On the potential for GWAS with phenotypic population means and allele-frequency data (popGWAS)

Pfenninger Markus[1,2,3*]

[1] Dept. Molecular Ecology, Senckenberg Biodiversity and Climate Research Centre, Georg-Voigt-Str. 14-16, D-60325 - Frankfurt am Main, Germany
[2] LOEWE Centre for Translational Biodiversity Genomics, Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, D-60325 – Frankfurt am Main, Germany
[3] Institute for Molecular and Organismic Evolution, Johannes Gutenberg University, Johann-Joachim-Becher-Weg 7, D-55128 – Mainz, Germany

*Corresponding author
Correspondence: Markus.Pfenninger@senckenberg.de

Abstract

This study explores the potential of a novel genome-wide association study (GWAS) approach for identifying loci underlying quantitative polygenic traits in natural populations. Extensive population genetic forward simulations demonstrate that the approach is generally effective for oligogenic and moderately polygenic traits and relatively insensitive to low heritability, but applicability is limited for highly polygenic architectures and pronounced population structure. The required sample size is moderate with very good results being obtained already for a few dozen populations scored. The method performs well in predicting population means even with a moderate false positive rate. When combined with machine learning for feature selection, this rate can be further reduced. The data efficiency of the method, particularly when using pooled sequencing, makes GWAS studies more accessible for research in biodiversity genomics. Overall, this study highlights the promise of this popGWAS approach for dissecting the genetic basis of complex traits in natural populations.

Keywords: biodiversity population genomics, molecular trait basis

Introduction

A major goal as well as a major challenge in evolutionary biology is to understand how genes influence traits, i.e. the genotype-phenotype link, (Brandes et al., 2022; Uffelmann et al., 2021). The difficulties in achieving this goal are primarily due to the fact that the heritable variation of many, if not most, relevant phenotypes is determined by small contributions from many genetic loci (Sella & Barton, 2019). Such complex traits are usually influenced by a few dozen genes that are mechanistically directly involved in their expression, but often also by numerous, if not almost all, other genes as well as the environment in the widest sense (Boyle et al., 2017).

Genome-wide association studies (GWAS) are commonly used to link complex phenotypic traits to their genomic basis (Brandes et al., 2022; Visscher et al., 2012). However, given the complexity of causal mechanisms and the small effects of individual loci, often only a small fraction of the genetic variation underlying phenotypic variance is often identified, despite the considerable logistic effort in terms of the number of phenotyped and genotyped individuals (Brandes et al., 2022; Visscher et al., 2017). As a result, accurate predictions of phenotypes from genomic data are still quite limited and there is currently no other strategy than to keep increasing sample sizes (Brandes et al., 2022). This is a problem in the medical sciences (Shendure et al., 2019), but the greater challenge for science and society probably lies in addressing the global biodiversity crisis. It would be highly desirable to have affordable methods to accurately understand the genomic basis of relevant traits and predict (non-model) species responses to all aspects of global change (Bernatchez et al., 2023; Waldvogel et al., 2020).

GWAS with wild populations has been advocated for some time (Santure & Garant, 2018). However, despite recent progress in high throughput, automated phenotyping (Dunker et al., 2022; Tills et al., 2023; Xie & Yang, 2020), the advances of biodiversity genomics in obtaining high quality reference genomes for almost every species (Exposito-Alonso et al., 2020; Formenti et al., 2022) and the possibility to gain cost-effective genome-wide population data (Czech, Peng, Spence, Lang, Bellagio, Hildebrandt, Fritschi, Schwab, Rowan, & Weigel, 2022; Schlötterer et al., 2014), relatively few empirical studies are currently available. This gap between the possibilities and actual practical application in biodiversity conservation (Heuertz et al., 2023; Hogg, 2023) is probably as much due to the still existing logistic and financial challenges as to a lack of data- and resource-efficient methods.

Here, I explore the potential of a new GWAS approach using phenotypic population means and genome-wide allele-frequency data. The rationale behind the approach is straightforward. If a quantitative polygenic trait has an additive genetic component, an individual's phenotypic trait value should at least roughly correlate with the number of trait-increasing alleles at the underlying loci (Uffelmann et al., 2021). Consequently, it was theoretically expected (Orr, 1998; Pritchard & Di Rienzo, 2010) and empirically shown (Turchin et al., 2012) that trait-increasing alleles will tend to have greater frequencies in the population with higher mean trait values, compared to the population with a lower trait mean. When examining populations with a range of different phenotypic trait means, we may therefore expect that the allele frequencies at the trait-affecting loci show a linear or at least steady relation with the observed trait means (Barton, 1999). I hypothesise here that this predicted relation can be exploited to distinguish potentially causal loci (and the linked variation) from loci not associated with the focal trait. In case of a successful evaluation, the major advantages of the proposed approach would be the reduced sequencing effort by the possibility to use pooled population samples (PoolSeq) and the opportunity to use bulk

phenotyping (e.g. by satellite imaging, flow-cytometry, etc.) on traits for which individual phenotyping is difficult or tedious.

The most important assumption for the approach is obviously that observed population differences in the focal trait means have at least partially a genetic basis. Since the environment has usually an effect on the phenotype (Sella & Barton, 2019), total phenotypic variance should be adjusted for known fixed environmental effects, because this increases the fraction of variance due to genetic factors (Visscher et al. 2008). Predicting additive genetic values with even higher accuracy can be achieved by taking into account GxE interactions through repeated phenotypic measurements of the same individuals under different environmental conditions, e.g. by time series (Visscher et al. 2008). I assumed therefore that environmental influence on the phenotypic trait variance among populations has been statistically removed as much as possible. Similarly important is the assumption that the genetic variance of the focal quantitative trait can be adequately described by an additive model. Both empirical and theoretical evidence suggests that this is indeed the case for most complex traits (Hill et al., 2008). Even though epistatic interactions are wide spread (Mackay, 2014), Sella and Barton (Sella & Barton, 2019) argue that the marginal allelic effects on quantitative traits are well approximated by a simple additive model.

The aims of this study were i) to understand whether and under which circumstances the hypothesised pattern of a linear relation between the population allele frequencies at causal loci and the phenotypic population means of the respective trait emerges, ii) to evaluate the influence of population genetic parameters of typical natural systems and the experimental design on the likelihood of identifying causal loci underlying an additive quantitative trait, in particular to elucidate the limits of the approach with regard to genetic architecture and population structure, iii) to explore the possibilities for statistical genomic prediction of phenotypic population means from the allele frequencies at the identified loci, and iv) to evaluate the statistical power of the method for a realistic range of effective genome sizes. I used individual-based population genomic forward simulations and machine learning approaches (minimum entropy feature selection) for prediction and utilised an information theory-based framework for evaluation of the proposed method.

Material and methods

Expectation of a positive correlation between quantitative trait loci allele frequencies and phenotypic population means.

Consider a biallelic, codominant system for the additively heritable component of a quantitative trait with n loci contributing to the trait. In this system, all loci contribute equally to the phenotypic trait, with one allele per locus making a greater contribution than the other. The phenotypic trait value x of an individual can then be determined by simply adding up the number of trait-increasing alleles (g with values of 0, 1 or 2) over all n quantitative trait loci (QTL) and multiplying this sum with a scaling constant k:

$$x = k \times (g_1 + g_2 + \cdots + g_n) \quad (1)$$

When adding more individuals, the phenotypic population trait mean is defined as the mean of the row sums:

123 (2)

	QTL_1	⋯	⋯	QTL_n	individual trait value
Ind_1	g_{11}	g_{12}	⋯	g_n	$x_1 = k \times \sum_{l=1}^{n} g_{1l}$
Ind_2	g_{21}	⋯	⋯	⋯	⋯
⋯	⋯	⋯	⋯	⋯	⋯
Ind_m	g_{m1}	⋯	⋯	g_{mn}	$x_m = k \times \sum_{l=1}^{n} g_{1l}$
					population trait mean
					$\bar{x} = \frac{\sum_{i=1}^{m} x_i}{m}$

124 The columns of this matrix can be used to calculate the population allele frequency (AF)
125 of the trait increasing allele for each QTL.

126 (3)

	QTL_1	⋯	⋯	QTL_n		$individual\ trait\ value$
Ind_1	g_{11}	g_{12}	⋯	g_n		$x_1 = k \times \sum_{l=1}^{n} g_{1l}$
Ind_2	g_{21}	⋯	⋯	⋯		⋯
⋯	⋯	⋯	⋯	⋯		⋯
Ind_m	g_{m1}	⋯	⋯	g_{mn}		$x_m = k \times \sum_{l=1}^{n} g_{1l}$
					$mean\ QTL\ AF$	$population\ trait\ mean$
AF	$AF_1 = \frac{\sum_{i=1}^{m} g_i}{2m}$	⋯	⋯	$AF_1 = \frac{\sum_{i=1}^{m} g_i}{2m}$	$\overline{AF} = \frac{\sum_{i=1}^{n} AF_i}{n}$	$\sim \quad \bar{x} = \frac{\sum_{i=1}^{m} x_i}{m}$

127 The mean population allele frequency at QTL loci is thus directly proportional to the
128 phenotypic population trait mean. This relationship remains unchanged even if the
129 individual locus contributions are not identical, with some loci contributing more or less to
130 the phenotypic trait value. In this case, a scaling vector is required to weigh the individual
131 locus contributions to individual trait values, and those of the AFs to the population trait
132 mean. Since the AFs are by definition bounded by zero and one, the population trait mean is
133 minimal when the allele frequencies of the trait-increasing allele at all QTL are zero and
134 maximal when all QTL AFs are one. This proportionality links the individual genotypes and
135 the AFs at the QTL linearly with the population trait mean.
136 If we extend this to a set of populations and order them with decreasing phenotypic
137 population means, we can be sure that the mean QTL AFs of the populations will also be
138 ordered in decreasing sequence:

139 (4)

	Pop_1	Pop_2	⋯	Pop_n
population trait mean	$\bar{x}_1 >$	$\bar{x}_2 >$	⋯	$\bar{x}_n$
	$\sim$	$\sim$	⋯	$\sim$
$mean\ QTL\ AF$	$\overline{AF_1} >$	$\overline{AF_2} >$	⋯	$\overline{AF_n}$
QTL_1	AF_{11}	AF_{21}	⋯	AF_{11}
QTL_2	AF_{12}	AF_{22}	⋯	⋯
⋯	⋯	⋯	⋯	⋯
QTL_m	AF_{1m}	⋯	⋯	AF_{nm}

140 The answer to whether the allele frequencies in every row i.e. at every contributing
141 locus can be used to predict the population trait mean depends on whether the expected
142 covariance between these two vectors is positive:

143 (5) $E[\mathrm{cov}(QTL_m, \text{population trait mean})] > 0$

144 where

$$\text{cov}[\text{QTL}_m, \text{population trait mean}] = E[\sum_{i=1}^{n}(AF_{im} - E[AF_m]) * (\bar{x}_i - E[\bar{X}])] \quad (6)$$

with $\bar{X}$ representing the grand mean over all populations. As all elements in the QTL matrix are positive, they inherently tend to contribute positively to their column means. Therefore, AF larger than the overall AF mean at his locus tend to be on the left side of the population closest to the overall phenotypic mean in the ordered matrix above. Conversely, AF smaller than the locus AF mean are rather on the right. This leads intuitively to a positive expected covariance between each row and the column mean, in particular if the number of populations becomes large. Conversely, the AF at (unlinked) loci not contributing to the phenotypic population trait mean have an expectation of zero.

I tested these general expectations and the effect of different scaling vectors for the effect size distribution of QTL with a first set of simulations. I generated a matrix of size *n x m* populated with random AF between zero and 1. To avoid stochastic effects due to sample size, the number of populations *n* was fixed at 10,000. The number of QTL *m* was varied from oligogenic to highly polygenic (10, 20, 50, 100, 200, 500, 1000, 2000, 5000). Three different distributions of loci effects were tested, i) a flat distribution with all loci contributing equally, ii) a mildly decreasing exponential function and iii) a steeply decreasing exponential function with few loci contributing much and many very little (Supplemental Figure 1).

Each of the *m* columns was used to calculate the phenotypic population mean of the respective population by adding up the AF multiplied with the respective locus weight. The resulting *n* phenotypic population means were then correlated to the *n* AF of each of the *m* loci and the resulting *m* Pearson correlation coefficients (i.e. the standardized covariance) recorded. From these, mean and standard deviation were calculated and tested, whether they conform to a normal distribution (scipy.stats.normaltest). Furthermore, a second matrix of identical size was populated with random AF, and the correlation of these non-contributing loci to the population means derived from the QTL matrix was computed. The simulations were repeated 10 times in every possible parameter combination and the results averaged (Supplemental Script 1).

Individual based Wright-Fisher forward model

A Wright-Fisher individual forward genetic simulation model was used to investigate the potential of a genome-wide association study based on the means of a population trait and population allele frequency data. In the simulation, all loci were assumed to be unlinked, thus representing haplotypes in LD rather than single SNPs. (Visscher et al., 2017). For each simulation run, the initial allele frequencies for all loci in the total population were randomly drawn from a range of 0.1 to 0.9. To generate a hermaphroditic and diploid individual, two alleles were randomly drawn with a probability based on their frequency at the respective locus, and the resulting genotype at this locus was recorded. This process was repeated for all loci. As a result, each individual was represented by a vector of biallelic genotypes ($AA = 0$, $Aa; aA = 1$, $aa = 2$). To model a quantitative, fully additive trait, a variable number of loci were assigned as quantitative trait loci (QTL). In addition, a much larger number of neutral loci was modelled.

Genetic architecture of the quantitative trait

The continuous trait value was measured in arbitrary units. The allele (A) at each QTL had no effect on the individual's trait value, resulting in a completely homozygous A

individual at the QTL having a phenotypic trait value of 0. The alternative allele (*a*) added a locus-specific value to the trait. Two distributional extremes have been considered for the allelic effects on the trait value: i) a uniform distribution where each locus contributes 0.5 units to the trait. An individual that is completely homozygous for the alternative allele *a* therefore had a trait value equal to the number of QTL, ii) an exponential distribution with few loci having large effects and many having very small effects, scaled such that the maximum possible trait value was also equal to the number of QTL (see Supplementary Figure 2A). To model the effect of phenotyping errors, unaccounted environmental influence (*i.e.* phenotypic plasticity), and/or the unspecific contribution of the genomic background to the trait, a random value drawn from a Gaussian distribution with a mean of zero and selectable standard deviation between 0.1 and 3 could be added to the genetically determined phenotype value of each individual. The phenotypic value of each individual's trait was determined by summing the allelic effects of all genotypes at all QTL loci plus the random value and the result recorded.

Reproduction and selection

Subpopulations in each run were created from the same initially drawn random allele frequency array, mimicking a common descent. Due to sampling variance, the realised allele frequencies and thus the mean subpopulation trait value differed from the initial frequencies of the total (ancestral) population. A subpopulation always comprised 500 adult individuals.

Each subpopulation was reproduced at least once to obtain a genotype distribution in Hardy-Weinberg equilibrium. For reproduction, two random individuals were chosen with replacement from the adult population. The genotype of an offspring individual at a locus was determined by randomly choosing one of the two alleles from each designated parent at this locus. Each parent fostered n_juv offspring; therefore, 2 x n_juv were produced in each mating. After reproduction, the parental generation was discarded to prevent overlapping generations. Each generation had N/2 matings, resulting in an offspring population of N*n_juv individuals.

Because the offspring population was much larger than the the size of the adult population, it was necessary to reduce it. This was achieved by a combination of 'hard' natural selection and random mortality. An individual's survival to the adult stage was determined by the absolute deviation of its phenotypic trait value from a pre-specified selective trait optimum for the respective subpopulation. This selective trait optimum for a subpopulation was determined by adding a random value taken from a Gaussian distribution with a mean of zero and a standard deviation of 2.5 to the initial population mean. An individual's survival probability was determined by an exponential decline function with strength *s* (the exponent of the function, see Supplementary Figure 2B). Individuals were randomly selected one by one from the offspring population and their survival probability calculated. A respectively biased coin was then tossed to determine their fate. This process was repeated until the adult population size was reached, and any remaining offspring individuals were indiscriminately discarded. If the phenotypic mean of the subpopulation was close to or at the selective optimum (see below), this process resulted in stabilising selection. If the population was away from the optimum, rapid directed selection towards the optimum was observed, depending on the strength of selection. For the assessment of the effect of population structure, the subpopulations could evolve in complete isolation from each other for a predetermined number of generations

(2-50). This introduced random genetic drift among the populations at both QTL and neutral loci. Both drift and selection towards different trait optima led to variation in population trait means among the subpopulations.

Simulation scenarios

I considered scenarios were subpopulations with quantitative phenotypic population differences in mean for the trait in question were screened from a larger total population. Although the population trait mean differences in the simulation of this scenario were created by drift and local adaptation, any other source of heritable phenotypic population differentiation, such as maladaptation, introgression, or e.g. in the case of managed species, human choice, may also be the reason for differentiation in population means. The range of phenotypic variation among the subpopulations was not predetermined, but an emergend feature of the simulation parameters.

After evolving the subpopulations for the desired number of generations, phenotypic trait means, and genome-wide allele frequencies were recorded. While the phenotypic means for each subpopulation was calculated over all individuals, the allele frequencies were estimated in a PoolSeq (Kofler et al., 2011) like fashion from subsamples of 50 individuals. The range of phenotypic trait means of the population sample was recorded. Trait heritability was determined in the last generation by regressing the phenotypic values of the offspring against the mean of their respective parents (Lynch & Walsh, 1998). As measure for population subdivision due to drift, F_{ST} among all subpopulations was calculated from the variance of the true allele frequencies (Wright, 1949).

Population GWAS

Assuming a linear relation between the phenotypic (sub)population means, and the population allele frequencies of the causal loci on the other, I calculated an ordinary linear regression between these two variables for all loci in the genome. I used the resulting $-\log_{10}$ p value as measure of regression fit and effect size. I recorded the number of true positive loci (TPL) among the loci beyond a predefined outlier threshold. As GWAS performance measures, the true positive rate (TPR = recall, sensitivity, discovered proportion of all QTL), positive predictive value (PPV = precision, proportion of TPL among outliers considered) and false discovery rate (FDR = proportion of false positive loci among outliers considered, type I error) were calculated.

Influence of natural system and experimental design factors

In a first set of simulations, I explored the influence of factors inherent to the natural system and the experimental design on population GWAS performance. As factors of the natural system, I assumed characteristics that are beyond control of the researcher, such as heritability of the trait and its genetic architecture (number of QTL, distribution of allele trait contribution). While the degree of population differentiation and range of phenotypic differentiation are also inherent to the organism studied, the choice of samples may allow a certain control over these parameters. The number of subpopulations screened is clearly a study design decision (Table 1).

Table 1. Simulation parameters, their abbreviations, values used in simulations, their biological meaning and whether the parameter is a feature of the natural system under scrutiny or under the control of the researcher.

Parameter	Abbreviation	Values in the simulation	Biological meaning	Degree of knowledge in natural systems/under the control of study design
Number of subpopulations scored for phenotypic population means and genome wide allele frequencies	n_pop	12, 24 36, 48, 60	-	Full control
Number of quantitative trait loci contributing to the focal trait	n_qtl	30, 50, 70, 110, 500	Genetic architecture of the trait	*A priori* unknown
Distribution of allelic effects on the focal trait	allelic_contr	Flat, exponential	Genetic architecture of the trait	*A priori* unknown
Standard deviation of random phenotypic variation added to individuals	pheno_plast	0.1, 1, 2, 3	Heritability of the trait	*A priori* unknown
Number of generations of independent evolution of the subpopulations	gen	2, 5, 10, 30, 50	Population structure	Partial control

A genetic trait architecture of 30, 50, 70, 110 and 500 loci, flat and exponential allelic effect distributions, as well as phenotypic plasticity coefficients of 0.1, 1, 2 and 3 were applied. Selection strength was fixed at 0.5 (Suppl. Fig. 1B). Simulations were run for 2, 5, 10, 30 and 50 generations among 12, 24, 36, 48 and 60 subpopulations of 500 individuals each. For this set of simulations, 1000 neutral loci and a fixed outlier threshold (upper 5% quantile, either 21 or 22 outlier loci, respectively) were applied. Each possible parameter combination was run in five replicates, resulting in 5000 simulation runs.

The effect of each parameter on PPV was assessed with ANOVA over all simulations, grouped after the respective parameter classes. The relative influence of the number of populations, QTL loci, distribution of allelic contributions, trait heritability, phenotypic range and population subdivision on the proportion of TPL among the outlier loci was determined with a General Linearized Model (GLM).

Genomic prediction and validation

The loci identified by GWAS were used to devise a statistical genomic prediction model to obtain a score that uses observed allele frequencies at the identified loci to predict the mean population phenotype of unmeasured populations. To remove remaining uninformative or redundant loci, I applied feature selection, which is particularly suitable for bioinformatic data sets that contain many features but comparatively few data points. The minimum entropy feature selection (MEFS) technique uses mutual information to measure the dependence between each feature and the target variable. For a given number of features (k), the data set of the allele frequencies at selected outlier loci and the respective phenotypic population means was repeatedly randomly divided in training (80%) and test set (20%), a multiple regression model fitted and the r^2-fit of the test sets to the predicted phenotypes recorded. The best model for the current k was recorded and the process repeated for all k in a range between 2 and the number of selected loci – 1. Finally, the best model (i.e. highest r^2) among all k was chosen as best prediction model. MEFS was implemented with the Python module scikit-learn 1.3.2 (Pedregosa et al., 2011)

The performance of the selected best prediction model for each run was tested with independent data. Ten additional populations were created under the same parameters as the initial set of populations and their mean population phenotypes calculated as described above. Then the allele frequencies at the predictive loci as identified by the best prediction model were extracted and phenotypic prediction scores according to the best prediction model calculated. The performance of the statistical genomic prediction was then evaluated

by calculating the Pearson correlation coefficient *r* between the observed mean population phenotypes and the phenotypic prediction scores for the ten validation populations (Supplemental Script 2).

Method performance with realistic genome sizes

Whether and which proportion of TPL, i.e. causal loci can be expected to be reliably identified with the proposed method depends crucially on the total number of loci screened as this number determines the length and size of the distributional tail of random associations of neutral loci with the mean population phenotypes. The number of effectively independently evolving loci in a population depends on genome size, effective population size (including all factors that affect it locally and globally) and LD structure (Chakraborty, 1981; Taylor & Higgs, 2000). There are hardly any empirical estimates in the literature, but dividing typical genome sizes by typical mean genome-wide LD ranges suggested that a few tens of thousands to a few hundreds of thousands of independent loci per genome is a realistic range for a large number of taxa (see Supplemental Table 1). I have therefore considered 1,000, 5,000, 10,000, 30,000, 50,000 and 100,000 independent neutral loci for samples of 12, 24, 36, 48 and 60 populations with a restricted set of parameters (number of QTL and allelic contribution). As the true number of QTL underlying a trait is rarely *a priori* known, I considered 10, 30, 50, 70 and 110 QTL loci in this analysis. I therefore recorded the number of TPL found in sets of loci with the absolutely highest 10, 30, 50 and 100 - $\log_{10}$p values, as well as outlier proportions of 0.0001, 0.001, 0.01, 0.02, 0.05 and 0.1 of the total number of loci in the respective simulation. As above, all simulations were run in all possible parameter combinations with five replicates each (Supplemental Script 3).

I analysed the performance of the method in an Area Under the Curve – Receiver Operator Curve (AUC-ROC) and – Precision, Recall (AUC-PR) framework as suggested by Lotterhos et al. (Lotterhos et al., 2022). For each combination of effective genome size and number of population scored, mean TPR, PPV and FDR were calculated over all replicates and parameter combinations for the respective set of simulations. The maximum F1 score (Rijsbergen, 1979) was used in addition to identify the optimal number of outliers to select.

All simulations were implemented in Python 3.11.7 (Van Rossum & Drake, 2009) and run under pypy 3.10 (Team, 2019), the respective scripts can be found in the Supplementary Material (Scripts 1-3). General statistical tests were performed with R (R Core Team, 2013).

Results

Allele frequencies at QTL loci co-vary positively with the population trait mean

The mean correlation coefficient between all QTL and the respective phenotypic population means was positive in every parameter combination and in every single simulation (Table 2).

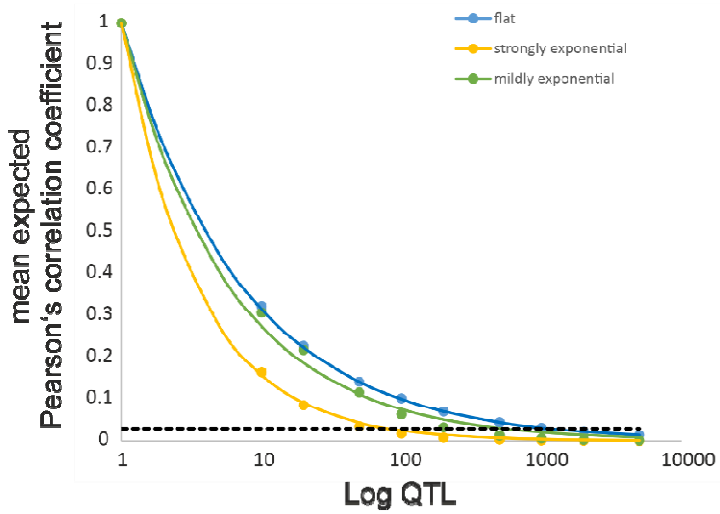
Table 2. Expected mean Pearson's correlation coefficients between QTL AF and phenotypic population means for three different locus contribution distributions and varying number of QTL.

n_QTL	Flat	Mildly exponential	Strongly exponential
1	1	1	1
10	0.322	0.307	0.165
20	0.227	0.218	0.085
50	0.142	0.116	0.035
100	0.101	0.064	0.017

200	0.070	0.033	0.008
500	0.045	0.013	0.003
1000	0.031	0.006	0.001
2000	0.023	0.003	0.001
5000	0.014	0.001	0.000

The expected mean correlation coefficient decreased with increasing number of contributing QTL (Figure 1). This decay was best described by a negative exponential function of the form *number of QTL*$^{-1/x}$ with x ranging from 1.26 in case of the strongly unbalanced locus contributions to 2 for the flat distribution.

Figure 1. Plot of the expected correlation coefficient between individual QTL loci and the population trait mean in dependence of the number of QTL loci.



The distribution of the correlation coefficients did not deviate from a normal distribution for the flat locus contribution distribution, while it did for all other parameters. The correlation coefficients for the non-contributing loci had an expectation of zero and a mean standard deviation of 0.01, regardless of the number of loci.

Influence of simulation parameters on parameters of the simulated populations

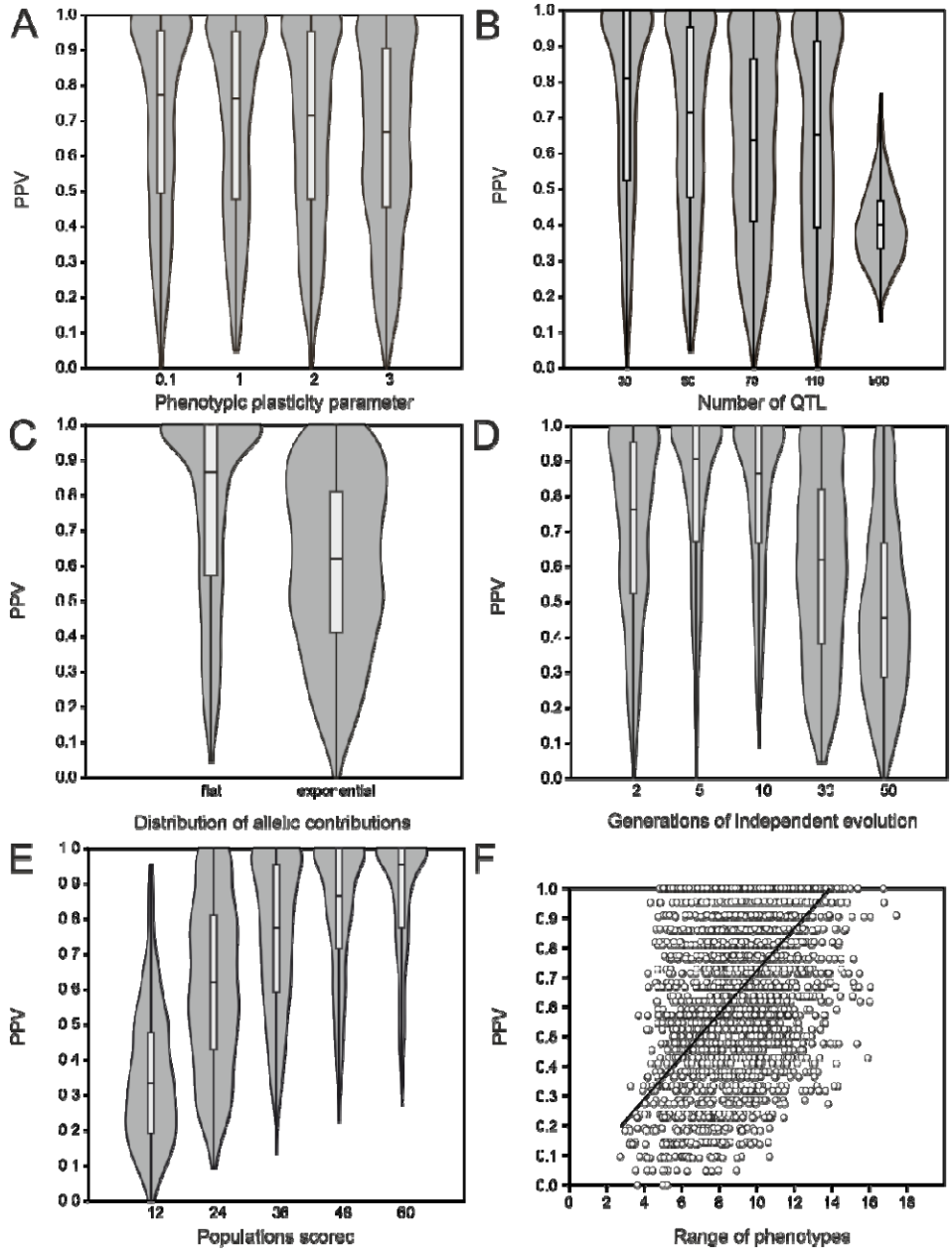
In the second set of simulations with 1000 neutral loci and an outlier threshold of the most extreme 5% -$\log_{10}$p values, the permutation of all parameters with five replicates yielded 4984 completed independent simulation runs. The 16 missing to the expected 5000 runs were due to one or more subpopulations going extinct during the simulation. In total, more than 30 billion individuals were simulated.

The number of independently evolving generations strongly influenced the population structure (r^2 = 0.996). F_{ST} estimates increased on average by 0.0018 per additional generation, with large variation. Resulting F_{ST} values ranged between 0.012 and 0.111 (Supplemental Figure 3A). Heritability of the trait depended strongly on the plasticity parameter (r^2 = 0.942, Supplemental Figure 3B). It decreased on average by 0.24 per unit standard deviation, with variations of up to 0.05 even among runs with identical parameters. Trait heritability estimates ranged from 0.16 to 1.02.

375 Factors influencing the proportion of detected true positive loci among outliers

376 Over all simulations in the first set with 1000 neutral loci, on average about 14.1 (mean
377 proportion 0.62) true positive loci (TPL) were among the highest 5% outliers. The TPL
378 values ranged between none (0) and 23 (1.0); the 25 percentile was 10 (0.38), the 75
379 percentile 20 (0.90). This exceeded in >95% of cases random expectations, when excluding
380 the highly polygenic case (n_qtl = 500), this proportion rose to more than 99%.

381 Figure 2. Effect of simulation parameters and emergent features on the proportion of
382 identified true positive loci. A) Phenotypical plasticity parameter as a proxy for
383 heritability. B) Number of QTL. C) Distribution of allelic contributions to phenotypic
384 trait. D) Generations of independent evolution as proxy for population structure. E)
385 Number of populations scored for population phenotypic means and allele
386 frequencies. F) Range of population phenotypic means as an emergent feature.



387

The phenotypic plasticity parameter had a significant effect on PPV ($F = 8.37$, $p = 1.51 \times 10^{-5}$), however, as the data plot already indicated (Figure 2A), this was due to the drop of the mean in the class with the lowest heritability only (from about 0.64 in the other classes to 0.59), as indicated by Dunn's post-hoc test (z statistic > 2.8 and p below 4.8×10^{-4} in all comparisons with class 3). The number of QTL showed a systematic effect on mean PPV (ANOVA $F = 226$, $p = 6.42 \times 10^{-177}$, Figure 2B). This was mainly due to the highly polygenic class; while the mean PPV for all different QTL numbers up to 110 was above 62.5%, it was as low as 39.8% for 500 QTL (z statistic > 17 and p below 2.1×10^{-69} in all comparisons). The distribution of allelic effects on the trait showed a moderate but highly significant effect on the mean PPV (mean equal contribution = 0.67, mean exponential = 0.57, $F = 154{,}5$ $p = 6.58 \times 10^{-35}$, Figure 2C). The relation of mean proportion of detected TPL and population structure was non-linear. Both very weak (2 generations) and strong population (30+ generations) structure led to a relatively lower proportion of TPL (0.62 and 0.46, respectively, Figure 2D), while for intermediate values TPL proportions of 0.72 (5 generations) and 0.70 (10 generations) were observed. The by far strongest effect on proportion of TPL among the selected loci had the number of populations screened ($F = 514$, $p = 0$). The values ranged from a mean PPV of 0.35 (s.d. = 0.18) with 12 populations to over 0.78 (s.d. = 0.23) with 60 populations. Given the chosen threshold, a diminishing return was observed above 36 populations sampled (Figure 2E). The phenotypic range in a simulation run had a moderate ($r = 0.42$, $p = 5.36 \times 10^{-203}$), yet significantly positive effect on detection of TPL. The realised range of population trait means in the simulations covered on average 15.4% (range = 0.1-56%) of the possible range. An increase of one unit in range increased the proportion of TPL by 0.07 (Figure 2F).

When jointly considering the effect of all parameters on PPV in a GLM, it turned out that all had a significant effect (Table 2). Their relative influence increased from F_{ST} ($r^2 = 0.008$) over distribution of allelic trait contribution ($r^2 = 0.013$), heritability ($r^2 = 0.024$), the number of populations ($r^2 = 0.100$), phenotypic range ($r^2 = 0.130$) to the number of QTLs, that had by far the greatest influence ($r^2 = 0.343$). In total, the parameters explained 61.8% of variance.

Table 2. Generalised Linear Model of factors influencing the proportion of TPL among outliers (PPV) in simulations.

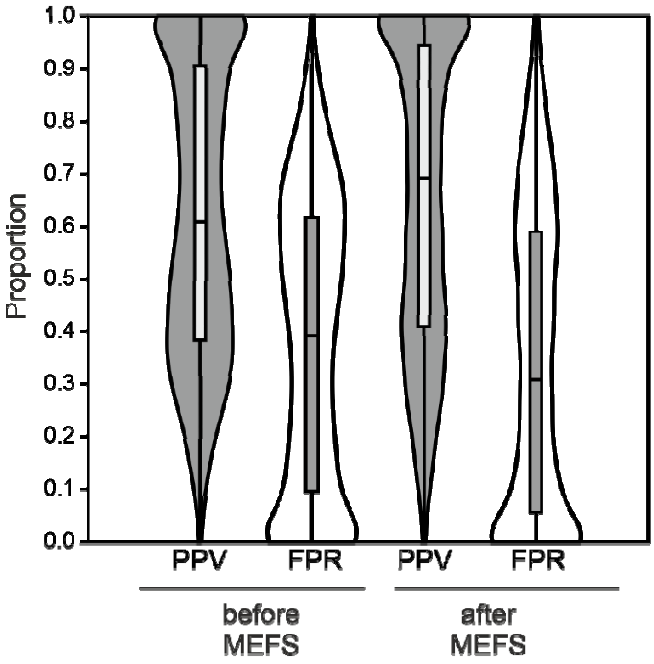
Factor	Coefficient	Std.err.	*t*	*p*	r^2
Constant	0.243	0.022	10,859	3.86E-23	
F_{ST}	-2.954	0.150	-19,749	2.21E-79	0.008
allelic_contr	0.153	0.009	16,122	6.46E-53	0.013
heritability	0.042	0.019	21,964	2.81E-02	0.024
n_pop	0.006	0.000	22,089	8.01E-99	0.100
range_pheno	0.037	0.002	19,492	2.33E-77	0.130
n_qtl	-0.001	0.304	-46,314	0.00E+00	0.343

Minimum Entropy Feature Selection and statistical phenotype prediction

Minimum Entropy Feature Selection (MEFS) removed on average 8.72 (range = 2-14, s.d. = 4.14) loci, corresponding to a proportion of 0.38 (s.d. = 0.19) from the statistically chosen initial outlier set. The procedure removed on average a larger proportion of FP than TPL (mean difference 0.14, $t = -19.9$, $p = 6.7 \times 10^{-79}$). This increased the proportion of TPL

in the final prediction set on average by 0.05 (range = -0.23-0.48, s.d. = 0.09) to a mean of 0.66 (range = 0-1, s.d. = 0.29, Figure 2).

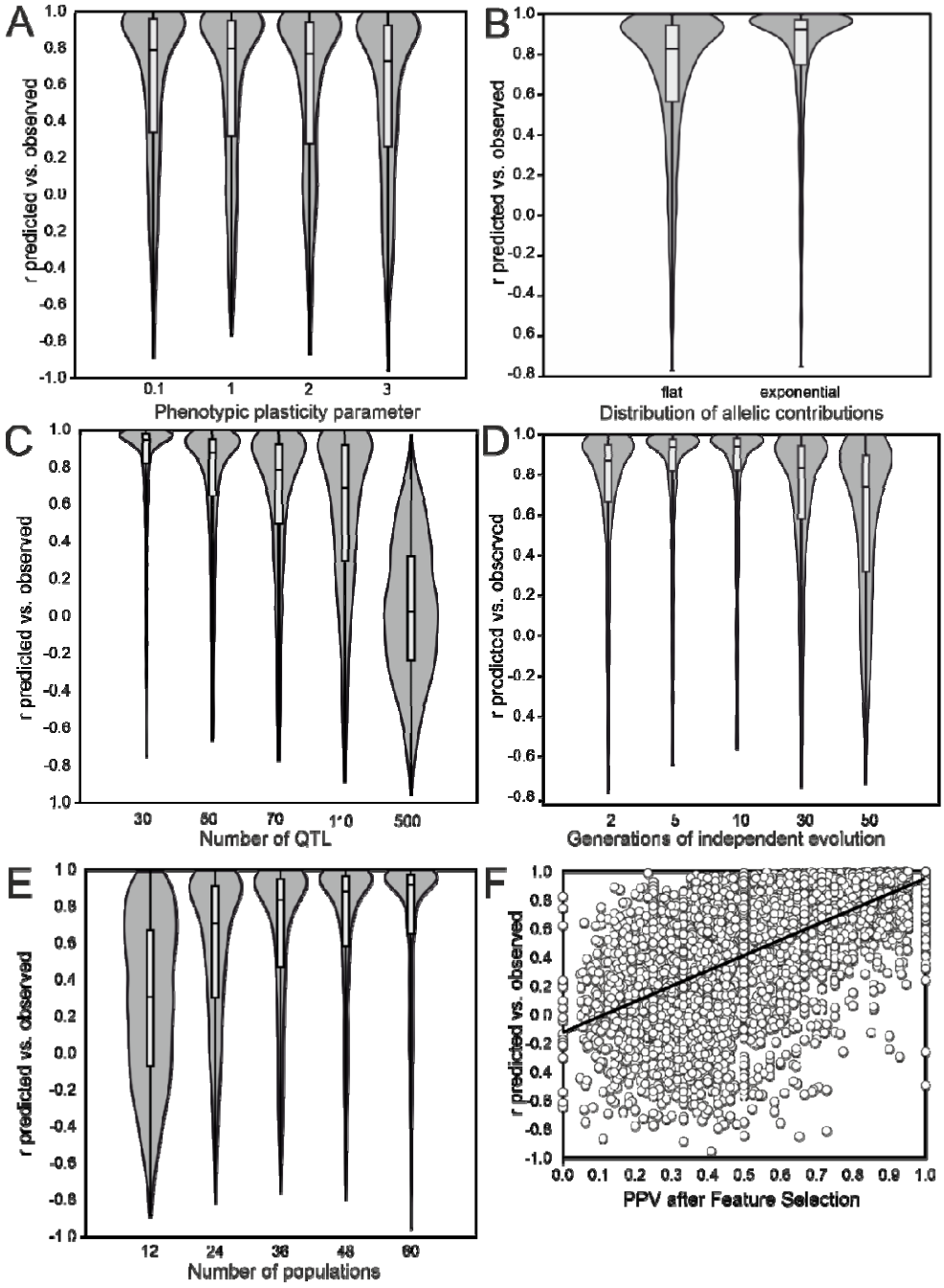
Figure 3. Effect of Minimum Entropy Feature Selection (MEFS) on the proportion of TPL and FP in the selected set.



The predictive accuracy of the SNP loci sets selected by MEFS was on average r = 0.58 (s.d. = 0.44). It ranged from -0.95 to 1.0. The distribution was highly skewed with 75% being higher than 0.30, the median was found at 0.76 and still 25% being higher than 0.94 (Supplemental Figure 3).

The accuracy of mean population phenotype prediction depended linearly on the number of TPL in the prediction set (r^2 = 0.28, p = 0), with any additional TPL increasing the correlation coefficient by 0.05 (Supplemental Figure 4A). Inversely, the accuracy of prediction decreased with a rising number of FP, but even with a considerable number of FP in the prediction set, accurate prediction was possible in a large number of cases (Supplemental Figure 4B). Overall, the prediction accuracy increased with increasing proportions of TPL among the prediction set, although even 100% TPL in the prediction set did not guarantee a highly accurate prediction (r > 0.8) in all cases (Supplemental Figure 4C).

Figure 4. Influence of simulation parameters on the accuracy of statistical population mean phenotype prediction. A) Phenotypic plasticity parameter as proxy for heritability. B) Distribution of allelic trait contributions. C) Number of trait-underlying QTLs. D) Generation of independent evolution as proxy for population structure. E) Number of populations scored. F) Proportion of TPL in the prediction loci set after MEFS.



As the prediction accuracy depended on the proportion of selected TPL, their relation to the individual simulation parameters was very similar to the results described in the previous section (Figure 4A-F). The number of populations screened was the most important factor. With 36 or more populations screened, 97.8% of simulations showed a prediction accuracy of 0.8 or better, independent of the other simulation parameters applied. In a GLM with all factors simultaneously considered, the proportion of TPL selected had the largest influence on prediction accuracy ($r^2 = 0.48$), followed by the number of QTL (0.34), the range of mean phenotypes (0.13), the number of populations screened (0.10). Heritability, distribution of allelic contributions and F_{ST} had only a minor influence on the prediction accuracy (<= 0.02, Table 3).

Table 3. Generalised Linear Model of factors influencing the accuracy of statistical phenotypic population mean prediction.

Factor	Coefficient	Std.err.	*t*	*p*	r^2
Constant	-0.0304	0.02	-14183	0.1562	
F_{ST}	-0.1981	0.16	-12778	0.2014	0.01
allelic_contr	0.1611	0.01	19	0.0000	0.01
heritability	0.0216	0.02	1265	0.2059	0.02
n_pop	0.0014	0.00	45782	0.0482	0.10
range_pheno	0.0104	0.00	56485	0.0002	0.13
n_qtl	-0.0009	0.03	-30692	0.0000	0.34
prop_TPL_FS	0.7761	0.02	34676	0.0000	0.48

Method performance with realistic effective genome sizes

The values for AUC-ROC ranged between 0.067 and 0.833, for AUC-PR between 0.013 and 0.730. There was an interaction between the effective genome size and number of populations scored. According to both AUC measures, the method performed best, when the number of populations scored was high and the genome small (Figure 5). An at least satisfactory (> 0.66 for AUC-ROC and > 0.53 for AUC-PR) overall performance was observed for 24 populations for the smallest genomes considered (1,000), for 36 populations up to 30,000 independent loci and for genome sizes up to 100,000 for 48 and 60. The similar values in both statistics and the plots suggested that there are diminishing returns for samples larger that about 48 populations. Moreover, closer inspection of the corresponding plots (Figure 6) suggested that for samples of 48 and 60 populations, an optimal ratio between TPL and FPL exists for approximately the 25 highest outlier loci, independent of genome size. For combinations with good performance, the maximum F1 score suggested that choosing the 30 highest outlier provided the optimal compromise between maximising TPR and minimising FPR (Supplemental Figure 6).

Figure 5. Heat-map of AUC-ROC (area under the curve – receiver operator characteristics) and AUC-PR (area under the curve – precision recall) in relation to effective genome size and number of populations scored.

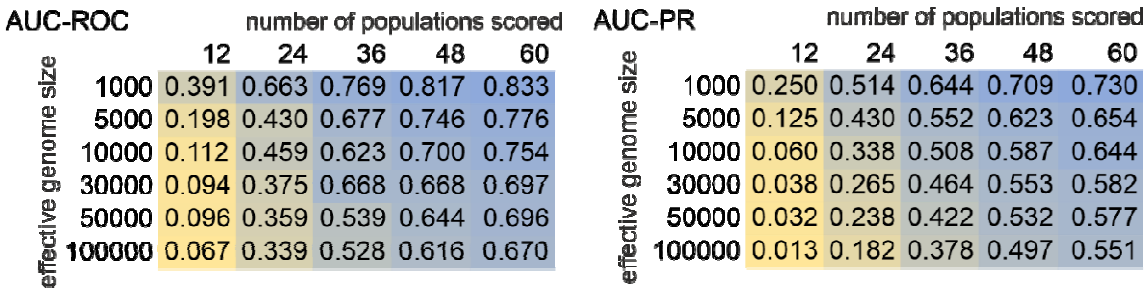
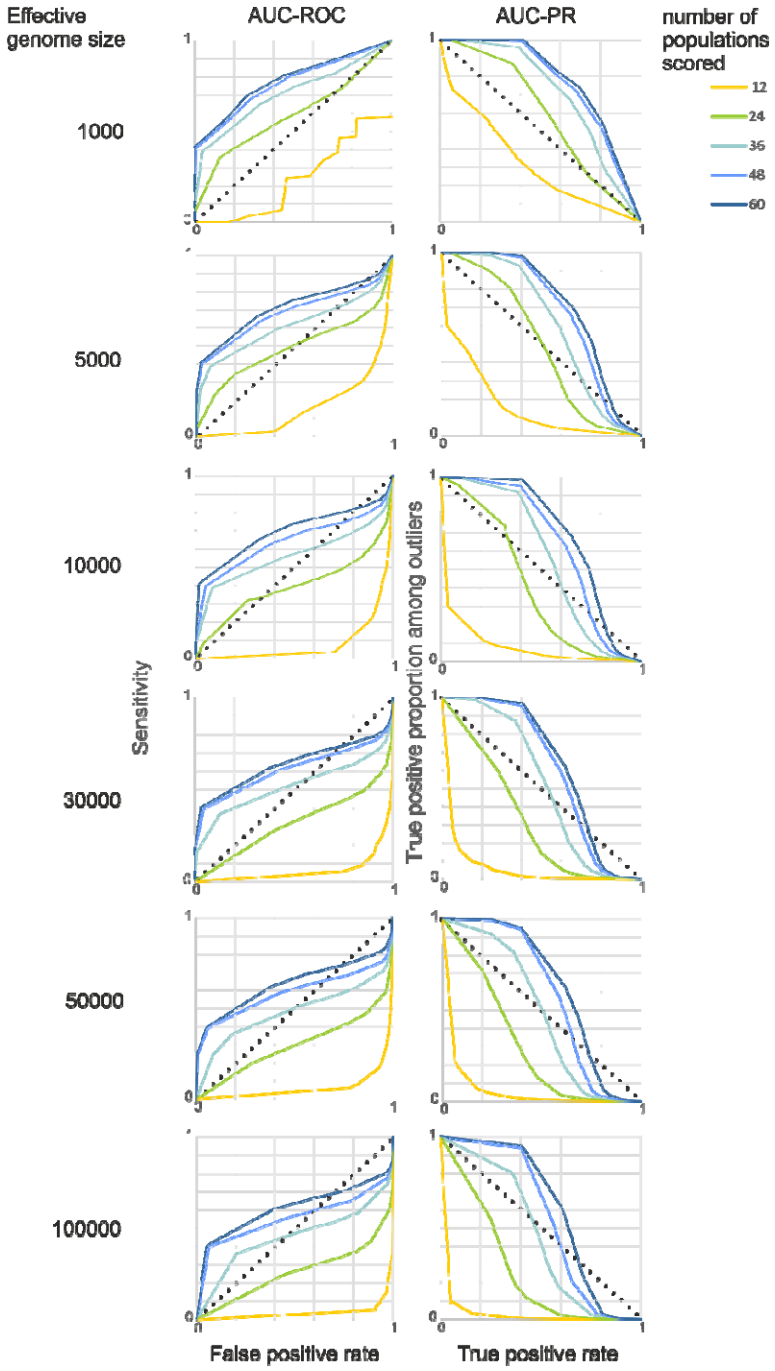


Figure 6. AUC-ROC and AUC-PR for a range of effective genome sizes. In the left column are the plots of AUC-ROC, i.e. FDR on the x-axis versus TPR on the y-axis. The right column shows AUC-PR plots, i.e. TPR on the x-axis versus PPV on the y-axis. The dotted lines indicate the threshold for a random effect.



Discussion

This study used extensive forward simulations to explore the potential of a novel GWAS approach utilising phenotypic population means and genome-wide allele-frequency data to identify loci potentially underlying quantitative polygenic traits. While the approach seems to be generally useful in a wide range of cases, there are also clear limits to its applicability.

General validity of the underlying assumptions

The initial simulations demonstrated that the expectation for the covariance of random population "allele frequencies" at contributing quantitative trait loci (QTL) and the respective population trait mean is consistently positive when an additive model applies. This is an inherent consequence of the common dependence of both variables on the QTL genotypes of the individuals in a population, as demonstrated in (3). Additive models seems to be an appropriate statistical approximation for most quantitative traits at population level (Hill et al., 2008), despite the description of many epistatic interaction on the molecular level (Moore & Williams, 2005).

The relation appeared to be largely independent of the distribution shape of locus contributions to the trait. While in the case of equal contributions i.e. a flat distribution, the correlation coefficients of individual loci are themselves a random variate, normally distributed around the expected mean. As the distribution becomes increasingly skewed, locus contribution becomes predictive of the correlation to the trait. Loci contributing more to the trait and thus accounting for more of the phenotypic variance will likely have a higher correlation of their allele frequencies to the population mean. Conversely, the expectation for non-contributing loci is zero. Therefore, it is principally possible to exploit the correlation between allele frequencies and population trait means for the identification of loci underlying an additive quantitative trait. However, some statistical limitations became obvious. Firstly, as the number of QTL increases, the expected mean correlation coefficients become so small that they are likely to be indistinguishable from the tail of the zero-centered normal distribution of non-contributing loci, even with an unrealistically high number of samples. Consequently, the method for identifying QTL by the positive covariance of their allele frequencies with the population trait means is *a priori* more suited for oligogenic to moderately polygenic traits. Secondly, the number of QTL and the distribution of locus contributions may influence the statistical identifiability of individual QTL. In particular, loci that contribute only minimally to the trait or that fall by chance below the expected mean correlation coefficient may overlap with the tail of the distribution of non-contributing loci.

These predictions remind of similar conditions for the contribution of different QTL architectures to phenotypic adaptation described by (Höllinger et al., 2023). They assert that phenotypic adaptation of oligogenic traits is achieved by detectable allele frequency shifts at some but not very many loci, while adaptation in highly polygenic traits is rather achieved by subtle perturbations of standing variation, with respective consequences for their detectability. Just as expected here, they stress the importance of stochastic effects that may lead to apparently heterogeneous locus contributions (Höllinger et al., 2023).

Limiting factors in natural settings

The Wright-Fisher forward simulations of a quantitative trait in a subdivided population with realistic properties and sample sizes largely confirmed the theoretical expectations. In particular when a sufficient number of populations was scored (>60), a large proportion of true positive loci could be reliably identified, with the exception of a few parameter combinations. The genetic architecture of the trait was an important predictor for the ability to identify causal loci. The most important other factor was the genetic trait architecture. While the loci underlying oligogenic and moderately polygenic traits could be fairly reliably identified, the highly polygenic scenario tested (500 loci) performed poorly. The difference between the two tested locus contribution distributions was not very

pronounced. This was likely due to the tendency of higher correlations between higher contributing loci and the trait, which ensured the inclusion of a substantial proportion of true positive loci in the selected outliers under a wide range of conditions.

The influence of mean heritability was similarly not marked. Even down to trait heritability estimates of 0.3, the success rate was only slightly reduced. This effect may be attributed to the averaging of phenotypes and genotypes across multiple individuals, which is likely to mitigate the inherent noise associated with individual data (Johri et al., 2022; Stinchcombe & Hoekstra, 2008). This finding is consistent with observations by (Zhang et al., 2018), who employed pooled data for GWAS. From a practical standpoint, the findings suggest that inevitable errors in phenotyping, which can compromise GWAS performance on individuals (Barendse, 2011), are likely to be less problematic when using the mean measured over many individuals. Furthermore, this finding indicates that the failure to entirely remove non-additive variance from the analysis does not necessarily compromise the method's ability to reliably identify trait-associated loci.

From a statistical perspective, it was anticipated that the range of phenotypic population means would influence the identification of true positive loci to some extent, given that a larger range of phenotypic means is inherently associated with on average larger allele-frequency differences among populations. The choice of populations with a large range of environmentally unexplained variance is therefore crucial. It is, however, important to emphasise that the underlying causes of the observed differences in trait means among populations are not of primary concern. These may be attributed to local adaptation, but also to maladaptation, human choice, or other factors. Likewise, increasing the number of populations screened increased the statistical power of the approach. However, it seemed that increasing the number of samples led to diminishing returns in statistical power gain beyond a certain threshold.

A pronounced population structure ($F_{ST} > \sim 0.07$) was a major factor impeding reliable identification of true positive loci, even with a high number of samples. This is probably due to distinct evolutionary trajectories in independently evolving populations. The genetic redundancy of polygenic traits can lead to evolution of the same phenotypes from different genomic bases(de Vladar & Barton, 2014; Kaneko & Furusawa, 2006), even if evolving from the same ancestral population (Barghi et al., 2019, 2020; Pfenninger et al., 2015). If different loci in different populations are causal for the observed phenotypic differences, a linear relation between population means and allele frequencies is not to be expected. It is therefore important that the allele frequencies in the studied populations are correlated either by recent common descent and/or recurrent gene-flow, i.e. that the population structure between the population scored is weak (Mathieson, 2021).

A situation where the overall genetic distance and the phenotypic differences are correlated, e.g. if an environmental gradient is correlated to the geographic distance between populations (IBD) and the trait value is an adaptation to this gradient, should as well be prone to produce false positives. To avoid such a situation, it is recommended to test for (the absence of) a correlation between genome-wide genetic distance and differences in phenotypic means (e.g. by a Mantel's test).

Accurate statistical genomic prediction in a wide range of conditions

Genomic prediction is deemed to be one of the major tools for the mitigation of climate change on biodiversity (Aguirre-Liguori et al., 2021; Bernatchez et al., 2023; Capblancq et al., 2020; Waldvogel et al., 2020). Contrary to its application in medicine or selective breeding

(Wray et al., 2019), however, accurate prediction of population responses is probably more important than the prediction of individual phenotypes. However, there is no theoretical obstacle, why the identified loci could not be used for individual genomic phenotype prediction, but this was not investigated here. Within the limits outlined above, the proposed method delivered very accurate predictions ($r > 0.8$) of population mean phenotypes. It should be noted, however, that the prediction is statistical in the sense that it produces a prediction score (de Los Campos et al., 2018) that correlates with the mean population phenotype and not the phenotype itself. Just like with any other genomic prediction (Kachuri et al., 2024), this limits the transferability of the prediction to other, more distantly related lineages or species.

Reducing the false positive rate is in any case advisable, as it proved to be the most important factor of prediction success with independent data. The application of a Machine learning approach, in this case Minimum Entropy Feature Selection (MEFS), prior to prediction reduced the already low false positive rate among the initially selected loci further. Other, comparable methods, such as e.g. , likely perform comparably or even better. Other factors influenced prediction success in a very similar fashion as the true positive rate. One notable exception was distribution of locus contributions. While true positive loci were more reliably identified from a flat distribution, prediction worked better when many loci of large effect were among the prediction set, most likely because these loci contribute more to phenotypic variance (Jain & Stephan, 2015).

Typical genome sizes of real species are no obstacle

The perhaps most important challenge was showing that the proposed method has enough statistical power to distinguish at least a part of the unknown, but likely relatively small number of QTL reliably from the large number of non-contributing loci in real genomes of real species. The evaluation of method performance with AUC-ROC and AUC-PR, as recommended recently (Lotterhos et al., 2022), showed a satisfactory performance even for genomes with moderately high effective sizes, provided a sufficiently high number of populations is screened. In particular restricting the selection of potentially causal QTL on a few dozen of the highest outliers promises to yield very low false positive rates. As shown above, already a limited number of true positive loci may be sufficient for reliable genomic prediction.

Practical considerations

The proposed method finds rather genomic regions or haplotypes associated to the trait in question than directly causal SNPs. However, this is true for most GWAS methods (Wang et al., 2010) and therefore fine-mapping and inference of causal processes remain to be done (Wallace, 2021). In practice, this requires that regions with high SNP outlier density need to be collapsed to haplotypes prior to further analysis. Knowledge on the local LD-structure, mean haplotype length, respectively recombination landscape can aid haplotype identification (Flister et al., 2013). Recently developed machine learning approaches makes such information available for pooled data (Adrion et al., 2020).

The possibly largest advantage of the proposed method is its data efficiency, if pooled sequencing is applied. Because the Pool-Seq approach (Schlötterer et al., 2014) yields highly accurate estimates of genome-wide allele frequencies at SNP sites (Czech, Peng, Spence, Lang, Bellagio, Hildebrandt, Fritschi, Schwab, Rowan, & consortium, 2022) the necessary

sequencing effort is marginal compared to individual based approaches (Ziyatdinov et al., 2021). This makes GWAS studies accessible for the usual funding in the field of biodiversity. Pooled sequencing for GWAS has been proposed (Yang et al., 2015) and applied (Giorello et al., 2023; Kumar et al., 2022; Pfenninger et al., 2021) with extreme phenotypes. What is now required is the application of the method to a real-world data set, a work which is in progress.

Conclusion

This study demonstrated the potential of the proposed GWAS approach for biodiversity genomics. By carefully considering the factors influencing its performance and addressing the limitations, this method can be a valuable tool for identifying the genetic basis of complex traits in natural populations.

Acknowledgements

The author wants to thank Bob O'Hara and Barbara Feldmeyer for comments on the manuscript.

Funding

The author declares that he has received no specific funding for this study.

Conflict of interest disclosure

The author declares that he complies with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Data, scripts, code, and supplementary information availability

Supplementary information including the Python code used for the simulations is available at https://**10.5281/zenodo.11562472**

References

Adrion, J. R., Galloway, J. G., & Kern, A. D. (2020). Predicting the landscape of recombination using deep learning. *Molecular Biology and Evolution*, *37*(6), 1790–1808.

Aguirre-Liguori, J. A., Ramírez-Barahona, S., & Gaut, B. S. (2021). The evolutionary genomics of species' responses to climate change. *Nature Ecology & Evolution*, *5*(10), 1350–1360.

Barendse, W. (2011). The effect of measurement error of phenotypes on genome wide association studies. *BMC Genomics*, *12*(1), 232. https://doi.org/10.1186/1471-2164-12-232

Barghi, N., Hermisson, J., & Schlötterer, C. (2020). Polygenic adaptation: A unifying framework to understand positive selection. *Nature Reviews Genetics*, *21*(12), 769–781.

Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., Dolezal, M., Taus, T., Kofler, R., & Schlötterer, C. (2019). Genetic redundancy fuels polygenic adaptation in Drosophila. *PLoS Biology*, *17*(2), e3000128.

Barton, N. H. (1999). Clines in polygenic traits. *Genetics Research*, *74*(3), 223–236.

663 Bernatchez, L., Ferchaud, A.-L., Berger, C. S., Venney, C. J., & Xuereb, A. (2023). Genomics for
664 monitoring and understanding species responses to global climate change. *Nature*
665 *Reviews Genetics*, 1–19.
666 Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From
667 polygenic to omnigenic. *Cell*, *169*(7), 1177–1186.
668 Brandes, N., Weissbrod, O., & Linial, M. (2022). Open problems in human trait genetics.
669 *Genome Biology*, *23*(1), 131. https://doi.org/10.1186/s13059-022-02697-9
670 Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020).
671 Genomic prediction of (mal) adaptation across current and future climatic landscapes.
672 *Annual Review of Ecology, Evolution, and Systematics*, *51*, 245–269.
673 Chakraborty, R. (1981). The distribution of the number of heterozygous loci in an individual
674 in natural populations. *Genetics*, *98*(2), 461.
675 Czech, L., Peng, Y., Spence, J., Lang, P., Bellagio, T., Hildebrandt, J., Fritschi, K., Schwab, R.,
676 Rowan, B., & Weigel, D. (2022). Efficient analysis of allele frequency variation from
677 whole-genome pool-sequencing data. *Population, Evolutionary, and Quantitative*
678 *Genetics Conference (PEQG 2022)*, 99.
679 https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3474009
680 Czech, L., Peng, Y., Spence, J. P., Lang, P. L., Bellagio, T., Hildebrandt, J., Fritschi, K., Schwab, R.,
681 Rowan, B. A., & consortium, G. (2022). Monitoring rapid evolution of plant populations at
682 scale with Pool-Sequencing. *BioRxiv*, 2022–02.
683 de Los Campos, G., Vazquez, A. I., Hsu, S., & Lello, L. (2018). Complex-trait prediction in the
684 era of big data. *Trends in Genetics*, *34*(10), 746–754.
685 de Vladar, H. P., & Barton, N. (2014). Stability and response of polygenic traits to stabilizing
686 selection and mutation. *Genetics*, *197*(2), 749–767.
687 Dunker, S., Boyd, M., Durka, W., Erler, S., Harpole, W. S., Henning, S., Herzschuh, U., Hornick,
688 T., Knight, T., Lips, S., Mäder, P., Švara, E. M., Mozarowski, S., Rakosy, D., Römermann, C.,
689 Schmitt-Jansen, M., Stoof-Leichsenring, K., Stratmann, F., Treudler, R., … Wilhelm, C.
690 (2022). The potential of multispectral imaging flow cytometry for environmental
691 monitoring. *Cytometry Part A*, *101*(9), 782–799. https://doi.org/10.1002/cyto.a.24658
692 Exposito-Alonso, M., Drost, H., Burbano, H. A., & Weigel, D. (2020). The Earth BioGenome
693 project: Opportunities and challenges for plant genomics and conservation. *The Plant*
694 *Journal*, *102*(2), 222–229. https://doi.org/10.1111/tpj.14631
695 Flister, M. J., Tsaih, S.-W., O'Meara, C. C., Endres, B., Hoffman, M. J., Geurts, A. M., Dwinell, M.
696 R., Lazar, J., Jacob, H. J., & Moreno, C. (2013). Identifying multiple causative genes at a
697 single GWAS locus. *Genome Research*, *23*(12), 1996–2002.
698 Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C.,
699 Crottini, A., Godoy, J. A., & Höglund, J. (2022). The era of reference genomes in
700 conservation genomics. *Trends in Ecology & Evolution*, *37*(3), 197–202.
701 Giorello, F. M., Farias, J., Basile, P., Balmelli, G., & Da Silva, C. C. (2023). Evaluating the
702 potential of XP-GWAS in Eucalyptus: Leaf heteroblasty as a case study. *Plant Gene*, *36*,
703 100430.
704 Heuertz, M., Carvalho, S. B., Galindo, J., Rinkevich, B., Robakowski, P., Aavik, T., Altinok, I.,
705 Barth, J. M., Cotrim, H., & Goessen, R. (2023). The application gap: Genomics for
706 biodiversity and ecosystem service management. *Biological Conservation*, *278*, 109883.
707 Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive
708 genetic variance for complex traits. *PLoS Genetics*, *4*(2), e1000008.
709 Hogg, C. J. (2023). Translating genomic advances into biodiversity conservation. *Nature*
710 *Reviews Genetics*, 1–12.

Höllinger, I., Wölfl, B., & Hermisson, J. (2023). A theory of oligogenic adaptation of a quantitative trait. *Genetics*, *225*(2), iyad139. https://doi.org/10.1093/genetics/iyad139

Jain, K., & Stephan, W. (2015). Response of polygenic traits under stabilizing selection and mutation when loci have unequal effects. *G3: Genes, Genomes, Genetics*, *5*(6), 1065–1074.

Johri, P., Aquadro, C. F., Beaumont, M., Charlesworth, B., Excoffier, L., Eyre-Walker, A., Keightley, P. D., Lynch, M., McVean, G., & Payseur, B. A. (2022). Recommendations for improving statistical inference in population genomics. *PLoS Biology*, *20*(5), e3001669.

Kachuri, L., Chatterjee, N., Hirbo, J., Schaid, D. J., Martin, I., Kullo, I. J., Kenny, E. E., Pasaniuc, B., Yuji 29, P. R. M. in D. P. (PRIMED) C. M. W. G. A. P. L. 20 C. M. P. 21 C. D. V. 22 23 D. Y. 24 W. Y. 19 25 26 Z. H. 27 28 Z., & Witte, J. S. (2024). Principles and methods for transferring polygenic risk scores across global populations. *Nature Reviews Genetics*, *25*(1), 8–25.

Kaneko, K., & Furusawa, C. (2006). An evolutionary relationship between genetic variation and phenotypic fluctuation. *Journal of Theoretical Biology*, *240*(1), 78–86.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlötterer, C. (2011). PoPoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PloS One*, *6*(1), e15925.

Kumar, S., Deng, C. H., Molloy, C., Kirk, C., Plunkett, B., Lin-Wang, K., Allan, A., & Espley, R. (2022). Extreme-phenotype GWAS unravels a complex nexus between apple (*Malus domestica*) red-flesh colour and internal flesh browning. *Fruit Research*, *2*(1), 1–14. https://doi.org/10.48130/FruRes-2022-0012

Lotterhos, K. E., Fitzpatrick, M. C., & Blackmon, H. (2022). Simulation Tests of Methods in Evolution, Ecology, and Systematics: Pitfalls, Progress, and Principles. *Annual Review of Ecology, Evolution, and Systematics*, *53*(1), 113–136. https://doi.org/10.1146/annurev-ecolsys-102320-093722

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits* (Vol. 1). Sinauer Sunderland, MA.

Mackay, T. F. (2014). Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nature Reviews Genetics*, *15*(1), 22–33.

Mathieson, I. (2021). The omnigenic model and polygenic prediction of complex traits. *The American Journal of Human Genetics*, *108*(9), 1558–1563.

Moore, J. H., & Williams, S. M. (2005). Traversing the conceptual divide between biological and statistical epistasis: Systems biology and a more modern synthesis. *BioEssays*, *27*(6), 637–646. https://doi.org/10.1002/bies.20236

Orr, H. A. (1998). Testing natural selection vs. Genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics*, *149*(4), 2099–2104.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Pfenninger, M., Patel, S., Arias-Rodriguez, L., Feldmeyer, B., Riesch, R., & Plath, M. (2015). Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic habitats of a neotropical fish (*Poecilia mexicana*). *Molecular Ecology*, *24*(21), 5446–5459. https://doi.org/10.1111/mec.13397

Pfenninger, M., Reuss, F., Kiebler, A., Schönnenbeck, P., Caliendo, C., Gerber, S., Cocchiararo, B., Reuter, S., Blüthgen, N., & Mody, K. (2021). Genomic basis for drought resistance in European beech forests threatened by climate change. *Elife*, *10*, e65532.

759 Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation–not by sweeps alone. *Nature Reviews*
760 *Genetics*, *11*(10), 665–667.
761 R Core Team, R. (2013). *R: A language and environment for statistical computing*.
762 Rijsbergen, C. van. (1979). *Information retrieval*. Butterworth-Heinemann.
763 https://dl.acm.org/doi/abs/10.5555/539927
764 Santure, A. W., & Garant, D. (2018). Wild GWAS—association mapping in natural
765 populations. *Molecular Ecology Resources*, *18*(4), 729–738.
766 https://doi.org/10.1111/1755-0998.12901
767 Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals—
768 Mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*,
769 *15*(11), 749–763.
770 Sella, G., & Barton, N. H. (2019). Thinking about the evolution of complex traits in the era of
771 genome-wide association studies. *Annual Review of Genomics and Human Genetics*, *20*,
772 461–493.
773 Shendure, J., Findlay, G. M., & Snyder, M. W. (2019). Genomic medicine–progress, pitfalls,
774 and promise. *Cell*, *177*(1), 45–57.
775 Stinchcombe, J. R., & Hoekstra, H. E. (2008). Combining population genomics and
776 quantitative genetics: Finding the genes underlying ecologically important traits.
777 *Heredity*, *100*(2), 158–170.
778 Taylor, C. F., & Higgs, P. G. (2000). A population genetics model for multiple quantitative
779 traits exhibiting pleiotropy and epistasis. *Journal of Theoretical Biology*, *203*(4), 419–
780 437.
781 Team, T. P. (2019, December 28). *PyPy*. PyPy. https://www.pypy.org/
782 Tills, O., Holmes, L. A., Quinn, E., Everett, T., Truebano, M., & Spicer, J. I. (2023). Phenomics
783 enables measurement of complex responses of developing animals to global
784 environmental drivers. *Science of the Total Environment*, *858*, 159555.
785 Turchin, M. C., Chiang, C. W. K., Palmer, C. D., Sankararaman, S., Reich, D., & Hirschhorn, J. N.
786 (2012). Evidence of widespread selection on standing variation in Europe at height-
787 associated SNPs. *Nature Genetics*, *44*(9), 1015–1019. https://doi.org/10.1038/ng.2368
788 Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C.,
789 Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature*
790 *Reviews Methods Primers*, *1*(1), 59.
791 Van Rossum, G., & Drake, F. L. (2009). *Introduction to python 3: Python documentation*
792 *manual part 1*. CreateSpace. https://dl.acm.org/doi/abs/10.5555/1592885
793 Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS
794 discovery. *The American Journal of Human Genetics*, *90*(1), 7–24.
795 Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J.
796 (2017). 10 years of GWAS discovery: Biology, function, and translation. *The American*
797 *Journal of Human Genetics*, *101*(1), 5–22.
798 Waldvogel, A.-M., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C., Kofler, R.,
799 Mock, T., Schmid, K., Schmitt, I., & Bataillon, T. (2020). Evolutionary genomics can
800 improve prediction of species' responses to climate change. *Evolution Letters*, *4*(1), 4–
801 18.
802 Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple
803 causal variants. *PLoS Genetics*, *17*(9), e1009440.
804 Wang, K., Dickson, S. P., Stolle, C. A., Krantz, I. D., Goldstein, D. B., & Hakonarson, H. (2010).
805 Interpretation of association signals and identification of causal variants from genome-
806 wide association studies. *The American Journal of Human Genetics*, *86*(5), 730–742.

807 Wray, N. R., Kemper, K. E., Hayes, B. J., Goddard, M. E., & Visscher, P. M. (2019). Complex
808 trait prediction from genome data: Contrasting EBV in livestock to PRS in humans:
809 genomic prediction. *Genetics*, *211*(4), 1131–1141.
810 Wright, S. (1949). THE GENETICAL STRUCTURE OF POPULATIONS. *Annals of Eugenics*,
811 *15*(1), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x
812 Xie, C., & Yang, C. (2020). A review on plant high-throughput phenotyping traits using UAV-
813 based sensors. *Computers and Electronics in Agriculture*, *178*, 105731.
814 Yang, J., Jiang, H., Yeh, C.-T., Yu, J., Jeddeloh, J. A., Nettleton, D., & Schnable, P. S. (2015).
815 Extreme-phenotype genome-wide association study (XP-GWAS): A method for
816 identifying trait-associated variants by sequencing pools of individuals selected from a
817 diversity panel. *The Plant Journal*, *84*(3), 587–596.
818 Zhang, W., Liu, A., Albert, P. S., Ashmead, R. D., Schisterman, E. F., & Mills, J. L. (2018). A
819 pooling strategy to effectively use genotype data in quantitative traits genome-wide
820 association studies. *Statistics in Medicine*, *37*(27), 4083–4095.
821 https://doi.org/10.1002/sim.7898
822 Ziyatdinov, A., Kim, J., Prokopenko, D., Privé, F., Laporte, F., Loh, P.-R., Kraft, P., & Aschard, H.
823 (2021). Estimating the effective sample size in association studies of quantitative traits.
824 *G3*, *11*(6), jkab057.
825