Detecting inbreeding depression in structured populations

Eléonore Lavanchy[a,b], Bruce S. Weir[c], and Jérôme Goudet[a,b]

[a]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
[b]Swiss Institute of Bioinformatics, University of Lausanne, 1015 Lausanne, Switzerland
[c]Department of Biostatistics, University of Washington, Seattle WA 98195, USA

August 14, 2023

Abstract

Measuring inbreeding as well as its consequences on fitness is central for many areas in biology including human genetics and the conservation of endangered species. However, there is no consensus on the most appropriate method, neither for quantification of inbreeding itself nor for the model to estimate its effect on specific traits. In this project, we simulated traits based on simulated genomes from a large pedigree and empirical whole-genome sequences of human data from populations with various sizes and structure (from the 1,000 Genomes project). We compare the ability of various inbreeding coefficients (F) to quantify the strength of inbreeding depression: allele sharing, two versions of the correlation of uniting gametes which differ in the weight they attribute to each locus and two identical-by-descent segments-based estimators. We also compare two models: the standard linear model and a linear mixed model including a genetic relatedness matrix (GRM) as random effect to account for the non-independence of observations. We find linear mixed models give better results in scenarios with population or family structure. Within the mixed models, we compare three different GRM matrices and show that in homogeneous populations, there is little difference among the different F and GRM for inbreeding depression quantification. However, as soon as strong population or family structure is present, the strength of inbreeding depression can be most efficiently estimated only if (i) the phenotypes are regressed on inbreeding coefficient based on a weighted version of the correlation of uniting gametes, which gives more weight to common alleles and (ii) with the GRM obtained from an allele sharing relatedness estimator.

Introduction

Inbreeding is the result of mating between relatives and is often associated with reduced fitness, a phenomenon called inbreeding depression (ID) and which was observed in many different species such as humans [7, 6], other animals [26, 12, 21], and plants [34].

Many different methods have been developed for inbreeding quantification and there is no consensus on which one is the best [1, 5, 11, 25, 33, 35]. The classical approach was first proposed by Sewall Wright in 1922 and makes use of pedigrees (called hereafter F_{PED}) [31]. With the advances in sequencing technologies, genomic-based inbreeding coefficients (hereafter called F_{genomic}) have been developed. Among these, some coefficients rely on the comparison between observed and expected heterozygosity such as F_{HOM} [8, 27], the expected allele sharing between individuals such as F_{AS} [35] or on the correlation between uniting gametes such as F_{UNI} [32]. In addition to estimating the realized inbreeding coefficient and requiring no prior knowledge of the mating behavior of the population, these genomic estimates are simple and straightforward to compute and do not require whole-genome sequencing (WGS) data; a few thousands SNPs are usually sufficient for reliable inbreeding estimation in humans [11]. However they also have a disadvantage: they usually rely on allelic frequencies (except for F_{AS}) and therefore if these frequencies have not been correctly estimated, this will affect the estimation of these coefficients. Another inbreeding coefficient was proposed by McQuillan *et al.* (2008): F_{ROH} uses runs of homozygosity (ROHs), long homozygous stretches as a proxy for IBD segments within individuals [22]. A model-based approach relying on hidden Markov models has also been developed for detecting IBD segments [19] by identifying homozygous-by-descent (HBD) segments. This model is the basis for many other model-based IBD segments detection methods such as `BCFTools` [24], `BEAGLE` [3] and `RZooRoH` [10]. The inbreeding coefficient estimated with these model-based approaches will be called F_{HBD} from now on. One advantage of these methods is that they do not depend on allelic frequencies which can be very valuable when only a few individuals are available. However, it has been shown that these coefficients, and especially F_{ROH} , are sensitive to SNP density and parameters used, and there is no consensus on what is the most suitable set of parameters at present [23, 18].

How to quantify ID, although central to conservation genetics for decades [16], is still debated. This debate includes two sub-questions: which statistical model should be employed ? And which F ? Regarding the model, the classical approach consisted of the use of linear regression of the phenotypes on the inbreeding coefficient. However, other models have been utilized, such as Generalized Linear models (GLMs) with various link functions. In 2019, Nietlisbach *et al.*. [25] compared different models and found that the common GLM models with logit link did not allow for accurate inbreeding depression strength estimation. They propose using GLM with logarithm link functions. Ultimately, the type of model is largely dependent on the distribution of the trait.

Regarding the choice of which F is more accurate for quantifying ID, many

studies have demonstrated that F_{genomic} yields better results than F_{PED} [17, 2, 13]. However, some studies found F_{UNI} to be more accurate than F_{ROH} [33], while others found that F_{ROH} provided the best estimates of ID [17, 13, 25]. In 2020, Caballero *et al.* [5] used simulations and included several populations with different histories: they found that the optimal F actually depends on how large the population is. F_{ROH} did a better job at quantifying ID in populations with small effective size while F_{UNI} was better at predicting ID estimates in populations with large effective sizes. This result was later confirmed by Alemu *et al.* [1] used SNP-array empirical cattle data for several groups of allelic frequencies and concluded that F_{UNI} and F_{GRM} (F_I and F_{III} respectively in [32]) are better at quantifying homozygosity at rare alleles while F_{ROH} and F_{HOM} are better for alleles at intermediate frequencies and correlate better with whole-genome homozygosity. Indeed, recessive deleterious alleles, which are thought to be responsible for inbreeding depression, should segregate at low frequencies in large populations as a result of negative selection. On the contrary, in small populations, drift can increase the frequency of deleterious recessive alleles to intermediate frequencies, making F_{ROH} and F_{HOM} more suitable for detecting ID. Indeed, in the simulations conducted by Yengo *et al.* [33], rare alleles always caused negative effects on fitness (referred to as DEMA, for Directional Effect of Minor Alleles). The authors showed that F_{HOM} (and thus F_{AS} since they have similar properties) is sensitive to DEMA while F_{UNI} and F_{ROH} are not. They also showed via simulations that all estimates of ID are somewhat sensitive to population structure, F_{UNI} being the least affected. They recommend estimating ID using Linkage Disequilibrium (LD) score and Minor Allele Frequency (MAF) bins, and summing the ID estimates from these bins as an overall estimate of ID for the trait.

In this paper we simulated traits based on both simulated and empirical WGS human data from populations with varying sizes and structure. We show that some F are more sensitive to population structure and DEMA than others. We confirm only some of Yengo *et al.* [33] results. Importantly, we show that accounting for the non-independence of observations with a mixed model via an allele sharing based genomic relationship matrix (GRM) (rather than the standard `GCTA` GRM) and using a modified version of F_{UNI} which gives more weight to common alleles resolves most of the issues raised by Yengo *et al.* [33].

Material and Methods

Simulated pedigrees

We simulated a polygamous pedigree from a dioecious population with overlapping generations (hereafter called PEDIGREE) using custom `R` scripts. The population started from 500 founders (equal numbers of males and females), and followed a polygamous mating system: female fertilities per time interval were drawn from a Poisson distribution with parameter $\lambda = 1$, mortality rate per time interval was set to 0.5, and only 10% of the males were allowed to

reproduce at each time step. Matings were recorded for 25 time steps, resulting in a pedigree of $11,924$ individuals (over 25 time steps).

In order to simulate the genotypes of the individuals, we proceeded in two steps. We used the `mspms` wrapper to the `msprime` software [15] to simulate the two haplotypes containing $L = 650,000$ loci for each founder individual. The L loci were uniformly distributed along a constant recombination map $20M$ long. For each reproduction event, the number of cross-overs was first drawn from a Poisson distribution and then randomly positioned along the genome. The non-founder genotypes were then obtained by drawing two gametes: one from each parent. For each gamete, the allele at the first locus is selected at random between the two alleles of the parent. The alleles at the next loci along the chromosome are copied from the chromosome with the chosen allele at the first locus until a recombination event occurs, at which point the alleles are copied from the other chromosome until the next crossing-over or the end of the chromosome.

In order to investigate the effect of using more realistic smaller sample sizes, we subsampled 2,500 individuals from the PEDIGREE population. We performed two types of sub-sampling: i) a random sub-sampling where individuals were subsampled completely randomly, ii) a stratified sub-sampling where we sought to retain the widest range of inbreeding coefficients in the sub-sampled population. Consequently, for this stratified sub-sampling individuals with $F^w_{\text{UNI}} \geq 0.2$ were always included and individuals with $F^w_{\text{UNI}} < 0.2$ were randomly selected until the population reached the desired size. 100 replicates were performed for each sub-sampling.

1000 Genomes

In order to extend our conclusions to even smaller sample sizes and populations with stronger structure (which are common in wild and/or endangered species), we used empirical data from phase 3 from the 1,000 Genomes project [28]. We considered i) a small sample from a homogeneous population with small effective size represented by 504 individuals from the super-population with East-Asian ancestry (EAS), ii) a small sample from a population with some admixture and larger effective population sizes represented by 661 individuals from the super-population with African ancestry and admixed individuals (AFR) and finally iii) a larger sample from a population with larger effective size and with genetic structure (global $F_{ST} = 0.083$) comprising all the 2,504 individuals (hereafter called WORLD) and represented by five super-populations: individuals with East-Asian ancestry (EAS), African ancestry (AFR), European ancestry (EUR), admixed American ancestry (AMR) and finally South-Asian ancestry (SAS). A more detailed description of the samples can be found at the 1,000 Genomes Project website.

Simulated traits

We simulated traits based on equation 1 following [33]: we consider a trait y whose phenotype is partly determined by the genotypes at Lc causal loci with $h^2 = 0.8$. We assume these loci to be bi-allelic, with one allele encoding for an increase in the trait value (the plus allele) and the other encoding for a decrease in trait value (the minus allele). Dominance was also considered since inbreeding depression (ID) occurs only if there is directional dominance: when heterozygotes at loci encoding for the trait are closer on average to the homozygote for the plus allele [20]. If gene effects are purely additive or if dominance is not directional, there is no ID. Finally, we assume no epistasis between loci, and no genotype-environment interaction.

For individual j, y_j is the individual trait value (its phenotype), calculated as the sum of allelic and genotypic effects over causal loci, an environmental effect and μ, the average trait value among all individuals. At locus l, x_{jl} is the minor allele count (MAC) $\in \{0, 1, 2\}$ of individual j. a_l represents the additive effect size of the alternate allele at locus l. d_l is the dominance effect size, the deviation of the heterozygous genotype from the mean of the two homozygotes. Finally, ϵ_j is the environmental contribution to the phenotype of individual j, drawn from a normal distribution.

$$y_j = \mu + \sum_{l=1}^{Lc} x_{jl} a_l + \sum_{l}^{Lc} x_{jl}(2 - x_{jl}) d_l + \epsilon_j \tag{1}$$

The strength of inbreeding depression b was set to -3 in all simulations, as in Yengo *et al.* [33]. We chose a value which was close to zero because if the the effect of inbreeding is too strong, it will always be detected. In addition, this value is in the range of observed inbreeding depression published estimates (for instance, table 10.4 from [20]).

We used equation 1 to simulate traits with varying architectures. To avoid causal markers with extremely low frequencies, we first excluded loci with $MAF \leq 0.01$ for both the EAS and AFR populations and loci with $MAF \leq 0.001$ for both the PEDIGREE and WORLD populations. We then simulated traits using 1,000 randomly chosen SNPs (after MAF filtering). We drew both the raw additive effect sizes of the alternate allele and the raw dominance effect sizes from a uniform $[0, 1]$ distribution (other distributions were explored with almost no effect on the results (results not shown)). As we expect alleles causing ID to be counter selected and thus removed or maintained at a low frequency (proportionally to their detrimental effect), the raw effect sizes were scaled inversely to MAF $a_j = raw_{aj}/p_j$ to mimic negative selection. We also scaled the dominance effects inversely to the locus expected heterozygosity $d_j = raw_{dj}/(2p_j(1-pj))$. In addition, we attributed the same sign to the effect sizes of all minor alleles in order to include what Yengo *et al.* [33] called Directional Effect of Minor Alleles (DEMA) [33]. However, in order to investigate the effect of the parameters mentioned above, we also simulated traits where the additive and dominance effect sizes were left unchanged $a_j = raw_{aj}$ and

$d_j = raw_{dj}$ and without DEMA. A summary of all the simulated scenarios can be found in table S1. In addition, graphical representation of the additive effect sizes and dominance coefficients distribution under these different scenarios can be found in Figure S1.

Individual inbreeding coefficients

We estimated individual inbreeding coefficients using several methods whose properties were recently described in detail in Zhang *et al.* [35]. Regarding the figures and tables presented in the main text, we do not filter on MAF for any of the Fs estimates. We use one allele-sharing-based estimator of inbreeding, hereafter called F_{AS} and described in [30, 35]:

$$F_{AS_j} = \frac{\sum_{l=1}^{L} A_{jl} - A_{Sl}}{\sum_{l=1}^{L} 1 - A_{Sl}} \tag{2}$$

where A_{jl} indicates the identity of the two alleles an individual j carries at locus l: one for homozygous and 0 for heterozygous and A_{Sl} is the average allele sharing proportion at locus l for pairs of individuals $j, k, j \neq k$.

Then, we compare two versions of F_{UNI} (initially described in [32]) and which measure the correlation between uniting gametes. The first version (hereafter called F^{u}_{UNI}) is the original F_{UNI} [32] measured as the average of ratios over SNPs (which attributes equal weight to all loci):

$$F^{u}_{\text{UNI}_\text{j}} = \frac{1}{L} \sum_{l=1}^{L} \frac{x_{jl}^2 - (1 + 2p_l)x_{jl} + 2p_l^2}{2p_l(1 - p_l)} \tag{3}$$

Similarly to equation 1, x_{jl} is the MAC of individual j at locus $l \in \{0, 1, 2\}$ and p_l is the derived allele frequency at locus l.

The second version (hereafter called F^{w}_{UNI}) is a modified version of F_{UNI} which measures the ratio of averages and thus gives more weight to loci with larger expected heterozygosity (i.e. with MAF close to 0.5). We are not aware of other investigations using the ratio of averages estimator F^{w}_{UNI} in the context of ID estimation.

$$F^{w}_{\text{UNI}_\text{j}} = \frac{\sum_{l=1}^{L} x_{jl}^2 - (1 + 2p_l)x_{jl} + 2p_l^2}{\sum_{l=1}^{L} 2p_l(1 - p_l)} \tag{4}$$

We also used four Identical-by-descent (IBD) segments based F. We called runs of homozygosity (ROHs) with `PLINK` [27] and default parameters. We also called Homozygous-by-descent (HBD) segments with `BCFTools` [24]. For both methods, we selected ROHs or HBD segments based on their size: either larger than 100Kb: $F_{\text{ROH}_{\text{100KB}}}$ and $F_{\text{HBD}_{\text{100KB}}}$ or larger than 1Mb: $F_{\text{ROH}_{\text{1MB}}}$ and $F_{\text{HBD}_{\text{1MB}}}$. For both methods the inbreeding coefficients were simply estimated as the fraction of genome falling within ROHs or HBD segments.

Finally, in the PEDIGREE population, we used the pedigree-based inbreeding coefficient: F_{PED} [31].

All inbreeding coefficients were estimated separately for each population of the 1,000 Genomes Project (EAS, AFR, WORLD) and with population specific SNPs and allelic frequencies (i.e. we removed monomorphic SNPs and estimated allelic frequencies in all the three populations). Consequently the same individual might have different F_{genomic} in the EAS and the WORLD population. This influenced only the IBD segments-based inbreeding coefficients (F_{ROH} and F_{HBD}) trivially but greatly influenced F_{AS} (though the rank of inbreeding among individuals was conserved) and both F_{UNI} (for which the rank of inbreeding among individuals was not conserved) since their formulae rely on allelic frequencies estimations. Comparison among the different inbreeding coefficients per population can be found in supplementary material (Figures S2 - S5). More details can be found in [35],

Estimation of Inbreeding Depression: *b*

We estimated the strength of ID (hereafter defined as b) using two different models. In the first model, b was estimated as the slope of regression of phenotypes on the different inbreeding coefficients with a classical linear model (LM):

$$\hat{b}_{LM} = Cov(Y, F)/Var(F)$$

where Y is the vector of trait values and F is the vector of individual inbreeding coefficients estimates.

In the second model, we estimate b as the fixed effect coefficient associated with the inbreeding coefficient in the following linear mixed model (LMM):

$$Y = bX + \omega + \epsilon$$

where Y is the vector of trait values, X is a matrix with two columns, the first containing ones and the second the individual inbreeding coefficients, ω is the random component of the mixed model with $\omega \sim N(0, \tau K)$, K being the genomic relationship matrix (GRM) and τ the additive variance component. Finally, ϵ is the individual residual variance and is defined as $\epsilon \sim \sigma^2 I_n$. From this, b is estimated as follows:

$$\hat{b}_{LMM} = (X'V^{-1}X)^{-1}X'V^{-1}Y$$

with $V = \tau K + \sigma^2 I_n$ [9]. We compare three different GRMs we estimated using all loci (no MAF filtering). The first mixed model included a GRM derived from allele sharing [11], hereafter called LMM_{AS} . We used the `R Hierfstat` package to estimate K and the `R gaston` package to estimate V and b. We could not use `GCTA` software to run the mixed model for this GRM because its leading eigenvalue is negative which the Choleski decomposition algorithm used for matrix inversion in `GCTA` cannot handle (it requires a positive definite matrix), while the Schur decomposition algorithm used in `gaston` can. We note that the standard GRM is not positive definite (one eigen value is 0), but the matrix to invert in the mixed model is not the GRM itself but $V = \tau K + \sigma^2 I_n$

which becomes positive definite and can be inverted if the heritability is smaller than one.

The second mixed model used the `GCTA` weighted GRM matrix [11, 29]. Similarly to F^{w}_{UNI} , this matrix uses the ratio of averages. For this model, we used `GCTA` and the `R SNPrelate` package to estimate V. We then used the `R gaston` package for estimating b with the LMM.

Finally, the third mixed model used the `GCTA` unweighted GRM matrix [32] which (similarly to F^{u}_{UNI}) utilizes the average of ratios and thus gives equal weight to all loci. For this model, we used `GCTA` to estimate V. We then estimated b with the LMM implemented in the `R gaston` package.

Note that the Average Information-Restricted Maximum Likelihood (AIREML) fitting method we used in the LMM is an iterative procedure, and should result in unbiased estimates. In some cases, the model did not converge, and gave highly biased b. For each scenario, regression model and population, the number of replicates which did not converge can be found in tables S6-S8.

Results

All the figures presented in the main text picture the scenario where alleles additive effect sizes and dominance coefficients are proportional to MAF and where there is a directional effect of minor alleles (DEMA) (i.e. the ADD & DOM & DEMA scenario from table S1) (see Figure S1). The results for the other scenarios are shown and discussed in supplementary material (Figures S8-S15, tables S2-S5).

Simulated pedigrees

Figure 1 presents the inbreeding depression (ID) strength estimates (b, see the methods section) for the different inbreeding coefficients (F), with two regression models in the PEDIGREE populations. The first column shows b estimated with the simple LM and the second column shows b estimated with LMM including the allele sharing GRM as random factor (LMM_{AS}). The first row shows results for the complete PEDIGREE population (n = 11,924). The second row shows results for a reduced sample size of the PEDIGREE population (n = 2,500, meant to match the size of the 1KG WORLD population) where sub-sampled individuals were chosen completely randomly. The third row also shows results for a reduced sample size of the PEDIGREE population (n = 2,500) but these individuals were selected to represent the entire spectrum of inbreeding values. The violin plots show b estimates distributions among the simulation replicates (100 replicates for the complete population, 10,000 replicates for both sub-sampled populations). The solid dark grey line is the true strength of ID (b = -3). The dashed red line represents the absence of ID (b = 0), indicating that ID was not detected in any replicate above this line. Root mean square error (RMSE) values associated with both models and populations are shown in table 1. Strikingly, in the PEDIGREE population, no F resulted

in a accurate estimation of b with the simple LM, whatever the sample size (Figure 1, panels A, C and E; Table 1). The inclusion of a GRM matrix as a random factor allowed for the correction of non-independence of observations and greatly improved b estimation (Figure 1, panels B, D, and F; table 1). In the complete PEDIGREE population, we see little difference between the three GRMs we tested (1, panel B vs Figure S8, panels A and B; table 1): all F yielded efficient (we use efficient to describe an estimate with low RMSE, thus which is unbiased and has low variance) estimates of b when used inside a LMM, except for F^{u}_{UNI} that slightly overestimates the strength of ID while F_{PED} slightly underestimates it. This suggests that large sample sizes (here 11,924 individuals) combined with a mixed model allow efficient ID estimation regardless of the F used. The three mixed models, however, perform less efficiently when the sample size is reduced, as we demonstrate with both subsampled PEDIGREE populations (n = 2,500): many replicates produced estimates above zero for b (Figure 1, panels D and F; Figure S8, panels C to F; table 1). RMSE were particularly large for F_{PED} , $F_{\text{HBD}_{100\text{KB}}}$ and $F_{\text{ROH}_{100\text{KB}}}$ with the mixed model using the unweighted `GCTA` GRM matrix ($\text{LMM}_{GCTA^{u}}$) (Figure S8, panel D; table 1). Additionally, increasing the variance of sub-sampled individuals' F (i.e. ranged sub-sampling) led to better estimates of b with reduced variance among replicates compared to random subsampling (Figure 1, panels D vs F: Figure S8, panels C vs E and D vs F, table 1).

1,000 Genomes Project

Figure 2 illustrates the estimates of ID strength (b) for the different inbreeding coefficients (F), when using either a LM or a LMM for two subsets of the 1,000 Genomes Project: EAS and AFR, as well as for the entire world population. It has the same structure as Figure 1. Root mean square error (RMSE) values associated with both models and populations can be found in table 2. Interestingly, we see little difference between LM and LMM and the different GRMs when there is no structure among the samples even with small sample sizes (**EAS**: Figure 2, panel A and B vs Figure S6, panels A and B; table 2; **AFR**: Figure 2, panel C and D vs Figure S6, panels C and D; table 2). Similarly to what was observed for the PEDIGREE population, when some structure exists (population structure in the WORLD population compared to family structure in the PEDIGREE population), the simple LM fails to accurately estimate the strength of ID, regardless of the F (Figure 2, panel E; table 2). In contrast to the pedigree population showing no difference between the three GRMs (Figure 1 and Figure S6), the most efficient estimates of b are obtained only with the LMM_{AS} model and with F^{w}_{UNI} in the highly structured WORLD population (Figure 2, panel F vs Figure S7 panels E and F; table 2). In fact, the models including the $GCTA^{w}$ and $GCTA^{u}$ matrices cannot efficiently estimate b with any of the inbreeding coefficients: even though F^{w}_{UNI} is unbiased, the variance is very large (panel F; Figure S7, table 2). In addition, several replicates did not converge when both $GCTA^{w}$ and $GCTA^{u}$ models were used which was never the case with the GRM_{AS} . Numbers of such replicates are indicated in the

Figures' legend and in supplementary tables S6-S8.

Comparing inbreeding coefficients

With both the LM and LMM_{AS} models in the three populations from the 1,000 Genomes Project (EAS, AFR and WORLD, panels A - F) and for the LM in the PEDIGREE population, F_{AS} is consistently underestimating the strength of ID, particularly when there is strong structure (WORLD: Figure 2, panels E and F). It is because DEMA is included in the model and strongly influences the quantification of ID by F_{AS} . In the absence of a DEMA, F_{AS} produces efficient estimates (Figures S10 - S13). In addition, F_{AS} is sensitive to the dominance effects being proportional to MAF but to a lesser extent and in the opposite direction (Figure S8 vs Figure S9). Concerning the other SNP-based F, F^{u}_{UNI} is constantly overestimating the strength of ID and is the most sensitive to population structure: its variance is much larger compared to F^{w}_{UNI} in the structured WORLD population and with all models (Figure 2, panel F; table 2). Interestingly, the variance of F^{u}_{UNI} is affected only when allele effect sizes and/or dominance coefficients are proportional to MAF, but not by DEMA (Figures S8-S15). In contrast, F^{w}_{UNI} is the least sensitive to allele effect sizes or dominance coefficients proportional to MAF and DEMA (Figures S8 – S15), which makes it the most appropriate F for estimating ID (Figure 2, panel F; table 2). Since the difference between F^{w}_{UNI} and F^{u}_{UNI} is the weight given to rare and common alleles, we conducted the same analyses (including the re-estimation of both F and GRMs estimation) on the WORLD population but excluding loci with MAF > 0.05 and showed that there is no difference between F^{w}_{UNI} and F^{u}_{UNI} when rare alleles are removed (Figure S16). Concerning the F calculated from ROHs and HBD segments, there is not much difference between `PLINK` and `BCFTools` except for the variance among b estimates, which is slightly smaller with `BCFTools` compared to `PLINK` (Figure 2, panels A - F; table 2). In addition, focusing on recent inbreeding by including only large segments (here larger than 1MB) yielded better results in the WORLD population (Figure 2, panel F). Since `BCFTools` is a model-based HBD approach, there is no mandatory length requirement. In light of this, we also estimated F_{HBD} based on HBD segments without any size restrictions, and the results are similar to those obtained using $F_{\text{HBD}_{100\text{KB}}}$ (Figure S17).

Comparing genetic relatedness matrices

Since we identified F^{w}_{UNI} as the best inbreeding coefficient, Figure 3 contrasts the four different models for this coefficient in the four populations: each panel corresponds to one population. As mentioned above, there is almost no difference among the different GRM matrices in the extremely large complete PEDIGREE population (Figure 3, panel A; table 1) and between any of the models in the two homogeneous populations (EAS and AFR) (Figure 3, panels B and C; table 2). However, in the highly structured WORLD population, LMM_{AS} gives the

most accurate result due to its smaller variance and RMSE (Figure 3, panel D; table 2).

Distribution of additive and dominance effects

We found a difference between the three linear mixed models only because the scenario presented in the main text includes effect sizes and dominance coefficients proportional to causal markers' MAF as well as DEMA. When none of these three parameters are included, there is little difference between the three linear mixed models (Figure S8, panels B, F, J, N vs panels C, G, K, O vs panels D, H, L, P; tables S2-S5). Additional simulations were conducted without additive and dominance coefficients proportional to loci's MAF and DEMA to assess their impact on ID detection. The individual and pairwise effects of additive and dominance coefficients being proportional to MAF and DEMA (the other scenarios of table S1) are explored and discussed in details in supplementary material and Figures S8-S15.

Finally, we also investigated i) the effect of the LDMS stratification method proposed by Yengo [33] (Figures S8-S15) but found that it only improves results with the simple LM and ii) the effect of using intermediate frequencies causal loci (Figure S18) which reduced the variance in b estimates for all inbreeding coefficients.

Discussion

By analyzing the phenotypes of a large simulated pedigreed polygamous population with high family structure as well as subsets of the 1000 genomes project [28], we demonstrated that, despite population or family structure, inbreeding depression estimates can be accurately measured if the data are analyzed with a mixed model using the genomic relationships among individuals as random effect. In comparison to the other genomic relationship matrices (GRMs), the one based on allele sharing provides the most consistent and accurate results, especially for smaller sample sizes and samples with a high family or population structure. And, among the several inbreeding estimators tested, F^{w}_{UNI} proved the most reliable to quantify inbreeding depression.

We observed trivial differences among the different models when there is no population structure (i.e. in the EAS and AFR populations). However, as soon as there is some structure (the WORLD and POLYPED populations) the classical linear model (LM) completely fails to estimate b regardless of the inbreeding coefficient used. This result is concordant with Yengo *et al.* (2017) [33] where the authors quantified ID using a simple linear model and demonstrated that F_{HOM} (whose properties are very similar to F_{AS}), F^{u}_{UNI} and two different F_{ROH} were sensitive to population structure. As for the comparison of three linear mixed models (LMM), they perform equally when there is no population structure (EAS and AFR) or very large sample sizes (11,924 individuals from the complete PEDIGREE population). Although samples of this

size are common for research on humans, they will seldom be found in wild populations. We therefore subsampled the PEDIGREE population to 2,500 individuals in order to investigate the effect of a smaller sample size. We used two types of sub-sampling: i) random sub-sampling where individuals were chosen completely randomly and ii) ranged sub-sampling where individuals where chosen to maximise the range of F in the sampled population. We stress that what we consider a small sample size (2,500 individuals) will not be found in many wild species, particularly for endangered populations, where monitoring inbreeding and inbreeding depression are critical. As expected, when we subsampled individuals from the PEDIGREE population, RMSE values associated with b estimation increased slightly for both LMM_{AS} and $\text{LMM}_{GCTA^{w}}$ mixed models and we failed to detect ID in some replicates. Accordingly, even with 2,500 individuals, we lack power and several thousands of individuals would be required to detect ID efficiently as Keller et al. and Caballero et al. previously pointed out [16, 4]. With the $\text{LMM}_{GCTA^{u}}$ mixed model, all inbreeding coefficients but F_{AS} and F_{UNI} had convergence issues, suggesting that the $\text{LMM}_{GCTA^{u}}$ mixed model is the least robust of the three mixed models. As expected, randomly sub-sampling individuals lead to a larger variance of b estimates compared to the ranged sub-sampling scheme, indicating that maximizing the variance of samples' F improves the estimation of b, although it is not obvious how such sampling could be done in non monitored natural populations. When we add strong population structure in addition to the small sample size (2,504 individuals from the highly structured WORLD population), we observe striking differences between the three different GRMs. The linear mixed model including the allele sharing based GRM (LMM_{AS}) resulted in the most efficient estimations of b. In addition, the mixed models with both $GRM_{\text{GCTA}^{\text{u}}}$ and $GRM_{\text{GCTA}^{\text{w}}}$ did not converge for high percentages of replicates (compared to 0% for LMM_{AS}) emphasizing that LMM_{AS} is the best model for quantifying inbreeding depression in highly structured populations (although the most used GRM is currently the one estimated from `GCTA`). This is because the allele sharing based GRM matrix is a better estimator of kinship compared to both `GCTA` matrices [11, 30]. Indeed what the GRM_{AS} estimates is the actual kinship in the population, based on how many alleles individuals share. In contrast, what both $GRM_{\text{GCTA}^{\text{w}}}$ and $GRM_{\text{GCTA}^{\text{u}}}$ estimate is a combination of individual kinship, their mean kinship with the other individuals and the overal mean kinship in the population (see eq. 3 in Goudet *et al.* [30]). Consequently, since the kinship itself is better estimated with GRM_{AS} , the non-independence of observations (and thus the population structure) is better accounted for with LMM_{AS} which leads to better b estimates. Importantly, the inclusion of a GRM in the ID estimation model is not limited to simple linear models. Even though we used only linear models in this study, any type of generalized linear model can incorporate a GRM as a random factor. Consequently this method can be applied to any trait distribution. Furthermore, by including the GRM-based random factor, the non-independence of observations is better accounted for than by including the population as a random factor, and no prior knowledge of the population structure is required.

Comparing F

Concerning the different inbreeding coefficients, we found F^{w}_{UNI} to be the best F for quantifying inbreeding depression. Indeed, F^{w}_{UNI} was the only coefficient we tested which was not sensitive to either additive and dominance effect sizes being proportional to MAF or DEMA resulting in the least biased estimation of b. On the contrary, we found that F^{u}_{UNI} was influenced by the dominance effect sizes being proportional to MAF and by population structure. Since F^{u}_{UNI} gives equal weight to all loci, the rare allele associated with large dominance coefficients add noise in the estimation of b. Similarly, when there is population structure, rare alleles which have strong influence on F^{u}_{UNI} are likely to be private alleles which will strongly bias population-specific allelic frequencies and eventually F^{u}_{UNI} estimation. Importantly, F^{u}_{UNI} performed as well as F^{w}_{UNI} when we filtered on $MAF > 0.05$ for F and all GRMs estimation. This is because F^{u}_{UNI} uses the average of ratios, which results in loci with small MAF strongly influencing the outcome. When these rare loci are filtered out, the estimated F is no longer biased. This explains why Yengo *et al.* [33] found that F^{u}_{UNI} was the best F for quantifying inbreeding depression with an homogeneous subset of the UK bio bank dataset: they filtered on $MAF > 0.05$ leading to F^{u}_{UNI} estimation not being influenced by rare alleles with strong additive and/or dominance effect sizes. Concerning F_{AS} , we found that it was very sensitive to DEMA. This result is also concordant with Yengo *et al.* [33] who found that F_{HOM} (with properties very similar to F_{AS}) was sensitive to DEMA. In this paper the authors explain that this sensitivity is due to F_{HOM} (and thus F_{AS}) correlating strongly with minor allelic count which will create a spurious association with inbreeding depression in the presence of DEMA. However, F_{AS} resulted in the most accurate estimates of b when DEMA was not included in the model, suggesting that it is the best F to estimate inbreeding for neutral regions. Finally, we found that ROHs and HBD segments based F, namely F_{ROH} and F_{HBD} , performed poorly: underestimating the strength of inbreeding depression (positive b) or displaying very large variance among replicates. This result is in contradiction with Kardos *et al.* [13, 14] and Nietlisbach *et al.* [25] who found that F_{ROH} and F_{HBD} were better at quantifying inbreeding depression compared to SNPs-independent based F. However, Alemu *et al.* [1] and Caballero *et al.* [5] showed the best F actually depends on the history of the population. Indeed, they showed that F_{ROH} and F_{HBD} and to a lesser extent F_{HOM} were better at quantifying homozygosity at loci with common alleles. On the contrary, F^{u}_{UNI} was better at quantifying homozygosity at rare alleles. Alemu *et al.* [1] and Caballero *et al.* [5] propose that, on the one hand, in populations with low effective sizes, selection is weaker and deleterious alleles may be able to reach intermediate frequencies as a result of drift. Therefore both F_{ROH} and F_{HBD} (and F_{HOM} in their analyses) should perform better in such populations. In our study, the standard scenario (with no ADD, no DOM and no DEMA) mimics what happens in such small populations and we found that F_{ROH} , F_{HBD} and F_{AS} (which has similar properties to F_{HOM}) performed better than F^{u}_{UNI} (which is the F_{UNI} they tested) in the highly structured WORLD population

and to a lesser extent in the family structured PEDIGREE population. With homogeneous populations, we do not observe any difference between these inbreeding coefficients. Nevertheless, this is consistent with Alemu [1] results, as they used families which consequently create structure. On the other hand, in populations with a large effective size, selection maintains deleterious alleles at low frequencies which explains why Yengo *et al.* (2017) found that F_{UNI} was the best F with the large UK biobank dataset and this is consistent with what we have found with the ADD & DOM & DEMA scenario which mimics what happens in populations with large effective sizes.

Conclusion

In this paper, we showed that the more accurate method for estimating inbreeding depression is to use a mixed model with an allele-sharing-based relatedness matrix as a random component but F^{w}_{UNI} as the inbreeding coefficient to predict inbreeding depression. The most commonly used GRM ($GRM_{\text{GCTA}^{\text{u}}}$) results in biased and highly variable estimates of b in structured populations. We stress that even if the results are greatly improved by using the allele-sharing GRM and F^{w}_{UNI} , the variance among replicates is still large and no inbreeding depression is detected in several replicates ($\hat{b} \geq 0$) in the highly structured WORLD population as well as in the small and slightly admixed AFR population. Therefore, detecting efficiently inbreeding depression of the magnitude commonly found and that we simulated requires very large sample sizes with several thousand individuals, particularly in structured populations. Unfortunately, this might be hardly feasible for wild and/or endangered populations.

References

[1] Setegn Worku Alemu, Naveen Kumar Kadri, Chad Harland, Pierre Faux, Carole Charlier, Armando Caballero, and Tom Druet. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. *Heredity*, 126(3):410–423, 2021.

[2] Camillo Bérénos, Philip A. Ellis, Jill G. Pilkington, and Josephine M. Pemberton. Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. *Molecular Ecology*, 25(13):3152–3168, 2016.

[3] Brian L. Browning and Sharon R. Browning. Detecting Identity by Descent and Estimating Genotype Error Rates in Sequence Data. *The American Journal of Human Genetics*, 93(5):840–851, 2013.

[4] Armando Caballero, Almudena Fernández, Beatriz Villanueva, and Miguel A. Toro. A comparison of marker-based estimators of inbreeding and inbreeding depression. *Genetics Selection Evolution*, 54(1):82, 2022.

[5] Armando Caballero, Beatriz Villanueva, and Tom Druet. On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. *Evolutionary Applications*, n/a(n/a), 2020.

[6] Francisco C. Ceballos, Scott Hazelhurst, David W. Clark, Godfred Agongo, Gershim Asiki, Palwende R. Boua, F. Xavier Gomez-Olive, Felistas Mashinya, Shane Norris, James F. Wilson, and Michele Ramsay. Autozygosity influences cardiometabolic disease-associated traits in the AWI-Gen sub-Saharan African study. *Nature Communications*, 11(1):5754, November 2020. Place: Berlin Publisher: Nature Research WOS:000594647500005.

[7] Francisco C. Ceballos, Scott Hazelhurst, and Michèle Ramsay. Assessing runs of Homozygosity: A comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics*, 19:106, 2018.

[8] Christopher C. Chang, Carson C. Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 2015.

[9] Claire Dandine-Roulland and Hervé Perdry. The Use of the Linear Mixed Model in Human Genetics. *Human Heredity*, 80(4):196–206, 2015.

[10] T. Druet and M. Gautier. A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular Ecology*, 26:n/a–n/a, 2017.

[11] Jerome Goudet, Tomas Kay, and Bruce S. Weir. How to estimate kinship. *Molecular Ecology*, 27(20):4121–4135, 2018.

[12] Jisca Huisman, Loeske E. B. Kruuk, Philip A. Ellis, Tim Clutton-Brock, and Josephine M. Pemberton. Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences*, 113(13):3585–3590, 2016.

[13] M. Kardos, G. Luikart, and F. W. Allendorf. Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. *Heredity*, 115(1):63–72, 2015.

[14] Marty Kardos, Pirmin Nietlisbach, and Philip W. Hedrick. How should we compare different genomic estimates of the strength of inbreeding depression? *Proceedings of the National Academy of Sciences*, 115(11):E2492–E2493, 2018.

[15] Jerome Kelleher, Alison M. Etheridge, and Gilean McVean. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. *PLOS Computational Biology*, 12(5):e1004842, 2016.

[16] Lukas F. Keller and Donald M. Waller. Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17(5):230–241, May 2002.

[17] Matthew C. Keller, Peter M. Visscher, and Michael E. Goddard. Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics*, 189(1):237–249, 2011.

[18] Eléonore Lavanchy and Jérôme Goudet. Effect of reduced genomic representation on using runs of homozygosity for inbreeding characterization. *Molecular Ecology Resources*, 23(4):787–802, May 2023. Publisher: John Wiley & Sons, Ltd.

[19] Anne-Louise Leutenegger, Bernard Prum, Emmanuelle Génin, Christophe Verny, Arnaud Lemainque, Françoise Clerget-Darpoux, and Elizabeth A. Thompson. Estimation of the Inbreeding Coefficient through Use of Genomic Data. *The American Journal of Human Genetics*, 73(3):516–523, September 2003.

[20] Michael Lynch and Bruce Walsh. *Genetics and Analysis of Quantitative Traits*. Sinauer, 1998.

[21] K. Martikainen, A. Sironen, and P. Uimari. Estimation of intrachromosomal inbreeding depression on female fertility using runs of homozygosity in Finnish Ayrshire cattle. *JOURNAL OF DAIRY SCIENCE*, 101(12):11097–11107, 2018.

[22] Ruth McQuillan, Anne-Louise Leutenegger, Rehab Abdel-Rahman, Christopher S. Franklin, Marijana Pericic, Lovorka Barac-Lauc, Nina Smolej-Narancic, Branka Janicijevic, Ozren Polasek, Albert Tenesa, Andrew K. MacLeod, Susan M. Farrington, Pavao Rudan, Caroline Hayward, Veronique Vitart, Igor Rudan, Sarah H. Wild, Malcolm G. Dunlop, Alan F. Wright, Harry Campbell, and James F. Wilson. Runs of homozygosity in European populations. *AMERICAN JOURNAL OF HUMAN GENETICS*, 83(3):359–372, 2008.

[23] R. Meyermans, W. Gorssen, N. Buys, and S. Janssens. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, 21(1):94, 2020.

[24] Vagheesh Narasimhan, Petr Danecek, Aylwyn Scally, Yali Xue, Chris Tyler-Smith, and Richard Durbin. BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, 32(11):1749–1751, 2016.

[25] Pirmin Nietlisbach, Stefanie Muff, Jane M. Reid, Michael C. Whitlock, and Lukas F. Keller. Nonequivalent lethal equivalents: Models and inbreeding metrics for unbiased estimation of inbreeding load. *Evolutionary Applications*, 0(0), 2018.

[26] Jennie E. Pryce, Mekonnen Haile-Mariam, Michael E. Goddard, and Ben J. Hayes. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *GENETICS SELECTION EVOLUTION*, 46, 2014.

[27] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, and Pak C. Sham. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3):559–575, 2007.

[28] The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526(7571):68–74, 2015.

[29] Paul M. Vanraden. Efficient methods to compute genomic predictions. *Journal of dairy science*, 91(11):4414–4423, 2008.

[30] Bruce S. Weir and Jérôme Goudet. A Unified Characterization of Population Structure and Relatedness. *Genetics*, 206(4):2085–2103, 2017.

[31] Sewall Wright. Coefficients of Inbreeding and Relationship. *The American Naturalist*, 56(645):330–338, 1922.

[32] Jian Yang, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics*, 88(1):76–82, 2011.

[33] Loic Yengo, Zhihong Zhu, Naomi R. Wray, Bruce S. Weir, Jian Yang, Matthew R. Robinson, and Peter M. Visscher. Detection and quantification of inbreeding depression for complex traits from SNP data. *Proceedings of the National Academy of Sciences*, 114(32):8602–8607, 2017.

[34] Chunzhi Zhang, Pei Wang, Die Tang, Zhongmin Yang, Fei Lu, Jianjian Qi, Nilesh R. Tawari, Yi Shang, Canhui Li, and Sanwen Huang. The genetic basis of inbreeding depression in potato. *Nature Genetics*, 51(3):374–378, March 2019. Number: 3 Publisher: Nature Publishing Group.

[35] Qian S. Zhang, Jérôme Goudet, and Bruce S. Weir. Rank-invariant estimation of inbreeding coefficients. *Heredity*, 128(1):1–10, 2022.

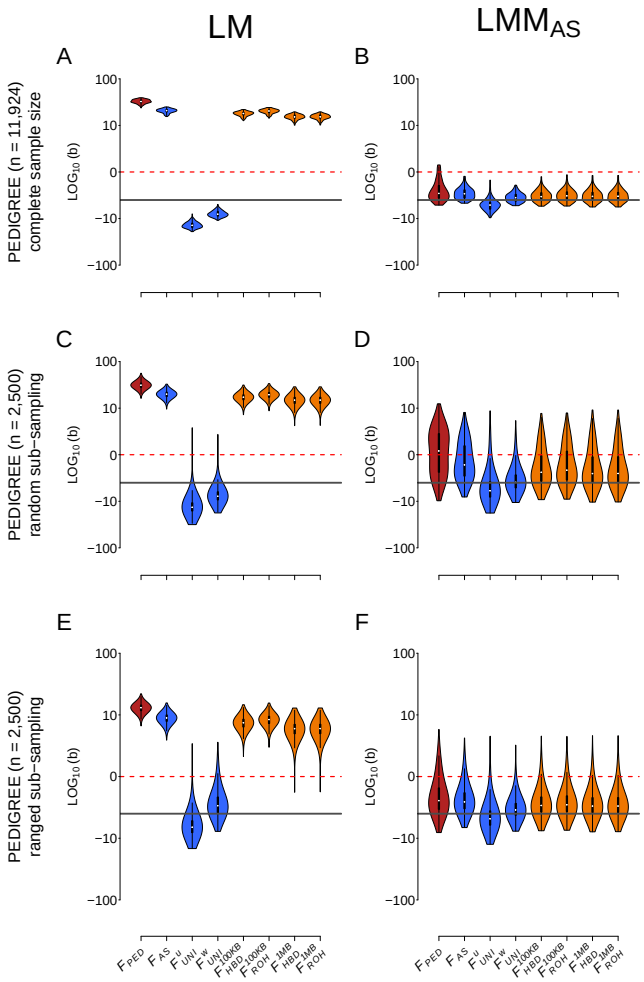


Figure 1: Comparison of the estimation of inbreeding depression strength (b) among different F estimates and two models in the PEDIGREE population. Each column represents a regression model. The first column depicts the simple linear regression (LM) (panels **A**, **C** and **E**) and the second column depicts the linear mixed model with the allele sharing relatedness matrix as a random component (LMM_{AS}) (panels **B**, **D** and **F**). The first row represents the complete simulated population (11,924 individuals, panels **A** and **B**). The second row shows the random subsampling (2,500 individuals, panels **C** and **D**). The third row shows the ranged subsampling (2,500 individuals, panels **E** and **F**). Inbreeding estimates presented in this graph are F_{PED} , F_{AS} , F^{u}_{UNI} , F^{w}_{UNI} , $F_{\text{HBD}_{100\text{KB}}}$, $F_{\text{ROH}_{100\text{KB}}}$, $F_{\text{HBD}_{1\text{MB}}}$ and finally $F_{\text{ROH}_{1\text{MB}}}$. For panels **A** and **B**, violin plots show the distribution of the inbreeding depression strength estimates (b) among the simulated 100 replicates. For panels **C** to **F**, violin plots represent the distribution of the inbreeding depression strength estimates (b) for the 10,000 simulated and sub-sampling replicates (100 sub-sampling replicates for each of the 100 simulation replicates). The solid dark grey line is the true strength of ID (b = -3). The dashed red line represents the absence of ID (b = 0), meaning that we failed to detect ID in any replicate above this line. Note that all panels are in $log10$ scale.

Model	Population	F_{PED}	F_{AS}	F^{u}_{UNI}	F^{w}_{UNI}	$F_{\text{HBD}_{\text{100KB}}}$	$F_{\text{ROH}_{\text{100KB}}}$	$F_{\text{HBD}_{\text{1MB}}}$	$F_{\text{ROH}_{\text{1MB}}}$
LM	PEDIGREE (complete)	34.82	22.71	10.17	4.17	19.93	22.22	17.4	17.44
LMM: AS	PEDIGREE (complete)	1.62	1.27	1.89	0.87	1.07	1.12	1.11	1.11
LMM: GCTA WE	PEDIGREE (complete)	1.62	1.27	1.89	0.87	1.07	1.12	1.11	1.11
LMM: GCTA UN	PEDIGREE (complete)	1.58	1.28	1.85	0.88	1.08	1.12	1.08	1.08
LM	PEDIGREE (random sub)	33.84	22.20	10.41	4.47	19.53	21.72	17.24	17.28
LMM: AS	PEDIGREE (random sub)	4.01	2.97	3.82	1.83	2.57	2.73	2.56	2.57
LMM: GCTA WE	PEDIGREE (random sub)	4.01	2.97	3.82	1.83	2.57	2.73	2.56	2.57
LMM: GCTA UN	PEDIGREE (random sub)	$> 1{,}000$	2.75	3.44	1.78	$> 1{,}000$	$> 1{,}000$	$> 1{,}000$	$> 1{,}000$
LM	PEDIGREE (ranged sub)	15.22	11.04	3.46	1.61	9.58	10.52	8.13	8.15
LMM: AS	PEDIGREE (ranged sub)	2.09	1.82	2.13	1.26	1.61	1.67	1.58	1.58
LMM: GCTA WE	PEDIGREE (ranged sub)	2.09	1.82	2.13	1.26	1.61	1.67	1.58	1.58
LMM: GCTA UN	PEDIGREE (ranged sub)	$> 1{,}000$	1.69	2.05	1.24	$> 1{,}000$	$> 1{,}000$	1.53	1.54

Table 1: **RMSE on b estimate in the PEDIGREE population** These values are for the complete ADD & DOM & DEMA scenario. See tables S2-S5 for the other scenarios

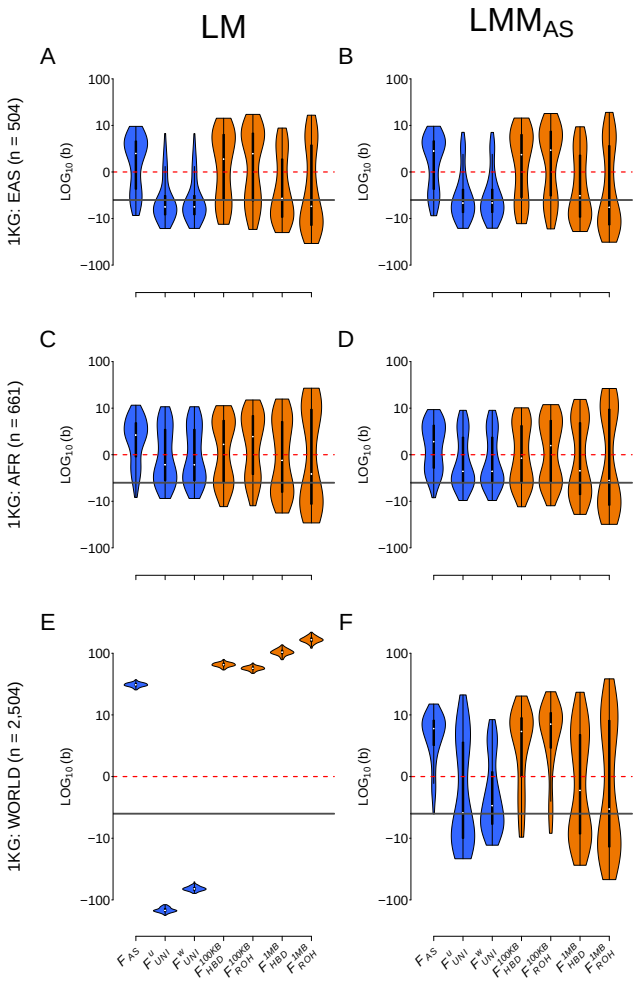


Figure 2: Comparison of the estimation of inbreeding depression strength (b) among different F estimates and models in four different populations. Each column represents a regression model. The first column depicts the simple linear regression (LM) (panels **A**, **C** and **E**) and the second column depicts the linear mixed model with the allele sharing relatedness matrix as a random component (LMM_{AS}) (panels **B**, **D** and **F**). The three rows correspond to the three populations from the 1,000 Genomes project: EAS on panels **A** and **B**, AFR on panels **C** and **D** and WORLD on panels **E** and **F**. Inbreeding estimates presented are F_{AS} , F^{u}_{UNI} , F^{w}_{UNI} , $F_{\text{HBD}_{100\text{KB}}}$, $F_{\text{ROH}_{100\text{KB}}}$, $F_{\text{HBD}_{1\text{MB}}}$ and finally $F_{\text{ROH}_{1\text{MB}}}$. Violin plots represent the distribution of the inbreeding depression strength estimates (b) among the 100 simulation replicates. The solid dark grey line is the true strength of ID (b = -3). The dashed red line represents the absence of ID ($b = 0$), meaning that we failed to detect ID in any replicate above this line. Note that all panels are in log_{10} scale.

Model	Population	F_{AS}	F^{u}_{UNI}	F^{w}_{UNI}	$F_{\text{HBD}_{100\text{KB}}}$	$F_{\text{ROH}_{100\text{KB}}}$	$F_{\text{HBD}_{1\text{MB}}}$	$F_{\text{ROH}_{1\text{MB}}}$
LM	EAS	5.55	4.9	4.86	7.14	7.93	6.19	10.58
LMM: AS	EAS	5.67	4.68	4.64	7.41	8.22	6.12	10.39
LMM: GCTA WE	EAS	5.67	4.68	4.64	7.28	8.06	6.11	10.39
LMM: GCTA UN	EAS	5.48	4.74	4.71	7.1	7.87	6.18	10.57
LM	AFR	5.93	4.81	4.81	6.03	7.21	7.21	13.12
LMM: AS	AFR	5.15	4.07	4.07	5.46	6.2	7.15	13.1
LMM: GCTA WE	AFR	5.15	4.07	4.07	$> 1{,}000$	$> 1{,}000$	7.16	13.1
LMM: GCTA UN	AFR	5.78	4.42	4.42	5.92	6.93	7.2	13.11
LM	WORLD	32.91	142.95	62.21	67.42	59.15	107.67	169.73
LMM: AS	WORLD	8.63	8.34	4.17	9.15	10.97	8.78	14.6
LMM: GCTA WE	WORLD	9.84	$> 1{,}000$	$> 1{,}000$	11.19	13.92	$> 1{,}000$	$> 1{,}000$
LMM: GCTA UN	WORLD	18.18	$> 1{,}000$	$> 1{,}000$	27.52	26.91	$> 1{,}000$	$> 1{,}000$

Table 2: **RMSE on b estimate in the three 1,000 Genomes Project populations: EAS, AFR and WORLD** These values are for the complete ADD & DOM & DEMA scenario. See tables S2-S5 for other scenarios

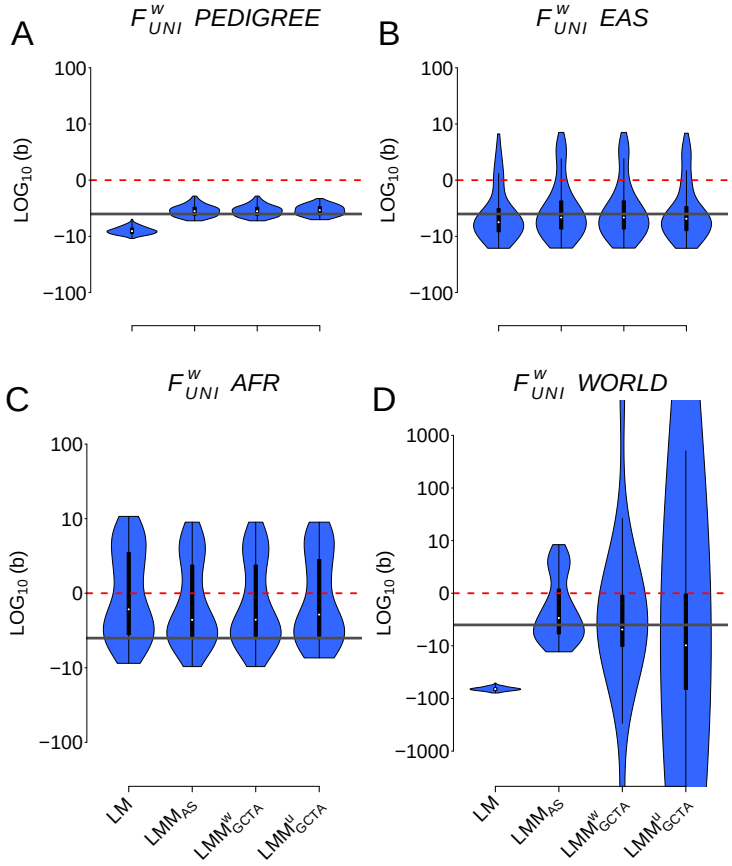


Figure 3: Comparison of the inbreeding depression strength estimates (b) with F^{w}_{UNI} in the four populations with four different models. The four models are: i) the simple linear regression (LM), ii) the linear mixed model with the allele sharing relatedness matrix as a random factor, iii) the linear mixed model with the weighted `GCTA` relatedness matrix as a random factor and iv) the linear mixed model with the unweighted `GCTA` relatedness matrix as a random factor. Panel **A** the simulated PEDIGREE population, panel **B** depicts the EAS population, panel **C** the AFR population and finally panel **D** the WORLD population. Note that all panels are in log_{10} scale. Also note that LMM did not converge for some replicates (yielding estimated b values above 1000 or below -1000, not shown in the graph). Percentages of replicates which did not converge: panel **D** (WORLD): 21% for $GRM_{\mathrm{GCTA^{w}}}$; 20% for $GRM_{\mathrm{GCTA^{u}}}$.

LM

LMM$_{AS}$

A

B

C

D

E

F

PEDIGREE (n = 11,924) complete sample size

PEDIGREE (n = 2,500) random sub−sampling

PEDIGREE (n = 2,500) ranged sub−sampling

LOG$_{10}$ (b)

100

10

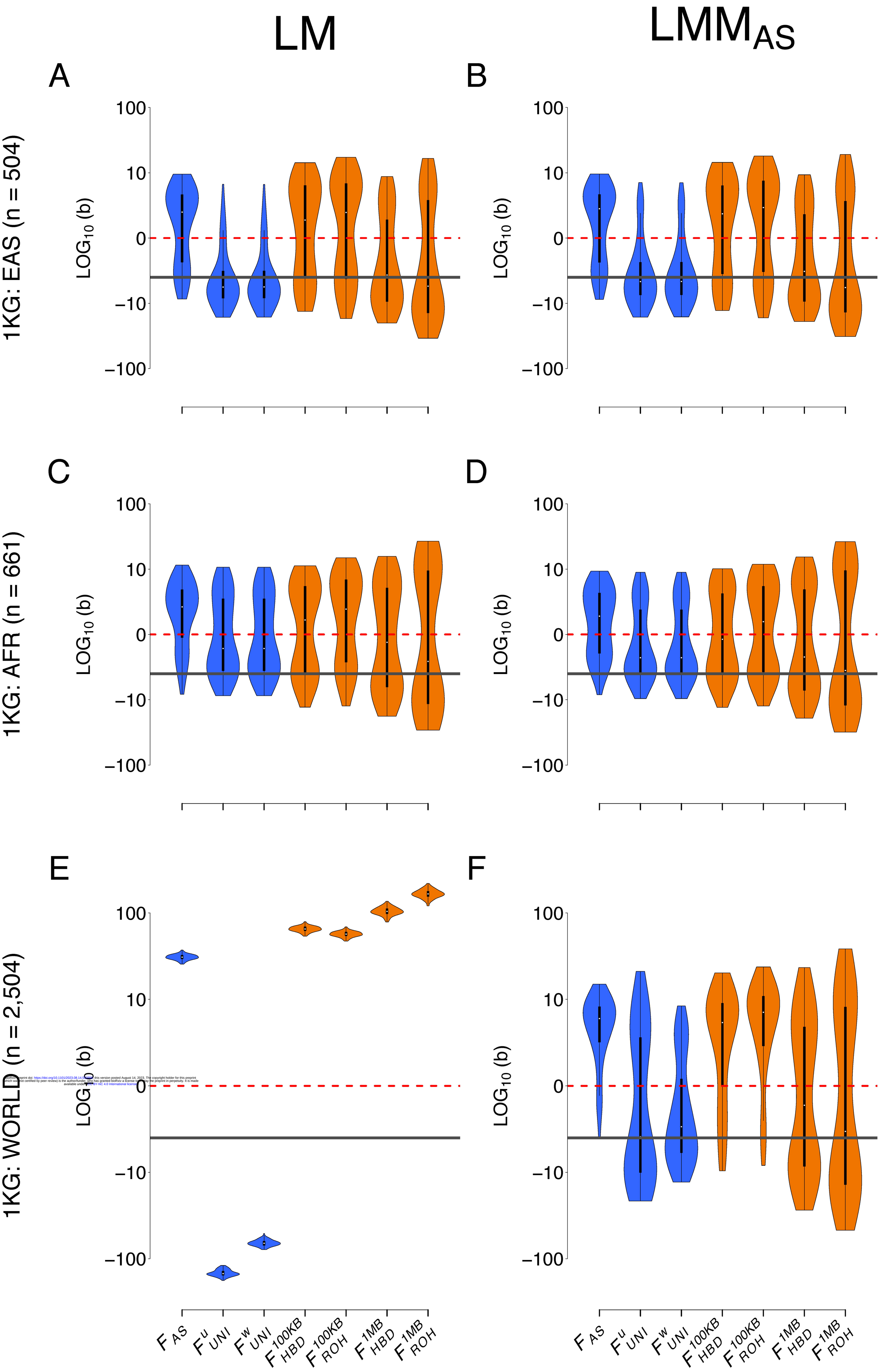
0

−10

−100

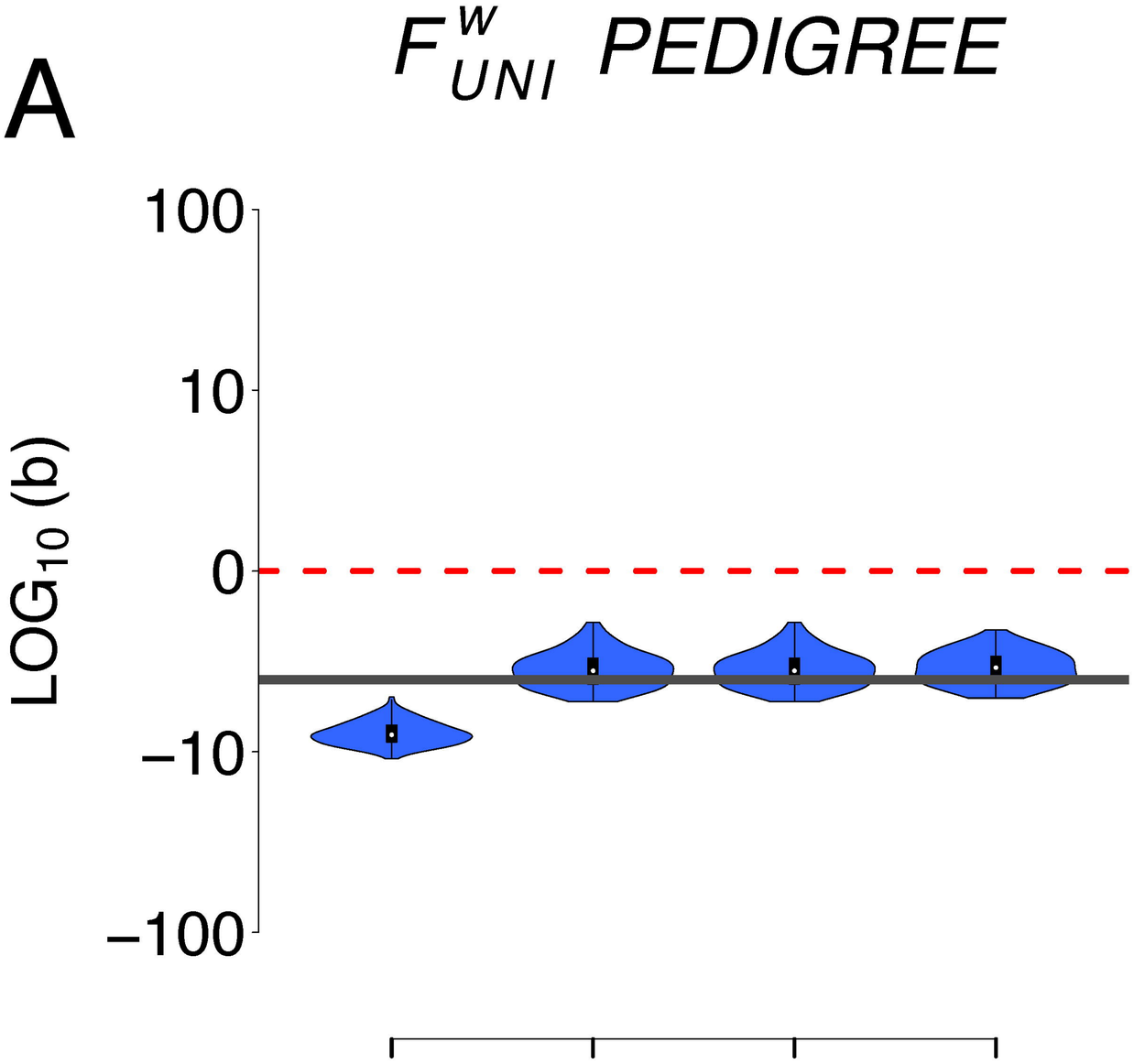
F_{PED} F_{AS} F^{u}_{UNI} F^{w}_{UNI} F_{HBD}^{100KB} F_{ROH}^{100KB} F_{HBD}^{1MB} F_{ROH}^{1MB}

Model	Population	F_{PED}	F_{AS}	F^{u}_{UNI}	F^{w}_{UNI}	$F_{\text{HBD}_{100\text{KB}}}$	$F_{\text{ROH}_{100\text{KB}}}$	$F_{\text{HBD}_{1\text{MB}}}$	$F_{\text{ROH}_{1\text{MB}}}$
LM	PEDIGREE (complete)	34.82	22.71	10.17	4.17	19.93	22.22	17.4	17.44
LMM: AS	PEDIGREE (complete)	1.62	1.27	1.89	0.87	1.07	1.12	1.11	1.11
LMM: GCTA WE	PEDIGREE (complete)	1.62	1.27	1.89	0.87	1.07	1.12	1.11	1.11
LMM: GCTA UN	PEDIGREE (complete)	1.58	1.28	1.85	0.88	1.08	1.12	1.08	1.08
LM	PEDIGREE (random sub)	33.84	22.20	10.41	4.47	19.53	21.72	17.24	17.28
LMM: AS	PEDIGREE (random sub)	4.01	2.97	3.82	1.83	2.57	2.73	2.56	2.57
LMM: GCTA WE	PEDIGREE (random sub)	4.01	2.97	3.82	1.83	2.57	2.73	2.56	2.57
LMM: GCTA UN	PEDIGREE (random sub)	$> 1{,}000$	2.75	3.44	1.78	$> 1{,}000$	$> 1{,}000$	$> 1{,}000$	$> 1{,}000$
LM	PEDIGREE (ranged sub)	15.22	11.04	3.46	1.61	9.58	10.52	8.13	8.15
LMM: AS	PEDIGREE (ranged sub)	2.09	1.82	2.13	1.26	1.61	1.67	1.58	1.58
LMM: GCTA WE	PEDIGREE (ranged sub)	2.09	1.82	2.13	1.26	1.61	1.67	1.58	1.58
LMM: GCTA UN	PEDIGREE (ranged sub)	$> 1{,}000$	1.69	2.05	1.24	$> 1{,}000$	$> 1{,}000$	1.53	1.54

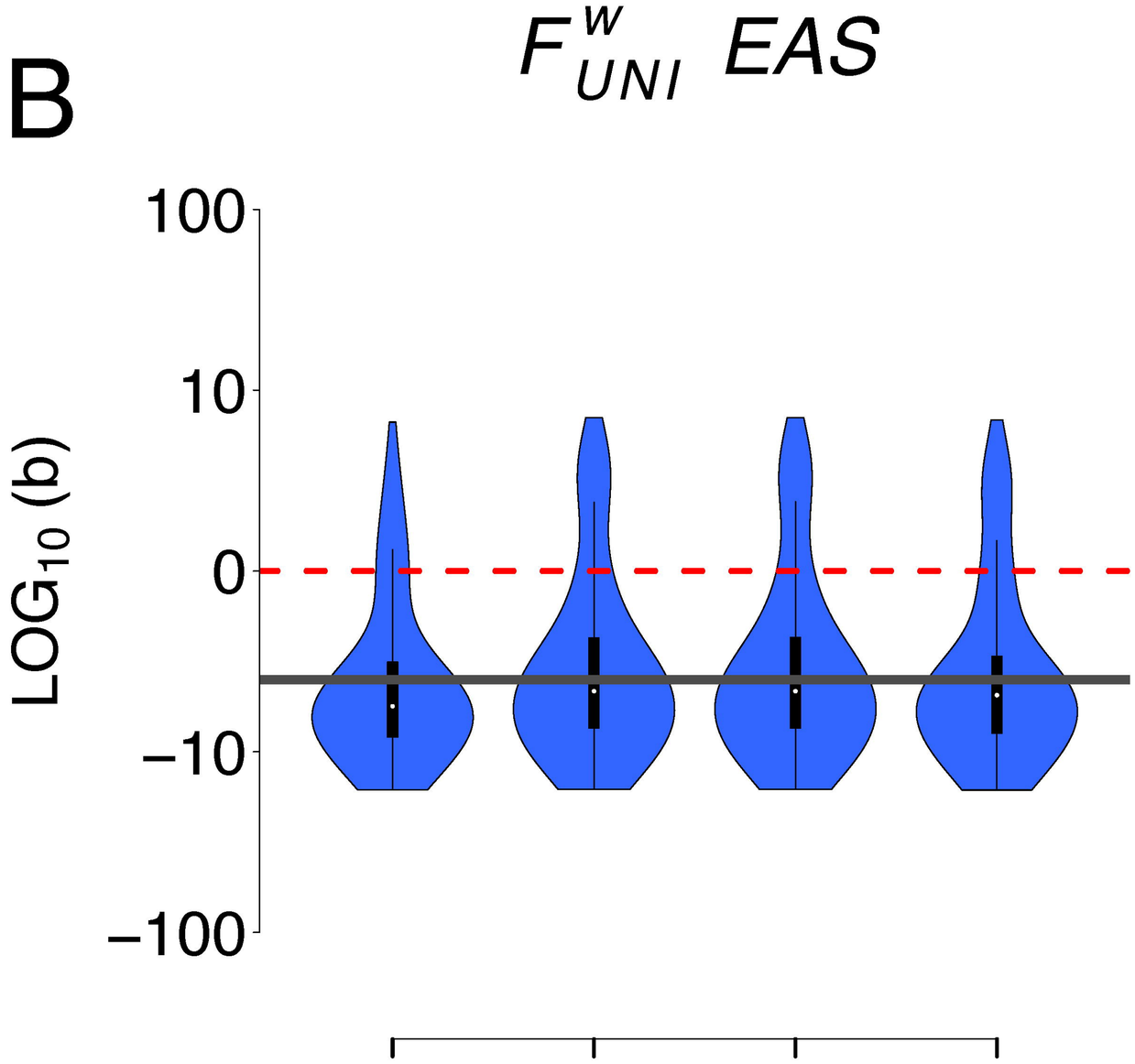


Model	Population	F_{AS}	F^{u}_{UNI}	F^{w}_{UNI}	$F_{HBD_{100KB}}$	$F_{ROH_{100KB}}$	$F_{HBD_{1MB}}$	$F_{ROH_{1MB}}$
LM	EAS	5.55	4.9	4.86	7.14	7.93	6.19	10.58
LMM: AS	EAS	5.67	4.68	4.64	7.41	8.22	6.12	10.39
LMM: GCTA WE	EAS	5.67	4.68	4.64	7.28	8.06	6.11	10.39
LMM: GCTA UN	EAS	5.48	4.74	4.71	7.1	7.87	6.18	10.57
LM	AFR	5.93	4.81	4.81	6.03	7.21	7.21	13.12
LMM: AS	AFR	5.15	4.07	4.07	5.46	6.2	7.15	13.1
LMM: GCTA WE	AFR	5.15	4.07	4.07	$> 1{,}000$	$> 1{,}000$	7.16	13.1
LMM: GCTA UN	AFR	5.78	4.42	4.42	5.92	6.93	7.2	13.11
LM	WORLD	32.91	142.95	62.21	67.42	59.15	107.67	169.73
LMM: AS	WORLD	8.63	8.34	4.17	9.15	10.97	8.78	14.6
LMM: GCTA WE	WORLD	9.84	$> 1{,}000$	$> 1{,}000$	11.19	13.92	$> 1{,}000$	$> 1{,}000$
LMM: GCTA UN	WORLD	18.18	$> 1{,}000$	$> 1{,}000$	27.52	26.91	$> 1{,}000$	$> 1{,}000$

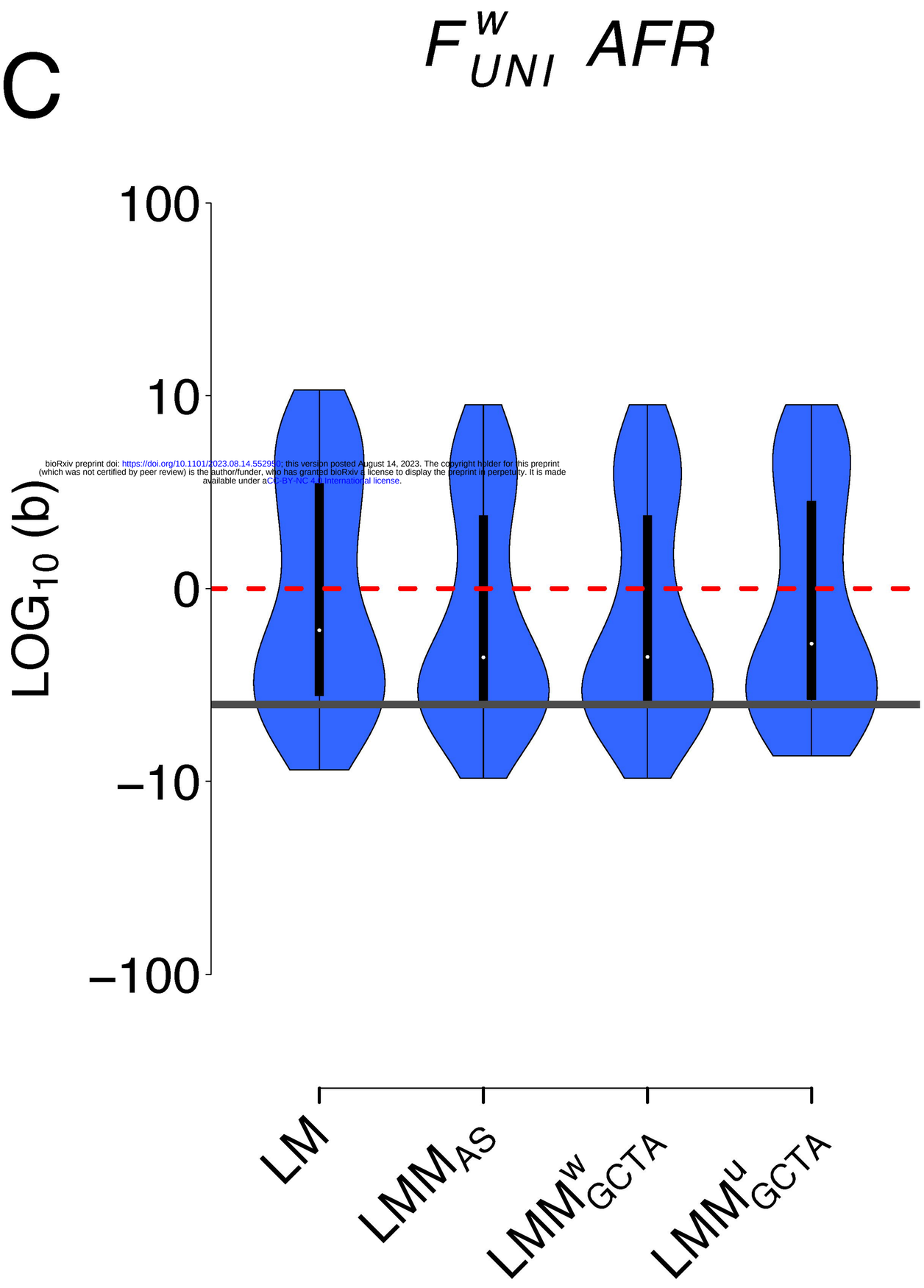
A



B



C



D