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

Cameron S. Taylor[1] and Daniel J. Lawson[12*]

2023

1: School of Mathematics, University of Bristol, Bristol, UK.
2: MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK.
*: Corresponding Author: dan.lawson@bristol.ac.uk. ORCID ID 0000-0002-5311-6213

1 Abstract

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 in order to properly quantify the history of selection on a complex trait.

Keywords

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 (Hivert et al., 2021, Wang and Huang, 2023, Canela-Xandri et al., 2018), 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 (Carmi et al., 2014), Icelanders (Rose, 2001), Finns (Kurki et al., 2014), and even experimentally explored in Bacteria (Wein and Dagan, 2019), whose relatively low genetic variation provides a unique opportunity to isolate SNPs associated with traits and disease (Hatzikotoulas et al., 2014). 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 (Bycroft et al., 2018) on which — arguably too much (Sirugo et al., 2019) 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 (Zeng et al., 2018). 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 (Johnson et al., 2021). 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 (Star and Spencer, 2013), 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 (Ashraf and Lawson, 2021), 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 bottleneck within a single ethnicity on the nature of a complex trait. One example is BRCA in Ashkenazi Jewish woman (Metcalfe et al., 2010) for which testing focuses on a single gene, although many genome-wide genetic effects are associated with breast cancer in the wider population (Couch et al., 2013). 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 (Lipson and Reich, 2017). As a consequence, Biobanks of primarily non-African Ancestry are not representative of the populations where the traits evolved. 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 (Wray and Visscher, 2008). 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 Figure 1a in which a bottleneck induces a change in the genetic architecture and a subsequent

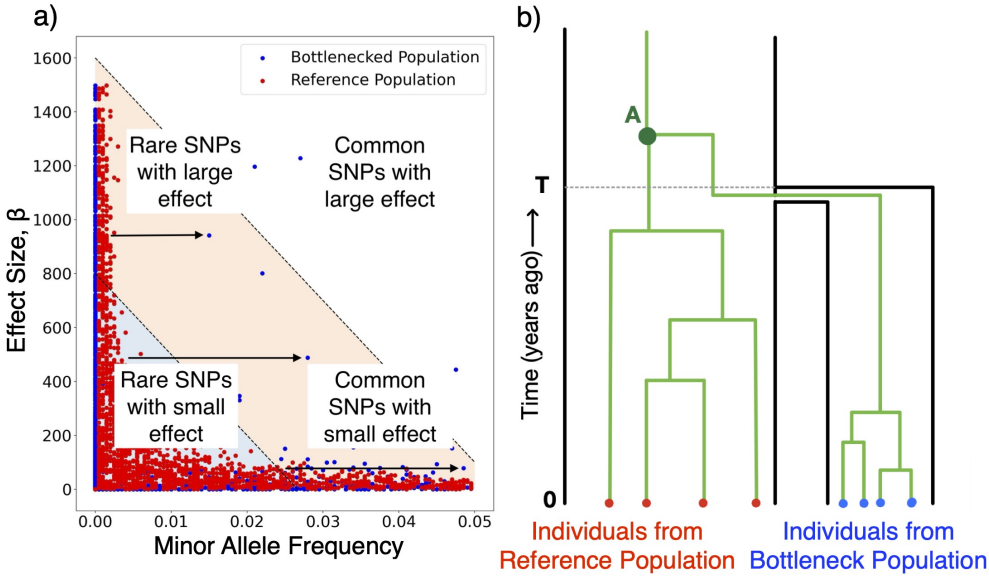

Figure 1: Illustration of the effect of genetic drift on complex trait architecture. a) The relationship between minor allele frequency and the effect size of SNPs in a population that has experienced a bottleneck (blue) and one that has not (red). The arrow shows the effect of drift, allowing rare causal 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.

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 (Ohta, 1992) which has been followed up experimentally (Lynch et al., 2016) still require work to translate into the polygenic case. Polygenic selection (Barghi et al., 2020) has been shown to change the distribution of private variants in populations (Durvasula and Lohmueller, 2021) and has been shown to be associated with heritability across traits (O'Connor et al., 2019), (Zeng et al., 2018) 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.

3 Results

We use a simulation-based approach to generate population histories accounting for genetic drift, mutation and recombination using the coalescent simulator msprime (Kelleher and Lohse, 2020). 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 figure 1b. Typically, the size of the Reference population, N_0, is large relative to the Bottlenecked population, N, taken to be 10000 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:

$$\begin{aligned} \beta_j &\sim N(0, \sigma_\beta^2), \\ \sigma_\beta^2 &= [f_j(1-f_j)]^s, \end{aligned} \tag{1}$$

where f_j is the frequency of the jth 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 $y_i = \sum_{j=1}^{L} x_{ij}\beta_j + \epsilon_i$ (Zhang et al., 2010). The relationship between effect size and frequency is determined using the parameterisation in Yang et al., 2011 which is equivalent to the model used in LDAK (Speed et al., 2012). Selection is captured by s, and we report results for negative selection ($s = -1$) so that common SNPs are unlikely to have large effect sizes. These effect sizes are transferred from the reference to the bottlenecked population, under the assumption of constant environmental variance. The environmental noise ϵ_i is chosen to control the heritability in the reference population (see Methods) allowing us to fix the reference population heritability 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. With the parameters described, a bottleneck typically leads to an increase in the heritability of complex traits. Several factors influence the extent to which heritability is impacted. A more severe bottleneck leads to higher levels of heritability on average, creating a left skew in the distribution. This is a result of amplified genetic drift due to fewer founding members of the sub-population, resulting in higher prevalence of rare SNPs with large effect. However, for the π values shown in this simulation (around 3500 causal SNPs) many complex traits can still have lower heritability, as large effect SNPs can be lost by chance.

The timing of the bottleneck can also impact how heritability is distributed. Whilst it minimally impacts the mean heritability of the complex trait in the bottlenecked population, traits in a recently bottlenecked population vary less than when the bottleneck was longer ago (Figure 2c-d). The increase in variance as the bottleneck occurred further in the past can be understood as a consequence of genetic drift, which takes time for frequencies of SNPs to change in the population. Over these time scales, the variance of heritability increases close to linearly with time in the past (Figure 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 15kya, around the time of the migration to the Americas; and the out-of-Africa bottleneck took place of the order of 2000 generations (60kya) ago.

Introducing migration rate, m, from the reference population to the bottlenecked population offsets the impact of the bottleneck on the heritability of complex traits (Alcala et al., 2013). There is no dependence between the time of the bottleneck and the impact of migration on mean heritability

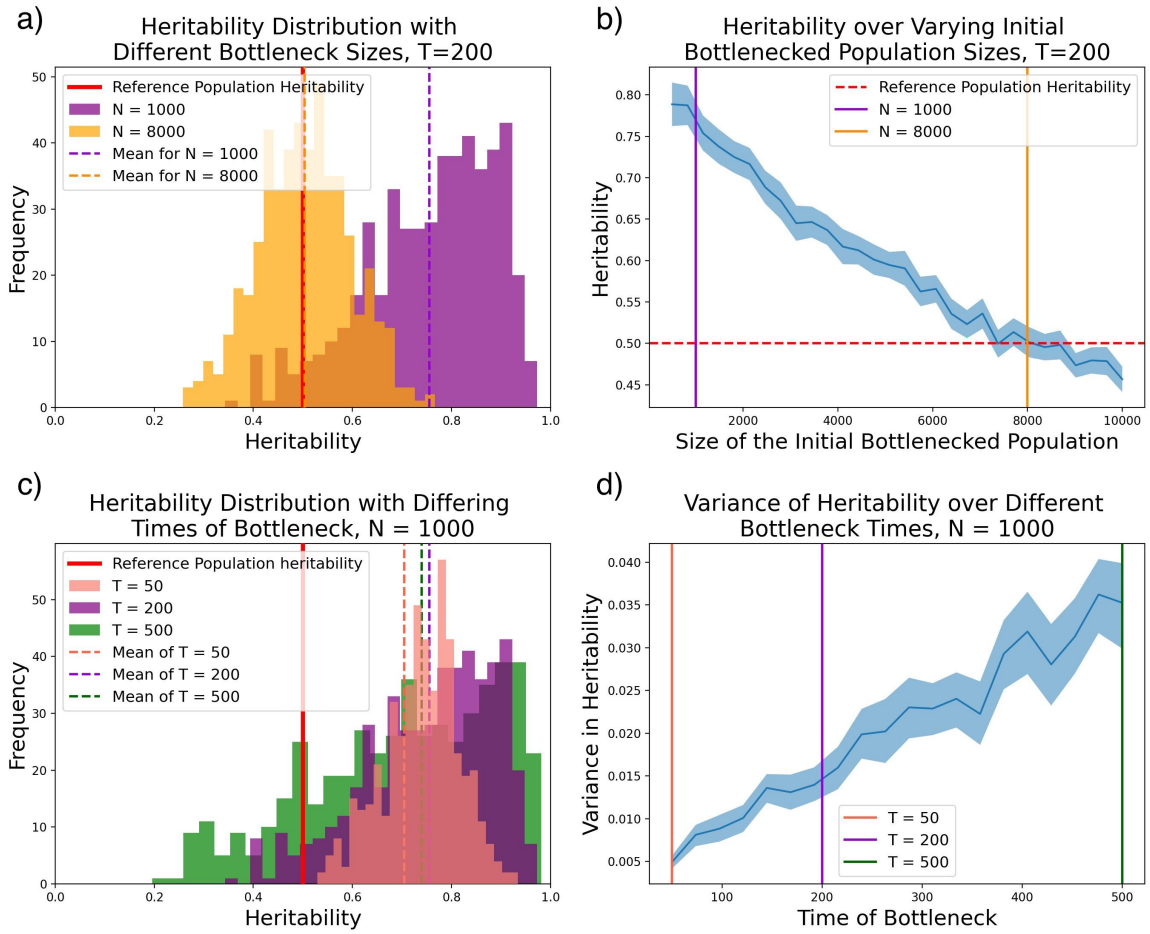

Figure 2: Impact of demographic factors of a bottleneck on heritability. a) Distribution of heritability at different sizes of bottleneck over 500 repeated simulations. Heritability distributions of populations with initial sizes $N = 1000$ and $N = 8000$ with a bottleneck that occurred at $T = 200$ generations ago. b) Mean heritability decreases as the size of the initial Bottlenecked population N increases. 30 samples were taken at each bottleneck size and the mean and error bars are plotted. c) Heritability is distributed with different variances but similar mean 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.

of traits (Figure 3). As migration is increased, heritability reverts to that in the Reference population. Whilst for some problems even 'one individual per generation' (Mills and Allendorf, 1996) is enough migration to offset drift, for complex traits there is no such simple threshold.

Population growth also impacts heritability change in the population after a bottleneck. As observed in figure 3b-c, there are complex non-linear relationships between bottleneck size and heritability in a growing population. At low growth rates or from modest bottlenecks, average heritability increases above the reference population's heritability, but very large growth rates,

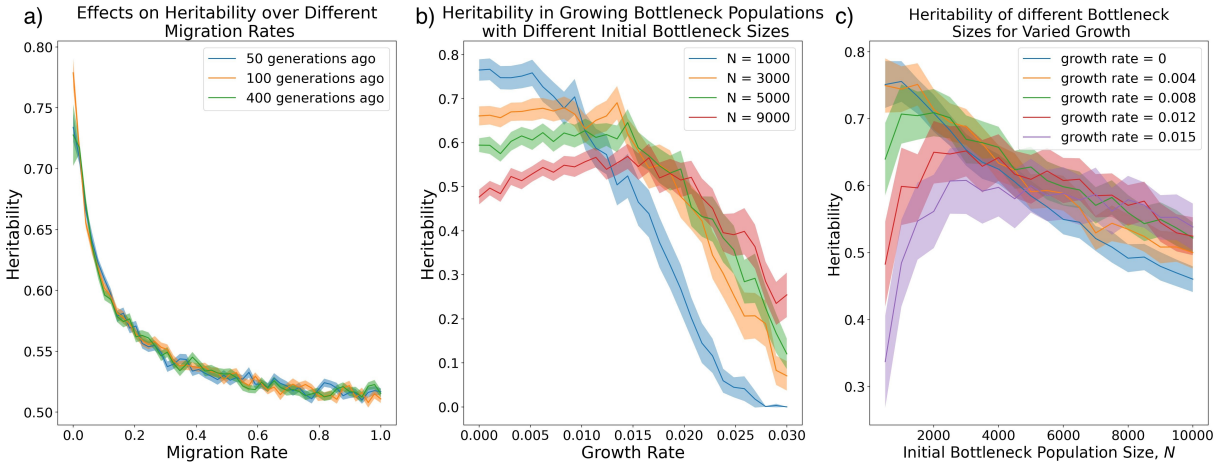

Figure 3: a) Mean heritability in the Bottlenecked population decreases as migration rate m is increased. The extent of the impact on heritability is independent of the time, T, at which the bottleneck occurred. 50 samples are taken at each time point and the mean with error bars is plotted. b) Growth in the Bottlenecked population reduces mean heritability. Heritability is measured in populations with different initial bottleneck sizes $N = 1000, 3000, 5000, 9000$ with varying growth rate (with T =200). Shown are the mean and standard errors based on 100 repeated simulations. c) As (b) but showing heritability as a function of the initial bottleneck size. The reference population is controlled to have heritability 0.5 throughout.

especially from severe bottlenecks, can dramatically reduce heritability. This is likely due to the fixation of large-effect alleles in the population, i.e. the mean trait value for the population may change, but the variation within it reduces. 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.

3.1 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 (Ashraf and Lawson, 2021). 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 (Malo et al., 2008, Thomas et al., 2011), 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 (Yang et al., 2011, Speed et al., 2012) 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, as measured by 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, the variance explained by a SNP must be larger than the top $\sim 3\%$ of SNPs when averaged over 200 traits. As shown in

Figure 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.

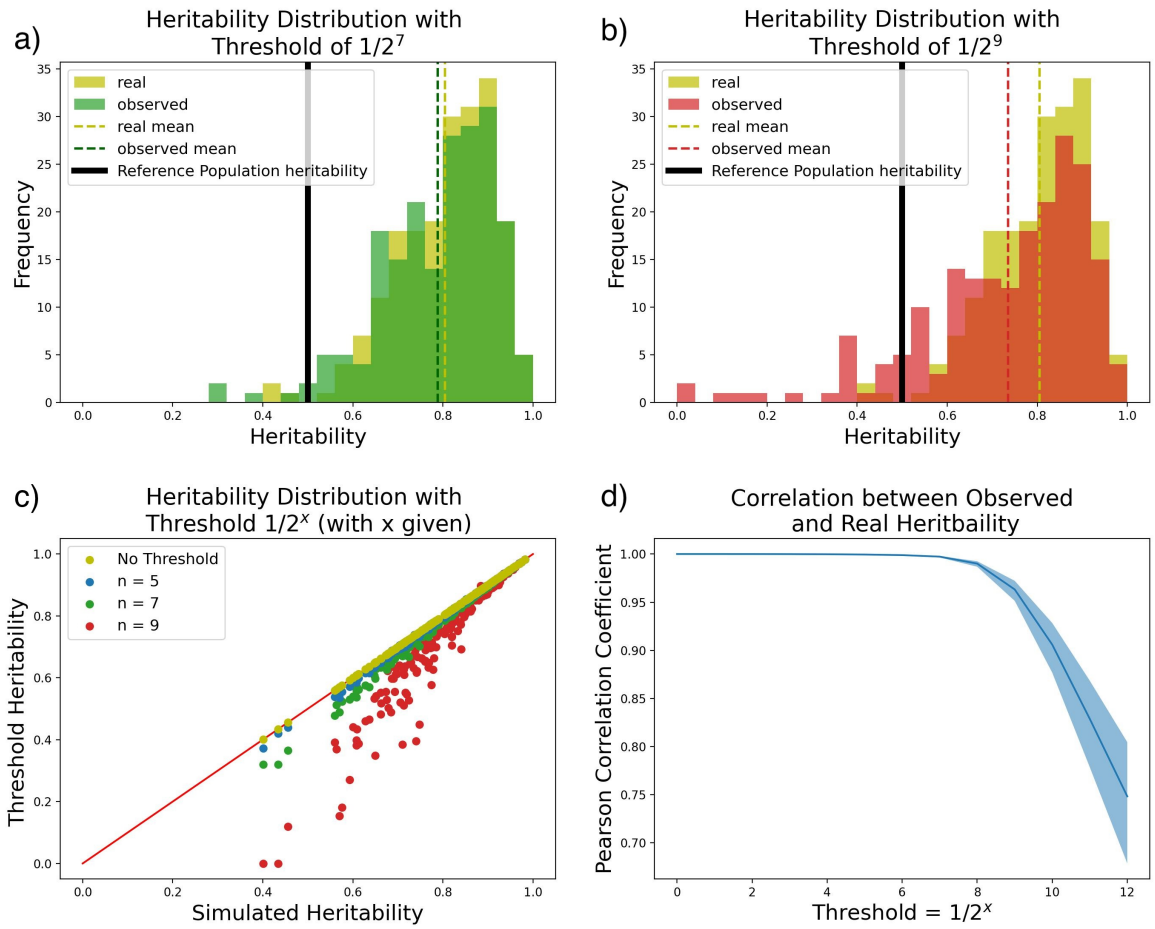

Figure 4: The distribution of heritability of a bottlenecked population as observable in a GWAS study against simulated heritability present, over 200 repeated experiments. (a) the simulated heritability distribution (yellow) and observed thresholded heritability (green) where only the $1/2^7$ percentile of most informative SNPs are observed. b) As in (a) but the observed heritability (red) only captures the $1/2^9$ percentile of most effective SNPs. c) Simulated heritability against observed heritability over 200 repeated experiments. The threshold heritability only captures the $t = \frac{1}{2^n}$ most informative SNPs. The red line depicts values such that simulated and threshold heritability are equal. d) Pearson correlation coefficient between the the real heritability and threshold heritability for thresholds $t = 1/2^n$ with standard error bars.

4 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 is not, however, sensitive to the time at which the bottleneck occurred, though the variability of heritability between equally selected traits increases as the bottleneck age increases. Therefore, immediately following a bottleneck, heritability is expected to have increased on average. 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.

However, the direction of the bias in heritability is hard to predict over a longer time period. Heritability changes nonlinearly with the entire population size history. The heritability of a complex trait is very sensitive to small levels of population growth, and severe bottlenecks coupled with large growth can reduce heritability towards 0. More work is needed to understand the mechanism behind this, to allow high quality prediction of the complex trait in an ancestral population, from only observations in contemporary populations.

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 (Lohmueller, 2014). Ancient DNA studies (Irving-Pease et al., 2021) 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 correlation structure of causal variants. Understanding selection for complex traits then is likely to remain a challenge (Kemper et al., 2014, Hallgrimsson et al., 2014). Quantifying these issues better is important for predicting future evolutionary responses, for example to climate (Wieczynski et al., 2021) and the Anthropocene (Zawada et al., 2019).

5 Methods

5.1 Generating Genealogies

Tree sequence data for individuals in the chosen demography is generated using msprime (Kelleher and Lohse, 2020). The demographic model chosen here is outlined in Figure 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 $N_0 = 10000$ 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).

5.2 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 Equation 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:

$$h^2 = \frac{V_g}{V_g + V_e} \tag{2}$$

This expression uses the genotype variance of the Reference population, $V_g = \mathrm{Var}(Y^0)$, 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

$$Y_i = \sum_{j=1}^{L} X_{ij}\beta_j + \epsilon_i. \tag{3}$$

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

In figures 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.

5.3 Thresholding

The variance explained by SNP j is

$$v_j = f_j(1 - f_j)\beta_j^2 \tag{4}$$

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 $\frac{1}{2^n}$ th quantile is used for $n = 1, 2, 3, \ldots$

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

Software Availability

The code used to obtain the results in this paper is available at

https://github.com/camerontaylor123/TaylorLawson2023

References

Ohta, T. (1992). The Nearly Neutral Theory of Molecular Evolution [Publisher: Annual Reviews]. *Annual Review of Ecology and Systematics*, *23*, 263–286. Retrieved September 22, 2023, from https://www.jstor.org/stable/2097289

Mills, L. S., & Allendorf, F. W. (1996). The One-Migrant-per-Generation Rule in Conservation and Management [Publisher: [Wiley, Society for Conservation Biology]]. *Conservation Biology*, *10*(6), 1509–1518. Retrieved June 25, 2023, from https://www.jstor.org/stable/2387022

Rose, H. (2001). The Commodification of Bioinformation: The Icelandic Health Sector Database.

Malo, N., Libiger, O., & Schork, N. J. (2008). Accommodating Linkage Disequilibrium in Genetic-Association Analyses via Ridge Regression. *The American Journal of Human Genetics*, *82*(2), 375–385. https://doi.org/10.1016/j.ajhg.2007.10.012

Wray, N. R., & Visscher, P. M. (2008). Estimating Trait Heritability [Cg_cat: Estimating Trait Heritability Cg_level: MED Cg_topic: Estimating Trait Heritability]. *Nature Education*, *1*(1), 29. Retrieved August 5, 2023, from https://www.nature.com/scitable/topicpage/estimating-trait-heritability-46889/

Metcalfe, K. A., Poll, A., Royer, R., Llacuachaqui, M., Tulman, A., Sun, P., & Narod, S. A. (2010). Screening for Founder Mutations in *BRCA1* and *BRCA2* in Unselected Jewish Women. *Journal of Clinical Oncology*, *28*(3), 387–391. https://doi.org/10.1200/JCO.2009.25.0712

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature genetics*, *42*(4), 355–360. https://doi.org/10.1038/ng.546

Thomas, A., Abel, H. J., Di, Y., Faye, L. L., Jin, J., Liu, J., Wu, Z., & Paterson, A. D. (2011). Effect of linkage disequilibrium on the identification of functional variants [_eprint: https://onlinelibrary.wiley.com/doi/p *Genetic Epidemiology*, *35*(S1), S115–S119. https://doi.org/10.1002/gepi.20660

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis [Publisher: Elsevier]. *The American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Speed, D., Hemani, G., Johnson, M. R., & Balding, D. J. (2012). Improved Heritability Estimation from Genome-wide SNPs [Publisher: Elsevier]. *The American Journal of Human Genetics*, *91*(6), 1011–1021. https://doi.org/10.1016/j.ajhg.2012.10.010

Alcala, N., Streit, D., Goudet, J., & Vuilleumier, S. (2013). Peak and Persistent Excess of Genetic Diversity Following an Abrupt Migration Increase. *Genetics*, *193*(3), 953–971. https://doi.org/10.1534/genetics.112.147785

Couch, F. J., Wang, X., McGuffog, L., Lee, A., Olswold, C., Kuchenbaecker, K. B., Soucy, P., Fredericksen, Z., Barrowdale, D., Dennis, J., Gaudet, M. M., Dicks, E., Kosel, M., Healey, S., Sinilnikova, O. M., Lee, A., Bacot, F., Vincent, D., Hogervorst, F. B. L., ... Cimba, O. B. O. (2013). Genome-Wide Association Study in BRCA1 Mutation Carriers Identifies Novel Loci Associated with Breast and Ovarian Cancer Risk [Publisher: Public Library of Science]. *PLOS Genetics*, *9*(3), e1003212. https://doi.org/10.1371/journal.pgen.1003212

Star, B., & Spencer, H. G. (2013). Effects of Genetic Drift and Gene Flow on the Selective Maintenance of Genetic Variation. *Genetics*, *194*(1), 235–244. https://doi.org/10.1534/genetics.113.149781

Carmi, S., Hui, K. Y., Kochav, E., Liu, X., Xue, J., Grady, F., Guha, S., Upadhyay, K., Ben-Avraham, D., Mukherjee, S., Bowen, B. M., Thomas, T., Vijai, J., Cruts, M., Froyen, G., Lambrechts, D., Plaisance, S., Van Broeckhoven, C., Van Damme, P., ... Pe'er, I. (2014). Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins [Number: 1 Publisher: Nature Publishing Group]. *Nature Communications*, *5*(1), 4835. https://doi.org/10.1038/ncomms5835

Hallgrimsson, B., Mio, W., Marcucio, R. S., & Spritz, R. (2014). Let's Face It—Complex Traits Are Just Not That Simple [Publisher: Public Library of Science]. *PLOS Genetics*, *10*(11), e1004724. https://doi.org/10.1371/journal.pgen.1004724

Hatzikotoulas, K., Gilly, A., & Zeggini, E. (2014). Using population isolates in genetic association studies. *Briefings in Functional Genomics*, *13*(5), 371–377. https://doi.org/10.1093/bfgp/elu022

Kemper, K. E., Saxton, S. J., Bolormaa, S., Hayes, B. J., & Goddard, M. E. (2014). Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*, *15*(1), 246. https://doi.org/10.1186/1471-2164-15-246

Kurki, M. I., Gaál, E. I., Kettunen, J., Lappalainen, T., Menelaou, A., Anttila, V., Hof, F. N. G. v. '., Fraunberg, M. v. u. z., Helisalmi, S., Hiltunen, M., Lehto, H., Laakso, A., Kivisaari, R., Koivisto, T., Ronkainen, A., Rinne, J., Kiemeney, L. A. L., Vermeulen, S. H., Kaunisto, M. A., … Jääskeläinen, J. E. (2014). High Risk Population Isolate Reveals Low Frequency Variants Predisposing to Intracranial Aneurysms [Publisher: Public Library of Science]. *PLOS Genetics*, *10*(1), e1004134. https://doi.org/10.1371/journal.pgen.1004134

Lohmueller, K. E. (2014). The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits [Publisher: Public Library of Science]. *PLOS Genetics*, *10*(5), e1004379. https://doi.org/10.1371/journal.pgen.1004379

Kelleher, J., Etheridge, A. M., & McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes [Publisher: Public Library of Science]. *PLOS Computational Biology*, *12*(5), e1004842. https://doi.org/10.1371/journal.pcbi.1004842

Lynch, M., Ackerman, M. S., Gout, J.-F., Long, H., Sung, W., Thomas, W. K., & Foster, P. L. (2016). Genetic drift, selection and the evolution of the mutation rate [Number: 11 Publisher: Nature Publishing Group]. *Nature Reviews Genetics*, *17*(11), 704–714. https://doi.org/10.1038/nrg.2016.104

Lipson, M., & Reich, D. (2017). A Working Model of the Deep Relationships of Diverse Modern Human Genetic Lineages Outside of Africa [Publisher: Oxford Academic]. *Molecular Biology and Evolution*, *34*(4), 889–902. https://doi.org/10.1093/molbev/msw293

Ogundijo, O. E., & Wang, X. (2017). Bayesian estimation of scaled mutation rate under the coalescent: A sequential Monte Carlo approach. *BMC Bioinformatics*, *18*(1), 541. https://doi.org/10.1186/s12859-017-1948-6

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data [Number: 7726 Publisher: Nature Publishing Group]. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Canela-Xandri, O., Rawlik, K., & Tenesa, A. (2018). An atlas of genetic associations in UK Biobank [Number: 11 Publisher: Nature Publishing Group]. *Nature Genetics*, *50*(11), 1593–1599. https://doi.org/10.1038/s41588-018-0248-z

Zeng, J., de Vlaming, R., Wu, Y., Robinson, M. R., Lloyd-Jones, L. R., Yengo, L., Yap, C. X., Xue, A., Sidorenko, J., McRae, A. F., Powell, J. E., Montgomery, G. W., Metspalu, A., Esko, T., Gibson, G., Wray, N. R., Visscher, P. M., & Yang, J. (2018). Signatures of negative selection in the genetic architecture of human complex traits [Number: 5 Publisher: Nature Publishing Group]. *Nature Genetics*, *50*(5), 746–753. https://doi.org/10.1038/s41588-018-0101-4

O'Connor, L. J., Schoech, A. P., Hormozdiari, F., Gazal, S., Patterson, N., & Price, A. L. (2019). Extreme Polygenicity of Complex Traits Is Explained by Negative Selection. *The American Journal of Human Genetics*, *105*(3), 456–476. https://doi.org/10.1016/j.ajhg.2019.07.003

Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell*, *177*(1), 26–31. https://doi.org/10.1016/j.cell.2019.02.048

Wein, T., & Dagan, T. (2019). The Effect of Population Bottleneck Size and Selective Regime on Genetic Diversity and Evolvability in Bacteria. *Genome Biology and Evolution*, *11*(11), 3283–3290. https://doi.org/10.1093/gbe/evz243

Zawada, K. J. A., Madin, J. S., Baird, A. H., Bridge, T. C. L., & Dornelas, M. (2019). Morphological traits can track coral reef responses to the Anthropocene [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1

2435.13358]. *Functional Ecology*, *33*(6), 962–975. https://doi.org/10.1111/1365-2435.13358

Barghi, N., Hermisson, J., & Schlötterer, C. (2020). Polygenic adaptation: A unifying framework to understand positive selection [Number: 12 Publisher: Nature Publishing Group]. *Nature Reviews Genetics*, *21*(12), 769–781. https://doi.org/10.1038/s41576-020-0250-z

Kelleher, J., & Lohse, K. (2020). Coalescent Simulation with msprime. In J. Y. Dutheil (Ed.), *Statistical Population Genomics* (pp. 191–230). Springer US. https://doi.org/10.1007/978-1-0716-0199-0_9

Ashraf, B., & Lawson, D. J. (2021). Genetic drift from the out-of-Africa bottleneck leads to biased estimation of genetic architecture and selection [Number: 10 Publisher: Nature Publishing Group]. *European Journal of Human Genetics*, *29*(10), 1549–1556. https://doi.org/10.1038/s41431-021-00873-2

Durvasula, A., & Lohmueller, K. E. (2021). Negative selection on complex traits limits phenotype prediction accuracy between populations. *American Journal of Human Genetics*, *108*(4), 620–631. https://doi.org/10.1016/j.ajhg.2021.02.013

Hivert, V., Sidorenko, J., Rohart, F., Goddard, M. E., Yang, J., Wray, N. R., Yengo, L., & Visscher, P. M. (2021). Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. *The American Journal of Human Genetics*, *108*(5), 786–798. https://doi.org/10.1016/j.ajhg.2021.02.014

Irving-Pease, E. K., Muktupavela, R., Dannemann, M., & Racimo, F. (2021). Quantitative Human Paleogenetics: What can Ancient DNA Tell us About Complex Trait Evolution? *Frontiers in Genetics*, *12*. Retrieved November 11, 2023, from https://www.frontiersin.org/articles/10.3389/fgene.2021.703541

Johnson, R., Burch, K. S., Hou, K., Paciuc, M., Pasaniuc, B., & Sankararaman, S. (2021). Estimation of regional polygenicity from GWAS provides insights into the genetic architecture of complex traits [Publisher: Public Library of Science]. *PLOS Computational Biology*, *17*(10), e1009483. https://doi.org/10.1371/journal.pcbi.1009483

Wieczynski, D. J., Singla, P., Doan, A., Singleton, A., Han, Z.-Y., Votzke, S., Yammine, A., & Gibert, J. P. (2021). Linking species traits and demography to explain complex temperature responses across levels of organization [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *118*(42), e2104863118. https://doi.org/10.1073/pnas.2104863118

Wang, M., & Huang, S. (2023). The collective effects of genetic variants and complex traits [Number: 4 Publisher: Nature Publishing Group]. *Journal of Human Genetics*, *68*(4), 255–262. https://doi.org/10.1038/s10038-022-01105-1

Supplementary info

The effective amount of genome simulated can be controlled by scaling the simulation parameters, though the scaling is not completely trivial. By changing the recombination rate, sequence length and mutation rate together, we can control the genome wide genomic architecture. For the purpose of this model it is equivalent to changing the polygenicity (i.e. the proportion of SNPs that have an effect on the trait) since in this model all SNPs that are simulated are treated as causal.

Following Kelleher et al., 2016 and Ogundijo and Wang, 2017, the expectation is that if the effective recombination rate, $sequence\ length\ \times recombination\ rate = constant$ and the relative mutation rate, $mutation\ rate/recombination\ rate = constant$, then the genomic architecture should be equivalent. This follows from the definitions of the scaled recombination rate $\rho = 4N_eL$ and the scaled mutation rate $\theta = 4N_e\mu$.

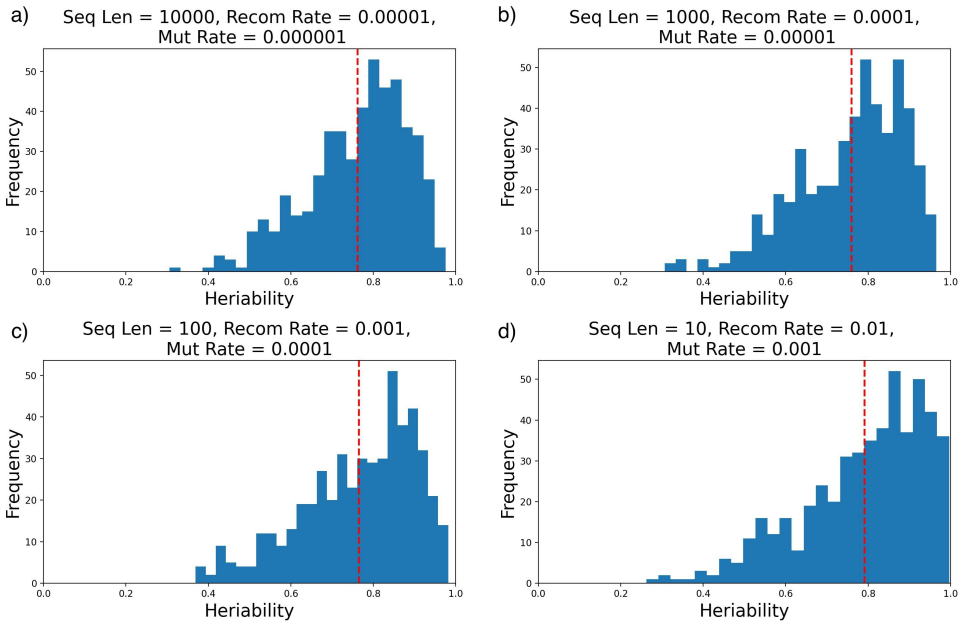

Figure S1: Heritability distribution across scaled sequence length, recombination rate and mutation rate that model the same genetic architecture. (a) sequence length = $1e4$, recombination rate $1e-5$, mutation rate = $1e-6$, (b) sequence length = $1e3$, recombination rate = $1e-4$, mutation rate = $1e-5$, (c) sequence length = $1e2$, mutation rate = $1e-3$, mutation rate = $1e-4$, (d) sequence length = 10, recombination rate = 0.01, mutation rate = 0.001.

Figure S1 shows 500 repeated experiments in four different combinations of heritability and sequence length with $L\rho = 0.1$ and $\mu/\rho = 0.1$. The distribution of heritability remains close to consistent in all four cases, though there may be increased volatility as the sequence length is decreased. However, Figure S2 shows the outcome of the experiments with different *recombination rate*, which despite sharing the same expected heritability, produces a different distribution. The variance in heritability increases from 0.0155 to 0.0384, as we have decreased the effective number of independent genome regions and hence increased autocorrelation along the genome. In this sense, both polygenicity and recombination rate are parameters of genomic architecture.

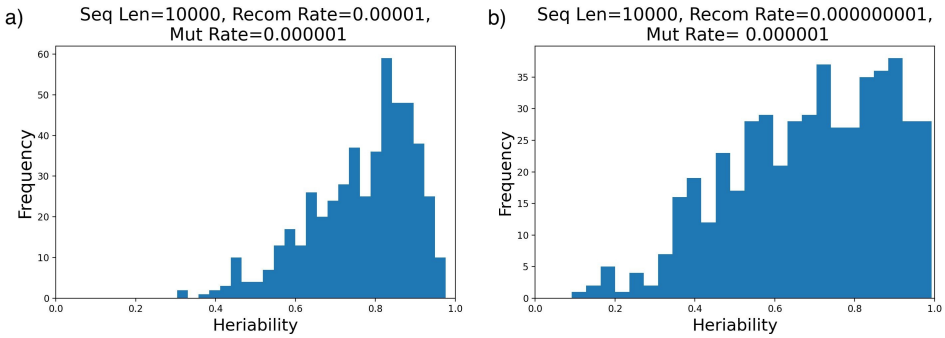

Figure S2: Heritability distribution in differing genetic architectures. (a) sequence length = $1e4$, recombination rate = $1e-5$, mutation rate = $1e-6$, (b) sequence length = $1e4$, recombination rate = $1e-8$, mutation rate = $1e-6$ (i.e. ρ is scaled by 0.001).