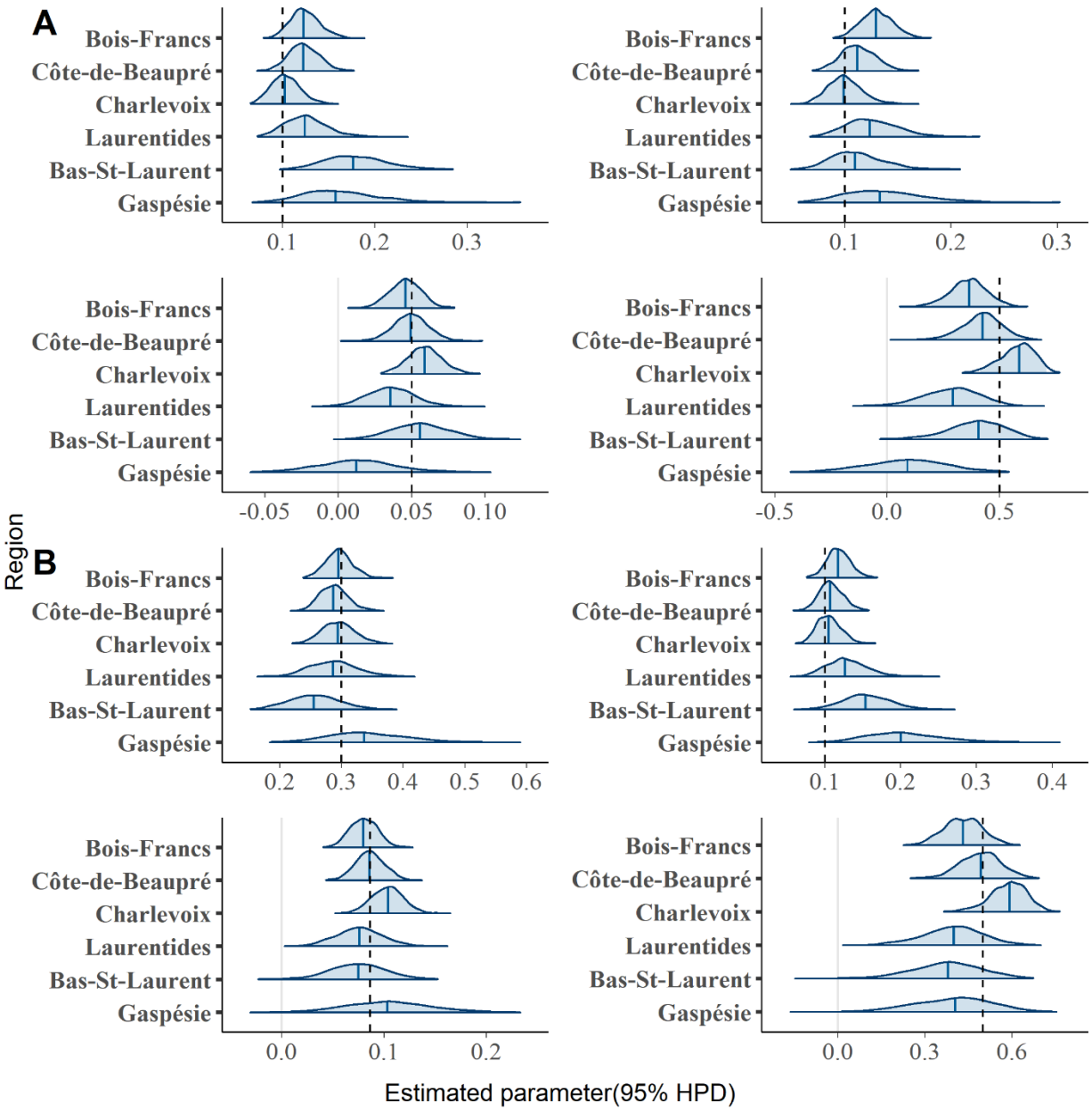

Figure 5. Posterior distributions of estimated quantitative genetics parameters (with the 95% HPD interval) for each of the six pedigrees based on Scenario 3 (bivariate simple phenotype with genetic correlation r_g= 0.5). Top left is additive variance of trait z_1, top right is additive variance of trait 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.

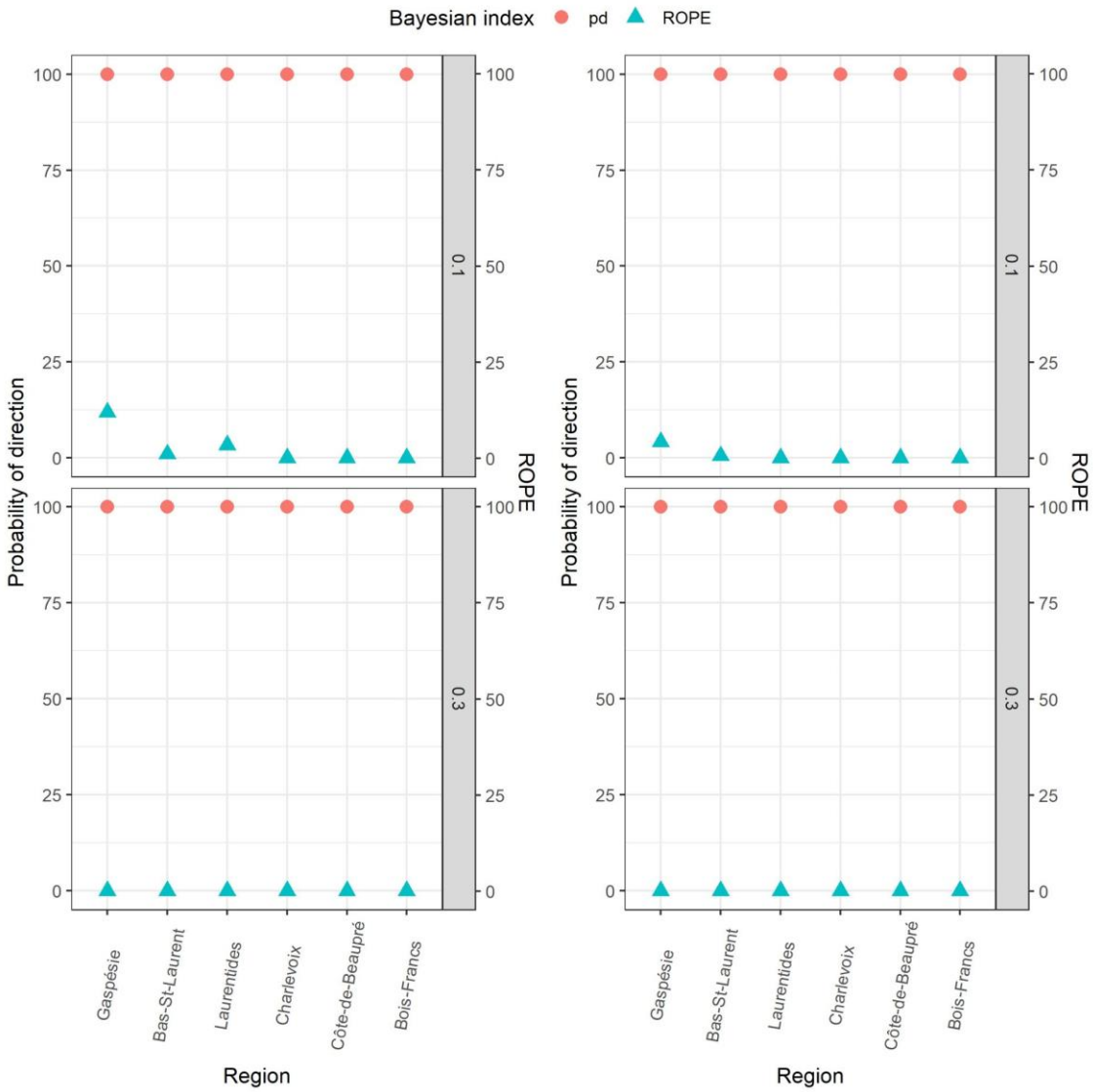

Figure 6. Estimates of Bayesian indices, posterior of direction (*pd*) and *ROPE*, for each of the six pedigrees based on five simulations. *Left*, Scenario 1 (univariate simple phenotype); *Right*, Scenario 2 (univariate complex phenotype). Dots represent the average *pd*; Triangles represent the average *ROPE*.

Supplementary information

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 millions individuals over the span of almost four centuries and covering 23 different subregions of French-Canada (Québec, see Figure S5). 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 visualise 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 S7 and S8 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: $C_x = \frac{number\ of\ known\ ancestors}{number\ of\ expected\ ancestors}$, 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 ψ 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 entropy* (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:

$$V = \sum_i p_i {log_2}^2(p_i) - S_B^2 = \sum_i {N_i}^2 \frac{1}{2^{N_i}} - S_B^2,$$

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:

$$S_B = \sum_{g=1}^{G} N_g \frac{\delta^g}{2^g}$$

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. δ 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 θ=0, where θ 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 indice to re-interpret 'animal model' results by considering, instead of θ=0, a range of θ 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 should the estimates of these parameters 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 ≥0.015 would be considered as non-negligeable, which might be viewed as liberal. We also applied a conservative 25% threshold, meaning that a substantial fraction (≥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 the average of 5 simulations.

Scenario	Gaspesie	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)	0.082 [0.001 - 0.198]*	0.130 [0.040 - 0.223]*	0.048 [0.009 - 0.139]*	0.072 [0.027 - 0.123]*	0.062 [0.021 - 0.107]*	0.070 [0.030 - 0.111]*
Scenario 2 (V_A=0.3)	0.063 [0 - 0.215]*	0.037 [0 - 0.169]*	0.066 [0.007 - 0.166]*	0.028 [0 - 0.089]*	0.052 [0.016 - 0.105]*	0.043 [0.003 - 0.094]*
Scenario 2 (V_A=0.1; improper param.)	0.154 [0.047 - 0.265]*	0.141 [0.078 - 0.238]*	0.123 [0.061 - 0.196]*	0.092 [0.054 - 0.130]*	0.106 [0.066 - 0.155]*	0.125 [0.089 - 0.166]*
Scenario 2 (V_A=0.3; improper param.)	0.145 [0.025 - 0.269]*	0.127 [0.044 - 0.222]*	0.120 [0.049 - 0.202]*	0.093 [0.042 - 0.143]*	0.098 [0.045 - 0.157]*	0.103 [0.059 - 0.151]*

Scenario 3 (V_A=0.1; 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; 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; 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; 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; 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; 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; 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; 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]
Average value (±SD)	0.0652 (±0.0540)	0.0625 (±0.0504)	0.0479 (±0.0466)	0.0378 (±0.0346)	0.0358 (±0.0407)	0.0440 (±0.0435)

Figures

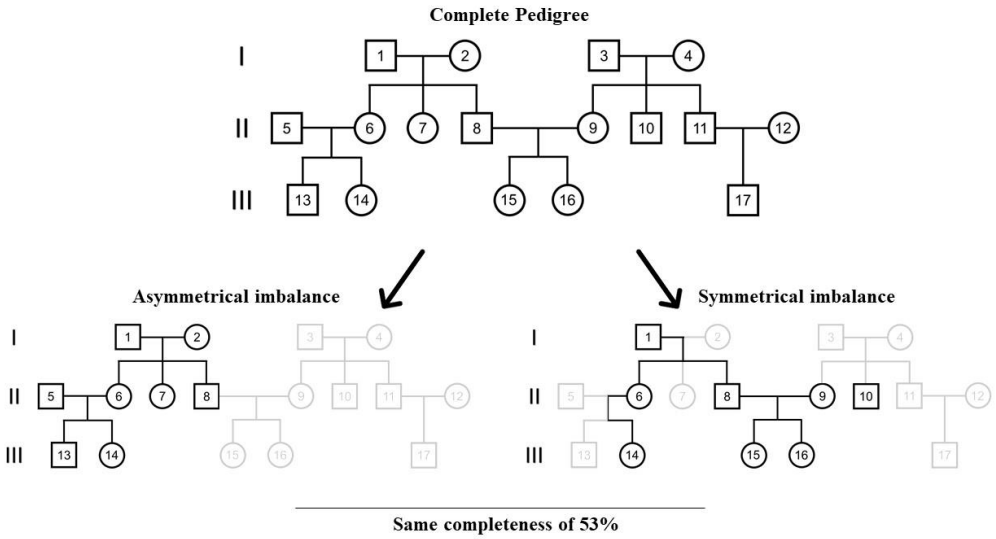

Figure S1. Illustration of the two types of pedigrees resulting, for the same degree of completeness, in either an asymmetrical or symmetrical incomplete pedigree.

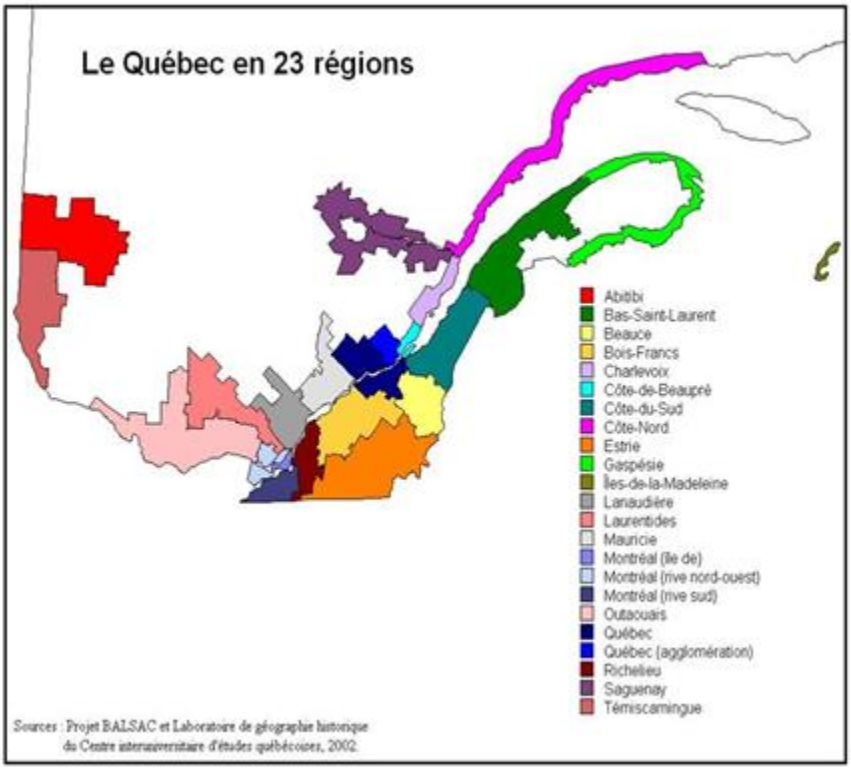

Figure S2. Map representing the 23 regions of French-Canada (Québec) for which the BALSAC project contains genealogical records. *Source: Project BALSAC and the laboratory of historical geography at the Inter-University Center for Québec Studies.*

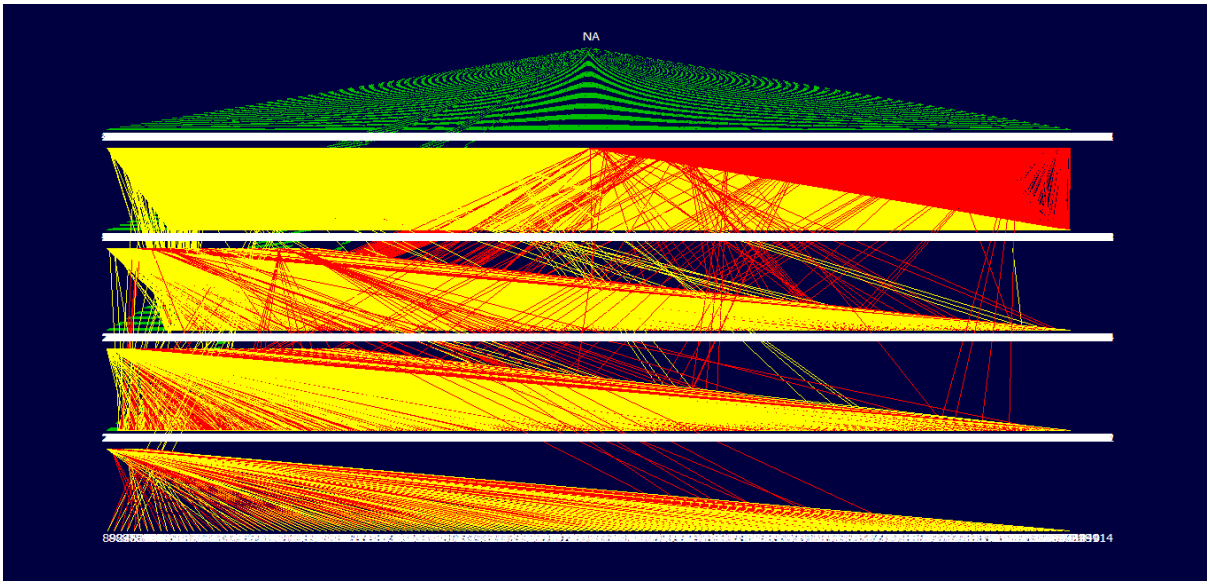

Figure S3. Visual representation of the pedigree of the subregion of *Gaspésie*. Red lines correspond to paternal links and yellow lines correspond to maternal links. Green lines correspond to individuals with no known parentage. Drawn using the program *Pedigree Viewer*

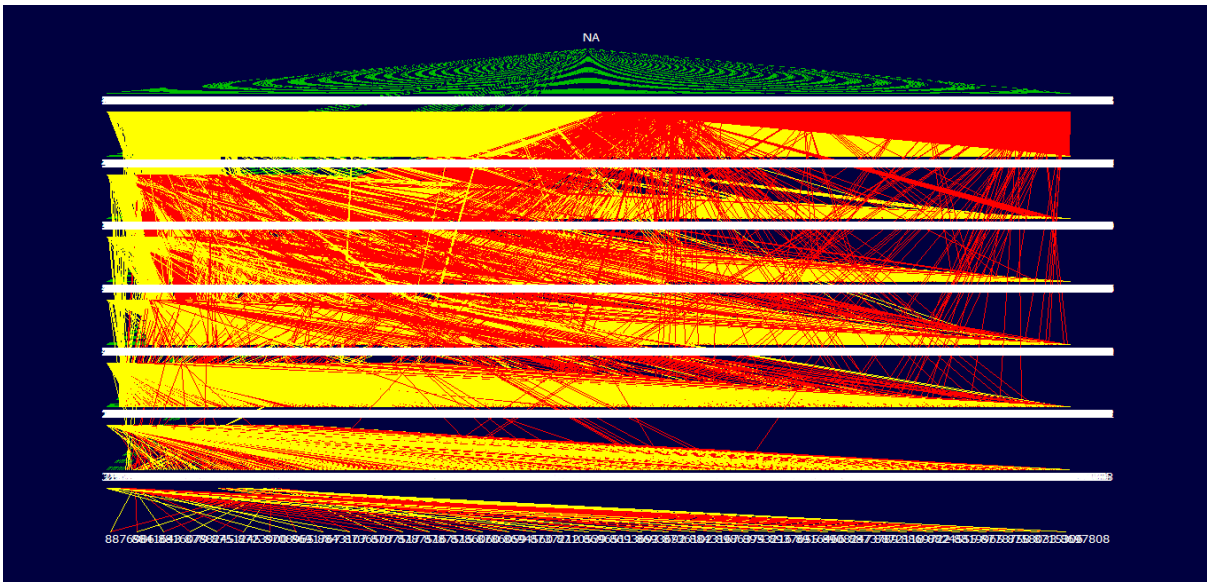

Figure S4. Visual representation of the pedigree of the subregion of *Bois-Francs*. Red lines correspond to paternal links and yellow lines correspond to maternal links. Green lines correspond to individuals with no known parentage. Drawn using the program *Pedigree Viewer*

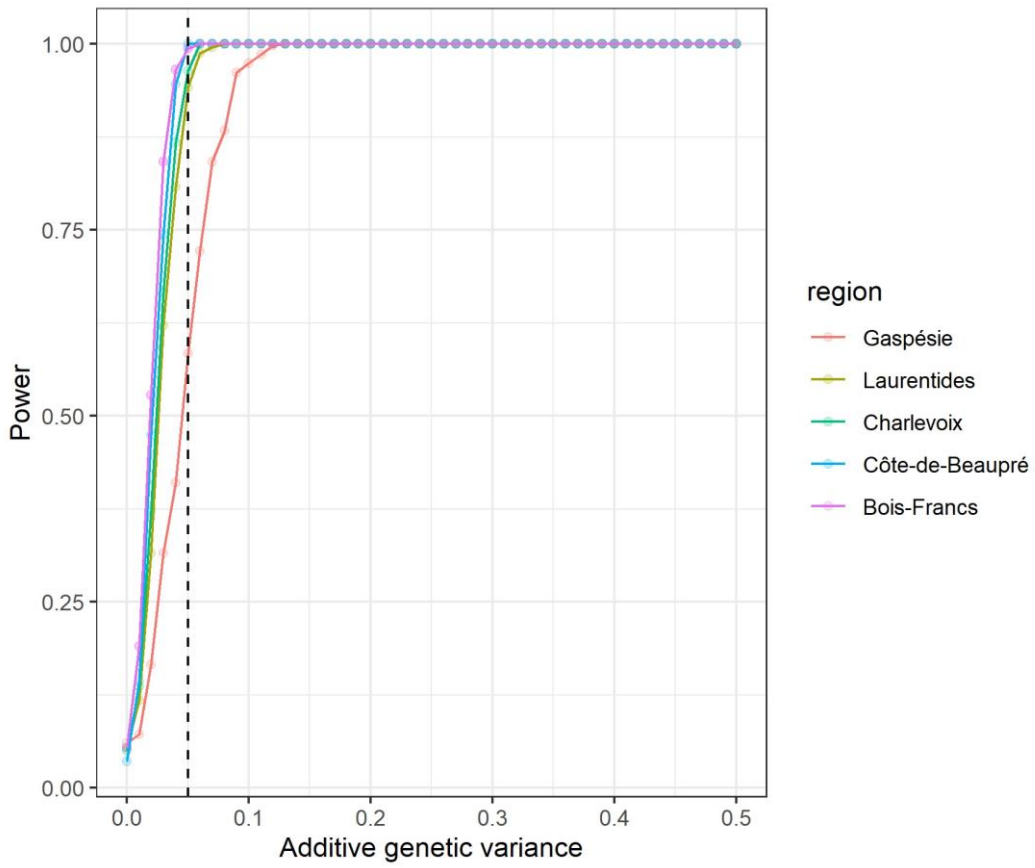

Figure S4. Power results based on a parent-offspring regression for each of the six pedigrees (univariate simple phenotype; V_A = [0 - 0.5]). Vertical dashed line represents the additive genetic value at which most pedigrees reach a 100% power.

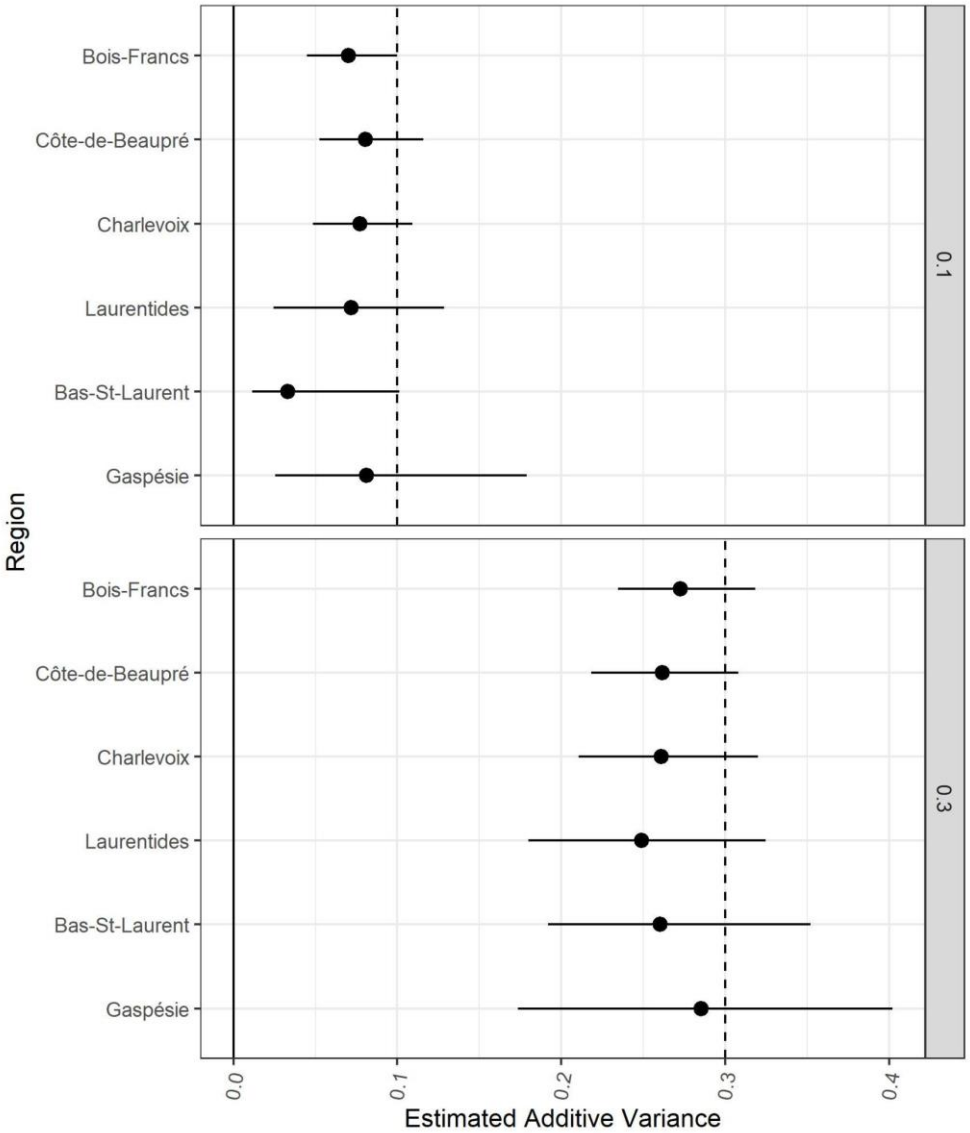

Figure S5. Estimates of additive genetic variance for each of the six pedigrees based on Scenario 1 (univariate simple phenotype; V_A = 0.1 or 0.3]) with a genealogical error rate increased to 1%. Each dot represents the average posterior mode-estimates, along with the average 95% credible intervals based on the five simulations per pedigree per parameter value. Vertical dashed lines represent the expected value of the variance.

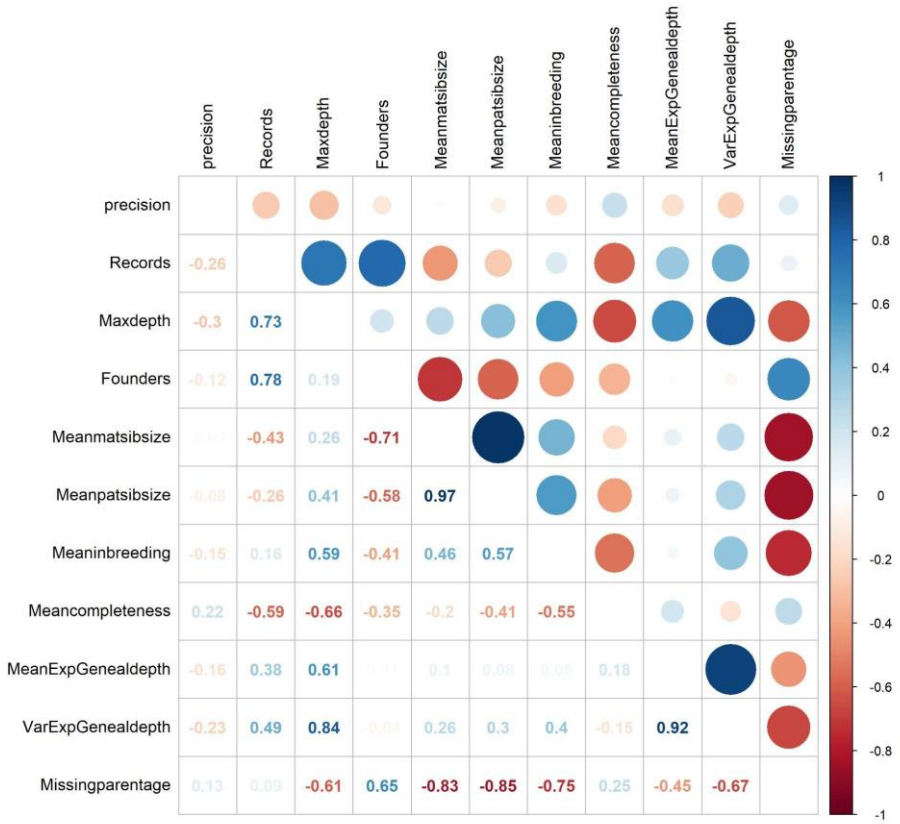

Figure S6. Correlation results between different pedigree and genealogical properties and the RMSE-based measure of precision and accuracy.

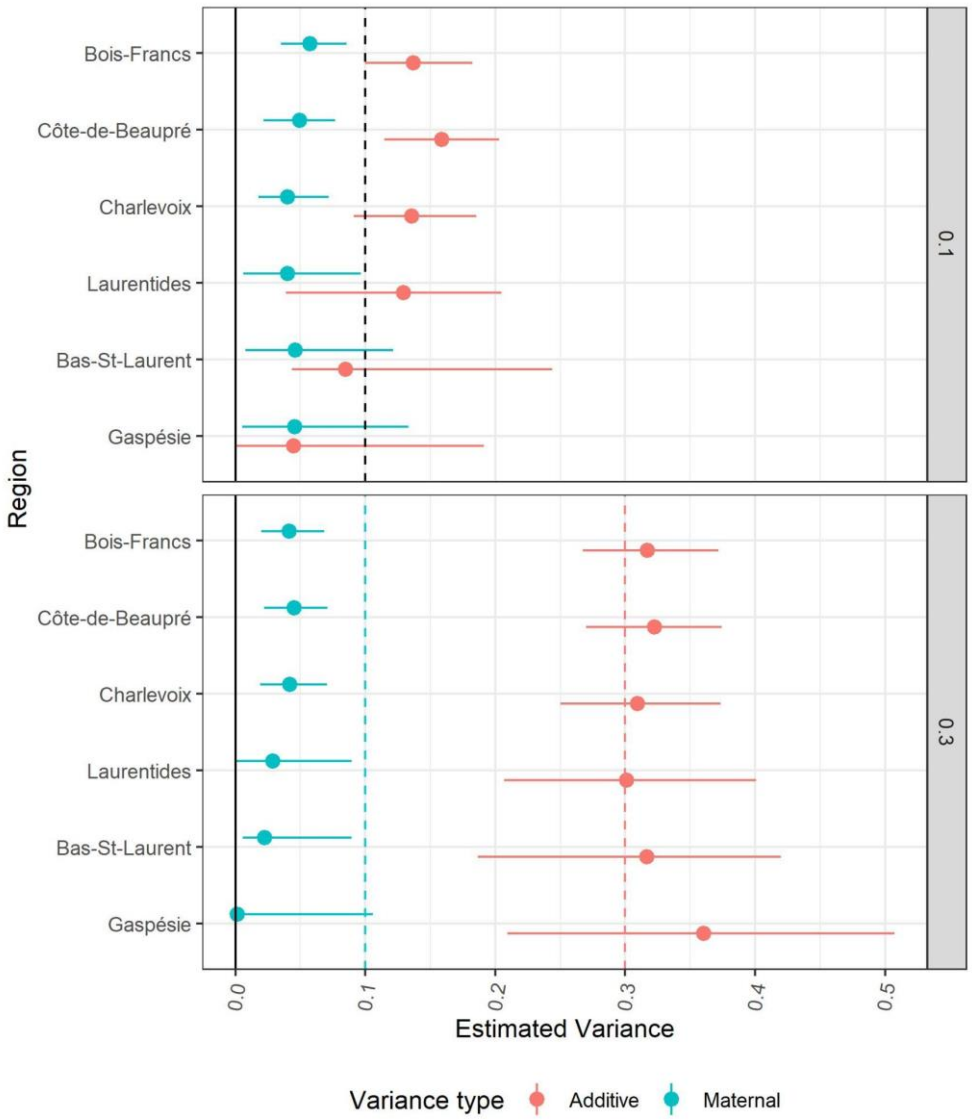

Figure S7. Estimates of additive genetic variance and maternal variance for each of the six pedigrees based on Scenario 2 (univariate complex phenotype; V_A = 0.1 or 0,3; V_M = 0.1) with a genealogical error rate increased to 1%. Each dot represents the average posterior mode-estimates, along with the 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.