Estimating narrow-sense heritability from genome-wide data in admixed populations

Georgios Athanasiadis,[1,]* Doug Speed,[2,3] Mette K. Andersen,[4] Emil V. R. Appel,[4] Niels Grarup,[4] Ivan Brandslund,[5] Marit Eika Jørgensen,[6,7,8] Christina Viskum Lytken Larsen,[8] Peter Bjerregaard,[8] Torben Hansen,[4] and Anders Albrechtsen[1,]*

[1]Department of Biology, Section for Computational and RNA Biology, University of Copenhagen, Copenhagen, Denmark

[2]Aarhus Institute of Advanced Studies, Aarhus University, Aarhus, Denmark

[3]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

[4]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[5]Department of Clinical Immunology and Biochemistry, Lillebaelt Hospital, Vejle, Denmark

[6]Steno Diabetes Center Copenhagen, Gentofte, Denmark

[7]Greenland University, Ilisimatusarfik, Nuuk, Greenland

[8]National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark

*Correspondence: yorgos.athanasiadis@gmail.com (G.A.), albrecht@binf.ku.dk (A.A.)

Abstract

Finding an efficient framework for estimating total narrow-sense heritability in admixed populations remains an open question. In this work, we used extensive simulations to evaluate existing linear mixed model frameworks in estimating total narrow-sense heritability in two population-based cohorts from Greenland and compared the results to data from unadmixed individuals from Denmark. When our analysis focused on Greenlandic sib pairs, the model with two relationship matrices, one capturing identity by descent and one capturing identity by state, returned heritability estimates close to the true simulated value, while using each of the two matrices alone led to downward biases. When phenotypes correlated with ancestry, heritability estimates were inflated. Based on these observations, we propose a post-estimation PCA-based adjustment that recovers successfully the true simulated heritability. We use this knowledge to estimate the heritability of ten quantitative traits from the two Greenlandic cohorts and report differences such as lower heritability for height in Greenlanders compared to Europeans. In conclusion, narrow-sense heritability in admixed populations is best estimated using a mixture of genetic relationship matrices on individuals with at least one first-degree relative included in the sample.

Introduction

Heritability is the fraction of phenotypic variance attributed to genetics. More specifically, assuming that the variance σ_P^2 of a phenotype equals the sum of its genetic σ_G^2 and environmental variance σ_e^2, heritability in its broad sense (H^2) is expressed by the ratio σ_G^2/σ_P^2.[1,2] The genetic variance σ_G^2 can be further broken down into its additive (σ_A^2), dominant (σ_D^2), and epistatic (σ_I^2) components.[3,4] Most of the existing methods focus on the fraction of phenotypic variance σ_P^2 owing to additive effects alone (σ_A^2/σ_P^2) – the so-called narrow-sense heritability (h^2).

Heritability is interesting in its own right, but it is also pivotal in quantitative genetic studies with many practical uses. Because, by definition, heritability measures the contribution of genetics to a phenotype, it allows us to gain insights into the genetic architecture of a trait. Moreover, knowledge about the heritability of a trait helps us evaluate the effectiveness of a genome-wide association study (GWAS), as the so-called single nucleotide polymorphism (SNP) heritability h_g^2 of a trait informs us about the maximum discovery potential of a given genotyping platform. Similarly, heritability estimates provide an upper bound to the accuracy of polygenic predictions with predictors having higher performance potential in traits with higher heritability.

There exist many ways to estimate narrow-sense heritability h^2 and they usually boil down to estimating the genetic variance owing to additive effects (σ_A^2). Assuming an additive model, a classical approach is to use the phenotypic correlation between related individuals,[1,2,5] which for a pair j and k is

$$\text{cor}(y_j, y_k) = \sigma_A^2 \mathbf{K_{causal}}_{[j,k]}.$$

$\mathbf{K_{causal}}$ is an idealized genetic relationship matrix (GRM) that reflects genetic relationships between individuals at an unknown set of causal variants.

Because the set of causal variants is unknown, $\mathbf{K_{causal}}$ has been approximated in the classical literature by the expected relatedness in the pedigree matrix $\mathbf{K_{PED}}$, which equals twice the kinship matrix $\mathbf{\Phi}$

$$\mathbf{K_{causal}} \approx \mathbf{K_{PED}} = E(2\mathbf{\Phi}).$$

The entries in the kinship matrix $\mathbf{\Phi}$ are known as kinship coefficients. Kinship coefficient φ is the probability that a random allele from subject j is identical by descent (IBD) to an allele at the same locus from subject k. As an example, for a pair of full siblings j and k, this expected probability equals ¼ and therefore $\mathbf{K_{PED[j,k]}} = \frac{1}{2}$. For a design that focuses exclusively on sib pairs and assuming no dominance contribution, an estimate of the additive genetic variance $\widehat{\sigma}_A^2$ is therefore two times the phenotypic correlation r_P of sib pairs

$$\widehat{\sigma}_A^2 = 2r_{P[\text{sibs}]}.$$

When extended pedigrees are available, the entire $\mathbf{K_{PED}}$ can be leveraged in a linear mixed model (LMM) framework. In this case, the phenotype vector $\mathbf{y}$ is modelled as

$$\mathbf{y} \sim N(\mathbf{\mu}, \sigma_A^2 \mathbf{K_{PED}} + \sigma_e^2 \mathbf{I})$$

and the genetic variance $\widehat{\sigma}_A^2$ is estimated with restricted maximum likelihood (REML).[6–9]

When genetic data are available, the total IBD fraction of the genome $\mathbf{K_{IBD[j,k]}}$, also referred to as $\widehat{\pi}$ ($\widehat{\pi} = 2\varphi$), can be estimated for j and k and used in the above LMM framework instead of its expected value from $\mathbf{K_{PED[j,k]}}$. In other words, instead of being approximated with $\mathbf{K_{PED}}$, $\mathbf{K_{causal}}$ is approximated with $\mathbf{K_{IBD}}$.[10,11]

The advent of SNP chips resulted in the use of thousands of markers in the computation of the GRM, thus allowing to estimate heritability in samples without pedigree information.[5,12] In this

case, assuming an N × M genotype matrix $\overline{\mathbf{G}}$ with zero-column mean and unit-column variance, $\mathbf{K_{causal}}$ is approximated with an identity-by-state (IBS) genetic covariance matrix $\mathbf{K_{IBS}}$

$$\mathbf{K_{IBS}} = \frac{1}{M}\overline{\mathbf{G}}\,\overline{\mathbf{G}}^{\mathbf{T}}.$$

Given that the set of M typed SNPs typically does not include all causal variants and/or it includes tag SNPs that are in imperfect linkage disequilibrium (LD) with said variants, the use of $\mathbf{K_{IBS}}$ can lead to an underestimation of the total heritability. For unrelated individuals, this estimate reflects only the proportion of phenotypic variance captured directly or indirectly by the typed SNPs – i.e. the so-called SNP heritability h_g^2 (note that $h_g^2 \leq h^2$). The GCTA software[13] uses $\mathbf{K_{IBS}}$ on unrelated individuals in order to estimate h_g^2. It has been shown that SNP heritability h_g^2 can vary with minor allele frequency (MAF), LD and genotype certainty.[14] The LDAK software[15] can be used in order to accommodate these parameters in the computation of $\mathbf{K_{IBS}}$.

More recently, it has been shown that $\mathbf{K_{IBD}}$ can be effectively substituted in the LMM by a truncated IBS genetic covariance matrix $\mathbf{K_{IBS>t}}$, in which all $\mathbf{K_{IBS}}$ entries below a threshold t are set to zero.[16] Moreover, the same authors introduced a method for the simultaneous estimation of SNP h_g^2 and total heritability h^2 by jointly fitting an LMM with $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ (or $\mathbf{K_{IBS>t}}$ instead of $\mathbf{K_{IBD}}$). This way, the authors provided heritability estimates with narrower confidence intervals and showed that total heritability estimates under this approach were very similar to those under $\mathbf{K_{IBD}}$ (or $\mathbf{K_{IBS>t}}$) alone.

In order for existing methods to produce meaningful heritability estimates, no population structure should be present in the studied samples.[4] Population structure can arise when individuals of different ancestry are found in the same sample and/or when individuals are admixed. Individuals from different populations tend to have different minor allele frequencies as well as different

5

environmental exposures.[4] Because population structure correlates with environmental structure, it can inflate heritability estimates. It has been shown that, for SNP heritability (h_g^2) estimates, inclusion of principal components (PCs) as fixed effects cannot fully account for the structure bias.[17] Moreover, the differences in ancestral allele frequencies affect the computation of $\mathbf{K_{IBS}}$, which ceases to be proportional to $\mathbf{K_{IBD}}$ – even for close relationships. To date, there is not clearly a best strategy for estimating and interpreting heritability in structured/admixed populations.

Nevertheless, an LMM that uses a relationship matrix $\mathbf{K_\gamma}$ based on local ancestry instead of genotypes has been proposed.[18] The authors of this method produced accurate estimates of total heritability h^2 for several phenotypes in admixed African American samples by fitting an LMM with $\mathbf{K_\gamma}$ and rescaling its regression coefficient accordingly. However, this method has limitations, as it relies on knowledge about local ancestry and assumes that samples are unrelated.

In light of the above, finding an efficient framework for total narrow-sense heritability estimates in admixed populations with high levels of relatedness remains an open question. Many experimental designs could benefit from a better understanding of how heritability estimates are affected by the simultaneous presence of population and family structure. In this work, we use extensive simulations to evaluate the performance of existing classical and LMM frameworks in estimating SNP (h_g^2) and total (h^2) heritability in two population-based cohorts from Greenland. The Greenlanders are an isolate with many unique characteristics, such as small population size, high degree of relatedness and extensive population structure with ancestry from both the Inuit and Europeans.[19,20] Our goal is to find a way to estimate and understand narrow-sense heritability in such populations, thus gaining valuable insights into the genetic architecture of complex traits in such populations.

Subjects and Methods

Samples

Greenlanders

The Greenlandic subjects (N = 4,659) came from two general health surveys. The first survey[21] consisted of Greenlanders living in Denmark (the BHH cohort; N = 546), recruited during 1998-1999, as well as Greenlanders living in Greenland (the B99 cohort; N = 1,328), recruited during 1999-2001 as part of a general population health survey. The second survey consisted of Greenlanders living in Greenland[22] (the IHIT cohort; N = 2,785), recruited during 2005-2010 as part of a population health survey.

Danes

The population-based Danish sample (N = 5,470) was obtained from Inter99,[23] a randomized intervention study collected at the Research Centre for Prevention and Health. In addition, 513 (N = 1,169) Danish sib pairs were identified across several available Danish cohorts, namely the population-based cohorts (i) Inter99[24] (N = 294); (ii) Helbred2006[25] (N = 121); (iii) Helbred2008[26] (N = 38); and (iv) Helbred2010[27] (N = 57), all recruited from the Research Centre for Prevention and Health, Glostrup Hospital, Denmark, as well as cohorts collected for the study of type 2 diabetes, namely (v) Vejle Diabetes Biobank[28] (N = 570), recruited at the Vejle Hospital, Denmark; (v) the ADDITION study[29] (N = 8), recruited at the Department of General Practice at the University of Aarhus, Denmark; and (vi) SDC[30] (N = 81) recruited at the outpatient clinic at Steno Diabetes Center, Denmark.

Ethics statement

The two Greenlandic surveys received ethics approval from the Commission for Scientific Research in Greenland (project 2011-13, ref.no. 2011-056978; project 2013-13, ref.no. 2013-090702), and the Danish studies were approved by the local ethics committees (protocol ref.no. H-3-2012-155). All studies were conducted in compliance with the Helsinki Declaration II and all participants gave their written consent after being informed about the study orally and in writing.

Genotyping and quality control

Both the Greenlandic and the unrelated Danish samples were typed on Illumina's Cardio-MetaboChip (Illumina, San Diego, CA, USA). The Cardio-MetaboChip includes 196,725 SNPs selected from genetic studies of cardiovascular, metabolic and anthropometric traits.[31] Moreover, the unrelated Danish samples and the Danish sib pairs were typed on Illumina's Infinium OmniExpress chip, which includes ~710,000 markers. Standard quality control was carried out with PLINK v1.9[32] and included filtering for *per-individual* (`--mind 0.01`) and *per-marker* (`--geno 0.01`) genotype missingness = 1%. The datasets passing quality control consisted of (i) 4,659 Greenlanders typed on 187,181 Cardio-MetaboChip autosomal SNPs; (ii) 5,470 unrelated Danes typed on 186,639 Cardio-MetaboChip and 618,037 OmniExpress autosomal SNPs; and (iii) 1,169 Danes forming sib pairs typed on 609,605 OmniExpress autosomal SNPs.

Phenotype simulations

We simulated 1,000 quantitative phenotypes with true total narrow-sense heritability $h^2 = \{ 0.4, 0.6, 0.8 \}$ using real genetic data from 4,659 Greenlanders and 5,470 + 1,169 Danes. We used either GCTA or our own scripts. Simulations were carried out as follows:

First, we defined an $N \times C$ causal genotype matrix $\mathbf{G_{causal}}$ by sampling C = 1,500 SNPs from a list of all available SNPs.

Second, we sampled SNP effects, represented by a $C \times 1$ effect vector **b**, from a standard normal distribution $N(0,1)$. In order to model the relationship between the effect size of SNP i and its allele frequency f_i, a genotype matrix **G** can be standardized according to the formula

$$\overline{\mathbf{G}}_{j,i} = (\mathbf{G}_{j,i} - 2f_i)[2f_i(1 - f_i)]^{\frac{\alpha}{2}}.[14]$$

GCTA assumes a specific inverse relationship between SNP effect size and allele frequency by converting the original matrix $\mathbf{G}_{\mathbf{causal}}$ into a zero column-mean and unit column-variance standard score matrix $\overline{\mathbf{G}}_{\mathbf{causal}}$. This can be seen as a special case of the above standardizing formula for $\alpha = -1$.

Third, we computed a vector of polygenic scores **S** by multiplying $\overline{\mathbf{G}}_{\mathbf{causal}}$ by the corresponding effect vector **b**

$$\mathbf{S} = \overline{\mathbf{G}}_{\mathbf{causal}}\mathbf{b}$$

and its additive genetic variance as

$$\sigma_A^2 = \mathrm{var}(\mathbf{S}).$$

Finally, we computed the phenotype vector **P** by adding an environmental vector ε of i.i.d. error terms to the polygenic score vector **S**

$$\mathbf{P} = \mathbf{S} + \varepsilon.$$

Error terms were sampled from the distribution $N\left(0, \mathrm{var}(\mathbf{S})\left(\frac{1}{h^2} - 1\right)\right).[13]$

In some simulations involving the Greenlanders, we also modelled the interaction between environment and ancestry by adding an interaction vector **E×Anc** to the sum

$$\mathbf{P} = \mathbf{S} + \mathbf{E \times Anc} + \varepsilon.$$

9

In particular, if $h^2_{E \times Anc}$ is the proportion of phenotypic variance explained by the interaction and $\mathbf{Q_{Inuit}}$ is the vector of proportion of Inuit ancestry,

$$\mathbf{E \times Anc} = \mathbf{Q_{Inuit}} \sqrt{\left(\frac{h^2_{E \times Anc}}{h^2}\right) \frac{var(\mathbf{S})}{var(\mathbf{Q_{Inuit}})}}.$$

As a consequence, the noise terms in ε are now sampled from $N\left(0, var(\mathbf{S})\left(\frac{1}{h^2} - 1\right) - var(\mathbf{E \times Anc})\right)$.

Linear mixed model

In an LMM, the phenotype $\mathbf{y}$ is modelled as a mixture of fixed and random effects (i.e. the effects of the causal variants)

$$\mathbf{y} = \boldsymbol{\mu} + \overline{\mathbf{G}}_{\mathbf{causal}}\mathbf{b} + \varepsilon.$$

Assuming that $\mathbf{b} \sim N(0, \frac{\sigma^2_A}{C})$ and $\varepsilon \sim N(0, \sigma^2_e \mathbf{I})$ under the GCTA model with $\alpha = -1$, $\mathbf{y}$ follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance

$$
\begin{aligned}
var(\mathbf{y}) \quad &= var(\boldsymbol{\mu} + \overline{\mathbf{G}}_{\mathbf{causal}}\mathbf{b} + \varepsilon) \\
&= var(\overline{\mathbf{G}}_{\mathbf{causal}}\mathbf{b}) + var(\varepsilon) \\
&= \overline{\mathbf{G}}_{\mathbf{causal}} var(\mathbf{b}) \overline{\mathbf{G}}^{\mathbf{T}}_{\mathbf{causal}} + var(\varepsilon) \\
&= \overline{\mathbf{G}}_{\mathbf{causal}} \frac{\sigma^2_A}{C} \overline{\mathbf{G}}^{\mathbf{T}}_{\mathbf{causal}} + \sigma^2_e \mathbf{I} \\
&= \sigma^2_A \mathbf{K_{causal}} + \sigma^2_e \mathbf{I}
\end{aligned}
$$

such that

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2_A \mathbf{K_{causal}} + \sigma^2_e \mathbf{I}).$$

Total narrow-sense heritability is then defined as

10

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2}.$$

Because **K**_{causal} is unknown, we used other GRMs instead, such as $\mathbf{K_{IBD}}$, $\mathbf{K_{IBS}}$, $\mathbf{K_{IBS>t}}$ etc.

Genetic relationship matrices

We computed the $\mathbf{K_{IBD}}$ matrix for the entire Greenlandic sample from pairwise kinship coefficients ($\widehat{\pi}_{j,k} = 2\varphi_{j,k} = \frac{1}{2}k_{1,j,k} + k_{2,j,k}$) using RelateAdmix[33] (or alternatively REAP[34]) on a genotype file with MAF cutoff = 0.01. We subsequently identified 1,465 Greenlandic sib pairs by use of empirical thresholding over k_1 (IBD1; $0.3 < k_1 \leq 0.7$) and k_2 (IBD2; $0.1 < k_2 \leq 0.5$) on the RelateAdmix output. We then re-computed the $\mathbf{K_{IBD}}$ matrix for the identified sib pairs using RelateAdmix. $\mathbf{K_{IBD}}$ for the Danish sib pairs was computed from the PLINK `--genome` output, using MAF cutoff = 0.01. For both the Greenlandic and Danish sib pairs, we also computed a truncated $\mathbf{K_{IBD>t}}$ matrix, in which all entries below a threshold t = 0.05 were set to zero, and a $\mathbf{K_{IBD}^0}$ matrix, in which all between-sib pair values were set to zero.

$\mathbf{K_{IBS}}$ and $\mathbf{K_{IBS>t}}$ (t = 0.05) for both Greenlanders and Danes were computed directly with GCTA using a MAF cutoff = 0.01. Causal variants were removed from the computation of $\mathbf{K_{IBS}}$ and $\mathbf{K_{IBS>t}}$. Because causal variants are selected from the entire list of available SNPs, we assume that they have a similar allele frequency distribution as the genotyped SNPs. We can explicitly control for this by adding `--grm-adj 0` to the GCTA command line, however, this setting had no effect on our estimates (data not shown). For some heritability estimations, we computed $\mathbf{K_{IBS}^*}$ after removing not only the causal variants, but also all variants in LD with the causal ones (i.e. by applying extreme LD pruning in the vicinity of causal variants). Finally, we also estimated heritability by use of a $\mathbf{K_{IBS}^c}$ matrix in which all causal variants were included.

Heritability estimation

Additive genetic variance $\hat{\sigma}_A^2$ and, subsequently, total narrow-sense heritability h^2 were estimated with the GRM-based restricted maximum likelihood (GREML) procedure implemented in GCTA and in LDAK for various GRMs. For the sib pairs in particular, we carried out total narrow-sense heritability estimations using (i) the IBD-based matrices ($\mathbf{K_{IBD}}$, $\mathbf{K_{IBD>t}}$ and $\mathbf{K_{IBD}^0}$) alone; (ii) the IBS-based matrices ($\mathbf{K_{IBS}}$, $\mathbf{K_{IBS}^*}$, and $\mathbf{K_{IBS}^c}$) alone; (iii) $\mathbf{K_{IBS}}$ together with $\mathbf{K_{IBD}}$, $\mathbf{K_{IBD>t}}$, $\mathbf{K_{IBD}^0}$ or $\mathbf{K_{IBS>t}}$; and (iv) the classical sib-pair approach. We also estimated $\hat{\sigma}_A^2$ after adjusting for the first 5, 10 or 20 PCs, or proportion of Inuit ancestry (where applicable).

Analysis settings

We ran phenotype simulations and heritability estimates on three groups of Greenlanders: (i) all samples (N = 4,659); (ii) sib pairs (N = 1,688); and (iii) unrelated individuals (N = 585), as well as two separate groups of Danes: (i) unrelated individuals (N = 5,470); and (ii) sib pairs (N = 1,169). In both populations, the $\hat{\pi}$ threshold for identifying unrelated individuals was 0.0625. We estimated total heritability on the above groups using different GRMs, without and with covariates (summarized in Table 1). For the sake of simplicity, we do not show results that are nonsensical (e.g. $\mathbf{K_{IBD}}$ for unrelated individuals).

Application to real data

We applied the best performing model to real phenotypic data from the two Greenlandic cohorts. All traits considered were quantitative and consisted of basic anthropometric traits (height, weight, body mass index, hip circumference, waist circumference and waist-to-hip ratio), as well as serum lipid levels (total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides). Data were rank-transformed to the quantiles of a standard normal distribution. Age and sex were included as covariates. We also carried out an empirical investigation of the impact of allele frequency and LD weighting – as defined in the LDAK model[14] – on heritability estimates. In particular, we estimated

total narrow-sense heritability for the ten available traits assuming seven different genotype standardizations by setting LDAK's parameter $\alpha = \{-1.25, -1, -0.75, -0.5, -0.25, 0, 0.25\}$, and accounting for LD weighting. In this context, the model used by GCTA can be seen as a special case of the LDAK model by setting $\alpha = -1$ and ignoring LD weighting.

Results

Admixture and relatedness in the Greenlandic and Danish data

Principal component analysis (PCA) and ancestry component analysis with ADMIXTURE[35] of 4,659 individuals showed that the general Greenlandic population is the result of admixture between Greenlandic Inuit and European populations and that there is high variance in the admixture profiles of the Greenlanders (Figure 1A).[19,20] Conversely, a sample of 5,470 Danish individuals appeared largely unstructured (Figure 1B).[36] We identified a large number of sib pairs in the Greenlandic sample (1,465 pairs; N = 1,688 individuals; Figure S1A). We also confirmed the lack of relatedness in the unrelated Danish samples (Figure S1B) and the presence thereof in the Danish sib pairs (Figure S1C).

Identity by state in the Greenlandic and Danish data

We illustrate the intrinsic differences of IBS between admixed and unadmixed populations by plotting the IBS-based genetic covariance against the IBD-based $\hat{\pi}$ estimates from the Greenlandic and Danish sib pairs, respectively (Figure 2). For a given kinship (e.g. full siblings), the corresponding IBS values were far more dispersed in the Greenlanders (Figure 2A) than in the Danes (Figure 2B). This is due to the heterogeneous admixture profiles in the Greenlanders (Figure 1A). In other words, whereas IBS is proportional to IBD in unadmixed related individuals, this does not hold for the admixed individuals.

Heritability estimates in phenotypes with no population-specific environmental effects

We explored a number of approaches for estimating narrow-sense heritability h^2 in the admixed Greenlandic and the unadmixed Danish population. We simulated quantitative traits by (i) randomly selecting 1,500 causal loci with effect sizes depending on the allele frequency, such the effect

sizes of the standardized genotypes are normally distributed as assumed in the GCTA software, and (ii) adding environmental noise so that the true simulated h^2 was 0.4, 0.6 or 0.8. In these simulations, all individuals were set to having the same environmental variance regardless of their ancestry. We then estimated h^2 in an LMM framework for different GRMs – namely $\mathbf{K_{IBD}}$, $\mathbf{K_{IBS}}$, $\mathbf{K_{IBS>t}}$ and combinations thereof (Figure 3; Figure S2; Figure S3).

Total heritability estimates in sib pairs using one GRM

The use of $\mathbf{K_{IBD}}$, which captures the fraction of the genome shared IBD ($\hat{\pi}$), resulted in underestimates of total heritability both in the Greenlandic and the Danish sib pairs (Figure 3). Total heritability in the Greenlandic sib pairs was also underestimated when we used $\mathbf{K_{IBD>t}}$ (Figure 3A) and $\mathbf{K_{IBD}^0}$ (Figure S2A). Conversely, using $\mathbf{K_{IBD>t}}$ or $\mathbf{K_{IBD}^0}$ in the Danish sib pairs did not result in any significant downward biases (Figure 3B; Figure S2B). These results were insensitive to the method of IBD inference: both our method of choice (RelateAdmix[33]) and an alternative (REAP[34]) returned similar results (data not shown).

For closely related unadmixed individuals, IBS is proportional to IBD and therefore estimates based on $\mathbf{K_{IBS}}$ will correspond to the total narrow-sense heritability as well.[37] Indeed, simulated h^2 was fully recovered in the Danish sib pairs when $\mathbf{K_{IBS}}$ was used (Figure 3B), but showed consistent downward biases across all simulated h^2 values in the Greenlandic sib pairs (Figure 3A). Removing all SNPs with any LD with the causal variants ($r^2 = 0$) from the GRM (i.e. using the $\mathbf{K_{IBS}^*}$ matrix) returned lower yet still comparable h^2 estimates in both the Greenlandic (Figure S2A) and Danish sib pairs (Figure S2B). Conversely, when causal variants were included in the computation of the GRM (i.e. using the $\mathbf{K_{IBS}^c}$ **matrix**), then $\mathbf{K_{IBS}^c} \cong \mathbf{K_{causal}}$ and, consequently, h^2 was recovered (Figure S4).

Total heritability estimates in sib pairs using two GRMs

15

The use of two GRMs (e.g. $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$) for heritability estimates is meant to leverage datasets in which both closely and more distantly related individuals are present.[16] However, when we applied this approach to the entire Greenlandic dataset (N = 4,659), we observed a downward bias in total heritability estimates for the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ model, while using $\mathbf{K_{IBS>t}}$ & $\mathbf{K_{IBS}}$ erroneously returned heritability estimates near 1.00 regardless of the true simulated value (Figure S5). Nevertheless, when we performed the two-GRM analysis on the 1,465 Greenlandic sib pairs alone, the true simulated h^2 was almost perfectly recovered for all GRM combinations, with estimates showing only a minor downward bias (Figure 3A; Figure S2A). Interestingly, these models outperform the classical sib-pair analysis as evidenced by their lower root-mean-square deviation (Table S1; Figure S3). Extending the sample to include more distant relatives ($\hat{\pi}$ = [0.15, 0.67]; N = 2,615) resulted in underestimates of the total heritability (Figure S6), implying that the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ model performs efficiently only on first-degree relatives when admixture is present. We therefore further examined the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ model in the following section, focusing our attention on the sib pairs.

SNP heritability estimates in unrelated individuals

The use of $\mathbf{K_{IBS}}$ on unrelated individuals yields SNP (h_g^2) rather than total heritability (h^2) estimates.[5] Bearing this in mind, we also estimated h_g^2 in both unrelated Greenlanders and unrelated Danes (Figure S7; Table S2). In all cases, we found that $h_g^2 < h^2$. For the Danish samples in particular, the MetaboChip h_g^2 was smaller than the OmniExpress h_g^2.

Heritability estimates in phenotypes with population-specific environmental effects

In all of the above simulations, we assumed that the environmental component was independent of ancestry. However, when we added an environmental component correlating with ancestry to the simulated phenotypes, the use of $\mathbf{K_{IBS}}$ in the unrelated Greenlanders led to overestimates of SNP heritability h_g^2, despite adjusting for population structure (Table S2).

16

Adjusting the $\mathbf{K_{IBD}}$ model for either the first 10 PCs or proportion of Inuit ancestry produced a consistent yet uninterpretable pattern along the $\frac{h^2_{E \times Anc}}{h^2}$ ratio (Figure S8). On the contrary, adjusting the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ model for the same covariates produced a predictable as well as interpretable pattern across all choices for the $\frac{h^2_{E \times Anc}}{h^2}$ ratio (Figure 4): even though the true simulated heritability $\frac{\sigma^2_A}{\sigma^2_A + \sigma^2_{E \times Anc} + \sigma^2_e}$ was not recovered, the resulting estimates corresponded to removing the environmental interaction component ($\sigma^2_{E \times Anc}$) from the denominator of the formula (i.e. $\frac{\sigma^2_A}{\sigma^2_A + \sigma^2_e}$) through adjustment for ancestry. We note that adjustment for 10 PCs was equivalent to adjustment for proportion of Inuit ancestry (Figure 4).

Thanks to the interpretability of the resulting "conditional" estimates, we were able to recover the true simulated – the "marginal" – heritability[38]. We achieved this by rescaling the resulting total heritability estimates by a factor of $1 - \frac{\hat{\sigma}^2_{E \times Anc}}{\sigma^2_P}$, where $\hat{\sigma}^2_{E \times Anc}$ is an estimate of the environmental interaction variance computed as $\hat{\sigma}^2_{E \times Anc} = \sigma^2_P - \sigma^2_{\hat{P}}$. $\hat{P}$ is the phenotype residuals after regressing out the effect of structure captured either by admixture proportions or the first two principal components.

Application to real phenotypes

We applied the best model (i.e. $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ & 10 PCs) and the follow-up PCA-based adjustment to ten quantitative traits in the 1,465 Greenlandic sib pairs (Table 2). Not all phenotypes were equally sensitive to the PCA-based adjustment of their estimated conditional heritability, implying trait-specific environment-by-ancestry interactions. The GCTA model accepts only one type of genotype standardization ($\alpha = -1$), resulting in strong assumptions about the distribution of effect sizes. We therefore also used the LDAK model[14] and found that the optimal α value for genotype

17

standardization varied across traits with most phenotypes supporting $\alpha \geq -0.5$ (Figure S9; Table S3). Total heritability estimates under the LDAK model[14] were generally higher than under GCTA (Table 2; Table S3), with the greatest difference observed for height (0.657±0.042 for GCTA against 0.786±0.041 for LDAK). In addition, heritability estimates in eight out of ten real phenotypes were smaller in the Greenlanders than in their European or Mexican counterparts estimated with similar models.

Discussion

In this work, we explored the performance of existing methods for heritability estimates in the admixed Greenlandic population. Our goal was to propose a framework for unbiased heritability estimates in datasets where both population and family structure are notably present, as well as a way to interpret the resulting estimates. Even though the main focus is on total narrow-sense heritability (h^2), we also report results for SNP heritability (h_g^2), a quantity that has gained a lot of attention in the past decade due to the availability of GWAS data.[5,12,17]

Through extensive simulations, we observed that all LMMs using one GRM led to downward biases in total heritability estimates when applied to family data from Greenland. Common choices of GRM, such as $\mathbf{K_{IBD}}$ and $\mathbf{K_{IBS}}$, led to underestimates of total heritability in Greenlandic sib pairs, whereas no such biases were generally observed for the Danish sib pairs, indicating that admixture exerts a biasing effect on both IBD- and IBS-based estimates. Even though this is not surprising for the IBS-based estimates, as the IBS ~ IBD assumption does not hold for the Greenlanders, it is not very clear why IBD-based heritability estimates are also affected by admixture.

We also observed that an LMM with two GRMs ($\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$), a method designed to work on data with notable presence of family structure,[16] also led to downward biases in total heritability estimates when applied to the entire dataset from Greenland. However, when the same analysis focused on the Greenlandic sib pairs, it returned nearly unbiased heritability estimates. This could be due to the fact that, by restricting the analysis to the sib pairs, we controlled more efficiently for the noise that comes from unmatched between-pair admixture proportions. We note that the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ model outperformed the $\mathbf{K_{IBD}}$ model in the Danes, rendering more advisable the use of two GRMs in total narrow sense heritability estimates in unadmixed populations too.

19

When there is no environmental correlation with ancestry, the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ (or any other combination of one IBD- and one IBS-based GRM) model provides a nearly accurate estimate of the true heritability matched only by the classical sib pair analysis. However, in reality we expect environmental structure, which correlates with genetic structure, to exert an inflating effect on true heritability estimates but found that adjusting for structure did not remove the inflation. Nevertheless, we provide a way to interpret the resulting total heritability estimates from the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ & 10 PCs (or $\mathbf{K_{IBS>t}}$ & $\mathbf{K_{IBS}}$ & 10 PCs) model, as well as a way to adjust for the inflation. In particular, this inflated quantity is referred to as "conditional heritability" in a recent paper,[38] resulting after adjustment for model covariates like in our case. We observed that, under the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ & 10 PCs (or $\mathbf{K_{IBS>t}}$ & $\mathbf{K_{IBS}}$ & 10 PCs) models, the resulting conditional heritability[38] estimate will be inflated by a factor of $1/(1-h^2_{E \times Anc})$ and we propose an adjustment that accounts efficiently for this inflation in order to retrieve the "marginal heritability".[38]

As for the total narrow-sense heritability estimates of the real phenotypes obtained with the best model ($\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ & 10 PCs), we observe that in some occasions these are lower for the Greenlandic population than for European populations. A notable example is height, for which total heritability in the Greenlanders was estimated to be 0.656±0.042 (0.611 after the PCA-based adjustment), whereas in unadmixed Europeans it was estimated at 0.860.[11] We believe that this could be due to the reduced genetic diversity observed in the Greenlanders as a consequence of their particular population history, which included an extreme and prolonged bottleneck in recent times,[19,20] even though we did observe a notable increase when LD weighting was included in the estimation model according to LDAK.[14]

Finally, our results from SNP heritability (h^2_g) estimates in unrelated Greenlanders were consistent with previous findings, whereby genetic structure was reported to inflate heritability estimates when the environment was not constant;[17] we found that accounting for structure would not always

suffice. In any case, SNP heritability estimates in the Greenlanders should be interpreted with caution because, as we saw, IBS is affected by admixture that can lead to artificially increased levels of LD between causal and typed markers.

In summary, we advise against the use of $\mathbf{K_{IBD}}$ or $\mathbf{K_{IBS}}$ alone for total narrow-sense heritability estimates in populations with substantial levels of population and family structure. Instead, $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ & 10 PCs on a subset with high relatedness (preferably sib pairs) is advisable given that $\mathbf{K_{IBD}}$ can now be efficiently computed for admixed populations.[33,34] As an alternative to $\mathbf{K_{IBD}}$, $\mathbf{K_{IBS>t}}$ can be used together with $\mathbf{K_{IBS}}$.[16] In any case, the resulting conditional h^2 estimates should be viewed as potentially inflated by a factor that we estimated at $1/(1-h^2_{E \times Anc})$ and an additional PCA-based adjustment should be carried out in order to recover the marginal total h^2 estimate.

Supplemental Data

Supplemental Data include 9 figures and 3 tables.

Acknowledgements

References

1. Fisher, R.A. (1918). The Correlation Between Relatives on the Supposition of Mendelian Inheritance. Trans. Roy. Soc. Edin. *52*, 399–433.

2. Wright, S. (1921). Systems of Mating. I. the Biometric Relations between Parent and Offspring. Genetics *6*, 111–123.

3. Visscher, P.M., Hill, W.G., and Wray, N.R. (2008). Heritability in the genomics era--concepts and misconceptions. Nat. Rev. Genet. *9*, 255–266.

4. Zaitlen, N., and Kraft, P. (2012). Heritability in the genome-wide association era. Hum. Genet. *131*, 1655–1664.

5. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.

6. Shaw, R.G. (1987). Maximum-likelihood approaches applied to quanti- tative genetics of natural populations. Evolution *41*, 812–826.

7. Lange, K. (2002). Mathematical and Statistical Methods for Genetic Analysis (New York: Springer-Verlag).

8. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. *42*, 348–354.

9. Blangero, J., and Almasy, L. (1997). Multipoint oligogenic linkage analysis of quantitative traits. Genet. Epidemiol. *14*, 959–964.

10. Visscher, P.M., Medland, S.E., Ferreira, M.A.R., Morley, K.I., Zhu, G., Cornes, B.K., Montgomery, G.W., and Martin, N.G. (2006). Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings. PLoS Genet *2*,.

11. Visscher, P.M., Macgregor, S., Benyamin, B., Zhu, G., Gordon, S., Medland, S., Hill, W.G., Hottenga, J.-J., Willemsen, G., Boomsma, D.I., et al. (2007). Genome partitioning of genetic variation for height from 11,214 sibling pairs. Am. J. Hum. Genet. *81*, 1104–1110.

12. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. *88*, 294–305.

13. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

14. Speed, D., Cai, N., UCLEB Consortium, Johnson, M.R., Nejentsev, S., and Balding, D.J. (2017). Reevaluation of SNP heritability in complex human traits. Nat. Genet. *49*, 986–992.

15. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. *91*, 1011–1021.

16. Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., and Price, A.L. (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet. *9*, e1003520.

17. Browning, S.R., and Browning, B.L. (2011). Population structure can inflate SNP-based heritability estimates. Am. J. Hum. Genet. *89*, 191–193; author reply 193-195.

18. Zaitlen, N., Pasaniuc, B., Sankararaman, S., Bhatia, G., Zhang, J., Gusev, A., Young, T., Tandon, A., Pollack, S., Vilhjálmsson, B.J., et al. (2014). Leveraging population admixture to characterize the heritability of complex traits. Nat. Genet. *46*, 1356–1362.

19. Moltke, I., Fumagalli, M., Korneliussen, T.S., Crawford, J.E., Bjerregaard, P., Jørgensen, M.E., Grarup, N., Gulløv, H.C., Linneberg, A., Pedersen, O., et al. (2015). Uncovering the genetic history of the present-day Greenlandic population. Am. J. Hum. Genet. *96*, 54–69.

20. Pedersen, C.-E.T., Lohmueller, K.E., Grarup, N., Bjerregaard, P., Hansen, T., Siegismund, H.R., Moltke, I., and Albrechtsen, A. (2017). The Effect of an Extreme and Prolonged Population Bottleneck on Patterns of Deleterious Variation: Insights from the Greenlandic Inuit. Genetics *205*, 787–801.

21. Bjerregaard, P., Curtis, T., Borch-Johnsen, K., Mulvad, G., Becker, U., Andersen, S., and Backer, V. (2003). Inuit health in Greenland: a population survey of life style and disease in Greenland and among Inuit living in Denmark. Int J Circumpolar Health *62 Suppl 1*, 3–79.

22. Jørgensen, M.E., Borch-Johnsen, K., Stolk, R., and Bjerregaard, P. (2013). Fat distribution and glucose intolerance among Greenland Inuit. Diabetes Care *36*, 2988–2994.

23. Glümer, C., Carstensen, B., Sandbaek, A., Lauritzen, T., Jørgensen, T., Borch-Johnsen, K., and inter99 study (2004). A Danish diabetes risk score for targeted screening: the Inter99 study. Diabetes Care *27*, 727–733.

24. Jørgensen, T., Borch-Johnsen, K., Thomsen, T.F., Ibsen, H., Glümer, C., and Pisinger, C. (2003). A randomized non-pharmacological intervention study for prevention of ischaemic heart disease: baseline results Inter99. Eur J Cardiovasc Prev Rehabil *10*, 377–386.

25. Thuesen, B.H., Cerqueira, C., Aadahl, M., Ebstrup, J.F., Toft, U., Thyssen, J.P., Fenger, R.V., Hersoug, L.-G., Elberling, J., Pedersen, O., et al. (2014). Cohort Profile: the Health2006 cohort, research centre for prevention and health. Int J Epidemiol *43*, 568–575.

26. Byberg, S., Hansen, A.-L.S., Christensen, D.L., Vistisen, D., Aadahl, M., Linneberg, A., and Witte, D.R. (2012). Sleep duration and sleep quality are associated differently with alterations of glucose homeostasis. Diabet. Med. *29*, e354-360.

27. Aadahl, M., Linneberg, A., Møller, T.C., Rosenørn, S., Dunstan, D.W., Witte, D.R., and Jørgensen, T. (2014). Motivational counseling to reduce sitting time: a community-based randomized controlled trial in adults. Am J Prev Med *47*, 576–586.

28. Petersen, E.R.B., Nielsen, A.A., Christensen, H., Hansen, T., Pedersen, O., Christensen, C.K., and Brandslund, I. (2016). Vejle Diabetes Biobank - a resource for studies of the etiologies of diabetes and its comorbidities. Clin Epidemiol *8*, 393–413.

29. Lauritzen, T., Griffin, S., Borch-Johnsen, K., Wareham, N.J., Wolffenbuttel, B.H., Rutten, G., and Anglo-Danish-Dutch Study of Intensive Treatment in People with Screen Detected Diabetes in Primary Care (2000). The ADDITION study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with Type 2 diabetes detected by screening. Int. J. Obes. Relat. Metab. Disord. *24 Suppl 3*, S6-11.

30. Andreasen, C.H., Stender-Petersen, K.L., Mogensen, M.S., Torekov, S.S., Wegner, L., Andersen, G., Nielsen, A.L., Albrechtsen, A., Borch-Johnsen, K., Rasmussen, S.S., et al. (2008). Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. Diabetes *57*, 95–101.

31. Voight, B.F., Kang, H.M., Ding, J., Palmer, C.D., Sidore, C., Chines, P.S., Burtt, N.P., Fuchsberger, C., Li, Y., Erdmann, J., et al. (2012). The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet. *8*, e1002793.

32. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

33. Moltke, I., and Albrechtsen, A. (2014). RelateAdmix: a software tool for estimating relatedness between admixed individuals. Bioinformatics *30*, 1027–1028.

34. Thornton, T., Tang, H., Hoffmann, T.J., Ochs-Balcom, H.M., Caan, B.J., and Risch, N. (2012). Estimating kinship in admixed populations. Am. J. Hum. Genet. *91*, 122–138.

35. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

36. Athanasiadis, G., Cheng, J.Y., Vilhjálmsson, B.J., Jørgensen, F.G., Als, T.D., Le Hellard, S., Espeseth, T., Sullivan, P.F., Hultman, C.M., Kjærgaard, P.C., et al. (2016). Nationwide Genomic Study in Denmark Reveals Remarkable Population Homogeneity. Genetics *204*, 711–722.

37. Hayes, B.J., Visscher, P.M., and Goddard, M.E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. Genet Res (Camb) *91*, 47–60.

38. Weissbrod, O., Flint, J., and Rosset, S. (2018). Estimating SNP-Based Heritability and Genetic Correlation in Case-Control Studies Directly and with Summary Statistics. Am. J. Hum. Genet. *103*, 89–99.

39. Mamtani, M., Kulkarni, H., Dyer, T.D., Almasy, L., Mahaney, M.C., Duggirala, R., Comuzzie, A.G., Blangero, J., and Curran, J.E. (2014). Waist circumference is genetically correlated with incident Type 2 diabetes in Mexican-American families. Diabet Med *31*,.

40. Vattikuti, S., Guo, J., and Chow, C.C. (2012). Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. PLoS Genet. *8*, e1002637.

41. van Dongen, J., Willemsen, G., Chen, W.-M., de Geus, E.J.C., and Boomsma, D.I. (2013). Heritability of metabolic syndrome traits in a large population-based sample. J. Lipid Res. *54*, 2914–2923.

Figure Titles and Legends

Figure 1: principal component analysis of (A) the entire Greenlandic (N = 4,659) and (B) unrelated Danish (N = 5,470) cohorts. Greenlandic individuals were colored according to their admixture proportions as estimated with ADMIXTURE assuming K = 2 ancestral components. Note that Greenlanders with high proportion of Inuit ancestry (blue points) present further population structure along the second principal component. Due to persisting batch effects, the Danish data were pruned with PLINK (window size = 50; step size = 5; $r^2 = 0.1$) before the analysis.

Figure 2: scatterplot of within-pair (blue) and between-pair (black) IBS-based genetic covariance against IBD-based $\hat{\pi} = 2\varphi$ estimates for the (A) Greenlandic (MetaboChip) and (B) Danish (OmniExpress) siblings. IBS values were computed with GCTA, whereas IBD values were computed with RelateAdmix for the Greenlanders and with PLINK for the Danes. A negative IBS value denotes that two individuals are less than average related to each other. An y = x dotted line is shown in red.

Figure 3: mean total heritability estimates from 1,000 simulated phenotypes and 95% confidence intervals of the sampling distribution in (A) Greenlandic and (B) Danish sib pairs under various LMMs and the classical approach with no ancestry-specific environmental interactions added to the phenotypes. Genotype scaling parameter α was set to -1 (GCTA's standard) for both phenotype simulation and heritability estimation. $\mathbf{K_{IBD}}$: IBD-based GRM; $\mathbf{K_{IBD>t}}$: IBD-based GRM in which all entries below t = 0.05 were set to zero; $\mathbf{K_{IBS}}$: IBS-based GRM.

Figure 4: mean total heritability estimates from 1,000 simulated phenotypes and 95% confidence intervals of the sampling distribution in Greenlandic sib pairs under the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ model, without (squares) and with (circles, triangles) population structure covariates, as well as after the additional PCA-based adjustment (diamonds). The relationship between the environment-by-

ancestry (E×Anc) interaction and the true simulated heritability is quantified by the $h^2_{E \times Anc}/h^2$ ratio on the x-axis. The models with the covariates (circles and triangles) correspond to estimates of conditional total heritability after adjusting for the E×Anc effect (dotted line). Further rescaling of the conditional heritability estimates by $1 - \frac{\hat{\sigma}^2_{E \times Anc}}{\sigma^2_P}$ returns the marginal simulated heritability (dashed line). $h^2_{E \times Anc}$: proportion of variance captured by the E×Anc interaction; Inuit adm.: proportion of Inuit admixture. Adjustment for 5 or 20 first PCs returned virtually identical results (data not shown).

Figure S1: plots of the estimated fraction of the genome where one allele is shared IBD (K1 or Z1) against the estimated fraction where two alleles are shared IBD (K2 or Z2) for (A) all pairs of Greenlandic individuals (Metabochip); (B) unrelated Danish individuals (OmniExpress); (C) and Danish sib pairs (Omniexpress).

Figure S2: mean total heritability estimates from 1,000 simulated phenotypes and 95% confidence intervals of the sampling distribution in (A) Greenlandic and (B) Danish sib pairs under various LMMs and the classical approach with no ancestry-specific environmental interactions added to the phenotypes. Genotype scaling parameter α was set to -1 (GCTA's standard) for both phenotype simulation and heritability estimation. $\mathbf{K_{IBD}}$: IBD-based GRM; $\mathbf{K_{IBD>t}}$: IBD-based GRM in which all entries below t = 0.05 were set to zero; $\mathbf{K^0_{IBD}}$: IBD-based GRM in which all between-sib pair entries were set to zero; $\mathbf{K_{IBS}}$: IBS-based GRM; $\mathbf{K^*_{IBS}}$: IBS-based GRM in which both causal SNPs and any SNPs in LD with them were omitted; $\mathbf{K_{IBS>t}}$: IBS-based GRM in which all entries below t = 0.05 were set to zero.

Figure S3: violin plot showing total heritability estimates from 1,000 simulated phenotypes in (A) Greenlandic and (B) Danish sib pairs under various LMMs and the classical approach with no ancestry-specific environmental interactions added to the phenotypes. The median values are shown

as white circles, the interquartile ranges as black bars, and the 5th and 95th percentiles as whiskers. Genotype scaling parameter α was set to -1 (GCTA's standard) for both phenotype simulation and heritability estimation.

Figure S4: mean total heritability estimates from 1,000 simulated phenotypes and 95% confidence intervals of the sampling distribution in Greenlandic and Danish sib pairs when the IBS-based GRM includes the causal variants ($\mathbf{K_{IBS}^c}$).

Figure S5: mean total heritability estimates from 1,000 simulated phenotypes and 95% confidence intervals of the sampling distribution in the entire Greenlandic sample under the (A) $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ and (B) $\mathbf{K_{IBS>t}}$ & $\mathbf{K_{IBS}}$ models.

Figure S6: mean total heritability estimates from 1,000 simulated phenotypes and 95% confidence intervals of the sampling distribution under the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ model in Greenlandic sib pairs (left); Greenlandic sib pairs including more distant relatives (middle); and the entire Greenlandic sample (right).

Figure S7: mean IBS-based heritability estimates from 1,000 simulated phenotypes and 95% confidence intervals of the sampling distribution in (A) unrelated Greenlandic and (B) unrelated Danish individuals. Estimations were carried out using both $\mathbf{K_{IBS}}$, returning the SNP heritability, and $\mathbf{K_{IBS}^c}$, returning the total heritability. SNP heritability estimation in Danes was carried using both the MetaboChip and the OmniExpress data. SNP heritability was lower for the MetaboChip than for the OmniExpress array data, because the MetaboChip SNPs capture a smaller fraction of the genome through LD compared to the denser OmniExpress array. When the causal variants were included in the GRM ($\mathbf{K_{IBS}^c}$), total narrow-sense heritability was almost perfectly recovered.

Figure S8: mean total heritability estimates from 1,000 simulated phenotypes and 95% confidence intervals of the sampling distribution in Greenlandic sib pairs under the $\mathbf{K_{IBD}}$ model, without (squares) and with (circles, triangles) population structure covariates. The relationship between the E×Anc interaction and the simulated total heritability is quantified by the $h^2_{E \times Anc}/h^2$ ratio on the x-axis. None of the models recover the true estimated heritability in a systematic and interpretable manner along the $h^2_{E \times Anc}/h^2$ ratio.

Figure S9: fitness comparisons of the genotype scaling parameter α under the LDAK model on the basis of likelihood, whereby higher likelihood indicates better-fitting α. Lines report log likelihoods from LDAK for seven values of α relative to the highest observed likelihood. Colored lines indicate the ten quantitative traits and the black line reports averages across all traits.

Tables

Table 1: overview of the different approaches for estimating heritability from simulated data. The linear mixed models (LMMs) were run with and without covariates. Only the meaningful combinations of methods and sample groups were considered

		Unrelated	**Sib pairs**	**All***
LMM	K_{IBD}	N/A	h^2	N/A
	K_{IBS} / K^*_{IBS}	h^2_g	h^2	N/A
	K_{IBD} & K_{IBS}	N/A	h^2 and h^2_g	h^2 and h^2_g
	$K_{IBS>t}$ & K_{IBS}	N/A	h^2 and h^2_g	h^2 and h^2_g
Classical sib-pair analysis		N/A	h^2	N/A

*analysis carried out in Greenlanders only; N/A: not applicable

Table 2: total narrow-sense heritability estimates for real phenotypes under the **K_{IBD}** & **K_{IBS}** & age & sex & 10 PCs model using (i) the GCTA model; (ii) GCTA followed by PCA-based rescaling; and (iii) the LDAK model with an optimal trait-specific parameter α and LD weighting, as well as comparison with (iv) published total heritability estimates in Europeans and Mexicans

	GCTA estimate (mean±s.e.)	**Rescaled GCTA point estimate**	**Best LDAK estimate (mean±s.e.)**	**Estimate from literature (mean±s.e.)**
Height	0.656±0.042	0.611	0.786±0.041	0.860±0.117[11]
Weight	0.533±0.044	0.504	0.536±0.049	0.630[39]
Body mass index	0.505±0.045	0.494	0.470±0.05	0.340±0.120[40]
Hip circumference	0.494±0.046	0.472	0.534±0.05	0.680[39]
Waist circumference	0.464±0.047	0.461	0.479±0.051	0.620[39]
Waist-to-hip ratio	0.352±0.050	0.348	0.385±0.054	0.280±0.120[40]
Total cholesterol	0.487±0.047	0.475	0.523±0.05	0.510[41]
HDL cholesterol	0.453±0.046	0.431	0.512±0.049	0.480±0.110[40]
LDL cholesterol	0.497±0.050	0.496	0.534±0.052	0.510[41]
Triglycerides	0.312±0.051	0.308	0.333±0.054	0.470±0.120[40]

Table S1: root-mean-square deviation of the total heritability estimates from their respective true simulated values for the K_{IBD} & K_{IBS} and the classical sib-pair analysis

Scaling parameter (α)	Sample	True simulated heritability	KIBD & KIBS	Classical sib pair analysis
$\alpha = -1$	Greenlandic siblings	0.4	0.044	0.062
		0.6	0.042	0.062
		0.8	0.037	0.070
	Danish siblings	0.4	0.080	0.085
		0.6	0.076	0.080
		0.8	0.070	0.075
$\alpha = 0$	Greenlandic siblings	0.4	0.050	0.062
		0.6	0.048	0.069
		0.8	0.038	0.074
	Danish siblings	0.4	0.078	0.083
		0.6	0.071	0.077
		0.8	0.065	0.071

Table S2: mean SNP heritability h_g^2 estimates and 95% confidence intervals (in square brackets) under the $\mathbf{K_{IBS}}$ model without and with covariates

$\frac{h^2_{E \times Anc}}{h^2}$ ratio	h^2	$\mathbf{K_{IBS}}$	$\mathbf{K_{IBS}}$ & 10 PCs	$\mathbf{K_{IBS}}$ & % admixture
0	0.4	0.382 [0.369 0.395]	0.384 [0.370 0.398]	0.386 [0.373 0.400]
	0.6	0.552 [0.539 0.564]	0.549 [0.535 0.563]	0.554 [0.541 0.567]
	0.8	0.733 [0.721 0.745]	0.728 [0.715 0.741]	0.735 [0.723 0.747]
0.05	0.4	0.445 [0.433 0.458]	0.387 [0.372 0.401]	0.386 [0.373 0.400]
	0.6	0.623 [0.610 0.636]	0.567 [0.553 0.581]	0.570 [0.556 0.584]
	0.8	0.800 [0.789 0.810]	0.754 [0.741 0.766]	0.758 [0.746 0.770]
0.1	0.4	0.499 [0.488 0.511]	0.389 [0.375 0.403]	0.391 [0.378 0.404]
	0.6	0.693 [0.681 0.704]	0.593 [0.579 0.606]	0.597 [0.584 0.610]
	0.8	0.855 [0.846 0.865]	0.778 [0.766 0.790]	0.784 [0.773 0.796]
0.15	0.4	0.550 [0.539 0.561]	0.404 [0.390 0.418]	0.401 [0.388 0.414]
	0.6	0.740 [0.729 0.750]	0.601 [0.587 0.615]	0.606 [0.593 0.619]
	0.8	0.908 [0.900 0.915]	0.815 [0.804 0.826]	0.819 [0.808 0.829]
0.2	0.4	0.589 [0.578 0.600]	0.406 [0.392 0.420]	0.405 [0.392 0.419]
	0.6	0.796 [0.786 0.806]	0.634 [0.620 0.648]	0.638 [0.625 0.652]
	0.8	0.941 [0.935 0.947]	0.846 [0.835 0.856]	0.850 [0.840 0.859]

One thousand simulations were carried out on 525 unrelated Greenlandic individuals for different values of the $\frac{h^2_{E \times Anc}}{h^2}$ ratio and total heritability (h^2).

33

Table S3: total heritability estimates of the ten quantitative traits from the Greenlandic sib pairs using the $\mathbf{K_{IBD}}$ & $\mathbf{K_{IBS}}$ & age & sex & 10 PCs model under the LDAK model assuming different values for the genotype scaling parameter α

Alpha (α)	Height	Weight	BMI	Hip	Waist	Waist-to-hip ratio	Cholesterol	HDL cholesterol	LDL cholesterol	Triglycerides
-1.25	0.766±0.041	0.501±0.049	0.438±0.049	0.522±0.049	0.463±0.050	0.377±0.053	0.512±0.049	0.469±0.049	0.524±0.052	0.317±0.053
-1	0.781±0.041	0.513±0.049	0.446±0.050	0.527±0.049	0.469±0.051	0.382±0.053	0.520±0.050	0.485±0.049	0.531±0.052	0.322±0.053
-0.75	0.788±0.041	0.523±0.049	0.454±0.050	0.531±0.050	0.474±0.051	0.384±0.054	0.523±0.050	0.498±0.049	0.535±0.052	0.326±0.053
-0.5	0.789±0.041	0.530±0.049	0.460±0.050	0.533±0.050	0.477±0.051	**0.385±0.054**	0.524±0.050	0.506±0.050	0.536±0.052	0.329±0.054
-0.25	**0.786±0.041**	0.533±0.049	0.465±0.050	0.533±0.050	0.478±0.051	0.383±0.054	**0.523±0.050**	0.510±0.050	0.535±0.052	0.331±0.054
0	0.781±0.041	0.535±0.049	0.468±0.050	0.534±0.050	0.479±0.051	0.382±0.053	0.521±0.050	0.511±0.049	**0.534±0.052**	0.333±0.054
0.25	0.776±0.041	**0.536±0.049**	**0.470±0.050**	**0.534±0.050**	**0.479±0.051**	0.380±0.053	0.519±0.050	**0.512±0.049**	0.533±0.052	**0.333±0.054**

Values in bold indicate the estimate with the highest likelihood.
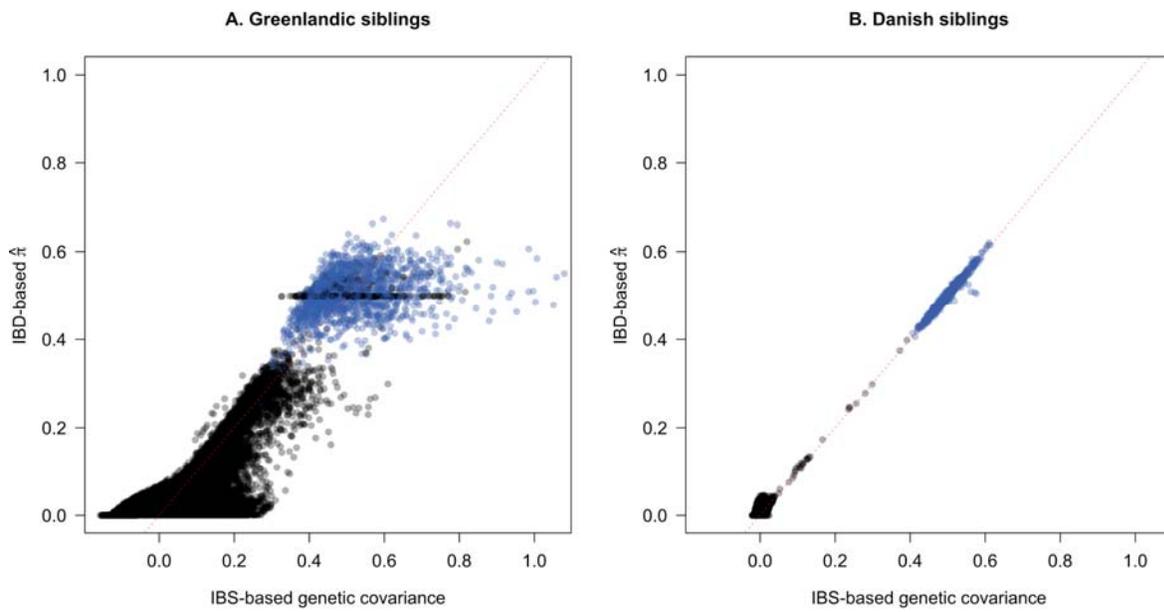
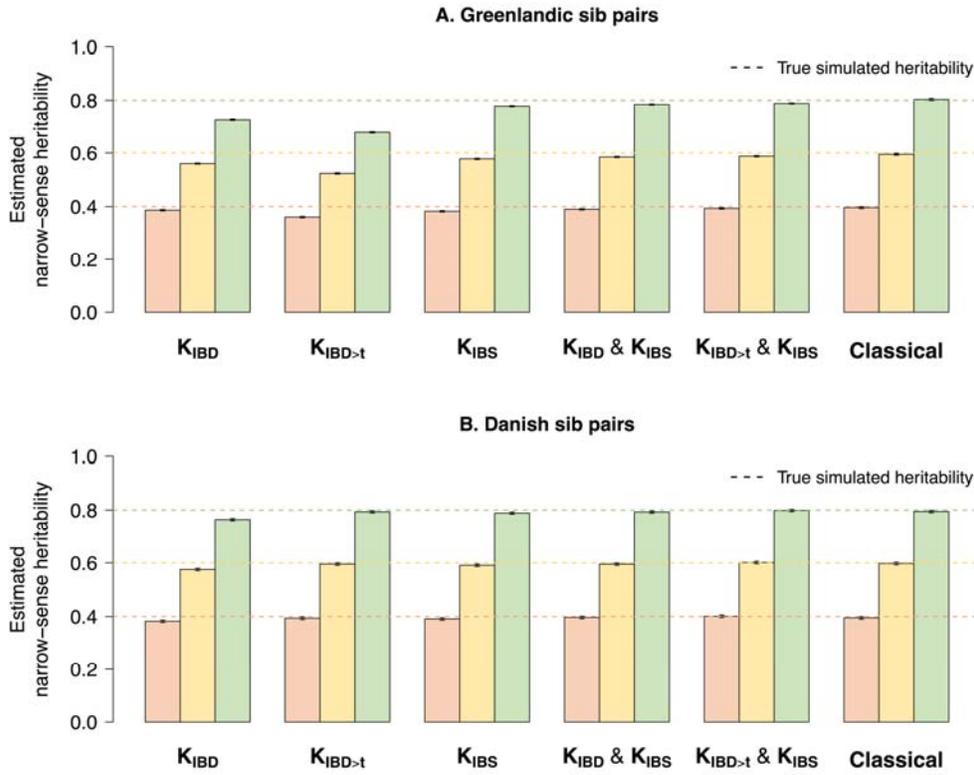# Figures

Figure 1
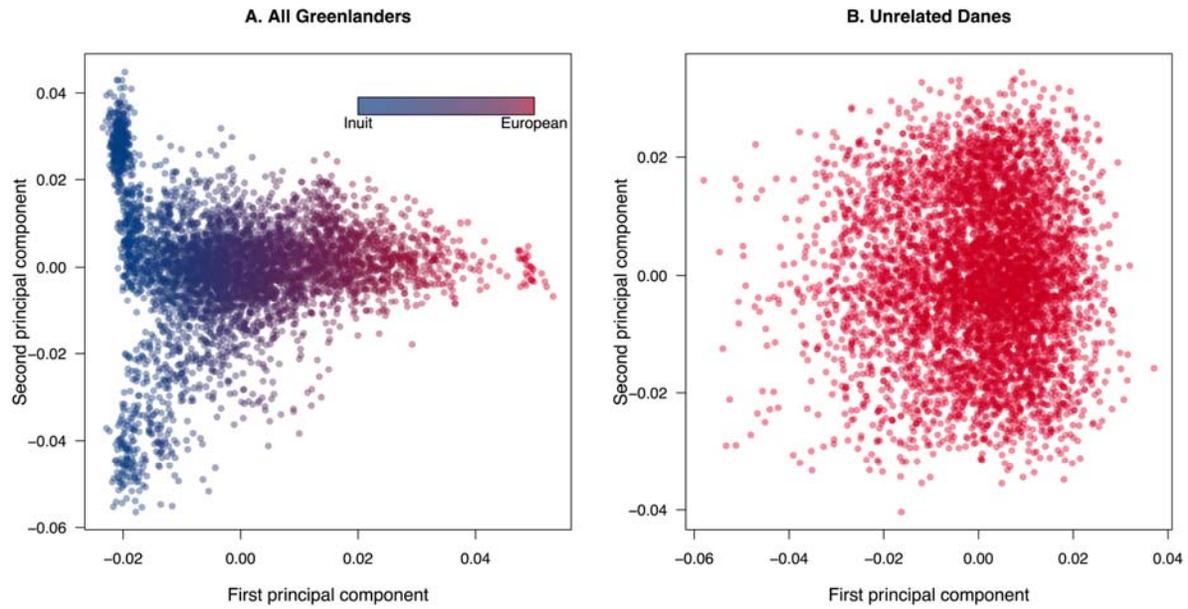
Figure 2

Figure 3

Figure 4
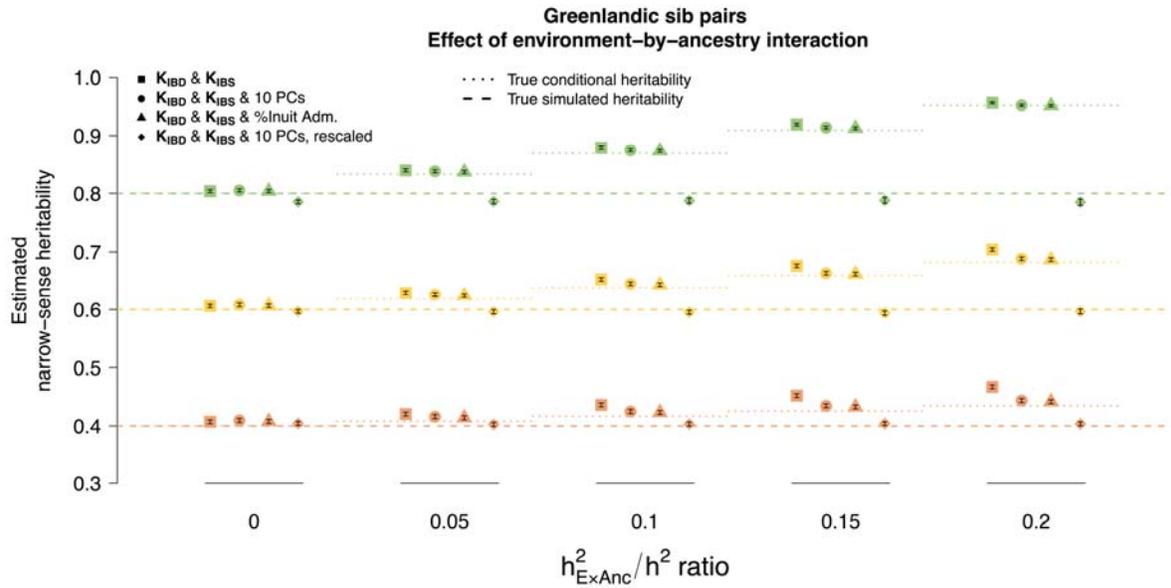
Figure S1
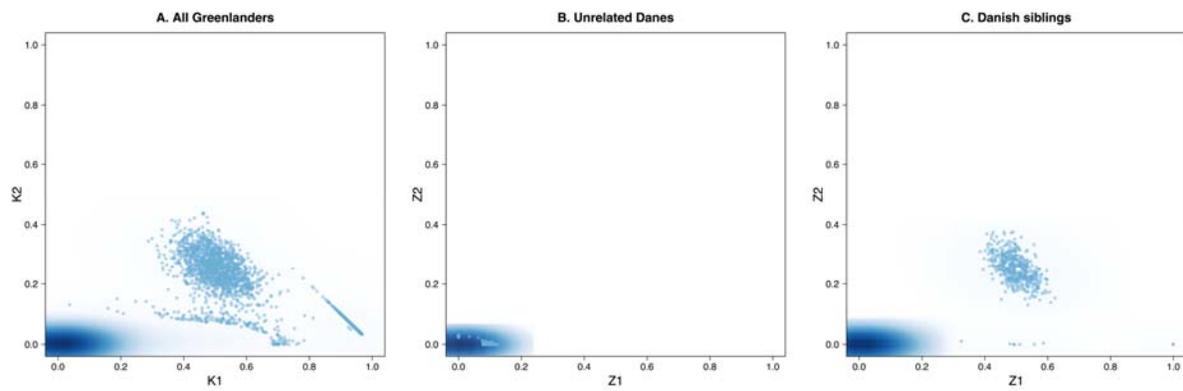
Figure S2
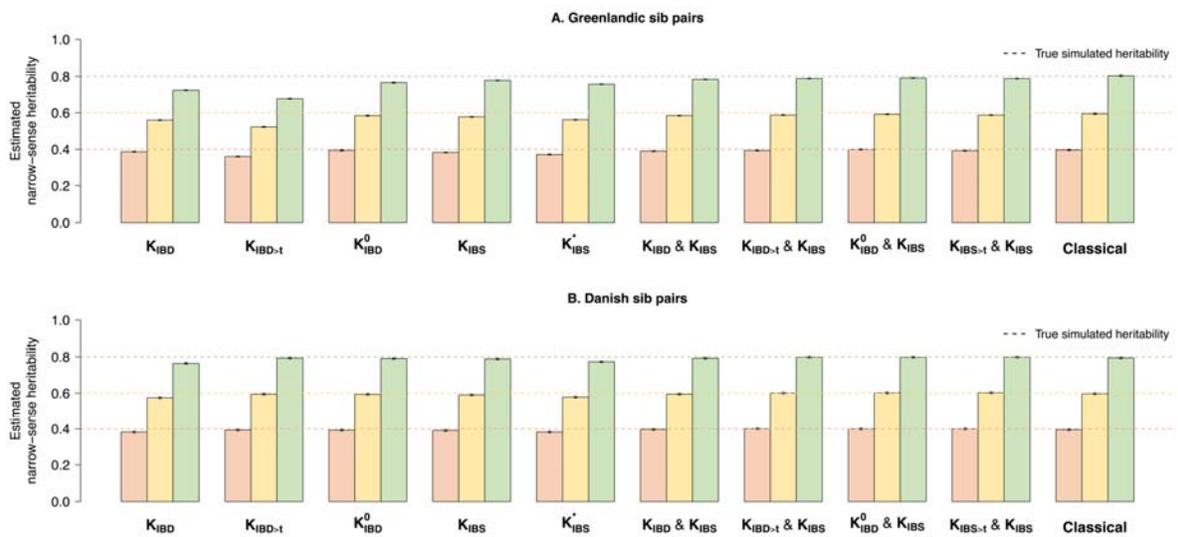
Figure S3
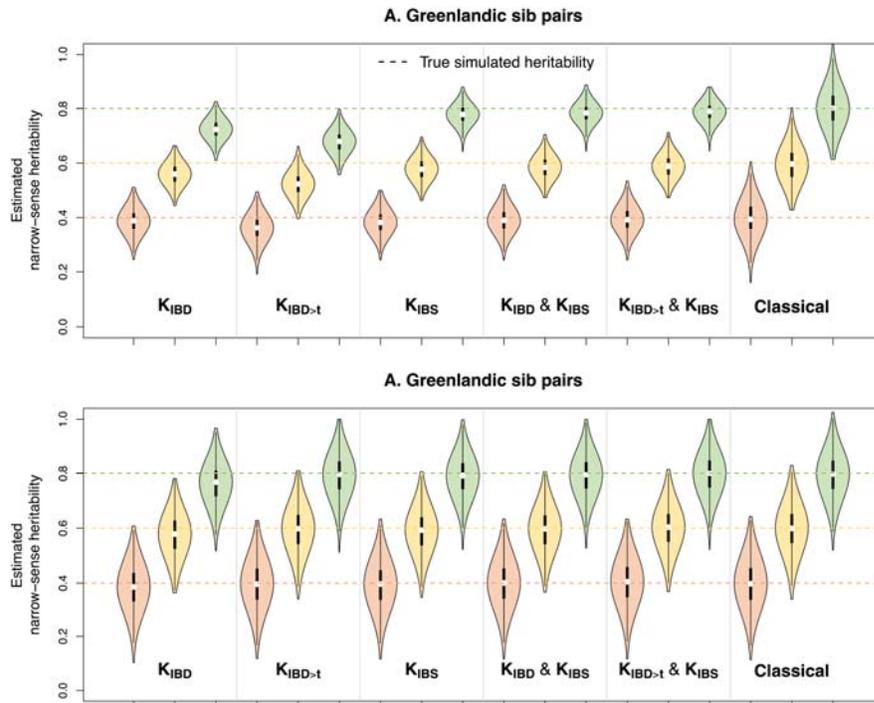
Figure S4
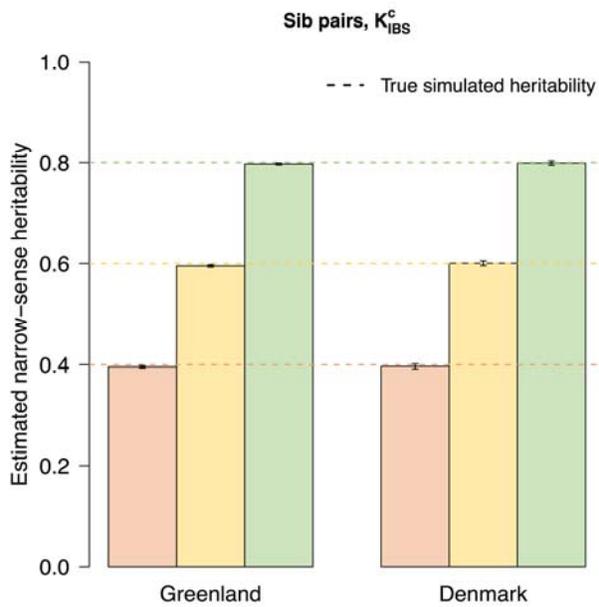
Figure S5
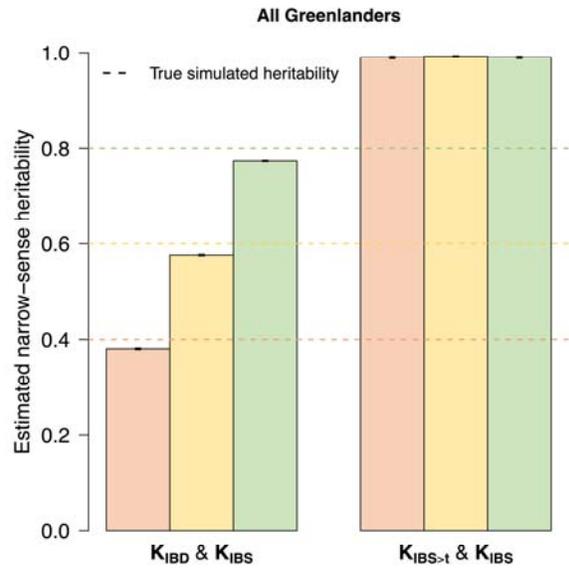
Figure S6

Figure S7

Figure S8

Figure S9