Association of adaptive phenotypic traits to causal variants using the extended genotype homozygosity matrix.

September 15, 2023

Abstract

TO DO ...

Introduction

The detection of the effect of adaptive phenotypic traits is fundamental for understanding the number of affected variants and the strength of selection for these adaptive traits. The detection of the association between an adaptive phenotype and the genotype can be difficult, depending on the effect of each associated variant and the effect of the adaptation over the positions around it. A number of studies, which the aim of understanding the patterns of frequencies for adaptive traits (*e.g.*, EYRE-WALKER, 2010; CABALLERO *et al.*, 2015; LOURENÇO *et al.*, 2011; CONNALLON and CLARK, 2015), shown that the genome architecture for a trait depends on the (mathematical) relationship of the trait versus fitness (that is, if the fitness is directly correlated with the trait or, alternatively, fitness and traits are correlated distributions). Simons et al. (2018) avoided this problem by inferring the relationship between trait and fitness from empirical data and estimated the genetic architecture of the trait, assuming pleiotropy and stabilizing selection. Also, Beissinger *et al.* (2018) developed a test to detect the effect of selection on a quantitative complex trait using the whole information on effect sizes.Uricchio *et al.* (2019) developed a new method to detect the signals of polygenic selection using the derived allele effect size from GWAS analysis and the derived allele frequency of the associated variants. All of these methods can be applied to observe the effects of environmental changes on the genome for specific traits related to adaptation. Recently Zeng *et al.* (2019) developed a bayesian algorithm (SBayesS) that requires GWAS summary statistics and functional genomic annotation to detect the effect of selection on complex polygenic traits.

Here, we are aimed to (i) discover if a given trait is under a recent adaptive process and (ii) to detect strong and medium adaptive effects with higher precision, based on the association of the trait versus phenotype. We propose to use the signals of linkage disequilibrium produced by causal adaptive variants to increase the precision in locating associate variants as well as distinguishing between adaptive versus non-adaptive variants. The reasoning is the following: the causal variants of an adaptive trait will left signals of adaptation (*e.g.*,, increase of homozygosity around the causal variants, more linkage disequilibrium, reduction of variation) around them. If a trait is under adaptive pattern, we expect that the associated variants exhibit signals of adaptation, while other non-adaptive traits will not have such signals on associated variants. Additionally, the signals of adaptation may increase the precision to locate the causal variants, together with the genotype data.

We propose to study the association between the phenotype and the genotype considering additionally (as additional genotype information) a matrix containing the extended genotype homozygosity. This matrix would be used in the analysis of association to detect the effect of adaptive selection of the studied phenotype. Our hypothesis is that this approach can have promising results to detect the causal mutations in panmictic populations that have experimented stronger selective events. A consistent simulation study is required in this project to evaluate different scenarios.

The resolution of this kind of statistics is expected to be in the range of genetic breeding improvement up to domestication process (approximately from some dozens to hundreds of generations (less than N_e generations, (SABETI *et al.*, 2002). This range may allow to detect critical traits and the effect of these traits on the genome under crucial events such as the domestication. In this work, the statistics used are calculated at genome level , which are analyzed with the phenotypes of individuals.

Methods

Considering a single population and polarized alleles (ancestral and derived)

Following SABETI *et al.* (2002), we define the statistic EHH_d as the proportion of homozygote individuals of the derived allele from position i (the target position) to position j, in relation to the number of homozygote individuals with derived variants at the position i (note that derived variants are obtained by comparison to an/several outgroup species). That is:

$$EHH_{d_{ij}} = \frac{\sum_{k=1}^{n} I_{d_{k,ij}}}{\sum_{l_d=1}^{n} I_{d_{l,i}}}, \tag{1}$$

where n is the number of individual samples, and $I_{d_{k,ij}}$ counts 1 if the individual k is homozygote for derived variant at position i and is still homozygote from position i to j, otherwise counts 0. In the same way, $I_{d_{l,i}}$ counts 1 if the individual l is homozygote for derived at position i, otherwise counts 0. The statistic EHH_a is calculated in the same way than EHH_d but considering ancestral variant instead of derived variant. Note that this value can not be calculated in case there are not derived (ancestral) homozygous individuals.

In order to quantify the effect of the neighbouring homozygosity at a given studied position (i), all values for EHH_{ij} are summed from the position i to their right and and to their left sides considering their distance (physical or recombinant), and until reaching a threshold arbitrary value of EHH_{ij} (let's say 0.1). That is:

$$iES_{d_i} = \sum_{j=x+1}^{y} \frac{(EHH_{d_{ij-1}} + EHH_{d_{ij})}}{2}(Pos_j - Pos_{j-1}), \tag{2}$$

x and y are the positions (at the right and at the left) where $EHH_{d_{ij}}$ becomes below the threshold or it is too far (by the presence of large gaps) from the central position (so the area out of x and y is considered unimportant), and Pos is the physical or the recombinant position (from a linkage map). The same is for the ancestral statistic (iES_{a_i}).

The integrated relative Extended homozygosity statistic is then calculated as:

$$iRES_{da_i} = \frac{iES_{d_i}}{iES_{a_i}}. \tag{3}$$

A value higher than 1 (if the genotype extension statistic for derived variants is larger than for ancestral variants) suggest the effect of positive selection favouring a derived variant at position i.

The homozigosity at each position per individual

In case of doing a GWAS analysis, usually each position and genotype is evaluated independently in relation to each individual phenotype. That means that it is crucial to distinguish the genetic information between individuals at this position. Here we propose to include the information concerning to the homogeneity of the individual at neighbouring regions from the focus position (that is, a way to consider the linkage disequilibrium between the two chromosome copies of an individual at this region) in addition to the genotype at the focus position. That is, it is calculated a statistic related to the neighbouring homogeneity for every position and individual. This statistic is obtained by slicing the $iRES_{da_i}$ statistic given the contribution of each individual (k) to the total:

$$iES_{dk_i} = \sum_{j=x+1}^{y} (EHH_{d_{k,ij-1}} + EHH_{d_{k,ij}})(Pos_j - Pos_{j-1})/2, \text{ and}$$
$$iRES_{da_{k,i}} = \frac{iES_{d,k_i}}{iES_a}, \tag{4}$$

where x and y are the positions (at the right and at the left boundaries for derived) and $EHH_{d_{k,ij}} = I_{d_{k,ij}}/(\sum_{l_d=1}^{n} I_{d_{l,i}})$.

Quantifying homogeneity with no outgroup and unphased data

To quantify the homogeneity with no outgroup and unphased data, it is convenient to use the framework developed by TANG *et al.* (2007) to study the degree of homogenization of a given position (using unphased genotype information) as is explained below.

Define candidate positions

In a first step, it is useful to define those candidate positions in relation to their maximum local extension of their homozygosity. This step is optional, as other criteria for defining candidate positions can be used. Following (TANG *et al.*, 2007), we define $EHHS_{ij}$ as the proportion of homozygote individuals from position i (the position of interest) to position j, in relation to the number of homozygote individuals at the position i. That is:

$$EHHS_{ij} = \frac{\sum_{k=1}^{n} I_{k,ij}}{\sum_{l=1}^{n} I_{l,i}}, \tag{5}$$

where n is the number of individual samples, and $I_{k,ij}$ counts 1 if the individual k is homozygote from position i to j (*i.e.*, using the genotype nomenclature, all variants have the values 0 or 2 at this region), otherwise counts 0. In the same way, $I_{l,i}$ counts 1 if the individual l is homozygote at position i, otherwise counts 0. This statistic is calculated from position i to any position (left or right) until this proportion becomes enough small to be considered negligible. This threshold value, although is somewhat arbitrary, it has been considered 0.1 in the original work (TANG *et al.*, 2007) and here it is used the same criteria. The $EHHS_{ij}$ values for the position i are kept and used to calculate the next statistic iES_i (TANG *et al.*, 2007).

The following calculation of the iES_i statistic pretends to quantify the effect of the neighbouring homozygosity at a given studied position. Having all values for $EHHS_{ij}$, we count the total area of homozigosity around the position i, that is, having the position i as the center, we sum, at their right and and their left, all the contributions of $EHHS_{ij}$, considering their distance (physical or recombinant). that is:

$$iES_i = \sum_{j=x+1}^{y} \frac{(EHHS_{ij-1} + EHHS_{ij})}{2}(Pos_j - Pos_{j-1}), \tag{6}$$

where x and y are the positions (at the right and at the left) where $EHHS_{ij}$ becomes bellow the threshold or it is too far (by the presence of large gaps) from the central position (so the area out of x and y is considered unimportant), where Pos may be the physical or the recombinant position (from a linkage map).

The homogeneity at each position per individual using unphased and no outgroup data

As explained above, for GWAS analyses it is convenient to have information of the homogeneity per individual (to be contrasted with phenotype data). The desired statistic is obtained by dividing the iES_i statistic given the contribution of each individual to the total:

$$iES_{k,i} = \frac{1}{2\sum_{l=1}^{n} I_{l,i}} \sum_{j=a+1}^{b} (Pos_j - Pos_{j-1})(I_{k,ij-1} + I_{k,ij}), \tag{7}$$

where $\sum_{k=1}^{k=n} iES_{k,i} = iES_i$ (considering the same threshold values -a and b- for the each of the samples, lke iES_i statistic). Here the only differential term between individuals (in relative terms) is the last sum.

Using a reference population: The quotient between the extension of homozygosity in target individuals from a population versus a reference population

Following the same reasoning, it is possible to estimate the effect of the extension of homozygosity per position and per individual in relation to the effect in a reference population (note that a reference population is not an outgroup species). This is useful in case we are considering the effect of selection in the target population while we assume no selection in the reference population. Those position that have high iES_i at both populations would be considered nuisance given by other factors, like genomic effect caused by the genetic architecture of the genome. Therefore, following (TANG *et al.*, 2007), we define the statistic derived from Rsb_i:

$$Rsb_{k,i} = \frac{iES_{k,i}}{iES_i^{popRef}} \tag{8}$$

where $\sum_{k=0}^{k=n} Rsb_{k,i} = Rsb_i$.

The study of association phenotype-genotype and the information related to the homogeneity of individuals in neighbouring regions

A multivariate regression model has been used to detect the association of a position with the phenotype, considering only, the genotype information, only the extended genotype homozygosity information or both. The probability of association is calculated with the linear model function in R (lm), considering the vector of phenotypes (Y), obtained by summing up (additive model) all the effects per variant position, considering their dominance parameter, in relation to the genotype matrix and/or the iES_k matrix (with a dimension of SNPs x individuals). All these vectors and matrix were normalised previously. The r^2 regression value was obtained for each SNP, and the p-value was estimated assuming a F distribution. False discovery rate and Bonferroni methods (BENJAMINI and HOCHBERG, 1995) that correct for multiple testing were applied to raw P-values.

To evaluate the capacity that we have to detect real beneficial mutations affecting the studied phenotype we measured the specificity (True Negative Rate) and sensitivity (True Positive Rate or statistical power) for the used matrices (the iES_k, the genotype counts and both together).

Simulations

Forward simulations of a population have been performed with Slim v3 (HALLER and MESSER, 2019). A population of 1000 diploid individuals (N_e) is simulated. We consider 10 chromosomes of 1Mb each (a total of 10Mb), where five chromosomes have no causal variants while each of the remained five chromosomes include 1000 coding regions of 1500bp evenly distributed across the total length. We consider mutation rate constant of 1.25e-7 per position and generation, and recombination rate variable following a curve with a minimum peak on the centre of chromosome (rate of 2.5e-7) and increasing logarithmically to achieve a maximum at telomeres (2.5e-5). The population is run for $10N_e$ generations and the entire populations is analyzed. Afterwards, a new environment is simulated in this population, and it is continued during 30 more generations. Then, again this new populations is also analyzed.

Functional positions (non-synonymous, defined as the two first positions of each codon) are under selective effects, while non-functional positions are considered neutral. Here, we model positive selection, as we are interested in detecting adaptive events. The new mutations falling in functional positions are under the effect of positive and negative selection: a proportion of mutants will be under positive selection (this ones following an exponential distribution), while the rest will be under negative selection (gamma distribution), according to (BOYKO *et al.*, 2008) results for human species.

We have looked at the differences in the distribution of selective and phenotype effects considering the Eyre-Walker proposal (EYRE-WALKER, 2010), here named

EW, and a gamma bivariant distribution (CABALLERO *et al.*, 2015) for negative values (using *simstudy* library in R). The EW distribution estimates the phenotype value from an exponential-like pattern and considering a nuisance parameter, while bivariant gamma distribution is more flexible in considering different patterns of phenotype effects. If we consider an exponential distribution for phenotype effects (shape=1), the distributions are quite similar (see Figure S1). The major difference is observed when comparing distributions with correlation 0, where the observed plot of the EW function is more spread. We decided to use the first option (EW model) because it is possible to estimate easily, by inversion, the selective effects for an environmental change.

Different scenarios of the relationship between the fitness effect and the phenotypic trait are considered, from no-correlated to highly correlated relationship, using the Eyre-Walker (EW) model (EYRE-WALKER, 2010), that is, the effect for a phenotype trait z is related to selective effects, $4Ns$, using the equation $z = \delta(4N_e s)^\tau(1 + \epsilon)$, where δ is a given random value (*e.g.*,-1 or 1), τ is related to the correlation of phenotypic trait with the s and ϵ is a normal distribution with mean 0 and standard deviation σ_ϵ. For positive selection, and in order to give phenotypic values close to optimal (zero) for high selective effects, the value of τ is transformed to negative. A normalization factor is included to achieve same limit values of z for all conditions. We considered that the effects of selection within genes are multiplicative, being 1, $1 + s$ and $1 + 2s$ $(+s^2)$ at ancestral homozygote, heterozygote and derived homozygote, respectively. Dominance values are calculated following WANG *et al.* (1998), in which the dominance becomes negatively correlated with selective effect (here for the absolute value of selective effect). The number of standing variants that contribute to the second population as positively selected become smaller when the displacement is larger, because the distribution of mutations are mostly concentrated on small selective and phenotypic effects, by definition of the model. Finally, environmental variance is added to the genotype effect in order to obtain scenarios with similar heritabilities, which should be around 50%.

The given modeled scenarios show stationary parametric conditions. In order to detect the effect of a higher proportion of variants under the effect of selective events, we have also included an additional step in simulations, where an environmental change occurs and modifies the selective effect of each phenotype, according to the correlation of phenotype with the fitness. In that case, the optimum phenotype is displaced from 0 (optimal at initial environment) to a given value, and the new selective effect is re-calculated by inverting the previous equation, Thus, in the new environment, standing variants change their selective effects to $s = \frac{|z-d|^{1/\tau}}{4*N_e}$, where d is the proportion of displacement of the optimal phenotype from previous optimal (zero). Phenotypic and selective effects for new variants at the new environment are following a new distribution with same shape (and mean for negative selective effects) but different mean and proportion of positive selective effects. As in standing variants, the optimal phenotypic effects are displaced (d).

To have clear patterns of strong selection in simulations, we included a parameter to force a number of selective sweeps. These selective sweeps are randomly located in nonsynonymous positions of the genome and are produced once the new environment occurs. The selective sweeps are generated using a number of derived mutations at frequencies higher than 0.05 but lower than 0.15, in which a strong selective effect is imposed. The phenotype pattern is modified to be the optimum value in this environment.

We collected in the output files all variants and their the frequencies, the selective effects for the two environments, the dominance, and the phenotypic effect, plus the complete sequence per individual. Here, the total population (1000 individuals) is collected. Permutation of phenotypes between individuals is performed to control the type I error rate for each scenario. As in (CABALLERO *et al.*, 2015), two different number of SNPs where chosen for each scenario: (i) the whole SNPs data and (ii) a subset of SNPs of 1% of the total.

Fifteen different scenarios were analysed, each one having two temporal samples, one before and one after an environmental change (Table 1). The population runs for $10N_e$ generations, then, the environmental change occurs and the population runs for $0.03N_e$ additional generations. The distribution of mutations for these scenarios are shown in Figures S2, S3 and S4 (note that the last 5 are not shown because they are equal to scenarios without forcing selection except for 5 mutations). Data is collected before the environmental change occurs and also at the end of simulation. With these conditions, we pretend to study the effect of positive selection (weak and pervasive versus strong and relatively rare) in a context of neutral deleterious mutations and, posteriorly, under a strong environmental change that modifies the fitness of the individuals. The effect of these environmental change is modulated, as well as the correlation of the studied trait in relation to the fitness. The first ten scenarios do not include deleterious mutations (only neutral positive variants), while the next 10 scenarios include deleterious and beneficial variants. The proposed new test is evaluated as well as other available tests (BEISSINGER *et al.*, 2018; URICCHIO *et al.*, 2019; ZENG *et al.*, 2019).

Once simulations are finished, two kind of output files are obtained: a *ms* file with the variants and positions and a table including the position, the selection coefficients for environment 1 and 2, the dominance, the type of mutation (neutral/functional) and the phenotype effect for each variant. *ms* format sequences were converted to genotype formatted file, where each individual can have 0,1 or 2 derived alleles per position. This file included the genotypes from the first and second environment, is filtered to erase SNPs wirh MAF<5% and it is used as input file for a software (Rbski) to calculate the matrix of statistics used for GWAS analysis.

Empirical data

To choose: rice, pigs, humans, bovine, drosophila?... depending on available data.

Scenario	τ	σ_ϵ	$\%(+)_2$	s_2 (+)	nsweeps_2
1	0.00	0.60	1.0	1.25e-2	0
2	0.00	0.60	10.0	1.25e-3	0
3	0.10	0.50	1.0	1.25e-2	0
4	0.10	0.50	10.0	1.25e-3	0
5	0.25	0.40	1.0	1.25e-2	0
6	0.25	0.40	10.0	1.25e-3	0
7	0.50	0.30	1.0	1.25e-2	0
8	0.50	0.30	10.0	1.25e-3	0
9	0.80	0.10	1.0	1.25e-2	0
10	0.80	0.10	10.0	1.25e-3	0
11	0.00	0.60	1.0	1.25e-2	5
12	0.10	0.50	1.0	1.25e-2	5
13	0.25	0.40	1.0	1.25e-2	5
14	0.50	0.30	1.0	1.25e-2	5
15	0.80	0.10	1.0	1.25e-2	5

Table 1: Conditions for the different simulated scenarios considered. Common parameters are those for the percentage of beneficial variants to $\%(+)_1 = 1.0$, the mean strength of beneficial variants in the first scenario to $s_1 = 1.25e - 2$ and the proportion of displacement ($d = 0.2$), except for conditions with $\tau = 0$ (no correlation phenotype-genotype), where no displacement was included. The first columns (τ, σ_ϵ) refers to the phenotypic trait, and the next two columns ($\%(+)_2$ and s_2) refers to the maximum proportion of beneficial variants and the mean strength of beneficial selective effects of variants at the second scenario (environmental change). For scenarios with deleterious variants, it is included a gamma distribution with a mean of $s = -0.03$ and shape of 0.2,

Results

Simulation of data under different scenarios and Validation

-Descriptive information about all the simulations. SNPs, Variation, SFS, patterns of selective sweeps. distribution of phenotypic trait per individual..

-GWAS analysis: Study of the effect of the different parameters on the detection of QTLs. Real versus detected QTLs. Effect of EGH matrix. Precision in the detection of the causal SNP.

-Comparison with the test of BEISSINGER *et al.* (2018). Detection of strong effects versus mild effects.

Real Data Analysis

TO DO

Discussion

TO DO

References

BEISSINGER, T., J. KRUPPA, D. CAVERO, N.-T. HA, M. ERBE, *et al.*, 2018 A simple test identifies selection on complex traits. Genetics **209**: 321–333.

BENJAMINI, Y., and Y. HOCHBERG, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B : 289–300.

BOYKO, A. R., S. H. WILLIAMSON, A. R. INDAP, J. D. DEGENHARDT, R. D. HERNANDEZ, *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet **4**: e1000083.

CABALLERO, A., A. TENESA, and P. D. KEIGHTLEY, 2015 The nature of genetic variation for complex traits revealed by gwas and regional heritability mapping analyses. Genetics **201**: 1601–13.

CONNALLON, T., and A. G. CLARK, 2015 The distribution of fitness effects in an uncertain world. Evolution **69**: 1610–1618.

EYRE-WALKER, A., 2010 Evolution in health and medicine sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc Natl Acad Sci U S A **107 Suppl 1**: 1752–6.

HALLER, B. C., and P. W. MESSER, 2019 Slim 3: Forward genetic simulations beyond the wright-fisher model. Mol Biol Evol **36**: 632–637.

LOURENÇO, J., N. GALTIER, and S. GLÉMIN, 2011 Complexity, pleiotropy, and the fitness effect of mutations. Evolution **65**: 1559–71.

SABETI, P. C., D. E. REICH, J. M. HIGGINS, H. Z. P. LEVINE, D. J. RICHTER, *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419**: 832–7.

TANG, K., K. R. THORNTON, and M. STONEKING, 2007 A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol **5**: e171.

URICCHIO, L. H., H. C. KITANO, A. GUSEV, and N. A. ZAITLEN, 2019 An evolutionary compass for detecting signals of polygenic selection and mutational bias. Evol Lett **3**: 69–79.

WANG, J., A. CABALLERO, P. D. KEIGHTLEY, and W. G. HILL, 1998 Bottleneck effect on genetic variance. a theoretical investigation of the role of dominance. Genetics **150**: 435–47.

ZENG, J., A. XUE, L. JIANG, L. R. LLOYD-JONES, Y. WU, *et al.*, 2019 Bayesian analysis of gwas summary data reveals differential signatures of natural selection across human complex traits and functional genomic categories. bioRxiv .

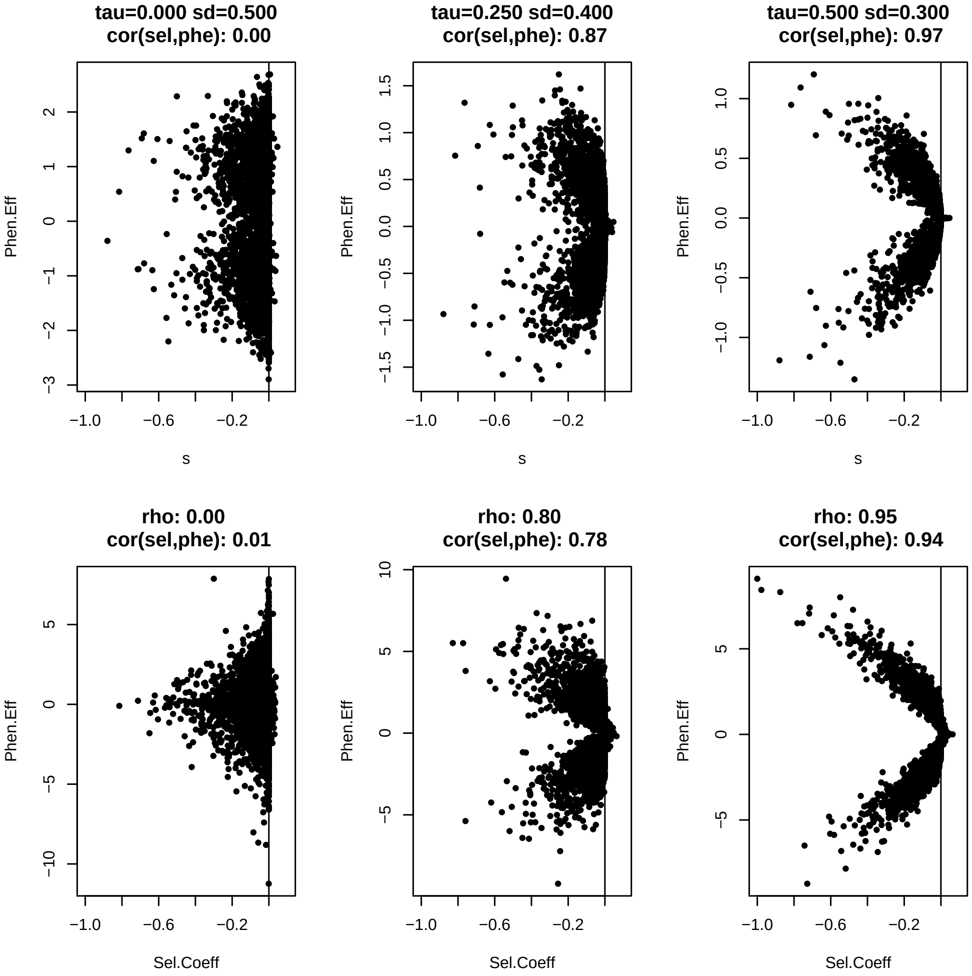


Figure S1: Comparison of distributions between Eyre-Walker (EW) function and a Gamma Bivariant correlated distribution. The selective effect follows a gamma distribution with shape=0.2 and mean=-0.03 for negative side and shape=1 and mean=0.01 for positive side and a proportion of positive variants of 5e-3. the first row show the distribution of Selective (x-axis) and Phenotype (y-axis) effects for three different conditions given the EW function (left: $\tau = 0.0$, $sd = 0.5$, center: $\tau = 0.25$, $sd = 0.4$, right: $\tau = 0.50$, $sd = 0.4$), The second row show the distribution of Selective and Phenotype effects for three different correlation conditions (left: $\rho = 0.0$, center: $\rho = 0.80$ and left $\rho = 0.95$) using the Gamma distribution.

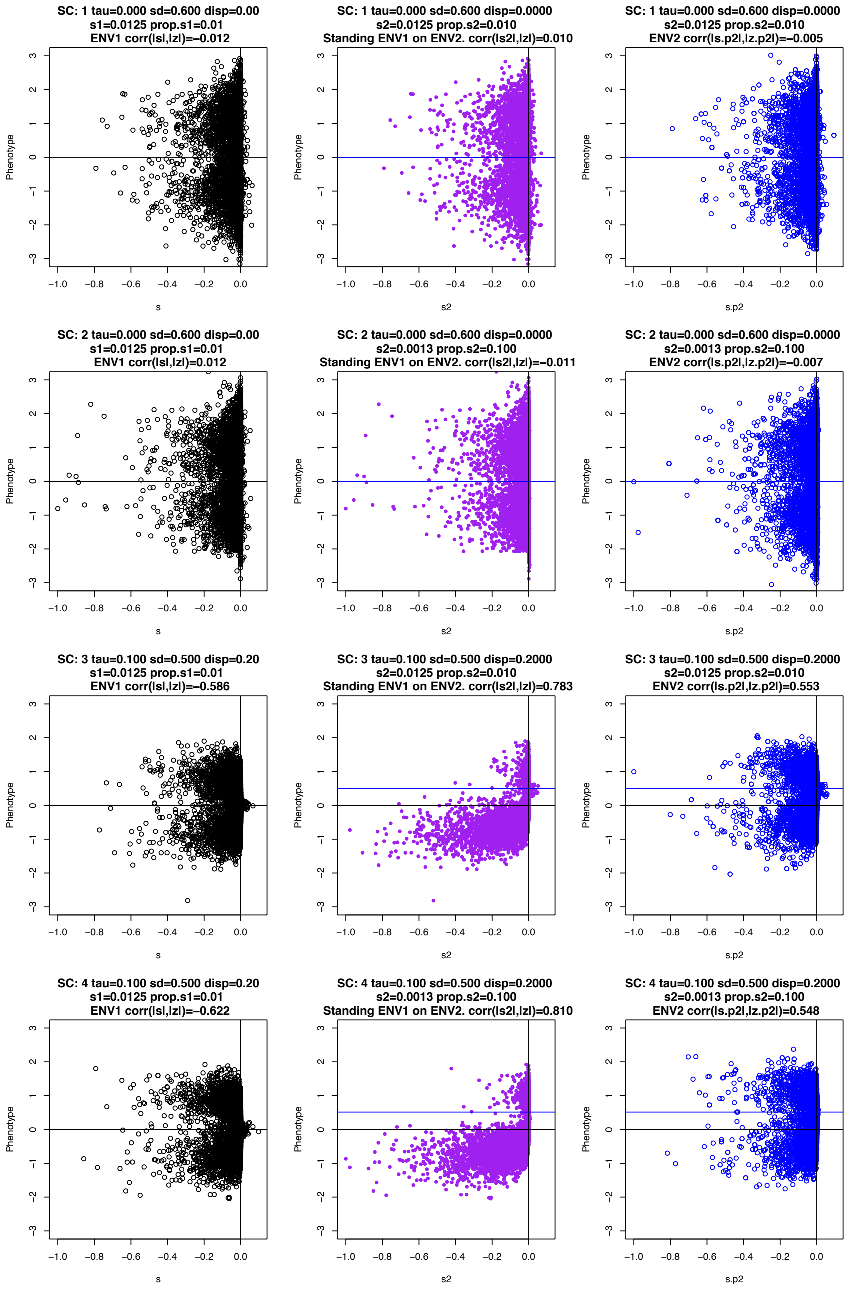
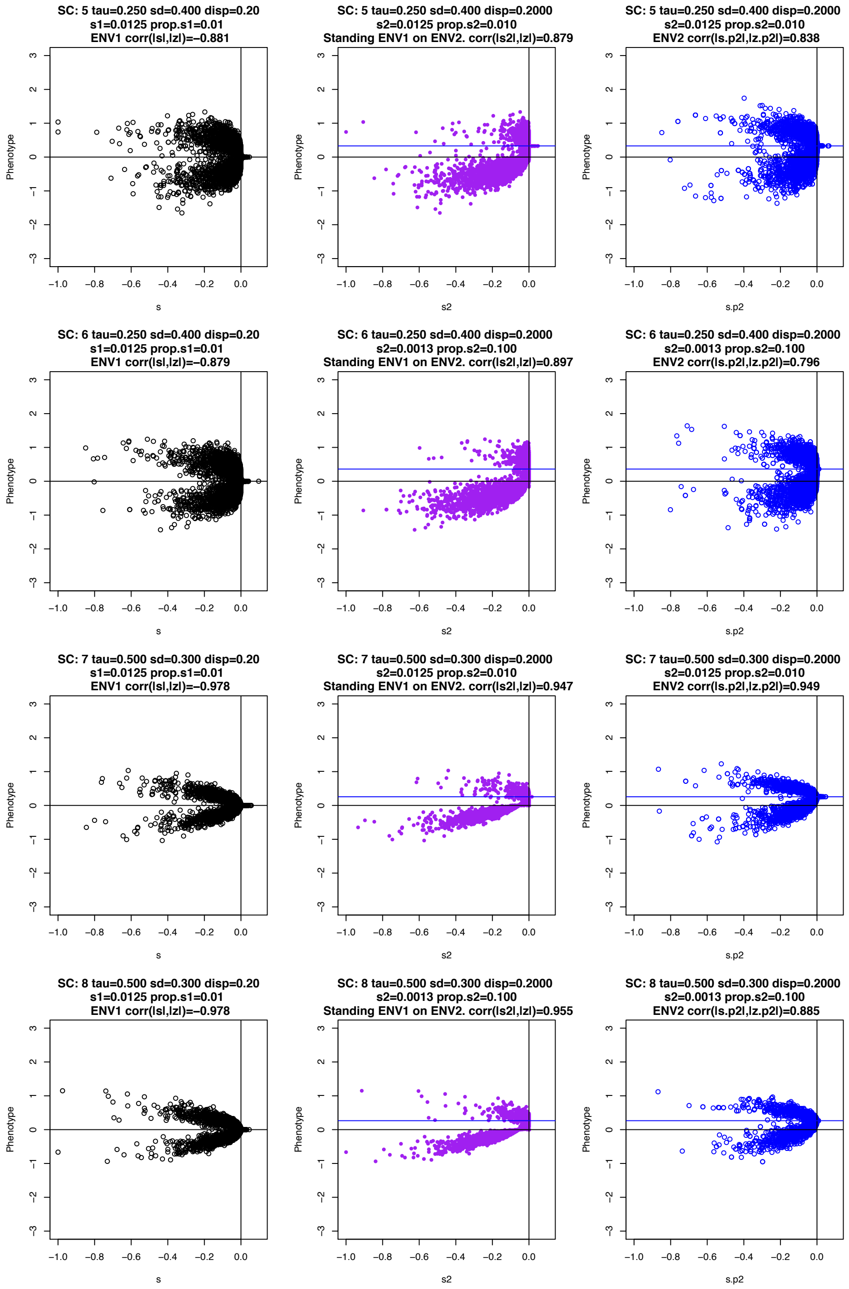


Figure S2: Distribution of the beneficial and deleterious mutations for scenarios 1,2,3, and 4. Each row indicates one scenario. Horizontal lines indicate the optimal phenotypic value for the first (black) and second scenario (blue). Vertical line separate deleterious from beneficial selective effects. At left side, Black circles indicate the mutations for the initial environment, at central, purple dots indicate the displacement of selective effects for standing mutations at the new second environment. At right side, the distribution of new mutations at the second environment (in blue) in relation to its selective and phenotype effect.



SC: 5 tau=0.250 sd=0.400 disp=0.20
s1=0.0125 prop.s1=0.01
ENV1 corr(|s|,|z|)=−0.881

SC: 5 tau=0.250 sd=0.400 disp=0.2000
s2=0.0125 prop.s2=0.010
Standing ENV1 on ENV2. corr(|s2|,|z|)=0.879

SC: 5 tau=0.250 sd=0.400 disp=0.2000
s2=0.0125 prop.s2=0.010
ENV2 corr(|s.p2|,|z.p2|)=0.838

SC: 6 tau=0.250 sd=0.400 disp=0.20
s1=0.0125 prop.s1=0.01
ENV1 corr(|s|,|z|)=−0.879

SC: 6 tau=0.250 sd=0.400 disp=0.2000
s2=0.0013 prop.s2=0.100
Standing ENV1 on ENV2. corr(|s2|,|z|)=0.897

SC: 6 tau=0.250 sd=0.400 disp=0.2000
s2=0.0013 prop.s2=0.100
ENV2 corr(|s.p2|,|z.p2|)=0.796

SC: 7 tau=0.500 sd=0.300 disp=0.20
s1=0.0125 prop.s1=0.01
ENV1 corr(|s|,|z|)=−0.978

SC: 7 tau=0.500 sd=0.300 disp=0.2000
s2=0.0125 prop.s2=0.010
Standing ENV1 on ENV2. corr(|s2|,|z|)=0.947

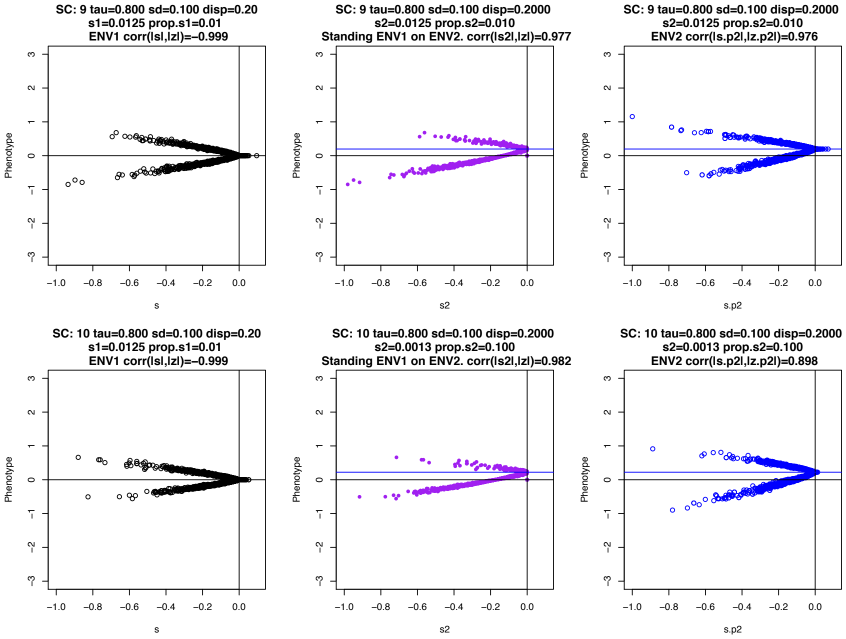
SC: 7 tau=0.500 sd=0.300 disp=0.2000
s2=0.0125 prop.s2=0.010
ENV2 corr(|s.p2|,|z.p2|)=0.949

SC: 8 tau=0.500 sd=0.300 disp=0.20
s1=0.0125 prop.s1=0.01
ENV1 corr(|s|,|z|)=−0.978

SC: 8 tau=0.500 sd=0.300 disp=0.2000
s2=0.0013 prop.s2=0.100
Standing ENV1 on ENV2. corr(|s2|,|z|)=0.955

SC: 8 tau=0.500 sd=0.300 disp=0.2000
s2=0.0013 prop.s2=0.100
ENV2 corr(|s.p2|,|z.p2|)=0.885

Figure S3: Distribution of the beneficial and deleterious mutations for scenarios 5,6,7, and 8. Each row indicates one scenario. Horizontal lines indicate the optimal phenotypic value for the first (black) and second scenario (blue). Vertical line separate deleterious from beneficial selective effects. At left side, Black circles indicate the mutations for the initial environment, at central, purple dots indicate the displacement of selective effects for standing mutations at the new second environment. At right side, the distribution of new mutations at the second environment (in blue) in relation to its selective and phenotype effect


SC: 9 tau=0.800 sd=0.100 disp=0.20
s1=0.0125 prop.s1=0.01
ENV1 corr(lsl,lzl)=–0.999

SC: 9 tau=0.800 sd=0.100 disp=0.2000
s2=0.0125 prop.s2=0.010
Standing ENV1 on ENV2. corr(ls2l,lzl)=0.977

SC: 9 tau=0.800 sd=0.100 disp=0.2000
s2=0.0125 prop.s2=0.010
ENV2 corr(ls.p2l,lz.p2l)=0.976

SC: 10 tau=0.800 sd=0.100 disp=0.20
s1=0.0125 prop.s1=0.01
ENV1 corr(lsl,lzl)=–0.999

SC: 10 tau=0.800 sd=0.100 disp=0.2000
s2=0.0013 prop.s2=0.100
Standing ENV1 on ENV2. corr(ls2l,lzl)=0.982

SC: 10 tau=0.800 sd=0.100 disp=0.2000
s2=0.0013 prop.s2=0.100
ENV2 corr(ls.p2l,lz.p2l)=0.898

Figure S4: Distribution of the beneficial and deleterious mutations for scenarios 9 and 10. Each row indicates one scenario. Horizontal lines indicate the optimal phenotypic value for the first (black) and second scenario (blue). Vertical line separate deleterious from beneficial selective effects. At left side, Black circles indicate the mutations for the initial environment, at central, purple dots indicate the displacement of selective effects for standing mutations at the new second environment. At right side, the distribution of new mutations at the second environment (in blue) in relation to its selective and phenotype effect