

# An Evolutionary Perspective on Epistasis and the Missing Heritability

Gibran Hemani<sup>1\*</sup>, Sara Knott<sup>2</sup>, Chris Haley<sup>1,3</sup>

**1** The Roslin Institute and Royal (Dick) School of Veterinary Science, University of Edinburgh, Edinburgh, United Kingdom, **2** MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, United Kingdom, **3** Institute for Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom

## Abstract

The relative importance between additive and non-additive genetic variance has been widely argued in quantitative genetics. By approaching this question from an evolutionary perspective we show that, while additive variance can be maintained under selection at a low level for some patterns of epistasis, the majority of the genetic variance that will persist is actually non-additive. We propose that one reason that the problem of the “missing heritability” arises is because the additive genetic variation that is estimated to be contributing to the variance of a trait will most likely be an artefact of the non-additive variance that can be maintained over evolutionary time. In addition, it can be shown that even a small reduction in linkage disequilibrium between causal variants and observed SNPs rapidly erodes estimates of epistatic variance, leading to an inflation in the perceived importance of additive effects. We demonstrate that the perception of independent additive effects comprising the majority of the genetic architecture of complex traits is biased upwards and that the search for causal variants in complex traits under selection is potentially underpowered by parameterising for additive effects alone. Given dense SNP panels the detection of causal variants through genome-wide association studies may be improved by searching for epistatic effects explicitly.

**Citation:** Hemani G, Knott S, Haley C (2013) An Evolutionary Perspective on Epistasis and the Missing Heritability. *PLoS Genet* 9(2): e1003295. doi:10.1371/journal.pgen.1003295

**Editor:** Trudy F. C. Mackay, North Carolina State University, United States of America

**Received:** July 21, 2012; **Accepted:** December 17, 2012; **Published:** February 28, 2013

**Copyright:** © 2013 Hemani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** GH would like to thank the BBSRC, Biosciences KTN, and Newsham Choice Genetics for the funding provided for his PhD. SK acknowledges funding from the BBSRC. CH is grateful for funding from the BBSRC and the MRC. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (<http://www.ecdf.ed.ac.uk/>). The ECDF is partially supported by the eDIKT initiative (<http://www.edikt.org.uk>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: [g.hemani@uq.edu.au](mailto:g.hemani@uq.edu.au)

## Introduction

There exists a paradox in evolutionary biology. Despite a near-ubiquitous abundance of genetic variation [1] traits under selection often evolve more slowly than expected and, contrary to expectation, genetic variation is maintained under selection. This problem is known as ‘stasis’ [2,3], and it is particularly evident in fitness-related traits where the genetic variation tends to be highest [4] yet there is commonly no observed response to selection at all [5–7]. There are a number of mechanisms by which this might arise, amongst which the most commonly cited are various forms of constraints [8,9] or stabilising selection [10]. Because stasis is widespread its properties may reveal insights into the genetic architecture of complex traits related to fitness and thus inform the strategies that are employed to detect their underlying genetic variants. After hundreds of genome-wide association (GWA) studies [11] a picture is emerging where the total genetic variation explained by variants that have been individually mapped to complex traits is vastly lower than the amount of genetic variation expected to exist as estimated from pedigree-based studies, a phenomenon that has come to be known as the problem of the ‘missing heritability’ [12]. Again, there are probably numerous contributing factors, and ostensibly the most parsimonious explanation is that complex traits comprise many small effects that GWA studies are underpowered to detect

[13,14], but whether this is the complete story deserves exploration.

With respect to the fields of both the aforementioned issues, it is typical to model genetic variation using an additive framework, such that each allele affecting a trait acts in an independent, linear, cumulative manner. For many practical applications this is a very useful approach (*e.g.* [15,16]), but there does exist a popular school of thought that suggests that the mechanisms of gene action, and the architecture of complex traits, are actually much more complex than the additive model allows (*e.g.* [17–20]). Epistasis, defined in functional terms as the event whereby the effect of one locus depends on the genotype at another locus, is one source of non-additive genetic variation. How it contributes to both the paradox of ‘stasis’ and the problem of the ‘missing heritability’ will be the focus of this study.

The importance of epistasis in complex traits has proven to be a particularly divisive issue throughout the history of quantitative genetics. Recently it has been suggested that epistasis might form part of the answer to the ‘missing heritability’ [21–24], but how this might manifest is not immediately obvious. When heritability estimates are reported for complex traits they typically pertain to the narrow-sense ( $h^2$ , the proportion of the phenotypic variance that is due to additive genetic effects). Calculating the broad-sense heritability ( $H^2$ , or the proportion of variance that is due to both additive and non-additive genetic effects), is an intractable

## Author Summary

In this study we have shown that two independent problems may have a common cause. Why do traits under selection exhibit additive genetic variance, and why is the proportion of the heritability explained by additive effects much smaller than the total heritability estimated to exist? Our results indicate that epistatic interactions can allow deleterious mutations to persist under selection and that these interactions can abate the depletion of additive genetic variation. Furthermore, a much larger element of non-additive genetic variance is maintained, which supports the notion that the heritability estimated from family studies could be a mixture of both additive and non-additive components. We show that searching directly for epistatic effects greatly improves the discovery of variants under selection, despite the multiple testing penalty being much larger. Finally, we demonstrate that common practices in genome-wide association studies could lead to both an ascertainment bias in detecting additive effects and a confirmation bias in perceiving that most of the genetic variance is additive.

problem for non-clonal populations [25], thus we have little knowledge of how much epistasis exists in human and animal traits. In this light one might suggest that we are actually dealing with two problems: the ‘missing heritability’, and the ‘unknown heritability’. By definition epistasis will form a part of the ‘unknown heritability’, but theory shows that epistatic interactions could also contribute to  $h^2$  estimates. This could arise through two possible mechanisms, firstly by generating real additive variation as marginal effects from higher order genetic interactions [26–29]; or secondly by creating a statistical illusion of additive variance through confounding between non-additive and common environment effects in twin study based estimates [24,30].

Beyond the realm of complex trait genetics it appears that epistasis does appear to be common. For example in molecular studies it is routine to observe ‘phenotypic rescue’ where the phenotypic effect of a gene knockout can be masked by a second knockout (e.g. [31]). Another commonly encountered form of epistasis is ‘canalisation’ [32], where phenotypes exhibit robustness to the knockout of one gene, requiring a second knockout to elicit a response (e.g. [33]). At the macroevolutionary scale, epistasis is also of central importance, for example it has recently been shown that an advantageous substitution in one species is often found to be deleterious in others, thus the substitution’s effect on fitness is dependent upon the genetic background in which it is found [34]. The mechanisms behind pathway-level [32,35,36] or species-level epistasis [20,34,37] are widely explored, and yet at the intermediate, within-population level there is a distinct lack of evidence for any widespread importance of epistasis arising from natural variation, and most genetic variation appears to be additive [28]. Nevertheless some convincing examples of epistasis have been reported, for instance there are a number of cases of canalisation in *Homo sapiens* [38,39], *Gallus gallus* [40], *Drosophila melanogaster* [41], *Saccharomyces cerevisiae* [42], and *Arabidopsis thaliana* [43] to name but a few.

At the statistical level, for a pair of single nucleotide polymorphisms (SNPs) that exhibit epistasis, in addition to interaction terms between the two loci, the total genetic effect is likely to also include marginal additive or dominance effects at each locus [28,44]. The proportion of additive to non-additive genetic variation will depend both on the genotype-phenotype map (G-P map), and the allele frequencies at each locus. In turn

these frequencies will depend on selection acting on the phenotype. Thus, if epistasis contributes towards fitness then how selection acts is highly dependent on the particular genotype-phenotype map in question [45]. Ostensibly, the additive framework that is used in GWA studies follows Occam’s razor, employing the hypothesis that introduces the fewest new assumptions (i.e. non-additive variation cannot be estimated, thus SNPs are not modelled to have non-additive effects). But whether the phenomenon of stasis can accommodate a purely additive genetic model remains an open question.

The premise of this study is centred around finding common ground between the problems of stasis and the missing heritability. Given that fitness related traits often exhibit stasis then the underlying genetic architecture may not solely comprise independent additive effects. Through theory and simulations we demonstrate that epistasis will maintain additive genetic variation under selection at higher levels than independent additive effects, and that by extending GWA studies to search for epistasis directly we could improve statistical power to detect additive genetic variation.

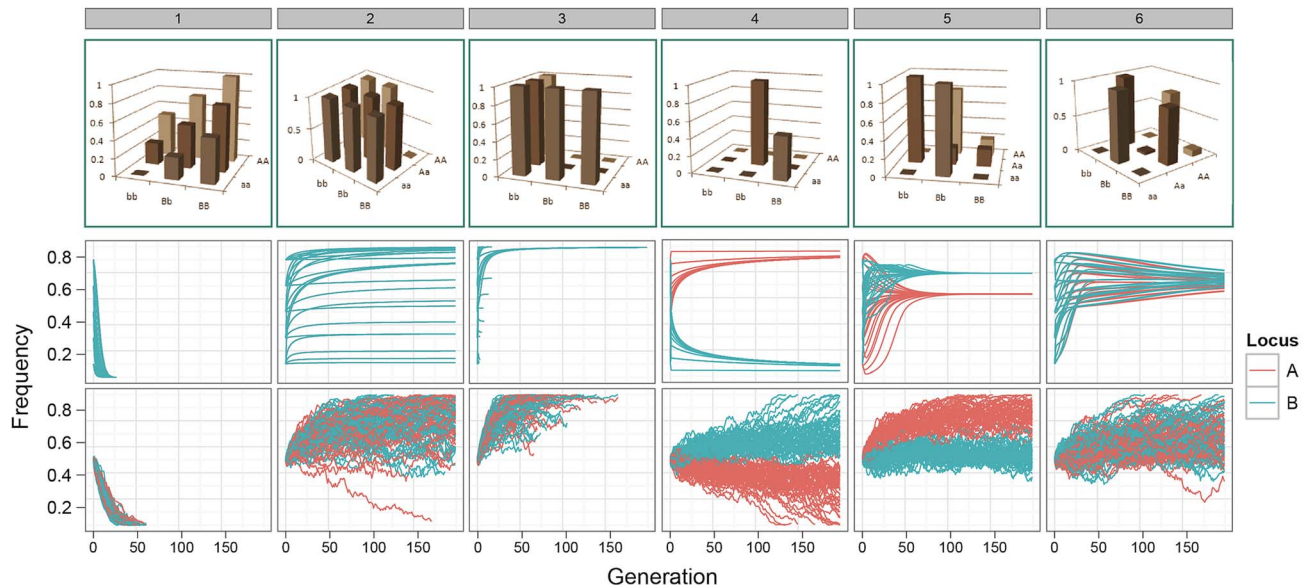
## Results

Selection rapidly drives deleterious additive effects to fixation, but how effective is selection at purging deleterious, non-additive effects? We simulated 56 G-P maps (including neutral, additive, dominant, and 51 epistatic two-locus patterns; Figure S1) and assuming that the phenotype had a direct effect on fitness we calculated their expected allele frequency trajectories over time. With these outcomes we were able to make inferences on i) the ability of epistasis to maintain genetic variation and the allele frequencies at which different G-P maps might stabilise, ii) the amount of genetic variation and the proportion of additive to non-additive variation that we would expect at frequencies that are evolutionarily stable, iii) the impact of incomplete linkage disequilibrium on our estimation of these G-P maps with SNP data, and iv) the relative performances of various GWA strategies at detecting additive genetic variation. The results are detailed below.

### Epistasis maintains genetic variation under selection

Our results demonstrate that for many of the patterns of epistasis that we assayed, deleterious effects can be maintained at intermediate frequencies over long evolutionary time periods (Figure 1 patterns 1–3, and Figures S1, S2 and S3). As might be expected, a number of G-P maps that maintained genetic variation at intermediate frequencies also exhibited over-dominance, or some form of heterozygote advantage (e.g. Figure 1 patterns 4–6). However, most patterns of epistasis that we assayed (Figure S1) do not exhibit heterozygote advantage, and these can also effectively temper the rate of extinction of deleterious alleles. Conversely, some level of under-dominance is required for variation to be maintained, for example although the classic  $A \times A$  pattern of epistasis (pattern 52, Figure S1) can theoretically avoid fixation when both loci are at allele frequencies of 0.5 (Figure S2), drift provides sufficient perturbation to prevent it from being maintained at equilibrium (Figure S3).

The consequences of these results are examined below. In summary, we show that a small amount of additive variation is maintained by epistasis but most genetic variation is non-additive; that there is a strong bias in GWA studies that lead to an overestimation of additive effects at QTLs; and that, perhaps counterintuitively, the most powerful way to uncover additive



**Figure 1. Allele frequency trajectories under selection.** Top row: G-P maps. 1. Independent additive effects at locus A and B; 2. Dominant pattern of canalisation; 3. Recessive pattern of canalisation; 4–6. Patterns generated by a genetic algorithm optimising for maximised additive variance and long-term survival at intermediate frequency. Middle row: Expected allele frequency trajectories for G-P maps under selection, as derived deterministically, with initial frequencies of 0.1, 0.3, 0.5, 0.7, and 0.9 enumerated over both loci. Frequencies on the y-axis correspond to alleles *a/b*. Only one colour appears for patterns 1–3 because the trajectories of both alleles are identical. Bottom row: The path of allele frequencies as observed through stochastic simulations of populations comprising 1000 individuals and  $H^2 = 10\%$  at generation 0, with initial allele frequencies at both loci of 0.5.

doi:10.1371/journal.pgen.1003295.g001

variation under selection is to parameterise the search to include epistatic effects using dense genotype information.

### Most genetic variation maintained under selection is non-additive

The genetic variance of a G-P map depends on the allele frequencies of the loci involved, and selection drives these allele frequencies to minimise the directional effect of each locus. From this we can calculate the expected changes in genetic variance over time. For many of the patterns of epistasis studied they maintain genetic variance over long evolutionary periods (Figure 2a and Figure S5), as often their allele frequencies can be maintained at intermediate levels. However the majority of this variance is non-additive, with almost all of the additive variance eventually disappearing (Figure 2b and Figure S6).

Although this study assesses a large number of G-P maps, because the parameter space of epistatic G-P maps is effectively infinite the question of how much additive variance can possibly be maintained under selection by a two-locus system still remains. To answer this we used a genetic algorithm to heuristically search the parameter space of the two-locus GP-map, with the objective of finding epistatic patterns that maintain high additive variance over evolutionary time. We should note that the purpose of this exercise is not to identify biologically feasible patterns *per se*, rather it is to assess the propensity for additive variance to be maintained under selection. The epistatic patterns that emerged with the highest level of maintained additive variance, as a proportion of total genetic variance, are shown in Figure 1 (patterns 4–6). The main feature of these patterns is that they exhibit overdominance, and that even in these extreme cases where the algorithm attempts to generate the G-P maps with the largest possible maintained additive variance, it is clear that the majority of genetic variance that is maintained will still be non-additive.

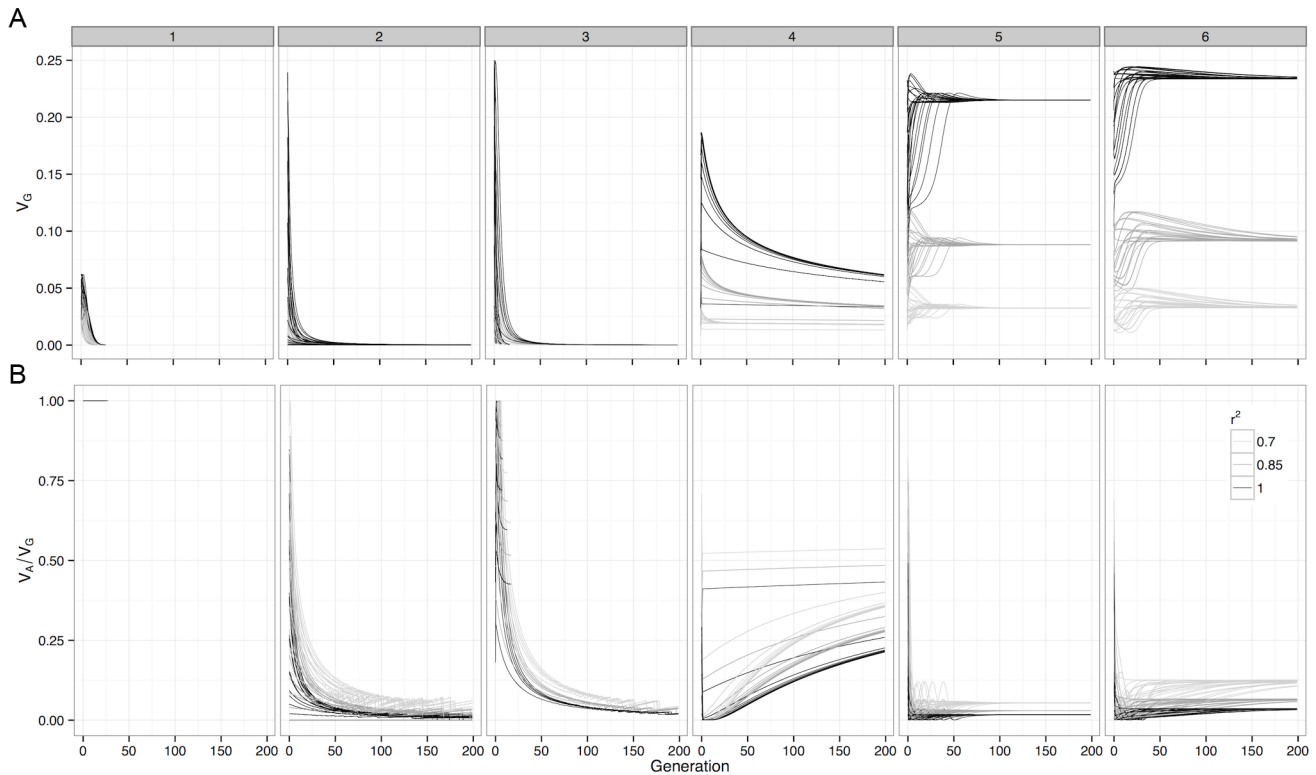
### Incomplete LD leads to the erroneous belief that genetic effects are mostly additive

While it appears that additive variance is difficult to maintain either through independent additive effects or through epistatic interactions, it is clear that most causal effects that have been discovered through GWA studies appear additive in nature [11]. This paradox may result from both ascertainment and confirmation biases that arise when there is incomplete linkage disequilibrium between underlying causal variants and the observed SNPs in a GWA study.

Figure 2b and Figure S6 show that although estimated genetic variance at observed SNPs decreases as LD with causal variants decreases, the estimated proportion of the variance that is additive actually increases. To illustrate this further Figure 3 shows how estimates of epistatic GP-maps change when LD is reduced, and two important biases can be shown.

Firstly, the higher order variance components (rows 3–5) rapidly haemorrhage genetic variance (see Figure S8), such that even at LD of  $r^2 = 0.5$  the genotype class means are close to identical. This means that detection is strongly dependent upon high or complete LD, even when effect sizes are large, so most epistatic mutations will remain undetected and their prevalence underestimated. Hence, because additive variance decays linearly with LD [46], at low LD they remain detectable leading to an ascertainment bias for additive vs non-additive effects.

Secondly, with the patterns of canalisation (rows 1–2), as LD reduces although some genetic variance is maintained, the G-P map appears entirely additive. Thus functional maps that confer epistatic effects that can be detected at relatively low LD are likely to be interpreted as being entirely additive. Thus, researchers who attempt to quantify the contribution of non-additive variance from SNPs associated with a trait are liable to incorrectly confirm that most variants act additively.



**Figure 2. Deterministic change in variance components of G-P maps under selection.** For (a) Genetic variance and (b) Additive variance as a proportion of genetic variation, with initial frequencies of 0.1, 0.3, 0.5, 0.7 and 0.9 enumerated over both loci. The variance decomposition was performed at the causal locus ( $r^2 = 1$ ), and at SNP pairs that were in incomplete LD with the causal loci. Boxes represent the different G-P maps from Figure 1.

doi:10.1371/journal.pgen.1003295.g002

### Searching for epistatic effects improves power to detect persistent additive variation

Typically GWA studies test each SNP one at a time for additive effects. To explicitly include interaction terms in the search this approach can be extended from one dimension (1D) to two dimensions (2D), where every pair of SNPs is tested jointly for an association [26,47].

Given that we know the trajectory of allele frequencies under selection, it is possible to ask what the best GWA strategy for detecting evolutionarily likely variants might be. Two main search methods were tested, 1D scans (as are typically performed in GWA studies), and exhaustive 2D scans (previously computationally unfeasible until the availability of more advanced software [47]). For each search method various different model parameterisations were also tested. We used a Bonferroni correction for all methods, so 2D scans had a much more severe multiple testing penalty than 1D scans.

Broadly, the results show two important points. Firstly, there is no single method that is always superior under the conditions that were tested. Secondly, it is very rare that parameterising for additive effects is the most powerful method (Figure 4a and Figure S7). Rather, if LD is no higher than, for example  $r^2 = 0.7$ , between causal variants and observed SNPs then one dimensional scans, although not particularly powerful in absolute terms, are most effective provided that both additive and dominance effects are modelled (2 d.f. test). In the case of very high LD (e.g. dense marker panels, imputed data, sequence data), a strong advantage in power to detect variants at evolutionarily likely frequencies can be

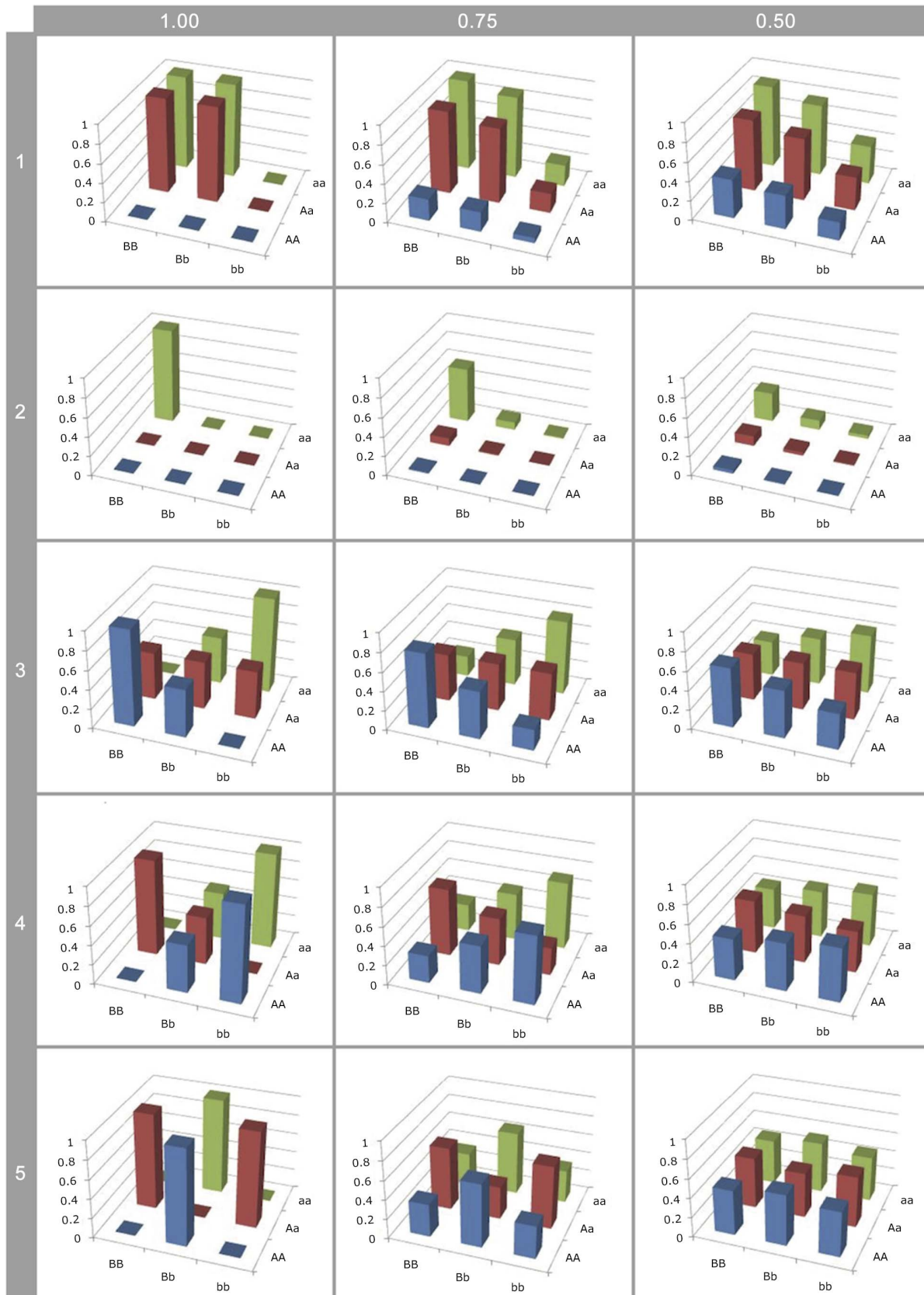
conferred by using exhaustive two-dimensional scans and modelling whole genotype effects (8 d.f. test).

A more detailed view of this relationship between LD and detectability is shown in Figure 4b (patterns 4–6). When LD between causal variants and observed SNPs is high, although additive variation exists, much greater power can be achieved in its detection if the search focuses on the larger non-additive variance components. However, as LD decreases, the proportion of the genetic variance that is additive increases, thus one dimensional scans gain an advantage. Nevertheless, as one might expect at low LD there is in general very little power from any method.

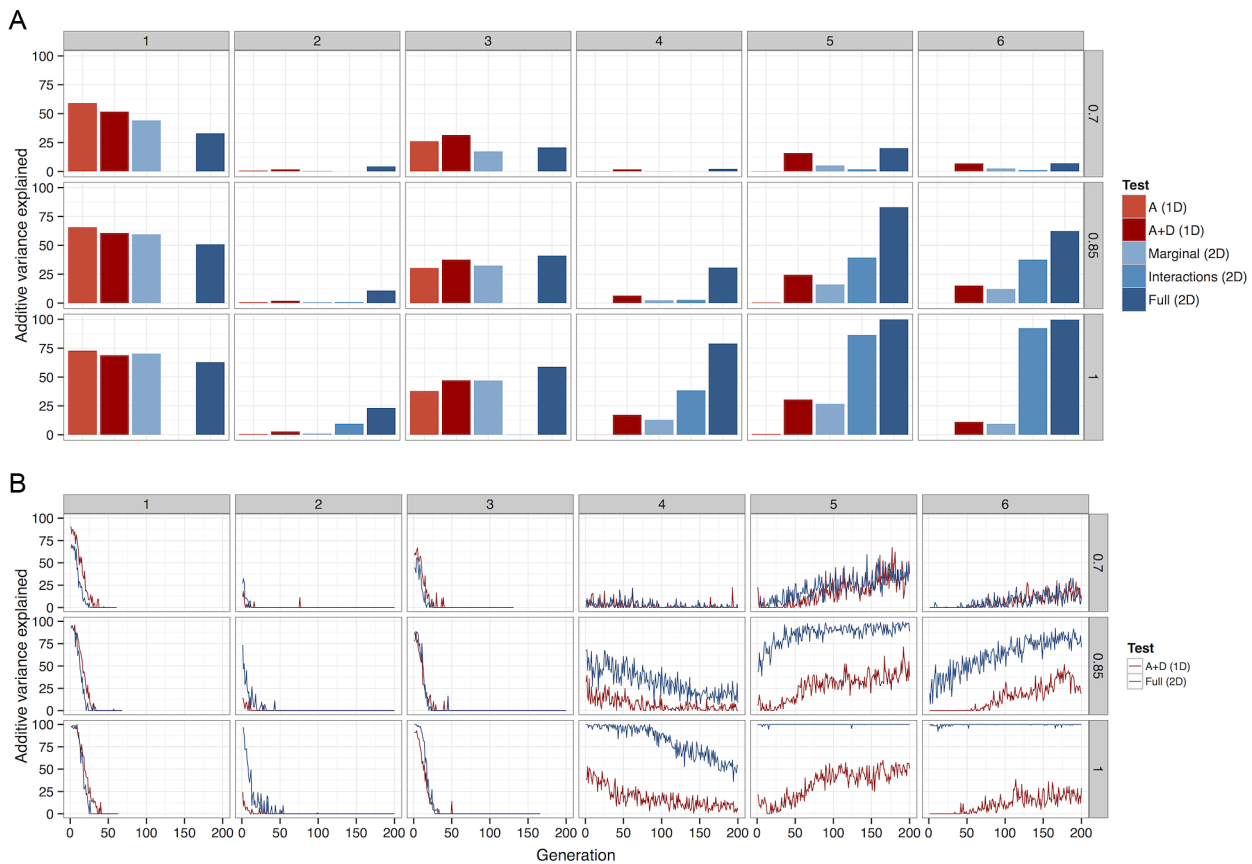
It may seem surprising that despite the much larger multiple testing penalty, the 2D scans perform well in terms of power. But there exists a trade-off between the extra variance explained by extending the search into higher dimensions, and the amount of variance required to be detected in order to overcome the multiple testing correction. The results show that because non-additive variance components can be maintained under selection the 2D strategy is conferred an advantage in this trade-off.

### Discussion

The architecture of genetic variation must be understood if we are to make progress in fields such as disease risk prediction, personalised medicine, and animal and crop breeding. This study sought to examine the potential for epistasis to maintain genetic variation under selection, and thus to inform GWA strategies based on these results.



**Figure 3. Effect of LD on G-P map estimation.** Different G-P maps of causal variants (rows of graphs) deterministically calculated from neighbouring SNPs in different levels of linkage disequilibrium (columns of graphs). All SNP and causal variant frequencies are set to 0.5. Rows 1–2: Canalisation; 3:  $A \times A$ ; 4:  $A \times D$ ; 5:  $D \times D$ . doi:10.1371/journal.pgen.1003295.g003



**Figure 4. The percentage of total additive variance detected by each of 5 different methods.** Columns of graphs refer to G-P maps (Figure 1), rows refer to  $r^2$  between causal variants and observed SNPs. (a) Deterministic calculations were performed 25 times, each with different initial allele frequencies. The percentage of additive variance explained is summed across all runs and generations. (b) The summed  $V_A$  detected at each generation as a percentage of the summed  $V_A$  simultaneously present in 50 populations. For clarity, only the most powerful 1D test (A+D) is compared against the most powerful 2D test (full parameterisation). A Bonferroni threshold was used,  $1.7 \times 10^{-7}$  for 1D strategies and  $1.1 \times 10^{-13}$  for 2D strategies.

doi:10.1371/journal.pgen.1003295.g004

We investigated to what extent deleterious mutations could be maintained as common polymorphisms under selection. A large sample of potential G-P maps were assayed [48] in order to develop a broad picture of the general behaviour of epistasis under selection, and this was extended further by heuristically searching through the parameter space of epistatic G-P maps. It was demonstrated that the maintenance of genetic variation at intermediate frequencies, for traits under selection over evolutionary time, could be achieved through a wide range of two-locus epistatic models. By definition, such is not the case for independent additive effects (Figures S2 and S3). Following on, it was demonstrated that even in the best case scenario, where G-P maps were generated to maximise additive variance, total genetic variance was mostly composed of non-additive components (Figure 2b and Figure S6). This finding is in disagreement with a recent study [28], which showed that for various two-locus epistatic models, the deterministic partitions of genetic variance calculated across different frequency distributions were largely comprised of the additive component. Here we show that those allele frequencies at which additive variance is high (a large proportion of the frequency spectrum), are evolutionarily unstable, thus should epistatic variants be affecting fitness traits then the majority of the variance will be non-additive. Ultimately there is no simple mechanism whereby two-locus epistasis will significantly contribute towards the missing heritability, unless  $h^2$  estimates

have been contaminated by non-additive genetic components or common environment effects. This is a well-known potential problem with full-sib based estimates and twin studies [30]. Indeed, a recent examination of this problem showed that additive variance estimates could be inflated significantly when complex traits are controlled by epistasis [24].

The results suggest that we should expect significant levels of non-additive variation to be maintained in fitness-related traits. While non-additive variance components are often considered to be nuisance terms in quantitative genetics [49], their existence can be levered to actually improve the detection of additive variance. Here the premise is that if additive variation is observed then there is likely to be an accompanying non-additive genetic component that allows it to persist in the population. Power comparisons were made between 1D and 2D scans, as well as different model parameterisations, with a view to testing the power to detect variants under selection at evolutionarily likely frequencies. Surprisingly, the simplest and most widely used parameterisation, modelling for additive effects in one dimension, was seldom the most powerful approach. On the contrary, because other forms of genetic variance are co-precipitated along with additive variance, by parameterising the tests to include them the power was seen to improve. However, it was observed that even with modest reductions in LD between causal variants and observed SNPs all testing strategies tended to decline in performance rapidly. This

leaves researchers in a difficult situation, where the strategy of increasing SNP panel densities as an intuitive response to improve LD coverage comes at a quadratic cost (in the two-locus case) in computation time and multiple testing penalties. An important outcome here is that there is no single test with consistently superior performance, and this resonates with the idea of the “no free lunch” theorem, which states that although competing algorithms will behave differently under different conditions, they will have identical performance when averaged across all conditions [50]. The key in such a situation is to know which conditions are most likely to manifest in the data, and here our argument is that for fitness traits non-additive effects are likely to exist at frequencies where additive variance is minimised.

Although the intention behind the use of the genetic algorithm in this study was to explore the potential for a two-locus system to maintain additive variance, rather than to necessarily identify biologically feasible maps, those maps that emerged did not appear biologically untenable. In fact they can be supported by reports in the literature due to their tendencies for exhibiting heterozygote advantage [51,52]. The example of the single locus case, overdominance, is central to processes of heterosis and inbreeding depression [52,53], and has been identified in molecular studies also [54,55]. Indeed, heterozygote advantage plays an important role in evolutionary theory, as it confers segregational load on a population, and this type of load cannot be purged due to balancing selection, potentially rendering populations susceptible to accumulating a critical mass of such polymorphisms [56]. The idea of a critical mass of deleterious mutations has been widely explored in amictic haploid populations, particularly in the context of Muller’s ratchet, and in this case synergistic epistasis has been suggested as a mechanism that could alleviate the problem in some situations [57,58]. This study may offer a similar answer for the analogous problem of segregational load in diploid populations, because it can be observed that while patterns of overdominance (Figure S3, pattern 55) form a stable equilibrium, small perturbations to this G-P map through the introduction of an interacting locus (*e.g.* patterns 45, 47, 53) could destabilise the equilibrium and lead to eventual fixation.

It is important to note that the processes underlying stasis and missing heritability are unlikely to be caused by any single factor. For example, a compelling argument is that though most traits exhibit genetic variation, selection acts upon multidimensional trait space in which there is no genetic variation [59], and this will hold under an additive model of genetic variation. It is also important to consider the manner in which traits of interests, such as human diseases, are involved in fitness. For example in an assessment of selection signatures on SNPs implicated in type 1 diabetes it has been shown that the causal alleles have undergone positive selection to a greater extent than protective alleles, while with Crohn’s disease the converse is true [60]. In the case of both diseases more variants are being discovered as sample sizes increase [61,62], but given that only a small proportion of the total heritability has so far been explained, and the search has concentrated on additive variants only, inferences about the genetic architecture cannot be made.

Occam’s razor can be invoked to justify the additive paradigm used in GWA studies [28]. But the analyses presented here demonstrate that perhaps rejecting more complex models in favour of simple ones should not always be the automatic choice. With sample sizes growing and with the tools now available to search for epistasis in a computationally efficient manner (*e.g.* [47,63–66]) it should be possible to explore the genetic architecture of complex traits in directions that were not previously possible.

## Methods

We were interested in simulating the behaviour of epistatic interactions under selection in order to ask the questions i) for how long and at which frequencies are deleterious mutations maintained under selection, ii) how much additive variance can this produce, and iii) what is the best strategy to identify evolutionarily persistent variation. We approached this problem through two methods, using stochastic simulations and by calculating the expected trajectories deterministically, and overall a range of different epistatic and dominant G-P maps were assessed. In addition, we heuristically searched the two-locus genotype-phenotype parameter space using a genetic algorithm to assess what the upper limit on how much additive genetic variation can be maintained under selection.

### Deterministic simulations

The evolutionary fate of an arbitrary two-locus epistatic fitness pattern can be characterised by the allele frequencies and recombination fraction of the two loci as a Markovian process. Therefore it is straightforward to calculate the trajectory of allele frequencies over evolutionary time for a wide range of epistatic patterns. For each G-P map, deterministic simulations were performed with varying conditions for initial allele frequencies (25 initial allele frequencies enumerating the set  $\{0.1, 0.3, \dots, 0.9\}$  over both loci) and linkage disequilibrium between the linked and causal SNPs ( $r^2 = \{1, 0.85, 0.7\}$ ). Variance components and expected test statistics for different parameterisations and under different assumed search strategies were calculated.

**Two-locus frequency calculations.** For a two-locus gametic fitness pattern  $G_{ij}$ , where each value of  $G$  represents the mean phenotypic value for individuals with haplotypes  $i$  and  $j$ ,

	<i>AB</i>	<i>Ab</i>	<i>aB</i>	<i>ab</i>
<i>AB</i>	$G_{11}$	$G_{12}$	$G_{13}$	$G_{14}$
<i>Ab</i>	$G_{21}$	$G_{22}$	$G_{23}$	$G_{24}$
<i>aB</i>	$G_{31}$	$G_{32}$	$G_{33}$	$G_{34}$
<i>ab</i>	$G_{41}$	$G_{42}$	$G_{43}$	$G_{44}$

assuming that  $G_{ij} = G_{ji}$ , this can be related to the two-locus G-P map  $W_{ij}$  as:

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
<i>BB</i>	$W_{11} = G_{11}$	$W_{12} = G_{13}$	$W_{13} = G_{33}$
<i>Bb</i>	$W_{21} = G_{12}$	$W_{22} = G_{14} = G_{23}$	$W_{23} = G_{34}$
<i>bb</i>	$W_{31} = G_{22}$	$W_{32} = G_{24}$	$W_{33} = G_{44}$

We can calculate the expected haplotype frequencies  $f_{AB}, f_{Ab}, f_{aB}, f_{ab}$  after one generation based on selection using [45] and [67]. Here, the haplotype frequencies of the current generation are represented as  $c_i$  where  $i = \{1, \dots, 4\}$  denotes each haplotype in the order listed above, and  $c'_i$  is the haplotype frequency of the next generation:

$$c'_i = (c_i G_i + \eta_i R G_{22} (c_2 c_3 - c_1 c_4)) / \bar{G}. \quad (1)$$

Here

$$G_i = \sum_j G_{ij}, \quad (2)$$

$\eta_1 = \eta_4 = 1$ ,  $\eta_2 = \eta_3 = -1$ ,  $R$  is the recombination fraction between the two loci ( $R=0.5$  denotes the two loci are effectively on separate chromosomes) and

$$\bar{G} = \sum c_i G_i. \quad (3)$$

We ran the simulations for 200 generations, or until one of the loci becomes fixed. If the minor allele from at least one locus  $l$  breaks the condition

$$1/2N \leq f_l, \quad (4)$$

where  $N$  is the population size (arbitrarily set to 1000 for these simulations), the epistatic pattern is considered fixed. While this condition is satisfied, expected variance decomposition and statistical power are assessed on the system at each generation.

**Variance decomposition.** As the allele frequencies change due to selection, while the functional epistatic pattern remains the same the variance components are liable to change. The following calculations, taken from [68], can be used to calculate the marginal additive variances at each locus in a pairwise epistatic interaction for populations at each generation of the simulations. Given marginal fitnesses at the three genotypes at locus A

$$u_i = f_B^2 W_{1i} + 2f_B f_b W_{2i} + f_b^2 W_{3i}, \quad (5)$$

and at locus B

$$v_i = f_A^2 W_{i1} + 2f_A f_a W_{i2} + f_a^2 W_{i3}, \quad (6)$$

the marginal additive variance at locus A is

$$2f_A f_a g_A^2 \quad (7)$$

and the marginal additive variance at locus B is

$$2f_B f_b g_B^2 \quad (8)$$

where

$$g_A = f_A u_1 + (1 - 2f_A) u_2 - f_a u_3 \quad (9)$$

and

$$g_B = f_B v_1 + (1 - 2f_B) v_2 - f_b v_3. \quad (10)$$

However, because linkage disequilibrium can be generated between interacting loci under selection (Figure S4) it is incorrect to quantify the additive variance as the sum of the two marginal variances. Instead, we use the decomposition method detailed in [69] and [70] to calculate the total additive genetic variance in a two-locus system as

$$V_A = 2(f_A f_a h_A^2 + 2h_A h_B d + f_B f_b h_B^2) \quad (11)$$

where

$$h_A = \left( g_A - \frac{d g_B}{f_A f_a} \right) \left( 1 - \frac{d^2}{f_A f_a f_B f_b} \right)^{-1}, \quad (12)$$

$$h_B = \left( g_B - \frac{d g_A}{f_B f_b} \right) \left( 1 - \frac{d^2}{f_A f_a f_B f_b} \right)^{-1}, \quad (13)$$

and

$$d = f_{AB} - f_A f_B. \quad (14)$$

It should be noted that these calculations assume Hardy-Weinberg equilibrium, and selection is likely to generate pseudo LD between unlinked markers, as well as favour certain genotypes over others which results in a violation of this assumption. However there is currently no known two-locus variance decomposition method that maintains orthogonality when the two loci are under linkage disequilibrium and Hardy-Weinberg disequilibrium [71], therefore correct estimates of variance components often cannot be made. However, given that current testing strategies still use the incomplete extant methods, we can examine their behaviour without the requirement of orthogonality between the non-additive components. We use the NOIA method of decomposition [71] to calculate total genetic variance ( $V_G$ ) and the 8 variance components,  $\{V_{A1}, V_{A2}, V_{D1}, V_{D2}, V_{AA}, V_{AD}, V_{DA}, V_{DD}\}$ .

**Detection of additive variance.** By specifying the broad-sense heritability  $H^2$  of a fitness trait at generation 0 for each simulation it is possible to calculate expected F-test performances under different parameterisations and scan strategies. During the simulation selection can modify  $V_G$  by changing allele frequencies, but the non-genetic variance,  $V_E$ , remains constant as a function of  $V_{G_0}$ , the genetic variance at initial allele frequencies:

$$V_E = \frac{V_{G_0}}{H^2} - V_{G_0}. \quad (15)$$

We wanted to find, given a GWAS testing strategy wherein a SNP's contribution to the narrow-sense heritability is only considered if the test statistic meets some significance threshold, how best to parameterise the hypothesis tests to maximise the expected amount of additive variance significantly identified for any given simulation time point. Using an F-test,

$$F = \left( \frac{V_{\text{explained}}}{k} \right) \left( \frac{V_E}{N - k + 1} \right)^{-1} \sim F(k, N - k + 1), \quad (16)$$

where  $N$  is the sample size and  $k$  is the number of parameters in the model, we compared different parameterisations of  $V_{\text{explained}}$  for exhaustive one and two dimensional scans by quantifying how much of the total additive variance in the two-locus system was detected using different GWAS strategies.

For the one dimensional strategy tests for purely additive effects ( $V_{\text{explained}} = V_{Ai}$ ;  $k=1$ ) or complete marginal effects ( $V_{\text{explained}} = V_{Ai} + V_{Di}$ ;  $k=2$ ) were performed at each locus  $i$ . A significance threshold of  $0.05/300000 = 1.7 \times 10^{-7}$  was set. If met at only one locus  $i$  then  $V_{Ai}$  additive variance was considered detected. If met at both loci then the total additive variance  $V_A$  was considered detected.

For the two dimensional strategy three different parameterisations were compared under the conditions of an exhaustive two dimensional scan. These were for purely marginal effects across both loci ( $V_{\text{explained}} = V_{A1} + V_{D1} + V_{A2} + V_{D2}$ ;  $k=4$ ), purely



epistatic effects ( $V_{explained} = V_{AA} + V_{AD} + V_{DA} + V_{DD}$ ;  $k=4$ ), and for total genetic variance ( $V_{explained} = V_G$ ;  $k=8$ ). The significance threshold was set at  $0.05/(300000^2/2 - 300000) = 1.1 \times 10^{-12}$ . If the pairwise test met this threshold then, for the purposes of understanding the efficacy of two dimensional strategies at detecting narrow sense heritability, the total additive variance  $V_A$  across both loci was deemed to have been detected.

**Incomplete LD between causal variants and observed SNPs.** We considered how variance decomposition and testing strategies were affected when the observed SNPs were at different levels of linkage disequilibrium with the causal variants ( $r^2 = \{1, 0.85, 0.7\}$ ). To do this, we transformed the G-P map of the causal loci ( $W_{ij}$ ), to the G-P map of the observed SNPs that have some level of LD with the causal loci ( $\tilde{W}_{ij}$ ). This is constructed by considering that when LD is reduced, the genotype class means of  $\tilde{W}_{ij}$  are a composite of not only  $W_{ij}$ , but also other genotype class means, and the expected contribution of the other class means depends on the level of recombination between the observed SNPs and the causal variants. We performed the above variance decomposition calculations on  $\tilde{W}_{ij}$ , assuming that

$$\tilde{c}_j = c_j \tag{17}$$

where  $\tilde{c}_j$  are the gametic frequencies of the observed SNPs. For simplicity, only the causal variants were inherited from one generation to the next, with new linked SNPs being composed at each new generation. The G-P map  $\tilde{W}_{ij}$  is calculated as

$$\tilde{W}_{ij} = \left( \sum_k \sum_l \mathbf{T}_{Aik} \mathbf{T}_{Bjl} W_{kl} \right) (f_{Ai} f_{Bj})^{-1}, \tag{18}$$

where the frequencies  $f_{mi}$  are the expected genotype frequencies for the  $m$  interacting causal variants  $A$  and  $B$ , such that

$$f_{Ai} = \begin{cases} f_A^2, & i=1 \\ 2f_A f_a, & i=2 \\ f_a^2, & i=3 \end{cases} \tag{19}$$

and

$$f_{Bi} = \begin{cases} f_B^2, & i=1 \\ 2f_B f_b, & i=2 \\ f_b^2, & i=3; \end{cases} \tag{20}$$

and matrix  $\mathbf{T}$  is 3-dimensional with dimensions  $2 \times 3 \times 3$ , where  $\mathbf{T}_{m..}$  is defined as

$$\begin{bmatrix} D_{m1}^2 & 2D_{m1}D_{m2} & D_{m2}^2 \\ 2D_{m1}D_{m3} & 2(D_{m2}D_{m3} + D_{m1}D_{m4}) & 2D_{m2}D_{m4} \\ D_{m3}^2 & 2D_{m3}D_{m4} & D_{m4}^2 \end{bmatrix} \tag{21}$$

where the four gametic frequencies ( $D_m$ ) for the interacting causal loci ( $m = \{A, B\}$ ), and their correlated observed SNPs were calculated as:

$$D_{m1} = r^2 f_m^2 (1 - f_m)^2 + f_m^2 \tag{22}$$

$$D_{m2} = D_{m3} = f_m(1 - f_m) - r^2 f_m^2 (1 - f_m)^2 \tag{23}$$

$$D_{m4} = r^2 f_m^2 (1 - f_m)^2 + (1 - f_m)^2. \tag{24}$$

**Genetic algorithm for generating epistatic patterns**

The purpose of genetic algorithms is to heuristically search a large solution domain for optimal model parameters whilst avoiding a computationally prohibitive exhaustive search [72]. In this case, the algorithm is used to search for two-locus epistatic fitness patterns  $W$  that simultaneously maximise additive genetic variance and avoid fixation through selection, where  $W$  is a  $3 \times 3$  G-P map whose values represent the fitness associated with each two-locus genotype. In this case the building blocks are the nine two-locus genotype class means that comprise the G-P map.

**Initialisation.** Initially a set  $\mathbf{W}$  of  $n_W$  randomly generated candidate patterns  $W$  are created by sampling values for each of the 9 cells from a uniform distribution, and then scaling all values so that the maximum and minimum values for each  $W$  are 1 and 0 respectively.

**Selection.** The candidate patterns are assessed based on two rounds of selection: expected time to fixation and expected level of additive variance generated. A set  $\mathbf{S}$  of simulations are initialised given sets of  $\Phi^A$  and  $\Phi^B$  initial allele frequencies for loci  $A$  and  $B$  respectively, such that  $n_\Sigma = n_{\Phi^A} n_{\Phi^B}$  (e.g. 25 simulations initialised by enumerating all combinations of the sets  $\Phi^A = \Phi^B = \{0.1, 0.3, 0.5, 0.7, 0.9\}$  across two loci). Allele frequency changes and fixation are measured as in equations (1) and (4) respectively at each generation  $\gamma$  for  $\Gamma$  generations. For the candidate pattern to be considered for selection at least  $v$  of its simulations must remain unfixated after  $\Gamma$  generations. For each candidate pattern its total additive genetic variance  $\sigma_{TA}$  over time is calculated by summing the joint additive variance for both loci ( $V_A$ ) for each generation starting after generation  $v$  and until either fixation or generation  $\Gamma$ , and across all initial allele frequency conditions:

$$\sigma_{TA} = \sum_i^{n_{\Phi^A}} \sum_j^{n_{\Phi^B}} \sum_{\gamma=v}^{\Gamma} \Theta(W, \Phi_i^A, \Phi_j^B, \gamma) \tag{25}$$

where  $\Theta(W, \Phi_i^A, \Phi_j^B, \gamma)$  is the additive variance at generation  $\gamma$  of simulation  $\Sigma_{ij}$ . The  $n_\Lambda$  candidate patterns with the largest total additive variances are selected for the next round, comprising the set  $\Lambda$ , or if no candidate patterns reach the threshold  $\tau$  then all patterns are randomly initialised again.

**Reproduction.** The set of  $\mathbf{W}'$  candidate patterns for the next round of selection is comprised of the  $\Lambda$  patterns selected from the current round, a set of  $\mathbf{M}_\Lambda$  mutations for each selected pattern  $\Lambda_i$ , and a set of  $\mathbf{P}$  random patterns produced as in the initialisation step, thus  $\mathbf{W}' = \{\mathbf{\Pi}, \Lambda, \mathbf{M}_1, \dots, \mathbf{M}_{n_\Lambda}\}$ . Mutation is performed by adjusting each element of the candidate pattern:

$$W'_{ij} = W_{ij} + \epsilon \tag{26}$$

where

$$\epsilon \sim N(0, \sigma) \tag{27}$$

and then scaling to the boundaries 0 and 1 as in the initialisation step.

**Termination.** The algorithm is performed for  $\rho$  rounds. Because the set  $\mathbf{L}$  candidate patterns from the previous round are always included in the following round, the maximum score will never decrease. Therefore the optimal epistatic pattern is the considered to be the highest scoring candidate pattern in the final round. Different patterns can be generated by rerunning the entire process with different random seeds.

The code for this algorithm is available at <https://github.com/explodecomputer/epiSpaces/>.

## Population simulations

To consider the potential impact of genetic drift and random noise on the conclusions from the deterministic simulations, similar conditions were recreated heuristically on randomly generated populations. For each epistatic pattern we generated 300 populations of 1000 individuals. Each individual has a two-locus genotype  $x_{ij}x_{ik}$  and a corresponding phenotype  $y_i$  such that

$$y_i = W_{x_{ij}x_{ik}} + \varepsilon \quad (28)$$

where

$$\varepsilon \sim N(0, V_E) \quad (29)$$

and  $x_{ij}$  and  $x_{ik}$  were the fitness values for individual  $i$  corresponding to the G-P map  $W_{jk}$ . The non-genetic variance of the trait was defined at generation 0 as in equation 15 and remained constant at each generation. The heritability,  $H^2$ , was set to 10% at generation 0. Each generation 500 individuals were sampled from a discrete probability distribution where the individual's phenotype was the relative probability of being sampled, and from these 250 random pairings were made to produce 1000 offspring for the next generation. Phenotypes for each new individual were created at each generation as in equation 28, and simulations continued until at least one locus reached fixation. The initial allele frequencies were 0.5 for each locus, and the simulations ran for 200 generations or until at least one locus became fixed.

The code for this algorithm is available at <https://github.com/explodecomputer/epiFit/>.

## Supporting Information

**Figure S1** G-P maps. 1 Neutral; 2–51 Enumeration of all binary trait patterns, excluding reflections, rotations and inversions, as derived by [48] (6 and 29 are non-epistatic); 52–56 Additive  $\times$  Additive, Additive  $\times$  Dominance, Dominance  $\times$  Dominance, Over-dominance, additive. (TIF)

**Figure S2** Deterministic trajectory of allele frequencies as in Figure 1 (row 2), but for an extended set of patterns (detailed in Figure S1) (TIF)

## References

- Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Associates, Inc.
- Bradshaw AD (1991) The Croonian Lecture, 1991. Genostasis and the limits to evolution. Philosophical transactions of the Royal Society of London Series B, Biological sciences 333: 289–305.
- Williams GC (1992) Natural Selection: Domains, Levels, and Challenges. Oxford: Oxford University press.
- Houle D (1992) Comparing evolvability and variability of quantitative traits. Genetics 130: 195–204.
- Hansen T, Houle D (2004) Evolvability, stabilizing selection, and the problem of stasis. In: The evolutionary biology of complex phenotypes, Oxford University press. pp. 130–150.
- Hall M, Lindholm A, Brooks R (2004) Direct selection on male attractiveness and female preference fails to produce a response. BMC Evolutionary Biology 4: 1–10.
- McGuigan K, Van Homrigh A, Blows MW (2008) An evolutionary limit to male mating success. Evolution; international journal of organic evolution 62: 1528–37.

**Figure S3** Simulated trajectory of allele frequencies as in Figure 1 (row 3), but for an extended set of patterns (detailed in Figure S1) (TIF)

**Figure S4** Quasi-LD generated by selection. For the 25 deterministic simulations the expected quasi-LD between the physically unlinked causal SNPs was calculated. It can be seen that significant levels are generated, such that orthogonal standard parameterisation methods would violate assumptions of independence. Boxes represent different G-P maps from Figure S1. (TIF)

**Figure S5** Deterministic change in genetic variance for loci under selection exhibiting various epistatic patterns (Figure S1), when LD between the causal variants and observed SNPs varies. For clarity, only the results from initial frequencies of 0.5 at both loci are shown. Boxes represent different G-P maps from Figure S1. (TIF)

**Figure S6** As in Figure S5, but this time showing the proportion of the genetic variance that is additive. (TIF)

**Figure S7** As in Figure 4a, but for only three tests - Additive in one dimension (A (1D)), genotype in one dimension (A+D (1D)), and full epistatic in two dimensions (F (2D)). Each box has the additive variance detected across all populations and generations as a proportion of the total additive variance that was created for each test when the observed SNPs were in varying levels of LD with the causal variants. For 44 patterns the full epistatic test is most powerful when  $r^2 = 1$ , but when  $r^2 = 0.7$  it is never the most powerful, rather 39 patterns are best detected by the one dimensional genotype parameterisation. (TIF)

**Figure S8** Relationship between genetic variance of observed SNPs ( $y$  axis) and their linkage disequilibrium with causal variants ( $x$  axis). Observed SNPs have the same allele frequency as their linked causal variants, and there is no linkage disequilibrium between causal variants or between observed SNPs. The blue line represents a purely additive G-P map, faint black lines each represent the 55 dominant or epistatic G-P maps in Figure S1, and the black dashed line represents the smoothed average of all black lines. Allele frequencies of G-P maps are represented by boxes, the frequency of locus A horizontally and locus B vertically. (TIF)

## Acknowledgments

We would like to thank the editors and anonymous reviewers who provided valuable advice on improving the manuscript.

## Author Contributions

Conceived and designed the experiments: GH SK CH. Performed the experiments: GH. Analyzed the data: GH. Wrote the paper: GH.

8. Hansen TF, Wagner GP (2001) Epistasis and the Mutation Load: A Measurement-Theoretical Approach. *Genetics* 158: 477–485.
9. Walsh B, Blows MW (2009) Abundant Genetic Variation+Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution, and Systematics* 40: 41–59.
10. Gould S (2002) *The Structure of Evolutionary Theory*. Cambridge, MA: Belknap Press, 880–885 pp.
11. Hindorf LA, Junkins HA, Hall PN, Mehta JP, Manolio TA (2010). A Catalog of Published Genome-Wide Association Studies, available at <http://www.genome.gov/gwastudies>. Accessed 12/10/2010.
12. Maher B (2008) The case of the missing heritability. *Nature* 456.
13. Eyre-Walker A (2010) Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proceedings of the National Academy of Sciences of the United States of America* 107 Suppl: 1752–6.
14. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42: 565–569.
15. Lande R (1975) The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical research* 26: 221–235.
16. Crow JF (2010) On epistasis: why it is unimportant in polygenic directional selection. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 365: 1241–4.
17. Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nature reviews Genetics* 5: 618–25.
18. Phillips PC (2008) Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9: 855–867.
19. Moore J (2003) The Ubiquitous Nature of Epistasis in Determining Susceptibility to Common Human Diseases. *Human Heredity* 56: 73–82.
20. Wright S (1931) *Evolution in Mendelian Populations*. *Genetics* 16: 97–159.
21. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461: 747–53.
22. Frazer KA, Murray SS, Schork NJ, Topol EJ (2009) Human genetic variation and its contribution to complex traits. *Nature reviews Genetics* 10: 241–51.
23. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews Genetics* 11: 446–50.
24. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*.
25. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era - concepts and misconceptions. *Nature Reviews Genetics* 9: 255–66.
26. Marchini J, Donnelly P, Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 37: 413–417.
27. Evans DM, Marchini J, Morris AP, Cardon LR (2006) Two-Stage Two-Locus Models in Genome-Wide Association. *PLoS Genet* 2: e157. doi:10.1371/journal.pgen.0020157
28. Hill WG, Goddard ME, Visscher PM (2008) Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genet* 4: e1000008. doi:10.1371/journal.pgen.1000008
29. Greene CS, Penrod NM, Williams SM, Moore JH (2009) Failure to Replicate a Genetic Association May Provide Important Clues About Genetic Architecture. *PLoS ONE* 4: e5639. doi:10.1371/journal.pone.0005639
30. Evans DM, Gillespie Na, Martin NG (2002) Biometrical genetics. *Biological psychology* 61: 33–51.
31. Arnoldo A, Curak J, Kittanakom S, Chevelev I, Lee VT, et al. (2008) Identification of small molecule inhibitors of *Pseudomonas aeruginosa* exoenzyme S using a yeast phenotypic screen. *PLoS Genet* 4: e1000005. doi:10.1371/journal.pgen.1000005
32. Waddington CH (1942). Canalization of development and the inheritance of acquired characteristics.
33. Masek J, Siegal M (2009) Robustness: mechanisms and consequences. *Trends in Genetics* 25: 395–403.
34. Breen MS, Kemena C, Vlasov PK, Notre Dame C, Kondrashov Fa (2012) Epistasis as the primary factor in molecular evolution. *Nature* 490: 535–538.
35. Kacser H, Burns J (1973) The control of ux. *Symp Soc Exp Biol* 27: 65–104.
36. Frankel N, Davis GK, Vargas D, Wang S, Payre F, et al. (2010) Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 466: 490–3.
37. Soylemez O, Kondrashov Fa (2012) Estimating the rate of irreversibility in protein evolution. *Genome Biology and Evolution*.
38. Strange A, Capon F, Spencer CCa, Knight J, Weale ME, et al. (2010) A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nature genetics* 42: 985–90.
39. Evans DM, Spencer CCa, Pointon JJ, Su Z, Harvey D, et al. (2011) Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nature Genetics* 43.
40. Carlborg O, Jacobsson L, Ahgren P, Siegel P, Andersson L (2006) Epistasis and the release of genetic variation during long-term selection. *Nat Genet* 38: 418–420.
41. Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. *Nature* 396: 336–342.
42. Fraser HB, Schadt EE (2010) The quantitative genetics of phenotypic robustness. *PLoS ONE* 5: e8635. doi:10.1371/journal.pone.0008635
43. Queitsch C, Sangster TA, Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. *Nature* 417: 618–624.
44. Culverhouse R, Suarez BK, Lin J, Reich T (2002) A perspective on epistasis: limits of models displaying no main effect. *American journal of human genetics* 70: 461–71.
45. Kimura M (1956) A Model of a Genetic System which Leads to Closer Linkage by Natural Selection. *Evolution* 10: 278–287.
46. Weir BS (2008) Linkage disequilibrium and association mapping. *Annual review of genomics and human genetics* 9: 129–42.
47. Hemani G, Theodoridis A, Wei W, Haley C (2011) EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics (Oxford, England)* 27: 1462–5.
48. Li W, Reich J (2000) A Complete Enumeration and Classification of Two-Locus Disease Models. *Human Heredity* 50: 334–349.
49. Fisher R (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52: 399–433.
50. Wolpert D, Macready W (1997) No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1: 67–82.
51. Comings DE, MacMurray JP (2000) Molecular heterosis: a review. *Molecular genetics and metabolism* 71: 19–31.
52. Luo LJ, Li ZK, Mei HW, Shu QY, Tabien R, et al. (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics* 158: 1755–71.
53. Moll R, Lonnquist J, Fortunato J, Johnson E (1965) The relationship of heterosis and genetic divergence in maize. *Genetics* 52: 139.
54. Chen C, Rainnie DG, Greene RW, Tongewa S (1994) Abnormal fear response and aggressive behavior in mutant mice deficient for alpha-calmodulin kinase II. *Science (New York, NY)* 266: 291–4.
55. Miskimins R, Ebato H, Seyfried TN, Yu RK (1986) Molecular basis for heterosis for myelin basic protein content in mice. *Proceedings of the National Academy of Sciences of the United States of America* 83: 1532–5.
56. Dobzhansky T (1970) *Genetics of the Evolutionary Process*. Columbia University Press.
57. Kondrashov A (1994) Mullers ratchet under epistatic selection. *Genetics* 136: 1469.
58. Butcher D (1995) Mullers Ratchet, Epistasis and Mutation Effects. *Genetics* 141: 431.
59. Hine E, McGuigan K, Blows MW (2011) Natural selection stops the evolution of male attractiveness. *Proceedings of the National Academy of Sciences of the United States of America* 108: 3659–64.
60. Corona E, Dudley JT, Butte AJ (2010) Extreme evolutionary disparities seen in positive selection across seven complex diseases. *PLoS ONE* 5: e12236. doi:10.1371/journal.pone.0012236
61. Franke A, McGovern DPB, Barrett JC, Wang K, Radford-Smith GL, et al. (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature genetics* 42: 1118–25.
62. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, et al. (2011) A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet* 7: e1002293. doi:10.1371/journal.pgen.1002293
63. Wan X, Yang C, Yang Q, Xue H, Fan X, et al. (2010) BOOST: A Fast Approach to Detecting Gene-Gene Interactions in Genome-wide Case-Control Studies. *American journal of human genetics* 87: 325–340.
64. Zhang X, Huang S, Zou F, Wang W (2010) TEAM: efficient two-locus epistasis tests in human genome-wide association study. *Bioinformatics* 26: i217–i227.
65. Kam-Thong T, Czamara D, Tsuda K, Borgwardt K, Lewis CM, et al. (2010) EPIBLASTERfast exhaustive two-locus epistasis detection strategy using graphical processing units. *European Journal of Human Genetics* 19: 465–471.
66. Hu X, Liu Q, Zhang Z, Li Z, Wang S, et al. (2010) SHEsisEpi, a GPU-enhanced genome-wide SNP-SNP interaction scanning algorithm, efficiently reveals the risk genetic epistasis in bipolar disorder. *Cell Research* 20: 854–857.
67. Lewontin R, Kojima K (1960) The Evolutionary Dynamics of Complex Polymorphisms. *Evolution* 14: 458–472.
68. Ewens WJ (2004) *Mathematical Population Genetics*. New York, New York, USA: Springer-Verlag, second edition.
69. Kojima K, Kelleher T (1961) Changes of mean fitness in random mating populations when epistasis and linkage are present. *Genetics* 46: 527.
70. Kimura M (1965) Attainment of Quasi Linkage Equilibrium When Gene Frequencies Are Changing by Natural Selection. *Genetics* 52: 875–90.
71. Alvarez-Castro JM, Carlborg O (2007) A Unified Model for Functional and Statistical Epistasis and Its Application in Quantitative Trait Loci Analysis. *Genetics* 176: 1151–1167.
72. Holland J (1975) *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.