


ARTICLE OPEN



Heritability of complex traits in sub-populations experiencing bottlenecks and growth

Cameron S. Taylor¹ and Daniel J. Lawson^{1,2}  

© The Author(s) 2024

Populations that have experienced a bottleneck are regularly used in Genome Wide Association Studies (GWAS) to investigate variants associated with complex traits. It is generally understood that these isolated sub-populations may experience high frequency of otherwise rare variants with large effect size, and therefore provide a unique opportunity to study said trait. However, the demographic history of the population under investigation affects all SNPs that determine the complex trait genome-wide, changing its heritability and genetic architecture. We use a simulation based approach to identify the impact of the demographic processes of drift, expansion, and migration on the heritability of complex trait. We show that demography has considerable impact on complex traits. We then investigate the power to resolve heritability of complex traits in GWAS studies subjected to demographic effects. We find that demography is an important component for interpreting inference of complex traits and has a nuanced impact on the power of GWAS. We conclude that demographic histories need to be explicitly modelled to properly quantify the history of selection on a complex trait.

Journal of Human Genetics; <https://doi.org/10.1038/s10038-024-01249-2>


INTRODUCTION

Recent advancements in Genome Wide Association Studies (GWAS) of large human genetic datasets facilitate the study of how genetic variation associates with phenotypes [1–3], offering insight into the causal genetic structure of complex traits. However, not all populations used in association studies are equivalent - some will contain a diversity of ancestries in their history whilst others may have been relatively isolated. Here we investigate the consequences of studying isolated subpopulations who experienced a population bottleneck on the association with complex traits. Examples include the Ashkenazi Jewish population [4], Icelanders [5], Finns [6], and even experimentally explored in Bacteria [7], whose relatively low genetic variation provides a unique opportunity to isolate SNPs associated with traits and disease [8]. The same principle applies to the ‘Out-of-Africa’ bottleneck that affected all non-Africans, in which includes the largest studies such as UK Biobank [9] on which — arguably too much [10] genetic inference is based. However, the genealogies of the ancestral populations in which complex traits evolved may not be well represented by the more homogeneous subgroup that was studied. Here, we explore the consequences of studying these population subgroups via simulation.

A number of Single Nucleotide Polymorphisms (SNPs) contribute to genetic variation of a single complex trait, with a distribution of the effect size β_i for each SNP i [11]. The frequency of the SNPs f_i can depend on the effect size, which (along with polygenicity and linkage disequilibrium between them) can jointly be considered as the ‘genetic architecture’ of the trait [12]. This relationship is linked to whether the trait is under selection, and

can be quantified by a selection parameter s that takes values $s < 0$ if variation in the trait is selected against (‘negative selection’ which reduces the frequency of large effect SNPs) and $s > 0$ if the trait is selected to change (‘positive selection’ which makes large effect SNPs more common). The complex traits we consider here evolved a long time ago and are now under negative selection. However, the frequency of SNPs is affected by genetic drift, an evolutionary mechanism that leads to changes in the frequency of different alleles between generations [13], and hence different apparent selection. The relationship between frequency and effect size can be amplified or reduced by drift depending on the genetic architecture of a trait, the history of the population [14], and which SNP frequencies are examined. Figure 1 illustrates this process.

Although increased genetic drift is well exploited to discover associations between SNPs and traits, the impact on the distribution of effects is less studied. We consider the impact of a population contraction creating a “bottlenecked population” (potentially considering recovery) within a single ethnicity on the nature of a complex trait. One example is BRCA in Ashkenazi Jewish women [15] for which testing focuses on a single gene, although many genome-wide genetic effects are associated with breast cancer in the wider population [16]. On a much larger scale, the same bottleneck process leads to benefits and limitations of European Biobanks. The bottleneck that took place as the ancestors of Eurasians left Africa means that these data contain a subset of African diversity, with a concentrated gene pool [17]. As a consequence, Biobanks of primarily non-African ancestry are not representative of the populations where the traits evolved.

¹School of Mathematics, University of Bristol, Bristol, UK. ²MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK. email: dan.lawson@bristol.ac.uk

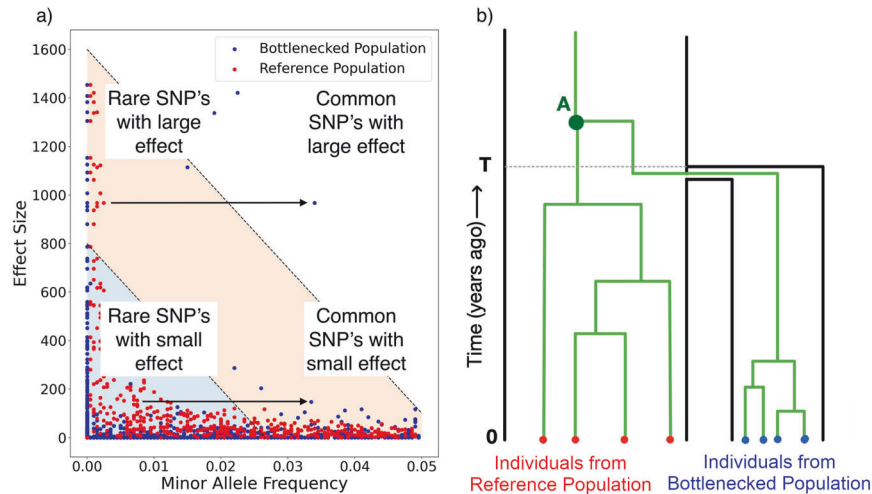


Fig. 1 Illustration of the effect of genetic drift on complex trait architecture. **a** The relationship between minor allele frequency and the effect size of SNPs in a population that has experienced a bottleneck (blue) and one that has not (red). The arrow shows the effect of drift, allowing rare causal SNPs with large effect size to become common, and changing heritability of complex traits after a population experiences a bottleneck. **b** A simple simulation scenario including a population separation time T generations in the past, leading to a Reference and Bottlenecked Population of size N_0 and N respectively. The green lines show the lineages of individuals for one locus that coalesce at the most recent common ancestor, which are relatively more diverse in the larger population

Genome-wide trait associations differ systematically in these populations, by the same isolation process described above, though varying in date, severity and migration history. We will investigate how these historical processes affect the association between genetic variation and complex traits.

A natural measure of the genome-wide relationship between genetic variation and a complex trait is heritability h^2 , that is, the proportion of phenotype variation due to genetic variation [18]. Aside from environmental variation, heritability of complex traits is determined by the number and frequency of SNPs and their respective effect sizes. The importance of the increase in frequency of rare SNPs with large effect size due to random drift is demonstrated in Fig. 1a in which a population reduction induces a change in the genetic architecture and a subsequent increase in heritability (from 0.5 to around 0.8). Complex traits are inherently polygenic so drift and selection acting on rare SNPs can be offset by that of common SNPs, and vice versa. Therefore, intuition on how genetic drift interacts with selection from well-studied single SNP models, based on the Nearly Neutral Theory of Molecular Evolution [19] which has been followed up experimentally [20] still require work to translate into the polygenic case. Polygenic selection [21] has been shown to change the distribution of private variants in populations [22] and has been shown to be associated with heritability across traits [11, 23] but without an explicit accounting for genetic drift.

We use a genome-wide simulation approach to explore how demographic structure, population separation and other factors characterise how a complex trait that has arisen in the past will evolve in future generations. We also consider how the realities of a detection threshold for associations interact with genetic drift. In only well-powered SNP subsets, we find that heritability is still strongly affected by genetic drift, implying that the patterns we observe will impact complex trait analysis from GWAS.

RESULTS

Our evolutionary framework regards selection on a complex trait considered to have been under selection in the past. Whilst this selection is strong, it is distributed over many loci leading to “nearly neutral” behaviour for each locus under a range of conditions (e.g. Eyre-Walker [24]). This facilitates a neutral model by assuming that selection has maintained SNPs at or below the

threshold where selection would efficiently reduce their frequency [19]. This model is practical for inference and widely used for inference of heritability (e.g. Yang et al. [25]), and is appropriate following a contraction if selection is either negligible (because it is less effective in a smaller or growing population) or absent (because the conditions in modern populations have changed).

We use a simulation-based approach to generate population histories accounting for genetic drift, mutation and recombination using the coalescent simulator msprime [26]. This generates (correlated) genealogies for every genetic loci for each individual of both a reference population and a Bottlenecked population. The population structure is simulated such that the ancestral population splits at some time T in the past, taken to be 200 generations in the base case, when some members form a sub-population that remains isolated from the Reference population, as shown in Fig. 1b. Typically, the size of the Reference population, N_0 , is large relative to the Bottlenecked population, N , taken to be 10,000 and 1000 respectively (unless otherwise stated).

To generate a phenotype, we choose a random sample of SNPs to be causal, and assume that the frequency of these SNPs has been under selection as a result of their effect on the phenotype. A proportion $1 - \pi$ of SNPs are assumed to have no causal effect whilst the proportion π (typically small) that contribute to phenotypic variance are distributed as:

$$\beta_j \sim N(0, \sigma_\beta^2),$$

$$\sigma_\beta^2 = [f_j(1 - f_j)]^s \quad (1)$$

where f_j is the frequency of the j th mutation in the sample. Because we are dealing with a simulation in which effect sizes are known, we can use a scaled mutation rate $\mu = \pi\mu_0$ as the mutation rate for causal SNPs leading to a random number of causal SNPs L with corresponding genotype values x_{ij} for individual i at SNP j , where frequencies f_j are for the reference population.

This is used to construct a simulated phenotype $y_i = \sum_{j=1}^L x_{ij}\beta_j + \epsilon_i$ [27]. The relationship between effect size and frequency is determined using the parameterisation in which is equivalent to the model used in LDAK [28]. Selection strength (on the trait) is parameterised by s in Eq. 1, and we report results for negative selection ($s = -1$) so that SNPs with large effect sizes have been selected to become rare. These effect sizes are transferred from the reference to the bottlenecked population,

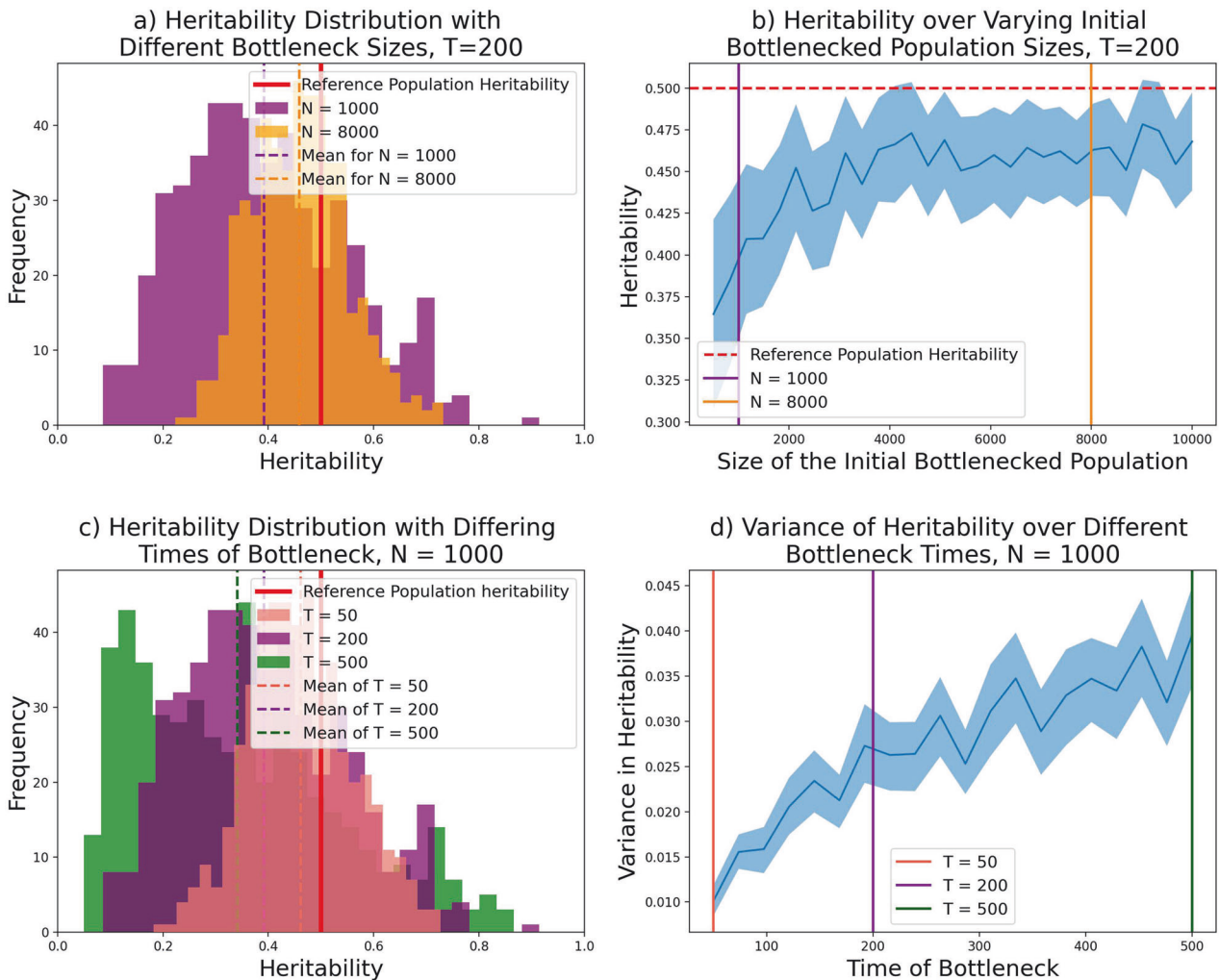


Fig. 2 Impact of demographic factors of a bottleneck on heritability. **a** Distribution of heritability at different sizes of bottleneck over 500 repeated simulations. Heritability distributions of populations with initial sizes $N = 1000$ and $N = 8000$ with a bottleneck that occurred at $T = 200$ generations ago. **b** Mean heritability decreases as the size of the initial Bottlenecked population N decreases. 30 samples were taken at each bottleneck size and the mean and error bars are plotted. **c** Heritability is distributed with similar tail but different mean and variance over 3 different Bottleneck times. 500 samples are taken for each bottleneck population with different times, $T = 50, 200, 500$. **d** There is a positive relationship between the time in the past that the bottleneck occurred, and the variance in heritability. 95% confidence intervals are shown for the 200 samples taken at each time point using an empirical bootstrap

under the assumption of constant environmental variance. The environmental noise ϵ_i is chosen to control the heritability in the reference population (see Methods) to $h^2 = 0.5$.

Figure 2a, b shows the distribution of heritability in simulated traits from this model for Bottlenecked Populations of different sizes (from $N = 500$ to 10,000) with $T = 200$. With the parameters described, the population contraction typically leads to an decrease in the heritability of complex traits. Several factors influence the extent to which heritability is impacted. A more severe contraction leads to lower levels of heritability on average. This is a result of amplified genetic drift due to fewer founding members of the sub-population, resulting in many SNPs being lost. Conversely, the variance increases with contraction size, as there is a possibility of rare SNPs with large effect being able to reach high prevalence.

The timing of the contraction can also impact how heritability is distributed. The mean heritability of the complex trait in the bottlenecked population decreases with age as more SNPs are likely to be lost. However, traits vary more when the contraction was longer ago (from $T = 50$ to 500, Fig. 2c, d) as the most extreme effect size SNPs can become more prevalent as a consequence of genetic drift. Over these time scales, the variance of heritability

increases close to linearly with time in the past (Fig. 2d), so heritability becomes increasingly varied for complex traits with equal genomic architecture. For reference, 200 generations corresponds to a split time of around 6000 years, similar in scale to the origin of the Ashkenazi Jewish population; 500 generations corresponds to 15 kya, around the time of the migration to the Americas; and the out-of-Africa bottleneck took place of the order of 2000 generations (60 kya) ago.

Introducing migration rate, m , from the reference population to the bottlenecked population offsets the impact of the contraction on the heritability of complex traits [29]. There is clear dependence between the time of the contraction and the impact of migration (from $m = 0-0.1$) on mean heritability of traits (Fig. 3a), with older events being more effected. Only modest migration ($m = 0.01$ with these parameters) is required to mitigate the effect of the contraction. For some problems even 'one individual per generation' [30] is enough migration to offset drift, for complex traits the threshold becomes age-dependent as drift interacts with replenishment of variability.

Population growth also impacts heritability change in the population after a contraction. As observed in Fig. 3b, whilst

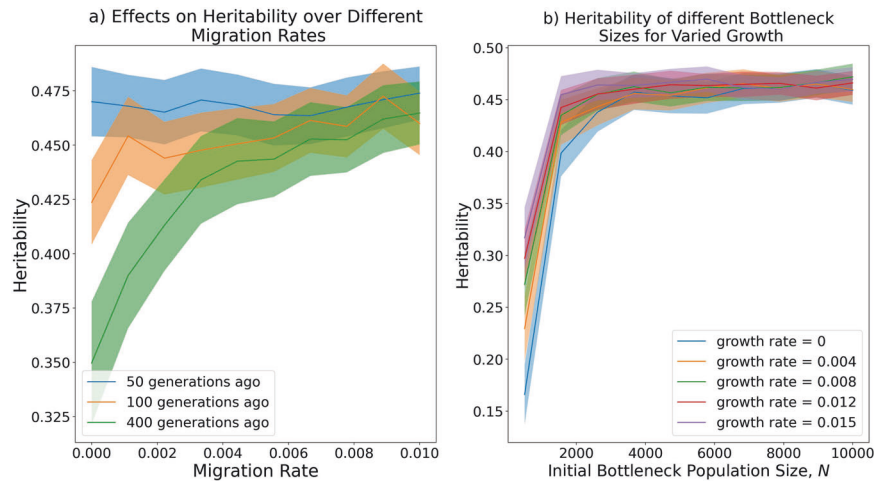


Fig. 3 Growth, migration and time since bottleneck change heritability. **a** Mean heritability in the Bottlenecked population as migration rate m is increased. 400 samples are taken at each time point and the mean with error bars is plotted. **b** Effect of bottleneck size and growth on heritability. Shown are the mean and standard errors based on 400 repeated simulations. The reference population is controlled to have heritability 0.5 throughout

growth following a bottleneck does recover some heritability, this is dominated by the strength of the contraction. It should be noted that the total bottlenecked population size can considerably exceed the reference population size in these figures - the population grows by a factor of nearly 20 in 200 generations with a growth rate of 0.015.

Impacts for GWAS inference

We have assumed in the simulation study that the effect size of SNPs for a particular complex trait are known. However, in reality effect sizes are inferred from SNP-trait association data using GWAS, which results in constraints being placed on what is known. It might therefore be feared that the outcomes of this paper cannot be directly applied to current complex traits, which we explored in [14]. There are 2 main issues that impact inference in real studies. The first of these is linkage disequilibrium which causes difficulties in isolating causal SNPs. We do not address this problem in this paper but it is explored elsewhere, for example by fine mapping, e.g. in [31, 32], and is unlikely to prevent the genetic drift we describe here.

The second issue pertains to the lack of statistical power of the data in real GWAS studies, weakening the power to resolve many traits. Whilst the models for which our genetic architecture was introduced [28] already considered missing heritability, these models assume selection acts to create the distribution of effect sizes relating to frequency that are present in the observed population. However, that distribution changes by genetic drift. To explore this further, we considered only SNPs that surpass a threshold of detectability in the bottlenecked population, either by conducting GWAS and ranking by p-value or using variance explained $v_j = f_j(1 - f_j) \beta_j^2$ in the trait. We considered only the top SNPs by choosing a variance explained threshold over a large number of simulations. For example, when we report the top $1/2^5$ SNPs, we consider the top $\sim 3\%$ of SNPs when averaged over 200 traits. As shown in Fig. 4, the heritability distribution can be effected by limited detection power. Specifically, for traits with low heritability in the target population, measured heritability for thresholded SNPs drops, as there are few (or even no) remaining SNPs that have sufficient explanatory power to be recorded. However, for traits with high observed heritability, the true heritability remains high. It follows that our observations on the impact of genetic drift on selection are valid for traits with high heritability, and increasingly for traits with lower heritability as GWAS sample size grows. This also implies that (as is widely

assumed) there will be a number of traits that are highly heritable, but that are more difficult to study in drifted (e.g. European) populations.

DISCUSSION

Population bottlenecks lead to genetic drift, which simultaneously reduces genetic variation whilst shifting the frequency of causal variants. Because large effect variants were selected to be rare, genetic drift will tend to increase the variance of these and hence can dramatically increase heritability. However, prediction of heritability is also affected by demographic factors, including growth and migration. Out of this, some general trends emerge.

Sharper bottlenecks have more dramatic impacts on the heritability of complex traits, likely due to a stronger 'founder effect'. The average heritability of a complex trait decreases for population contractions that occurred long ago, as the total amount of genetic drift determines mean heritability. However the variability of heritability between equally selected traits increases as the bottleneck age increases. Therefore some traits increase in heritability and some large effect SNPs rise to high frequency. Migration has a relatively simple impact on heritability as a regression to the mean, by mixing the two populations and maintaining all SNPs at intermediate frequency.

The discoverability of SNPs in GWAS does not appear to strongly interact with the relationship between heritability estimated in a bottlenecked population, and the heritability evolved in the reference population, as long as the heritability was not too low. Therefore inferences from GWAS studies are likely to be effected by the genetic drift process as we described here, even when only a small fraction of the effective SNPs are observed.

The simulation models used here imply that whatever genomic architecture evolved in ancestral populations, the complex process of demography will have changed it. We have demonstrated that it is relatively straight forward to perform inference forward in time, but there are currently no tools to perform the inverse inference: given an observed genomic architecture, and putative population history, can we infer how genomic architecture looked in the past? It seems unlikely that it will resemble what we see today. Further, it would seem to be sensitive to demographic parameters that remain difficult to infer [33]. Ancient DNA studies [34] provide constraints on what ancient populations might have looked like, but without massive samples we are not likely to obtain high-quality measurements of the frequency and

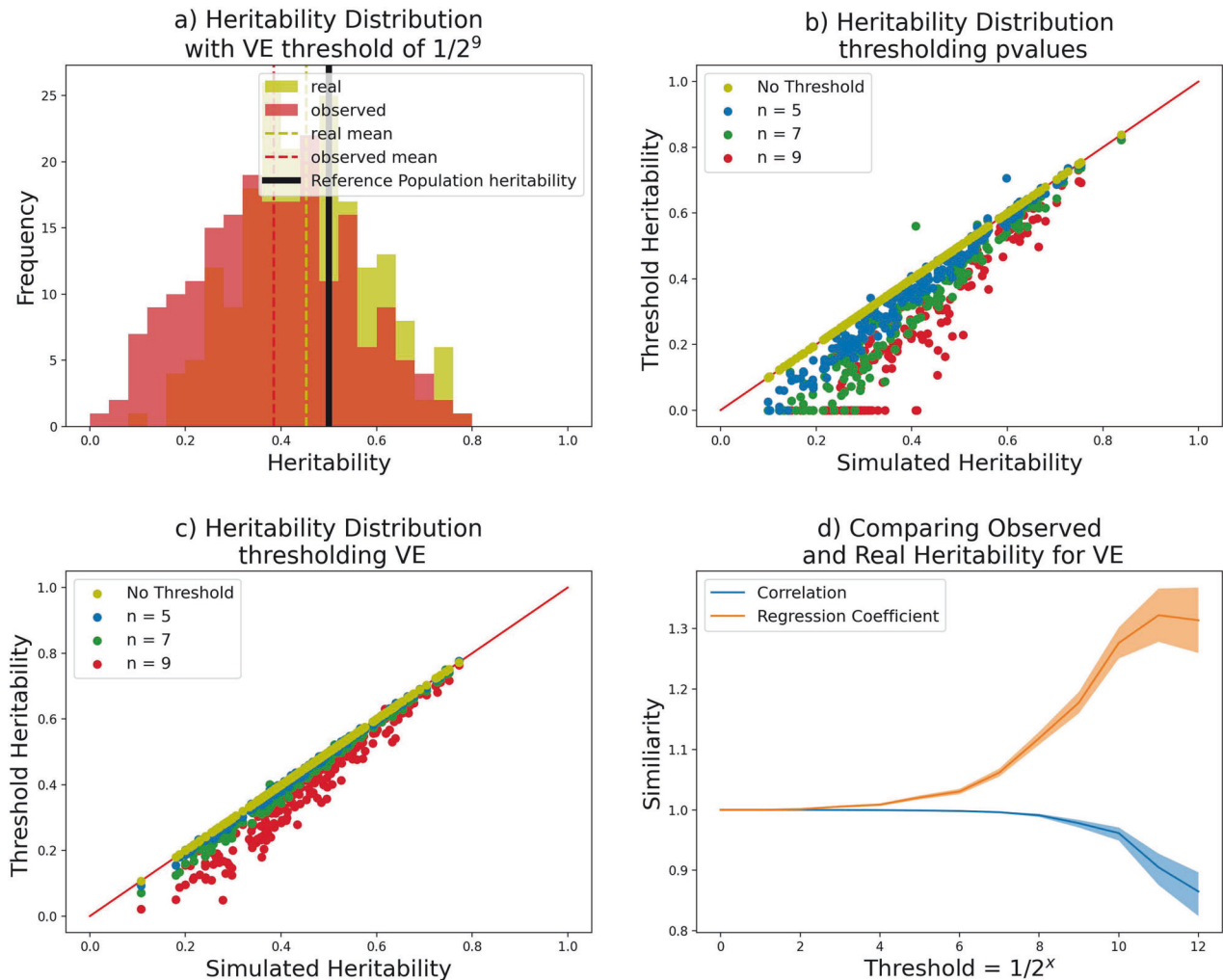


Fig. 4 The distribution of heritability of a bottlenecked population as observable in a GWAS study against simulated heritability present, over 200 repeated experiments. **a** the simulated heritability distribution (yellow) and observed thresholded heritability (green) where only the $1/2^9$ percentile of most informative SNPs (by VE = variance explained) are observed. **b** Comparing heritability when thresholding on p -values over 200 repeated experiments, where the threshold heritability only captures the $t = 1/2^t$ most extreme SNPs. The red line depicts values such that simulated and threshold heritability are equal. **c** as b, but where SNPs are ranked by variance explained. **d** Pearson correlation and regression coefficient between the real heritability and threshold heritability for thresholds $t = 1/2^n$ with standard error bars

correlation structure of causal variants. Understanding selection for complex traits then is likely to remain a challenge [35, 36]. Quantifying these issues better is important for predicting future evolutionary responses, for example to climate [37] and the Anthropocene [38].

MATERIALS AND METHODS

Generating genealogies

Tree sequence data for individuals in the chosen demography is generated using msprime [26, 39]. The demographic model chosen here is outlined in Fig. 1b, where two populations are derived from an ancestral population that experiences a split at some point in the past.

For a single simulation, the genealogies are created on 0.1 Morgans of genome. This is kept short for computational reasons, but the 200 simulations we average over can equivalently be thought of as generating 20 Morgans, with no change to the conclusions. However, it is important to emphasise that polygenicity and recombination rate are parameters of genomic architecture (See Supplement on scaling genomes). The time at which this split occurred and size of the initial bottleneck population are parameters that are explored throughout this paper, although a standard of $N_0 = 10,000$ initial people in the Reference population, $N = 1000$ in the

Bottleneck population and a split that occurred $T = 200$ generations ago is assumed by default.

Genetic variation is added to the model by superimposing mutations under the infinite sites model, with no time restriction as to when mutations may occur. Other complexities such as a population growth rate and continuous per-generation migration rate are included using the corresponding msprime parameters (see Software Availability for exact details).

Complex trait simulation

A complex trait is generated by statistically imposing the effect sizes of each SNP present in the reference population. Using the frequency of each SNP in the reference population, f_i , the effect sizes β_i are calculated by Eq. 1. In this paper the mutation rate is scaled to only include causal SNPs, so all SNPs have non-zero effect sizes.

In order to calculate the phenotype of an individual i , we need the contribution of environmental noise, ϵ_i . ϵ_i is distributed according to the environmental variance, V_e , of the trait, $\epsilon_i \sim N(0, V_e)$. We simulate ϵ_i with variance chosen such that the heritability in the Reference population remains at 0.5. Heritability is calculated as:

$$h^2 = \frac{V_g}{V_g + V_e} \quad (2)$$

This expression uses the genotype variance of the Reference population, $V_g = \text{Var}(Y^0)$, where $Y_i^0 = \sum_{j=1}^L X_{ij}^0 \beta_j$ is the genetic component of the phenotype. Here, X_{ij}^0 is the genotype of individual i at SNP j in the Reference population.

Constant environmental variance is assumed in this model, therefore we can formulate the phenotype of for an individual Y_i from their genome \mathbf{X}_i , as well as the V_e and β calculated in the reference population. This leads to the mixed linear model

$$Y_i = \sum_{j=1}^L X_{ij} \beta_j + \epsilon_i \quad (3)$$

By recalculating the genotype variance V_g in the bottleneck population, the heritability may be calculated using the equation above.

In Figs. 2–4 new genealogies with a different complex trait (i.e. new effect sizes) were created in each simulation. Figure 2d uses an empirical bootstrap to create error bars for the variance data.

Thresholding

The variance explained by SNP j is

$$v_j = f_j(1 - f_j)\beta_j^2 \quad (4)$$

which determines the power to detect the SNP in a GWAS study. We aim to mimic the findings that would realistically be found by thresholding the data to only include SNPs that surpass a certain power threshold.

The power threshold is determined by 200 repeated simulations that share the same demography, but are given unique population genealogies thus different complex trait effect sizes and frequencies. For each simulation the power of each SNP is calculated and the power threshold of the chosen quantile is determined from their combined distribution. In this case the $\frac{1}{200}$ th quantile is used for $n = 1, 2, 3, \dots$. The complex trait simulation procedure is then repeated, but SNPs that don't surpass the power threshold in the bottleneck population are discarded. Heritability is then recalculated with the reduced number of effective SNPs.

For Fig. 1d, Pearson correlation coefficients between the simulated and thresholded observed heritability values for each threshold level are calculated, as are regression coefficients. Standard error bars are added using Fishers z-transformation.

We further repeated the analysis thresholding not on variance explained, but on p -values, keeping only the smallest.

Software availability

The code used to obtain the results in this paper is available at <https://github.com/camerontaylor123/TaylorLawson2023>.

REFERENCES

- Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet.* 2018;50:1593–9. <https://doi.org/10.1038/s41588-018-0248-z>
- Hivert V, Sidorenko J, Rohart F, Goddard ME, Yang J, Wray NR, et al. Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. *Am J Hum Genet.* 2021;108:786–98. <https://doi.org/10.1016/j.ajhg.2021.02.014>
- Wang M, Huang S. The collective effects of genetic variants and complex traits. *J Hum Genet.* 2023;68:255–62. <https://doi.org/10.1038/s10038-022-01105-1>
- Carmi S, Hui KY, Kochav E, Liu X, Xue J, Grady F, et al. Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat Commun.* 2014;5:4835. <https://doi.org/10.1038/ncomms5835>
- Rose H. (2001). The commodification of bioinformation: the Icelandic health sector database. <https://wellcomecollection.org/works/kxenq2ue>.
- Kurki MI, Ga'ál El, Kettunen J, Lappalainen T, Menelaou A, Anttila V, et al. High risk population isolate reveals low frequency variants predisposing to intracranial aneurysms. *PLoS Genet.* 2014;10:e1004134. <https://doi.org/10.1371/journal.pgen.1004134>
- Wein T, Dagan T. The effect of population bottleneck size and selective regime on genetic diversity and evolvability in bacteria. *Genome Biol Evol.* 2019;11:3283–90. <https://doi.org/10.1093/gbe/evz243>
- Hatzikotoulas K, Gilly A, Zeggini E. Using population isolates in genetic association studies. *Brief Funct Genom.* 2014;13:371–7. <https://doi.org/10.1093/bfpg/elu022>
- Brocft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562:203–9. <https://doi.org/10.1038/s41586-018-0579-z>
- Sirugo G, Williams SM, Tishkoff SA. The missing diversity in human genetic studies. *Cell.* 2019;177:26–31. <https://doi.org/10.1016/j.cell.2019.02.048>
- Zeng J, de Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet.* 2018;50:746–53. <https://doi.org/10.1038/s41588018-0101-4>
- Johnson R, Burch KS, Hou K, Pacius M, Pasaniuc B, Sankaraman S. Estimation of regional polygenicity from GWAS provides insights into the genetic architecture of complex traits. *PLoS Comput Biol.* 2021;17:e1009483. <https://doi.org/10.1371/journal.pcbi.1009483>
- Star B, Spencer HG. Effects of genetic drift and gene flow on the selective maintenance of genetic variation. *Genetics.* 2013;194:235–44. <https://doi.org/10.1534/genetics.113.149781>
- Ashraf B, Lawson DJ. Genetic drift from the out-of-Africa bottleneck leads to biased estimation of genetic architecture and selection. *Eur J Hum Genet.* 2021;29:1549–56. <https://doi.org/10.1038/s41431-021-00873-2>
- Metcalfe KA, Poll A, Royer R, Llacuachqui M, Tulman A, Sun P, et al. Screening for founder mutations in *BRCA1* and *BRCA2* in unselected Jewish women. *J Clin Oncol.* 2010;28:387–91. <https://doi.org/10.1200/JCO.2009.25.0712>
- Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbaecker KB, et al. Genome-wide association study in *BRCA1* mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet.* 2013;9:e1003212. <https://doi.org/10.1371/journal.pgen.1003212>
- Lipson M, Reich D. A working model of the deep relationships of diverse modern human genetic lineages outside of Africa. *Mol Biol Evol.* 2017;34:889–902. <https://doi.org/10.1093/molbev/msw293>
- Wray NR, Visscher PM. Estimating trait heritability. *Nat Educ.* 2008;1:29 <https://www.nature.com/scitable/topicpage/estimating-trait-heritability-46889/>
- Ohta T. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst.* 1992;23:263–86. <https://www.jstor.org/stable/2097289>
- Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, et al. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet.* 2016;17:704–14. <https://doi.org/10.1038/nrg.2016.104>
- Barghi N, Hermisson J, Schl'otterer C. Polygenic adaptation: a unifying framework to understand positive selection. *Nat Rev Genet.* 2020;21:769–81. <https://doi.org/10.1038/s41576-020-0250-z>
- Durvasula A, Lohmueller KE. Negative selection on complex traits limits phenotype prediction accuracy between populations. *Am J Hum Genet.* 2021;108:620–31. <https://doi.org/10.1016/j.ajhg.2021.02.013>
- O'Connor LJ, Schoech AP, Hormozdiari F, Gazal S, Patterson N, Price AL. Extreme polygenicity of complex traits is explained by negative selection. *Am J Hum Genet.* 2019;105:456–76. <https://doi.org/10.1016/j.ajhg.2019.07.003>
- Eyre-Walker A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc Natl Acad Sci USA.* 2010;107 Suppl 1:1752–6.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88:76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Kelleher J, & Lohse K. (2020). Coalescent Simulation with msprime. In JY Duthel (Ed.), *Statistical Population Genomics* (pp. 191–230). Springer US. <https://doi.org/10.1007/978-1-0716-0199-09>
- Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet.* 2010;42:355–60. <https://doi.org/10.1038/ng.546>
- Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet.* 2012;91:1011–21. <https://doi.org/10.1016/j.ajhg.2012.10.010>
- Alcala N, Streit D, Goudet J, Vuilleumier S. Peak and persistent excess of genetic diversity following an abrupt migration increase. *Genetics.* 2013;193:953–71. <https://doi.org/10.1534/genetics.112.147785>
- Mills LS, Allendorf FW. The one-migrant-per-generation rule in conservation and management. *Conserv Biol.* 1996;10:1509–18. <https://www.jstor.org/stable/2387022>
- Malo N, Libiger O, Schork NJ. Accommodating linkage disequilibrium in genetic association analyses via ridge regression. *Am J Hum Genet.* 2008;82:375–85. <https://doi.org/10.1016/j.ajhg.2007.10.012>
- Thomas A, Abel HJ, Di Y, Faye LL, Jin J, Liu J, et al. Effect of linkage disequilibrium on the identification of functional variants. *Genet Epidemiol.* 2011;35:S115–S119. <https://onlinelibrary.wiley.com/doi/p>
- Lohmueller KE. The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet.* 2014;10:e1004379. <https://doi.org/10.1371/journal.pgen.1004379>
- Irving-Pease, EK, Muktupavela, R, Dannemann, M, & Racimo, F, (2021). Quantitative human paleogenetics: what can ancient DNA tell us about complex trait evolution? *Front Genet.* 2023;12. <https://doi.org/10.3389/fgene.2021.703541>

35. Hallgrímsson B, Miao W, Marcucio RS, Spritz R. Let's face it—complex traits are just not that simple. *PLoS Genet.* 2014;10:e1004724. <https://doi.org/10.1371/journal.pgen.1004724>
36. Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME. Selection for complex traits leaves little or no classic signatures of selection. *BMC Genom.* 2014;15:246. <https://doi.org/10.1186/1471-2164-15-246>
37. Wieczynski DJ, Singla P, Doan A, Singleton A, Han ZY, Votzke S, et al. Linking species traits and demography to explain complex temperature responses across levels of organization. *Proc Natl Acad Sci.* 2021;118:e2104863118. <https://doi.org/10.1073/pnas.2104863118>
38. Zawada KJA, Madin JS, Baird AH, Bridge TCL, Dornelas M. Morphological traits can track coral reef responses to the Anthropocene. *Funct Ecol.* 2019;33:962–75. <https://doi.org/10.1111/1365-2435.13358>
39. Kelleher J, Etheridge AM, McVean G. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput Biol.* 2016;12:e1004842. <https://doi.org/10.1371/journal.pcbi.1004842>

ACKNOWLEDGEMENTS

This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - <http://www.bristol.ac.uk/acrc/>.

COMPETING INTERESTS

None of the authors declared Financial and Non-Financial Relationships and Activities, and Conflicts of Interest regarding this manuscript.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s10038-024-01249-2>.

Correspondence and requests for materials should be addressed to Daniel J. Lawson.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024