The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses

Armando Caballero,*,[1] Albert Tenesa,[†,‡] and Peter D. Keightley[§]

*Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Universidad de Vigo, Vigo 36310, Spain, [†]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, Scotland, United Kingdom, [‡]Medical Research Council (MRC) Human Genetics Unit (HGU) at the MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, Scotland, United Kingdom, and [§]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, Scotland, United Kingdom

ABSTRACT We use computer simulations to investigate the amount of genetic variation for complex traits that can be revealed by single-SNP genome-wide association studies (GWAS) or regional heritability mapping (RHM) analyses based on full genome sequence data or SNP chips. We model a large population subject to mutation, recombination, selection, and drift, assuming a pleiotropic model of mutations sampled from a bivariate distribution of effects of mutations on a quantitative trait and fitness. The pleiotropic model investigated, in contrast to previous models, implies that common mutations of large effect are responsible for most of the genetic variation for quantitative traits, except when the trait is fitness itself. We show that GWAS applied to the full sequence increases the number of QTL detected by as much as 50% compared to the number found with SNP chips but only modestly increases the amount of additive genetic variance explained. Even with full sequence data, the total amount of additive variance explained is generally below 50%. Using RHM on the full sequence data, a slightly larger number of QTL are detected than by GWAS if the same probability threshold is assumed, but these QTL explain a slightly smaller amount of genetic variance. Our results also suggest that most of the missing heritability is due to the inability to detect variants of moderate effect (~0.03–0.3 phenotypic SDs) segregating at substantial frequencies. Very rare variants, which are more difficult to detect by GWAS, are expected to contribute little genetic variation, so their eventual detection is less relevant for resolving the missing heritability problem.

KEYWORDS quantitative trait variation; complex traits; missing heritability; fitness; additive genetic variance

STUDY of the nature of variation for quantitative traits, also known as complex traits, is one of the most active areas of research in genetics. This is particularly the case in human genetics because many common genetic diseases, including cancers, obesity, heart disease, and stroke, are complex traits controlled by many loci and influenced by multiple environmental factors. Large numbers of genetic loci influencing complex traits have been discovered using genome-wide association studies (GWAS) by associating putatively neutral SNPs with variation for the trait. For example, analysis of 160 complex disease phenotypes and quantitative traits has revealed associations with more than 2000 SNPs in humans (Visscher *et al.* 2012). A key general finding from GWAS is that the significantly associated SNPs account for only a small proportion of the trait's genetic variation, the so-called missing heritability problem (Manolio *et al.* 2009). For example, based on the resemblance between relatives, narrow-sense heritability for human height has been estimated to be as high as $h^2 = 0.7$–0.8 (Visscher 2008; Zaitlen *et al.* 2013), but the 679 variants identified in the latest GWAS meta-analysis of 79 studies analyzing more than 250,000 individuals accounted for only 16% of that heritability (Wood *et al.* 2014). Even when the information contained in all the SNPs analyzed is used, only about 50–60% of the variance is explained in the case of human height (Yang *et al.* 2010, 2011, 2015; Golan *et al.* 2014; Wood *et al.* 2014) and about 27% in the case of body mass index

doi: 10.1534/genetics.115.177220
Manuscript received April 10, 2015; accepted for publication October 9, 2015; published Early Online October 16, 2015.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177220/-/DC1.
[1]Corresponding author: Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Universidad de Vigo, Vigo 36310, Spain. E-mail: armando@uvigo.es

(Yang *et al.* 2015). This large gap between the heritability explained by the loci identified in GWAS and the trait's heritability is observed for most human disease traits (Visscher *et al.* 2012), although increasingly refined analyses are succeeding in explaining more variation (Yang *et al.* 2015).

A number of nonexclusive arguments have been proposed to explain missing heritability (Maher 2008; Manolio *et al.* 2009; Gibson 2012). First, it has been argued that many alleles causing variation for quantitative traits may be either rare or have small effects, and GWAS lacks power to detect them (Yang *et al.* 2010; Thornton *et al.* 2013). This is supported by the fact that increasingly large sample sizes tend to identify an increasingly large number of QTL (Wood *et al.* 2014). Second, a substantial part of the heritability estimated by resemblance between relatives could be due to epistatic interactions between loci (Zuk *et al.* 2012; Bloom *et al.* 2013), which cannot be captured in (single-locus) GWAS, suggesting that epistatic effects should be explicitly incorporated in GWAS (Hemani *et al.* 2013). On theoretical grounds, it has been argued, however, that most of the genetic variation for quantitative traits is likely to be additive (Hill *et al.* 2008; Hill 2010) and that epistatic interactions cannot be the main factor explaining missing heritability. Third, it also has been argued that synthetic associations between several rare causal variants and common tag SNPs are a common feature of GWAS (Dickson *et al.* 2010; Wang *et al.* 2010) and that the effect associated with a tag SNP underestimates the effects of the causal variants. Although several examples of synthetic associations have been found (Dickson *et al.* 2010; Wang *et al.* 2010), theoretical arguments and simulation studies suggest that this may not be a common feature of GWAS. For example, Wray *et al.* (2011) showed that the distribution of frequencies of the most strongly associated SNPs under a synthetic association model would be highly skewed toward low frequencies, an expectation that is inconsistent with empirical GWAS results. In addition, it has been suggested there are few scenarios that give rise to synthetic associations that would not be detected by linkage analyses (Orozco *et al.* 2010). Finally, simulation studies indicate that significantly associated markers tend to be in strong linkage disequilibrium with only a single causal QTL (Thornton *et al.* 2013).

If we accept that most missing heritability is explained by a lack of power to detect minor or rare variations, a question then arises as to whether a full genome sequence rather than a set of tag SNPs selected by frequency and linkage relationships (as in SNP chips) would increase the amount of genetic variation explained. Thornton *et al.* (2013) performed simulations to investigate the power of GWAS and other methods for the detection of genetic variants affecting complex traits. They compared the power of traditional SNP chips *vs.* resequencing studies and found that complete resequencing yielded only a modest improvement. For example, in their simulation model, the highest power, defined as the number of simulations in which at least one marker in a gene region exceeded the genome-wide significance threshold, peaked at 28% for GWAS but only increased to 38% for full resequencing. Spencer *et al.* (2009) also suggested that full resequencing is unlikely to improve the performance of GWAS using traditional SNP chips. However, Thornton *et al.* (2013) did not attempt to quantify the amount of genetic variance explained by full sequencing compared with that explained by GWAS.

Other QTL detection methods, assumed to be more powerful than traditional GWAS, have been developed that consider a composite *P*-value generated by a group of SNPs within a segment of the genome. These include the sequence kernel association test (SKAT) (Wu *et al.* 2011) and multiple-SNP methods such as regional heritability mapping (RHM), a variance-component approach based on restricted maximum-likelihood (REML) estimation of additive variance explained by genomic regions of a given size (Nagamine *et al.* 2012). A comparison of the power of these two methods and other composite *P*-value methods and single-SNP methods has been made recently by Uemoto *et al.* (2013). This study showed that RHM is more powerful for detecting rare variants than the other methods, particularly when applied to small genomic regions (containing 10–20 SNPs). However, the analysis was based on information from genotyped SNPs rather than on full sequence data. Thus, it is unknown whether RHM on full sequence data substantially improves the heritability explained compared to that explained by GWAS with full sequencing.

The preceding studies have compared the power of detection of genetic variants using full sequence information, single-SNP GWAS, and power composite *P*-value methods, but none have attempted to quantify the genetic variation revealed using these more powerful methods. This depends on the nature of genetic variation for complex traits. Several theoretical and simulation studies have recently investigated the genetic architecture of complex traits and its consequences for the nature of missing heritability. Eyre-Walker (2010) proposed a pleiotropic quantitative trait model in which mutations affecting a trait may have a range of relationships with fitness but assume a direct proportionality between trait and fitness effects. The model predicts that unless the trait and fitness effects are uncorrelated or the strength of selection acting on deleterious mutations is weak, most of the variance for the trait is contributed by rare mutations of large effect. For example, for scenarios that assume an intermediate correlation between the trait and fitness, and considering levels of selection as deduced from nonsynonymous mutations (Boyko *et al.* 2008), the model predicts that most variation is contributed by mutations segregating at frequencies below 0.01 (Eyre-Walker 2010; Figure 3). Thus, under this model, it is expected that a very substantial part of variation for complex traits correlated with fitness is due to mutations with frequencies so rare that they may pass undetected by GWAS.

A number of simulations and analyses have been carried out following the pleiotropic model of Eyre-Walker (2010). Agarwala *et al.* (2013) used this model, assuming a range of correlations between trait and fitness, and compared the results with empirical data for type 2 diabetes. These authors

found that scenarios where the trait and fitness are either uncorrelated or fully correlated were incompatible with the empirical data because there were too many or too few GWAS hits, respectively, relative to empirical observations. However, scenarios in which there was an intermediate correlation could not be ruled out, and these included a case in which rare variants [minor allele frequency (MAF) < 0.05] explain very little (<25%) heritability and another in which they explain most (>80%) heritability. Lohmueller (2014) and Simons *et al.* (2014) also used Eyre-Walker's model or an analogous one to investigate the impact of population bottlenecks and other demographic changes on the nature of deleterious mutation load. Lohmueller (2014) considered two scenarios in which the trait and fitness effects are uncorrelated or have an intermediate level of correlation. Simons *et al.* (2014) used a model of mutations with fixed weak or strong effect on fitness and effects on a quantitative trait that were either unrelated or proportional to the fitness effects. In both cases, the models assumed linkage equilibrium among mutations. Both studies produced consistent results, in agreement with Eyre-Walker (2010), suggesting that when trait and fitness effects are uncorrelated, common variants can account for most of the genetic variation regardless of demography. However, when there is proportionality between trait and fitness effects, rare deleterious mutations contribute much of the genetic variation, and this becomes more important if there are demographic fluctuations, particularly population expansions. Finally, North and Beaumont (2015) investigated the distribution of a complex trait under the Eyre-Walker (2010) model and concluded that it predicts trait distributions that are unrealistically leptokurtic when the correlation of mutational effects between the trait and fitness is high, suggesting that low or intermediate values for the correlation parameter are more realistic.

The possibility that most of the genetic variation for a quantitative trait (other than fitness itself) can be accounted for by mutations at low frequencies (say, below 0.01), as suggested by the Eyre-Walker (2010) model, could help to explain missing heritability. However, most SNPs found in GWAS segregate at intermediate to high frequencies (Yang *et al.* 2010; Lee *et al.* 2011) and explain about 50% of the heritability (Zaitlen *et al.* 2013). This makes it unlikely that most significant SNPs are tagged to rare causal variants (Yang *et al.*, 2011). Note that, as shown by Wray *et al.* (2011), the variance explained by rare causal variants is expected to be several-fold higher than the variance explained by their associated SNPs. This implies that the heritability explained would be much larger than the real one if most common SNPs found in GWAS are due to single rare causal variants.

Here we investigate a different pleiotropic model in which the relationship between fitness effects and trait effects is not directly proportional, as in the Eyre-Walker (2010) model. The objective is to construct realistic models of quantitative variation by individual-based forward-in-time simulation to investigate whether single-SNP GWAS or RHM applied to full sequence data substantially increases the number of quantitative trait variants detected and the additive genetic variance explained relative to GWAS applied to traditional SNP chips. We also investigate whether our pleiotropic model of variation captures the main features of real GWAS and the extent of missing heritability and provides information on the nature of variants missed in association studies.

Materials and Methods

We carried out simulations using knowledge of the distribution of annotated coding and noncoding regions in the genome, empirical estimates of mutation and recombination rates in humans, and the empirically inferred distribution of effects for deleterious nonsynonymous mutations. The protocol involved the simulation of a complex trait correlated with fitness in a large population subject to mutation, drift, selection, and recombination. GWAS analysis of the complex trait was carried out using information on typical SNP chips or full sequence data. RHM of genetic variants also was applied to the whole genome sequence. The numbers, effects, and genetic variance explained by the QTL detected were compared, allowing an understanding of the nature of missing heritability and the possible advantages of using full sequence data and multiple-SNP methods. The model and simulation protocol are described in what follows, and more details are provided in an Appendix in the Supporting Information.

Quantitative trait model and simulation parameters

We assumed the organization (coding and noncoding regions) of a 10-Mb genome sequence in a randomly chosen piece of a mammalian chromosome (in particular, the region between nucleotides 14,000,000 and 24,000,000 of *Mus musculus* chromosome *19*). This sequence includes 72 protein-coding genes and is partitioned into 971 alternating coding and noncoding regions in which only 77,522 positions (<1%) are in coding regions.

SLiM software (Messer 2013), modified to provide mutations with pleiotropic effects on a quantitative trait and fitness, was used for the simulations. A constant unstructured population of size $N = N_e = 1000$ individuals was run for 10,000 generations. This burn-in period ensured that allele frequencies were close to mutation-selection equilibrium. In the final burn-in generation, the population was expanded to 10,000 individuals to simulate a frequency distribution of genetic variants corresponding to an unscaled population size that was 10 times larger.

The per-nucleotide mutation rate u and recombination rate r were assumed to be equal to 10^{-7}, implying values of $N_e u = N_e r = 10^{-4}$, which are appropriate for human populations (Li and Sadler 1991; Kong *et al.* 2002). Thus, because we used $N_e = 1000$ in the simulations and effective sizes for human populations are an order of magnitude larger (see, *e.g.*, Charlesworth 2009), we increased the mutation and recombination rates by an order of magnitude to simulate the genetic variation corresponding to a population that is 10 times larger.

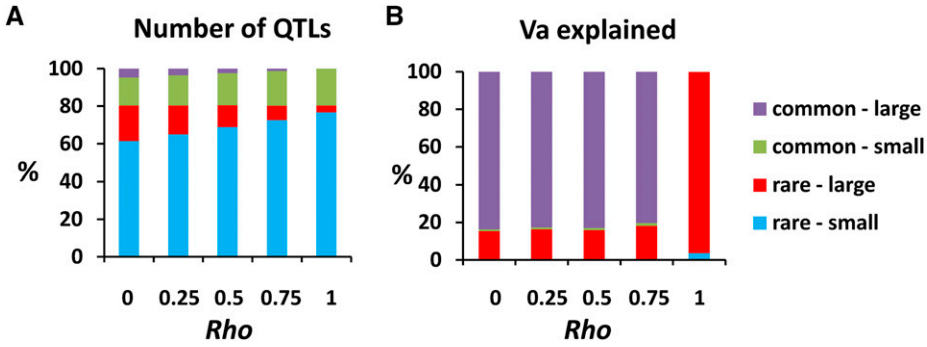

Figure 1 Distribution of number of QTL segregating in the population and their contribution to the additive genetic variance V_A. For simplicity, QTL are classified as rare (gene frequency $q \leq 0.05$), common ($q > 0.05$), small effect ($a \leq 0.02$ and 0.04 pSD, for ρ between 0 and 0.75 and $a \leq 0.18$ pSD for ρ = 1), and large effect (larger a). (A) Number of QTL for each class as a proportion of all QTL. (B) Proportional contribution to the total additive variance of each class of QTL.

The scaled recombination rate is consistent with an average value of 1 cM/Mb in the genome.

In noncoding regions of the sequence, mutations were assumed to be neutral for fitness and to have no effect on the trait. Synonymous sites, comprising 25% of sites in coding regions, were also assumed to evolve neutrally. The remaining 75% of mutations in coding regions were assumed to have a homozygous effect a on the quantitative trait and a correlated homozygous effect s on fitness. Values of a and s were obtained from a bivariate gamma distribution with shape parameter $\beta = 0.2$ and mean s such that $4N_e s = -4000$, and all mutations were deleterious, additive within loci, and had multiplicative gene action between loci. The distribution and mean effect of mutational effects for fitness are consistent with the empirical results of Boyko *et al.* (2008) for amino-acid-changing (nonsynonymous) mutations. For the quantitative trait, the mean absolute value of $a = 1.0$, positive or negative with equal probability, with additive gene action within and between loci. The correlation ρ between a and s was varied, taking values of 0, 0.25, 0.5, 0.75, and 1, but the marginal distributions of a and s were the same for all values of ρ.

Genotypic values for the quantitative trait were obtained by adding the effects of the mutations carried by the individual. Phenotypic values were obtained by adding a normal deviate $N(0,V_E)$ to the genotypic value, where V_E is the environmental variance. The equilibrium additive genetic variance for the quantitative trait varied for each correlation scenario because larger values of ρ imply a closer relationship between trait and fitness effects and thus lower mean values of mutations for the trait. Although traits more closely related with fitness have lower heritabilities (Mousseau and Roff 1987), the intention here was to investigate a trait with a given heritability under different genetic correlations with fitness. Thus, to have the same heritability value ($h^2 = 0.8$) for the trait in all ρ scenarios, the additive variance was obtained in each scenario, and the appropriate environmental variance V_E was calculated accordingly and applied for the calculation of phenotypic values.

To obtain quantities and statistics corresponding to the equivalent of a whole human genome sequence, the results of 300 simulation runs (with 10 Mb each) were added or averaged to produce results for a 3-Gb genome. This process was replicated six times for each of the five ρ values considered.

GWAS analysis

One thousand individuals were randomly sampled from the population and analyzed with the GenABEL package (Aulchenko *et al.* 2007). Two types of analyses were carried out, one for a typical SNP chip (referred to as *Chip*) and the other involving all SNPs simulated, thus incorporating the full genome sequence (referred to as *Sequence*). For the Chip analysis, the following steps were carried out to model the discovery of SNPs in the HapMap Project and subsequent selection of tag SNPs: first, among all SNPs simulated (around 9.7 million per genome), a random sample was obtained so that an approximately uniform distribution of SNP frequencies was achieved. This was done by sampling SNPs with a probability equal to $4q(1 - q)$, where q is the SNP frequency, so that all SNP frequency classes were chosen with approximately equal probability. Next, the program PLINK (Purcell *et al.* 2007) was used to prune out groups of SNPs with extremely high linkage disequilibrium. A threshold value of $r^2 = 0.9$ was considered using the indep-pairwise function with window size 1000 bp and step 500. The resulting SNP chip was composed of ~1.36 million SNPs. Estimates of the additive variance explained by the whole SNP chip were obtained with the software REACTA (Cebamanos *et al.* 2014), a version of GCTA (Yang *et al.* 2011).

The single-SNP GWAS analysis was performed with GenABEL. For this analysis, SNPs were "cleaned" to remove those with a MAF < 0.05, leaving about 1.2 million SNPs to be analyzed in the Chip, *i.e.*, about one-tenth the total number of segregating SNPs. Other cleaning options were implemented, such as exclusion of SNPs that deviate from Hardy-Weinberg equilibrium (HWE) [false discovery rate (FDR) $<$ 0.2] and exclusion of individuals with unusually high heterozygosity (FDR $< 1\%$), but this led to the exclusions of no further SNPs. Regression of the individual phenotypic values for the trait on single-marker frequencies was carried out with the function gtscore on the cleaned data. The threshold unadjusted probability P_{1df} used to consider significance was $\alpha = 10^{-8}$, a typical cutoff value used in GWAS (Dickson *et al.* 2010).

After detection of significant SNPs in a given simulation run, we carried out a procedure, following Dickson *et al.*

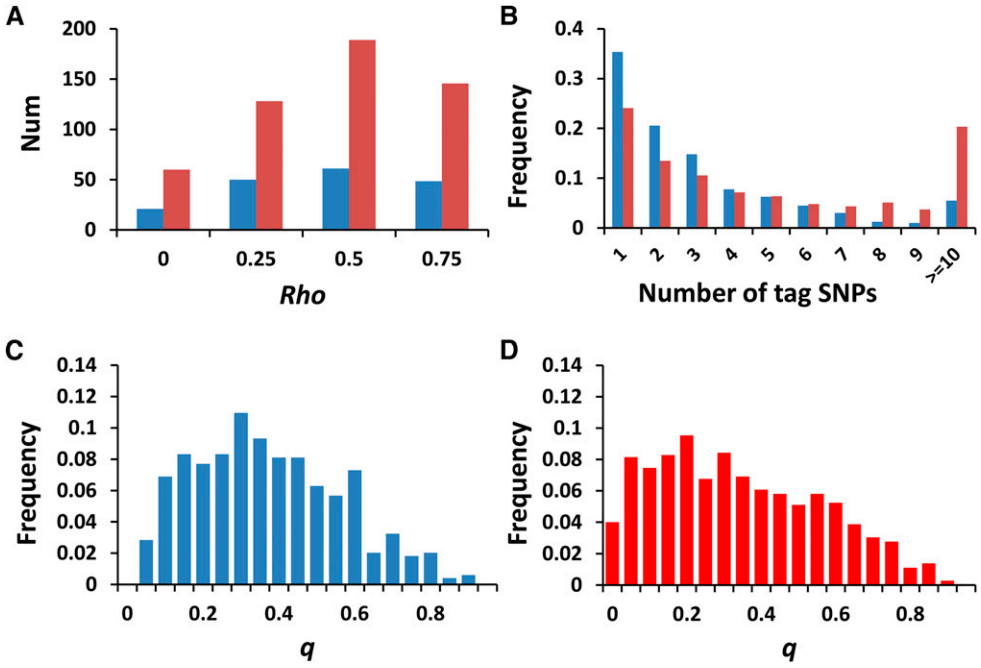

Figure 2 Numbers and effects of significant SNPs detected in the Chip analysis (blue) or the Sequence analysis (red). (A) Average number of significant SNPs per genome for different values of the correlation between mutational effects on fitness and the quantitative trait (ρ). (B) Number of significantly associated tag SNPs for each QTL. (C and D) Distribution of frequencies of significant SNPs found in the Chip analysis (C) or the Sequence analysis (D). B–D combine results for all ρ values. For the Chip analysis, MAF > 0.05 was assumed for all ρ values.

(2010), to ascertain which causal SNPs (QTL) were responsible for the significance of each significantly detected SNP. For each significant SNP found in the analysis, GWAS was again carried out fitting every QTL in the model, one by one. For example, if there were, say, 80 QTL involved in a given run (whether or not included in the Chip), GWAS was carried out 80 times, fitting each QTL in the model in turn. Fitting a given QTL in the model implies that its effect is removed in the regression analysis. If the significance of a given SNP is lost after the fitting of a given QTL, this therefore implies that the QTL was responsible for the significance of the SNP. The loss of significance was assessed using a conservative *P*-value of $\alpha' = 10^{-7}$. When several significant SNPs were associated with (caused by) a given QTL, the most associated SNP was considered as the one with the lowest *P*-value.

For the Sequence analyses, the procedure was the same as described earlier, except that all SNPs simulated were analyzed rather than only those in the Chip. To check for type I error in QTL detection, all simulations for the scenario with $\rho = 0.5$ were repeated 100 times using random permutations of the individual phenotypes.

RHM analysis

RHM analysis was carried out for the whole sequence with the software REACTA (Cebamanos *et al.* 2014) using windows of 20 consecutive SNPs and an overlap of 10 SNPs. Thus, in a single 10-Mb run, ~2400 regions were analyzed, *i.e.*, about 720,000 windows per full genome. Estimates of the additive variance were assumed to be significant if the *P*-value α was less than a critical value ranging from 10^{-4} to 10^{-12}. All analyses were repeated using randomization of individual phenotypes to check for type I error in QTL detection.

Results

Simulated quantitative trait variation

The quantitative trait and fitness were simulated using a range of values for the genetic correlation parameter ρ from 0 (no correlation between the absolute mutational effects a on the trait and fitness s) to 1 (the trait is fitness itself) (Supporting Information, Figure S1). Our model is different from the pleiotropic model of Eyre-Walker (2010) in the sense that it allows for mutations with large effects on the trait and low fitness effects unless $\rho = 1$. To illustrate this difference, Figure S1F compares the joint distribution of trait and fitness effects assumed by Eyre-Walker (2010) [see also Figure 1 of Eyre-Walker (2010)] with that for the corresponding intermediate ρ value under our model (Figure S1C).

The distribution of the number of QTL segregating in the population for the different ρ values is shown in Figure 1A. For simplicity, the figure arbitrarily classifies the QTL as rare (with frequency $q \leq 0.05$) or common ($q > 0.05$) and as having small effects on the trait [with absolute effect a between 0 and 0.02–0.04 phenotypic SD (pSD) for ρ between 0 and 0.75 and between 0 and 0.18 pSD for $\rho = 1$] and large effects (the remainder). Note that ~80% of QTL segregating in the population are rare, and the stronger the correlation between trait and fitness effects, the lower is the proportion of major QTL. However, for values of $\rho < 1$, there is a small proportion of mutations of large effect for the trait that are common in the population. For $\rho = 1$ (the trait is fitness itself), no common QTL of large effect segregate in the population, as expected. Figure 1B shows the proportion of additive genetic variance explained by the QTL. For $\rho < 1$, ~85% of the variance is explained by common QTL of large

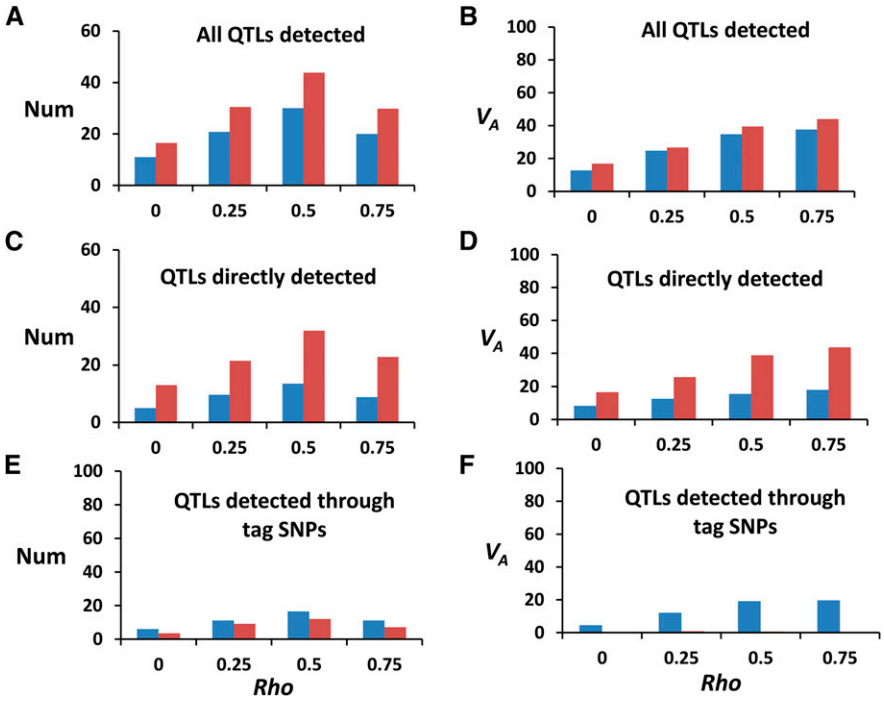

Figure 3 Average number of QTL detected (responsible for the significance of the SNPs shown in Figure 2) per genome in the Chip analysis (blue) or the Sequence analysis (red) and percentage of total additive variance V_A explained by them for different values of the correlation between mutational effects on fitness and the trait (ρ). (A and B) Results for all QTL detected. (C and D) Results for QTL directly detected (*i.e.*, the QTL itself was significant in the analysis). (E and F) Results for QTL detected indirectly through a significant neutral tag SNP.

effect, and most of the remaining variance is explained by rare QTL, also of large effect. For fitness itself (ρ = 1), however, all the variance is explained by rare, mostly large-effect alleles.

Significant SNPs found with GWAS

The average allele frequency of all SNPs simulated was 0.12, and the distribution of MAFs was highly leptokurtic (Figure S2). For the SNPs belonging to the Chip, the frequency distribution was approximately uniform (Figure S2), as in typical SNP chips. The average number of significant SNPs detected per genome was about three times larger with the full Sequence than with the Chip (Figure 2A and Table S1), and there were generally more significant SNPs tagged to a given QTL in the Sequence analysis (mean 6.5 SNPs *vs.* 3.3 SNPs for Chip) (Figure 2B and Table S2). Permutation of individual phenotypes for the case of ρ = 0.5 showed that only 0.07 SNPs, on average, would be detected per genome with $\alpha = 10^{-8}$, suggesting that there is a very low type I error rate in the analysis. The average estimated effect for significant SNPs for the trait was around 0.15 pSD (range 0.11–0.18 pSD) for the different ρ scenarios (Table S1). No significant SNPs were detected for fitness itself (ρ = 1). The full Sequence analysis detected a substantially larger number of SNPs segregating at lower frequencies (Figure 2D), which were undetectable in the Chip analysis (Figure 2C).

Detection of QTL with GWAS

The causal QTL associated with the significant SNPs found in the Chip and Sequence analyses were identified and their contributions to the additive genetic variance calculated (see *Materials and Methods*). Figure 3 shows the average number of QTL detected per genome for the Chip and Sequence analyses, along with the proportion of the total additive genetic variance V_A explained by them. The number of QTL detected was larger for cases with intermediate values of ρ (Figure 3A), corresponding to the larger number of significant SNPs (Figure 2A), but the proportion of additive variance explained by them increased approximately linearly with the value of ρ (Figure 3B). The number of detected QTL was substantially larger (47% on average) with the Sequence analysis than with the Chip analysis (Figure 3A). In total, over all the genomes analyzed, 404 QTL were detected by both analyses, 89 were detected only by the Chip analysis and 314 only by the Sequence analysis (Figure S3).The distribution of allele frequencies for the QTL detected by the two analyses is shown in Figure 4 (all ρ scenarios are combined because average allele frequencies were similar for all of them) (see Table S1). Clearly, the Sequence analysis detected a substantially larger number of QTL segregating at low frequencies, but these contributed little additive variance. Thus, the proportion of V_A explained by the QTL detected was always lower than 50%, with only a modest improvement (<20%) with the Sequence analysis (Figure 3B).

When the additive variance was estimated using the information contained in the whole chip SNP, this was close to 100% for all values of ρ except ρ = 1; *i.e.*, the estimated V_A was 102, 98, 90, 89, and 1% of the true genic variance for ρ = 0, 0.25, 0.5, 0.75, and 1, respectively.

Figure 3, C–F, differentiates between the QTL detected directly in the analysis (*i.e.*, the QTL itself is significant) and the QTL detected indirectly via a neutral significant tag SNP. Note that the increase in the number of QTL detected by the Sequence analysis was mainly due to an increase in the

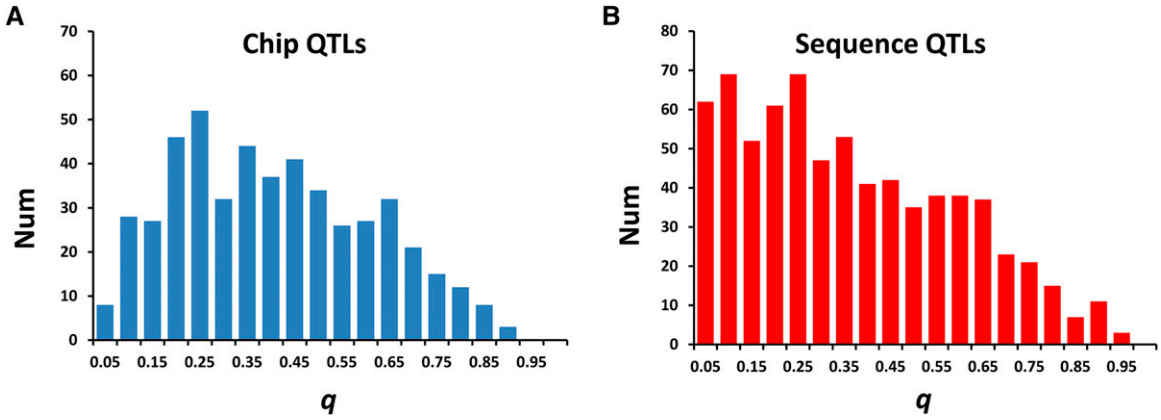

Figure 4 Distribution of frequencies q of all detected QTL (combining all genomes and values of ρ) by (A) the Chip analysis and (B) the Sequence analysis.

number of QTL detected directly (Figure 3, C–F). Thus, the Sequence analysis detected QTL with a higher precision than the Chip analysis.

Regarding QTL detected indirectly via a tag SNP, estimated effects and frequencies of neutral tag SNPs in the Chip analysis were in good agreement with the effects of their corresponding causal QTL (Figure S4). The mean physical distance between tag SNPs and their causal QTL was 15.9 kb (median 4.3 kb), with an average linkage disequilibrium r^2 value of 0.75 (Figure S4 and Table S2). Thus, the most associated tag SNPs were generally close to their causal QTL, but in some cases the distance could be rather large (tens or even hundreds of kilobases) (Figure S4C).

We also addressed the issue of whether or not synthetic associations are a general feature of GWAS. Figure 5 presents the number of significant SNPs (mean numbers for each ρ scenario are given in Table S2) for which significance is due to one, two, three, or more causal QTL. The results are similar for both the Chip and Sequence analyses. About 65% of significant SNPs could be explained by a single causal QTL, about 25% by two QTL, and the remainder by three or more QTL. Therefore, synthetic associations between causal QTL are not found in most cases. In addition, in the cases where two or more QTL caused the significance of a given tag SNP, one of the QTL had a much larger effect on the trait and stronger linkage disequilibrium with the tag SNP than the others (Figure S5). Thus, in these situations, one of the causal major QTL had the leading signal, and the other causal QTL had a secondary minor role.

Detection of QTL with RHM

Analyses of the full genome sequence also were carried out with the RHM method for simulations with $\rho = 0.5$, *i.e.*, the scenario with the largest number of QTL detected by GWAS. We assumed a range of significance thresholds ($\alpha = 10^{-4}$ to 10^{-12}) to detect significant genomic regions of 10 SNPs each. As expected, the number of QTL detected increased as α increased (Table S3), but the proportion of significant regions actually including QTL was correspondingly smaller, as expected, indicating that higher type I errors increased with increasing α (see also Rowe *et al.* 2013). The contribution to the total additive variance of the QTL detected was not dramatically increased by increasing α, *i.e.*, 46% with $\alpha = 10^{-4}$ and 31% with $\alpha = 10^{-8}$ (Table S3). Analyses carrying out random permutations of individual phenotypes showed that no significant regions were found if $\alpha = 10^{-8}$ (Table S3). Thus, we used this threshold value for further analysis, which is the same as that used for GWAS.

Characterization of the missing genetic variance

Table 1 shows the total number of QTL segregating in all simulations for $\rho = 0.5$, distinguishing between rare ($q \leq 0.05$) and common ($q > 0.05$) QTL and between different ranges of effects on the trait. The table also shows the total number of QTL detected by GWAS (with the full Sequence), those detected by RHM, and those detected by both methods. The contributions of these detected QTL to the additive variance are also shown in the right-hand part of the table.

Regarding GWAS results, it can be seen that although 39% of the additive variance is recovered by the QTL detected, a large number of them are missed, particularly those that are rare and have small effects, but many common QTL of moderate effect are also missed. It is precisely these QTL that contribute the bulk of genetic variation. Therefore, most of missing heritability is due to the inability to detect these QTL with moderate effects (0.03–0.28 pSD) segregating at substantial frequencies rather than very rare ones, which contribute very little to the genetic variation. This is illustrated in Figure 6, which shows the proportion of the total additive variance detected (Figure 6A) or missed (Figure 6B) by single-SNP GWAS as a function of the QTL effects and frequencies. Figure 6A shows that most of the variance detected (39.4% of the total) is explained by QTL of effect larger than 0.14 pSD segregating at intermediate frequencies. The missing additive variance (60.6%) is mostly due to QTL of minor to moderate effect ($0.03 < a < 0.28$ pSD) segregating at intermediate frequencies (Figure 6B). The proportion of variance missed from QTL with frequency lower than 0.05 is only 15%.

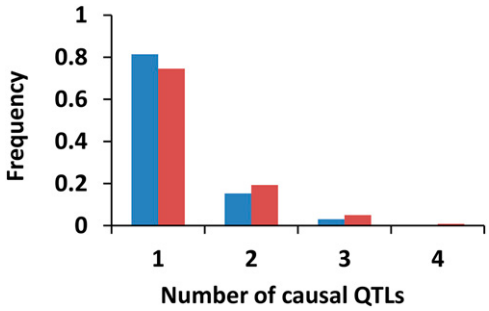

Figure 5 Distribution of the frequency of significant SNPs for which detection is due to a given number of causal QTL for the Chip analysis (blue) and the Sequence analysis (red) (results combining all simulations and values of ρ).

With the same threshold probability value, RHM detected about 13% more QTL than GWAS (Table 1). From the 263 QTL detected by GWAS and the 298 QTL detected by RHM, only 135 were coincident. However, most of the additional QTL detected by RHM were rare and of minor effect, and some more common major QTL were missed with RHM that were detected with GWAS. Thus, the proportion of V_A detected by RHM (31%) was lower than that detected by GWAS (39%). The combination of both methods (GWAS + RHM) detected up to 425 QTL but gave only a marginal increase in the proportion of variance detected (41%). A close inspection of the QTL detected by RHM shows that almost all QTL of small effect were close to a single major QTL in a given genomic region. Thus, RHM detected major rare QTL, and because their signals were extended to nearby regions, other minor QTL were detected by chance (see Figure S6 and Table S4 for illustrations of this observation).

Discussion

Nature of quantitative trait variation for complex traits

Our simulation model assumes an equilibrium among mutation, negative selection, and genetic drift. We based our distribution of fitness effects on that inferred by Boyko *et al.* (2008) from segregating amino acid polymorphisms in humans. According to this study, the best-fitting distribution of fitness effects among newly arising amino-acid-changing mutations is a log-normal or gamma distribution with a rather leptokurtic shape (corresponding to a shape parameter of around 0.2 for the gamma distribution) that peaked near neutrality and with a mean effect of mutations implying $4N_es$ of about −3,000. Lohmueller (2014) and North and Beaumont (2015) also assumed the parameters inferred by Boyko *et al.* (2008) in their simulations. We assumed this distribution shape both for the quantitative trait and for fitness using a bivariate distribution with correlation ρ. With this distribution, the average fitness effect for QTL segregating in the population was about −0.1, but the mean fitness effects of QTL detected in the Sequence analysis were −0.0009, −0.0014, −0.0070, and −0.0033 for ρ = 0, 0.25, 0.5, and 0.75, respectively. No QTL were detected with ρ = 1 (the trait is fitness itself) because major QTL were highly deleterious. These fitness effects are broadly consistent with the range of estimated fitness effects for rare segregating human amino-acid-changing alleles (−0.003 to −0.001) estimated by Kryukov *et al.* (2007).

In our simulations, we made some simplifying assumptions. Mutations with effect on the quantitative trait (and fitness) were assumed to occur only in coding regions because the data used for the distribution of mutational effects for fitness refer to amino-acid-changing (nonsynonymous) mutations (Boyko *et al.*, 2008). Many mutations for quantitative traits presumably also occur in noncoding regions, but their fitness effects are more poorly understood (Torgerson *et al.* 2009; Halligan *et al.* 2013). The effective population size considered in the simulations was of 1000 individuals, which was expanded to 10,000 individuals in the last simulated generation. Given that the recent effective size of human populations is an order of magnitude larger than we simulated (see, *e.g.*, Charlesworth 2009), we scaled the mutation and recombination rates, increasing them by an order of magnitude. This is, however, a common practice, and the scaling guarantees that the amount of genetic variance achieved for the quantitative trait is approximately that for a population that is 10 times larger than that simulated. We checked this by performing simulations with a population size five times larger, *i.e.*, $N = 5000$, assuming $\rho = 0.5$ and scaled mutation and recombination rates and mean fitness effects to keep the same previous values of $N_eu = N_er = 10^{-4}$ and $4N_es = -4000$. Six hundred simulations of 10-Mb sequences were run for 50,000 generations. The total number of SNPs found per genome was 11.4 million, close to the 9.7 million found with $N = 1000$, and the total additive genetic variance per genome was not significantly different from that obtained assuming $N = 1000$ (1089 ± 35 for $N = 1000$ *vs.* 1037 ± 20 for $N = 5000$). Note, in addition, that although the genetic variance increases indefinitely with increasing population size under a neutral model, under our pleiotropic model of mutations, this increase is very slow unless the correlation between trait and fitness is close to zero [see Keightley and Hill (1990) and Figure 6 of Caballero and Keightley (1994)].

The pleiotropic model proposed by Eyre-Walker (2010) and assumed by Agarwala *et al.* (2013), Lohmueller (2014), and North and Beaumont (2015) differs from ours in that it implies a direct proportionality between trait and fitness effects for all mutations. A proportional relationship between fitness s and trait a values also was assumed by Simons *et al.* (2014) such that $a = cs$, where s can be 0.0002 (weak fitness effect) or 0.01 (strong effect), and c is a constant between 0 and 1. With these models, nearly neutral mutations with large effects on the trait are not generally possible (see, *e.g.*, Figure S1F), except when the trait and fitness effects are uncorrelated or fitness effects are weak. However, mutational effects on fitness deduced from empirical data seem to be rather large (mean $4N_es \sim -3000$) (Boyko *et al.* 2008). Thus, a strong relationship between s and a implies that genetic

Table 1 Number of QTL and their contribution to the additive genetic variance V_A (in percent) for all QTL segregating in the population and those detected by GWAS with the full sequence as a function of their gene frequencies (rare: $q \leq 0.05$; common: $q > 0.05$) and absolute effects on the quantitative trait (a, in pSDs)

		Number				%V_A			
Frequency	Effect a (pSDs)	All QTL	GWAS	RHM	GWAS + RHM	All QTL	GWAS	RHM	GWAS + RHM
Rare, $q \leq 0.05$	0–0.03	99,801	14	127	138	0.2	0.0	0.0	0.0
	0.03–0.14	13,530	0	18	18	3.6	0.0	0.0	0.0
	0.14–0.28	2,504	0	1	1	4.9	0.0	0.0	0.0
	0.28–0.56	758	2	1	3	5.1	0.1	0.0	0.2
	>0.56	71	4	3	4	1.9	0.8	0.7	0.8
Common, $q > 0.05$	0–0.03	24,637	52	38	64	1.2	0.0	0.0	0.0
	0.03–0.14	2,727	14	5	14	19.5	0.2	0.1	0.2
	0.14–0.28	555	73	26	76	30.7	8.6	3.6	8.9
	0.28–0.56	127	92	67	95	24.9	21.8	18.8	22.7
	>0.56	12	12	12	12	7.9	7.9	7.9	7.9
		144,722	263	298	425	100.0	39.4	31.1	40.6

The results correspond to the six whole genomes simulated under the scenario with $\rho = 0.5$. The threshold value for detection is a probability of $\alpha = 10^{-8}$ both for GWAS and for RHM.

variation is mostly explained by rare variants of large effect. For example, Figure 3A of Eyre-Walker (2010) shows that when $4N_e s = -3000$ and $\tau = 0.5$ (implying an intermediate correlation between trait and fitness) (see Figure S1F), the peak for the density of variance occurs for mutations with frequencies between 10^{-3} and 10^{-4}. Mutations segregating at such low frequencies are expected to go undetected by GWAS but might contribute to explaining missing heritability. North and Beaumont (2015) showed that assuming intense selection, the model of Eyre-Walker (2010) is compatible with the observed distribution of quantitative traits only if the correlation between trait and fitness is intermediate or low. In addition, Agarwala *et al.* (2013) found that by assuming the model of Eyre-Walker (2010), intermediate values of the correlation between trait and fitness also would be compatible with data from type 2 diabetes, but models assuming no correlation or a complete correlation could be ruled out.

In contrast, in our pleiotropic model, nearly neutral mutations of large effect can occur irrespective of their effect on the trait, except for the case of $\rho = 1$ (see Figure S1, A–E), even though we assumed a conservatively large mean fitness effect of mutations ($4N_e s = -4000$). It can be argued that even if a quantitative trait is strongly correlated with fitness, this should not necessarily imply that all mutations of substantial effect on the trait also must have a substantial negative fitness effect. In fact, most of the SNPs detected by GWAS are common (Yang *et al.* 2010; Lee *et al.* 2011; Zaitlen *et al.* 2013), and these contribute about 50% of the heritability of the traits (Zaitlen *et al.* 2013). Although common SNPs could be associated with either rare or common causal QTL, it seems unlikely that all common SNPs found by GWAS are associated with rare QTL (Lee *et al.* 2011; Wray *et al.* 2011), implying that common QTL of large effect are also possible. Thus, with the pleiotropic model assumed here, if $\rho = 1$, all variance is due to rare QTL (see Figure 1B), in agreement with the model of Eyre-Walker (2010), and no significant SNPs are detected with GWAS. For other ρ values, however, we found that a substantial amount of variance is contributed by common QTL of substantial effect (Figure 1B), and thus, these QTL can be detected for all values of ρ (except $\rho = 1$). Even so, for intermediate values of ρ, our model predicts that variants of large effect are more likely at lower frequencies, in accordance with empirical results, as will be shown later.

Our results suggest that most of missing heritability is due to the inability to detect QTL of moderate effect segregating at substantial frequencies, perhaps because their effects are masked by the simultaneous segregation of other major QTL. This is clear from Table 1 and Figure 6B, which show that many QTL with a range of frequencies and effects between 0.03 and 0.28 pSD are missed, and these account for 41.5% of the total variance. An increasing number of these mutations with moderate effect and segregating at high frequencies are expected to be found as sample sizes increase and detection methods become more powerful, reducing the missing heritability. However, many other variants segregating at low frequencies ($q < 0.05$) across the full distribution of effects or segregating at intermediate frequencies ($q > 0.05$) but with very minor effects are also missed, and these will be more difficult to detect. Because they are expected to contribute very little to the genetic variation (Table 1), their eventual detection will not contribute substantially to the general understanding of variation for the trait.

Significant SNPs found by GWAS

The genomic mutation rate assumed produced a total of about 9.7 million SNPs. For the Chip analysis, we considered a typical genotyping panel in which the distribution of SNP frequency was approximately uniform (Figure S2) and in which SNPs segregating at low frequencies (MAF < 0.05) were excluded from the analysis. The number of significant SNPs found with the Chip analysis (of the order of a few dozen SNPs) (Figure 2A) is consistent with typical GWAS results. Likewise, the distribution of frequencies of significant SNPs is also consistent with that observed in real GWAS

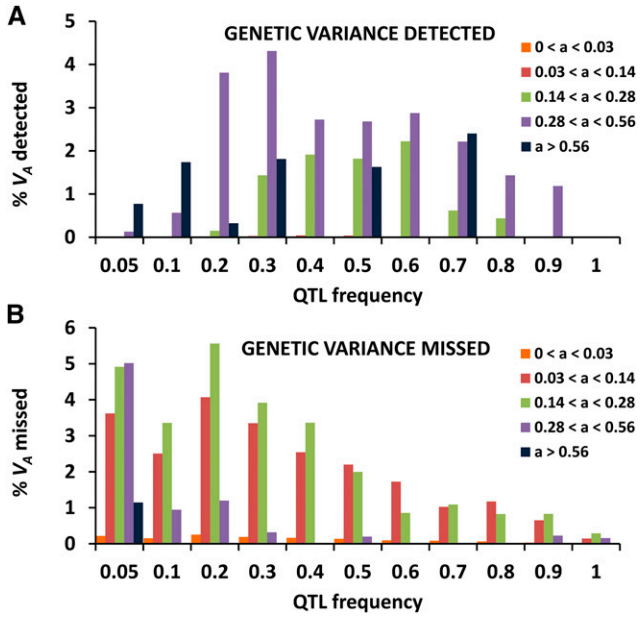

Figure 6 Proportion of additive variance detected (A) and missed (B) by single-SNP GWAS (scenario ρ = 0.5) for different classes of QTL as a function of their effects (*a*, in pSDs) and frequencies.

[compare our Figure 2C with Figure 2A of Wray *et al.* (2011), obtained from the GWAS catalog]. In addition, the average estimated effect of SNPs detected in the Chip analysis was in the range 0.11–0.18 pSD (Table S1), in broad agreement with the average effects of SNPs detected in GWAS, *e.g.*, affecting human height (0.8 cm between the two homozygous genotypes, or about 0.12 pSD) (Visscher 2008). With the Sequence analysis, the number of SNPs detected increased approximately threefold (Figure 2A) compared to the Chip analysis, and this increase was manifest particularly for SNPs segregating at low frequencies (Figure 2D). Models assuming an intermediate correlation ρ between effects on the quantitative trait and fitness produced larger numbers of significant SNPs than lower or higher values of ρ. This is explained by a combination of two opposing factors. First, the larger the value of ρ, the lower is the mean trait effect for segregating mutations because mutations with large effect on the trait tended to have more deleterious effects on fitness. This factor increases the chances of QTL detection for low values of ρ. However, because we assumed a constant heritability for all ρ scenarios, the environmental variance assumed was necessarily lower with increasing values of ρ, so the power to detect significant SNPs of large effect was increased for high values of ρ, except for ρ = 1, where no significant SNPs were found because of their low frequency.

Recently, results from the UK10K Project (UK10K Consortium 2015), based on an analysis of whole genome or exome sequences of nearly 10,000 individuals, have allowed for the identification of a large number of variants affecting different disease traits. One of the main findings from this study is the allelic spectrum for single-marker association variants. This spectrum shows that variants segregating at high frequencies tend to have smaller effects on the trait than variants segregating at low frequencies [see Figure 3 of UK10K Consortium (2015)]. A similar negative relationship between allelic effects and allelic frequencies has been shown by Yang *et al.* (2015) (see their Figure 4) regarding human height and body mass index. These observations are compatible with a pleiotropic model of mutations between the trait and fitness, as investigated in this paper.

We made a comparison between this empirical allelic spectrum and that obtained from our simulations. Figure 7 shows the predicted effects $E(a)$ obtained for the 263 most associated tag SNPs obtained by GWAS with full sequences for the simulated scenario of ρ = 0.5 plotted against their corresponding SNP's MAF. Note that a large number of these tag SNPs, 191 of 263, are actually QTL. Figure 7 also presents empirical results obtained from data in Table S5 of UK10K Consortium (2015), considering the 39 variants from the table with genome-wide significance and the values of β (estimated heterozygous effect) and frequencies obtained with whole genome sequenced samples and genome-wide analysis (WGS + GWA). Values of β from this table are multiplied by 2 to make them comparable with simulated homozygous effects. The figure clearly shows that simulated data from our pleiotropic model assuming an intermediate correlation between the trait and fitness are highly compatible with the empirical results for a variety of disease traits. Note that in concordance with empirical results, the simulation results show that rare tagged SNPs tend to have larger estimated effects than more common tag SNPs. However, it also may be noticed (see Figure S7) that the contribution to the predicted additive variance from common tagged SNP is much larger than that from rare SNPs, in agreement with the results shown in Table 1.

QTL detected by GWAS

The number of QTL detected in the Sequence analysis was about 50% larger, on average, than in the Chip analysis (Figure 3A). This improvement is explained by an increase in the number of QTL directly detected (*i.e.*, the QTL itself is detected as a significant SNP) (Figure 3, C and D), whereas the number of QTL detected via linkage to a significant neutral SNP was actually smaller than in the Chip analysis (Figure 3, E and F). The full Sequence analysis therefore gives a substantial increase in the number of QTL detected, and this detection is made with higher precision, *i.e.*, more frequent detection of the QTL than a nearby tag SNP. This 50% increase in the number of QTL detected with the Sequence analysis is compatible with the simulation results of Thornton *et al.* (2013). These authors carried out simulations of a 100-kb sequence in which mutations affect a quantitative trait subject to stabilizing selection. Neutral mutations occurred 10 times more frequently than those affecting the trait, and these latter mutations had exponentially distributed effects with a variable mean. The main objective of that work was to investigate the power of detection of the trait variants under a number of methods. The authors concluded that traditional

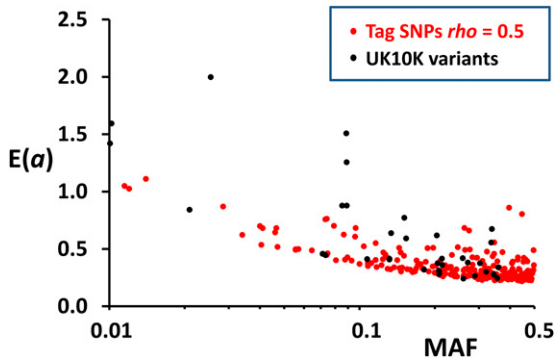

Figure 7 Allelic spectrum showing the simulated estimated homozygous effects *E*(*a*) of significant SNPs plotted against their MAF. Red dots refer to significant SNPs (263) found in simulations for the scenario ρ = 0.5. Black dots refer to the 39 variants (for different disease traits) with genome-wide significance from Table S5 of UK10K Consortium (2015). Homozygous effects *E*(*a*) are obtained by multiplying the β values given in the table (in SDs) by 2 under the columns for whole-genome sequenced samples and genome-wide analysis [WGS + GWA-based (four-way)]. MAF values are obtained from the corresponding column of EAF (Effects allele frequency).

marker-by-marker GWAS has a maximum power of detection (proportion of simulation replicates with at least one significant SNP detected) of 28%, and this increased to 38% by making use of the full sequence data. This is a somewhat smaller improvement that we observed, perhaps reflecting differences in the models assumed, such as the shape of the distribution and mean effects of mutations considered, our joint distribution of effects on fitness and the trait rather than a trait subject to stabilizing selection, and differences in the heritability for the trait.

We further addressed the issue of the proportion of the additive genetic variance explained by the Chip and Sequence analyses. We found that the amount of variance explained increases with the value of ρ (Figure 3B). The higher its value, the lower is the number of major QTL that segregate at high frequencies in the population (Figure 1), susceptible to detection by GWAS, and thus the larger is their relative contribution to the total additive variance. The proportion of the total additive genetic variance explained by the QTL detected in the Chip analysis was always between 10 and 40% (and even lower if expressed as a fraction of the phenotypic variance). This is in broad agreement with the percentages of variation explained by significant SNPs in a range of studies (Visscher *et al.* 2012; Wood *et al.* 2014) with very large sample sizes. We found, however, that the whole SNP chip variation accounts for all or almost all (except for ρ = 1) of the additive variance in the population. Yang *et al.* (2010) showed that empirical SNP chips account for about 50% of the total additive variance for human height (similar proportions are also seen for other traits; Zaitlen *et al.* 2013) and that the missing 50% may be due to incomplete linkage disequilibrium between SNPs and QTL. Because our whole SNP chip explained close to 100% of V_A, the missing heritability found in our simulated GWAS is likely to be due to SNPs whose effects are not large enough or that segregate at too low frequencies to reach statistical significance rather than due to incomplete linkage disequilibrium between causal QTL and their associated SNPs. This occurs because, in our simulations, sampling of individuals is made from a large part of the whole population, and the simulated chip is expected to provide precise information on most segregating variants. Real chips are expected to be less perfect than simulated ones, and real sample sizes are much smaller than the actual population size and also may be biased by population structure and other possible sampling complications. Thus, our results refer to an ideal scenario and should be considered as conservative.

The Sequence analysis detected about 50% more QTL than the Chip analysis, but these additional QTL were generally of small effect and produced an increase of less than 20% in the proportion of additive genetic variance explained (Figure 3B). This implies that increasing the efforts devoted to detecting new QTL of minor effect is likely to increase their number but will only marginally increase the additive variance explained. In fact, increasing sample size makes it possible to identify more causal variants, suggesting that many variants of small effect are yet to be found. For example, a GWAS of human height with a discovery sample of 130,000 individuals identified 180 loci, accounting for 10% of the phenotypic variation (Lango Allen *et al.* 2010). The latest GWAS meta-analysis of 79 studies analyzing more than 250,000 individuals identified 679 variants (Wood *et al.* 2014), the effects of which range from very minor (<1 mm, ~0.01 pSD) to vary large (>300 mm, >3 pSDs). However, this threefold increase in the number of identified variants implied a modest increase (up to 16%) in the proportion of phenotypic variance explained. In fact, when the number of variants considered was extended to the ~9500 SNPs with the lowest *P*-value, the proportion of phenotypic variance explained rose only to 29%. This suggests that much missing heritability is due to a lack of power to detect variants of relatively small effect and/or frequencies (Lango Allen *et al.* 2010; Yang *et al.* 2010; Bloom *et al.* 2013) and that increasing the size of the GWAS will only decrease the missing heritability modestly.

For the Chip analysis, a large number of QTL were detected via linkage to a neutral tag SNP. The estimated effects and frequencies of these tag SNPs were generally highly correlated with the corresponding true effects and frequencies of the QTL (Figure S4). However, the genetic distances between tag SNPs and causal QTL for the Chip analysis were typically large (mean 15.5 kb, median 5.2 kb) (Figure S4). Thus, although most tag SNPs are close to and have high linkage disequilibrium with their corresponding causal QTL, some could be tens or even hundreds of kilobases apart. These results are consistent with a number of observations indicating that most GWAS-hit SNPs are within tens of or a few hundred kilobases of candidate QTL (see Orozco *et al.* 2010).

We also addressed the issue of whether synthetic associations, as proposed by Dickson *et al.* (2010), are a general feature of GWAS analysis. Under this hypothesis, the effect

sizes of causal variants are much larger that the effects associated with the common tag SNPs. We found that about 35% of significant SNPs detected in the Chip analysis were caused by two or more QTL (Figure 5). However, these causal variants were common rather than rare. In addition, in these cases, one of the QTL had a leading effect on the tag SNP, whereas the other(s) had a secondary role, having a much smaller effect on the trait and much lower linkage disequilibrium with the tag SNP than the main causal QTL (Figure S5). Therefore, although a substantial proportion of the significant SNPs can be associated with more than one QTL, only one of them has a main role in the signal, the other contributing little to the variation for the trait. We thus conclude that synthetic associations are not expected to be a major explanation for missing heritability. This is in concordance with the simulation results of Thornton *et al.* (2013) and the arguments of Orozco *et al.* (2010) and Wray *et al.* (2011) regarding the disagreement between some of the predictions from the synthetic association model and the empirical GWAS observations.

QTL detection by multiple-SNP methods

Our results thus suggest that even with knowledge of full sequence data, single-SNP GWAS has power only to detect variants contributing at most 50% of the additive genetic variance. Other methods combining the *P*-values for multiple SNPs (*e.g.*, SKAT) (Wu *et al.* 2011) or analyzing the amount of variance contributed by specific segments of the genome (*e.g.*, RHM) (Nagamine *et al.* 2012) have been proposed as more powerful alternatives to single-SNP GWAS. For example, Thornton *et al.* (2013) found that SKAT and a new method developed by them [excess of significant markers (ESM)] based on the excess of marginally significant markers in a given genomic region could increase power substantially relative to that of single-SNP GWAS. Thus, for the most successful mutational scenario assumed by Thornton *et al.* (2013), the power of detection was shown to be increased from 38% by single-SNP GWAS to 55 and 77% by SKAT and ESM, respectively. Uemoto *et al.* (2013) found that RHM had greater power to detect rare variants in a genomic region than the other single- or multiple-SNP methods, in some scenarios giving double the power of single-SNP GWAS.

Our simulations are analogous to those of Thornton *et al.* (2013) and Uemoto *et al.* (2013). However, there are a number of differences regarding the mutational models considered. The simulations of Uemoto *et al.* (2013) assumed a fixed number of 1, 5, or 10 QTL in a given genomic region. Thornton *et al.* (2013) assumed mutation and recombination rates analogous to ours, but the number and distribution of effects of mutations with effects on the trait were somewhat different. They assumed that 10% of mutations had an effect on the trait, while we assumed a much lower proportion ($<$1%). In addition, they assumed that mutation effects were sampled from an exponential distribution with mean effect from 0 up to 0.5, whereas our assumed distribution was very much leptokurtic. Despite these differences between the models simulated, we also found that RHM detects about 13% more QTL, on average, than GWAS on full sequence data, in agreement with the preceding studies, suggesting that multiple-SNP or genomic region analyses are more powerful than single-SNP GWAS. Our results further show, however, that the additional QTL detected by RHM contribute very little to the overall additive genetic variance (Table 1) mainly because many QTL of small effect are captured by the signal of a QTL of strong effect in a nearby region (Figure S6 and Table S4). In fact, some major QTL segregating at substantial frequencies detected by GWAS were missed by RHM if the same threshold probability value was assumed (Table 1), implying that the amount of variance explained by the QTL detected by RHM was lower than that explained by those detected by GWAS (Table 1).

Conclusions

The pleiotropic simulation model considered suggests that most variation is due to common QTL of substantial effect, except when the trait under study is fitness itself, in which case most variation is due to rare QTL. In this model, the missing heritability is explained by the inability to detect a substantial number of QTL of moderate effect (0.03–0.3 pSD) segregating at substantial frequencies rather than rare ones because these generally contribute little to genetic variation. The use of full sequence data and regional heritability mapping can improve the detection rate of variants, but they seem to be of limited efficiency in reducing missing heritability.

Acknowledgments

We thank Adam Eyre-Walker for helpful discussions on early stages of this work and two anonymous referees for useful comments on the manuscript. This work was funded by the Ministerio de Economía y Competitividad (CGL2012-39861-C02-01), Xunta de Galicia (GPC2013-011), and Fondos Feder: "Unha maneira de facer Europa."

Literature Cited

Agarwala, V., J. Flannick, S. Sunyaev GoT2D Consortium, and D. Altshuler, 2013 Evaluating empirical bounds on complex disease genetic architecture. Nat. Genet. 45: 1418–1427.

Aulchenko, Y., S. Ripke, A. Isaacs, and C. Van Duijn, 2007 GenABEL: an R library for genorne-wide association analysis. Bioinformatics 23: 1294–1296.

Bloom, J., I. Ehrenreich, W. Loo, T. Lite, and L. Kruglyak, 2013 Finding the sources of missing heritability in a yeast cross. Nature 494: 234–237.

Boyko, A., S. Williamson, A. Indap, J. Degenhardt, R. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4: e100083.

Caballero, A., and P. D. Keightley, 1994 A pleiotropic model of variation in quantitative traits. Genetics 138: 883–900.

Cebamanos, L., A. Gray, I. Stewart, and A. Tenesa, 2014 Regional heritability advanced complex trait analysis for GPU and traditional parallel architectures. Bioinformatics 30: 1177–1179.

Charlesworth, B., 2009 Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10: 195–205.
Dickson, S., K. Wang, I. Krantz, H. Hakonarson, and D. Goldstein, 2010 Rare variants create synthetic genome-wide associations. PLoS Biol. 8: e1000294.
Eyre-Walker, A., 2010 Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc. Natl. Acad. Sci. USA 107: 1752–1756.
Gibson, G., 2012 Rare and common variants: twenty arguments. Nat. Rev. Genet. 13: 135–145.
Golan, D., E. Lander, and S. Rosset, 2014 Measuring missing heritability: inferring the contribution of common variants. Proc. Natl. Acad. Sci. USA 111: E5272–E5281.
Halligan, D. L., A. Kousathanas, R. W. Ness, B. Harr, L. Eory *et al.*, 2013 Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. PLoS Genet. 9: e1003995.
Hemani, G., S. Knott, and C. Haley, 2013 An evolutionary perspective on epistasis and the missing heritability. PLoS Genet. 9: e1003295.
Hill, W., 2010 Understanding and using quantitative genetic variation. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 73–85.
Hill, W., M. Goddard, and P. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4: e1000008.
Keightley, P. D., and W. G. Hill, 1990 Variation maintained in quantitative traits with mutation-selection balance: pleiotropic side effects on fitness traits. Proc. Biol. Sci. 242: 95–100.
Kong, A., D. Gudbjartsson, J. Sainz, G. Jonsdottir, S. Gudjonsson *et al.*, 2002 A high-resolution recombination map of the human genome. Nat. Genet. 31: 241–247.
Kryukov, G., L. Pennacchio, and S. Sunyaev, 2007 Most rare missense alleles are deleterious in humans: Implications for complex disease and association studies. Am. J. Hum. Genet. 80: 727–739.
Lango Allen, H., K. Estrada, G. Lettre, S. Berndt, M. Weedon *et al.*, 2010 Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832–838.
Lee, S. H., N. R. Wray, M. E. Goddard, and P. M. Visscher, 2011 Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. 88: 294–305.
Li, W., and L. Sadler, 1991 Low nucleotide diversity in man. Genetics 129: 513–523.
Lohmueller, K. E., 2014 The impact of population demography and selection on the genetic architecture of complex traits. PLoS Genet. 10: e1004379.
Maher, B., 2008 Personal genomes: the case of the missing heritability. Nature 456: 18–21.
Manolio, T., F. Collins, N. Cox, D. Goldstein, L. Hindorff *et al.*, 2009 Finding the missing heritability of complex diseases. Nature 461: 747–753.
Messer, P., 2013 SLiM: simulating evolution with selection and linkage. Genetics 194: 1037–1039.
Mousseau, T. A., and D. A. Roff, 1987 Natural selection and the heritability of fitness components. Heredity 59: 181–197.
Nagamine, Y., R. Pong-Wong, P. Navarro, V. Vitart, C. Hayward *et al.*, 2012 Localising loci underlying complex trait variation using regional genomic relationship mapping. PLoS One 7: e46501.
North, T. L., and M. A. Beaumont, 2015 Complex trait architecture: the pleiotropic model revisited. Sci. Reports 5: 9351.
Orozco, G., J. Barrett, and E. Zeggini, 2010 Synthetic associations in the context of genome-wide association scan signals. Hum. Mol. Genet. 19: R137–R144.
Purcell, S. B., K. Neale, L. Todd-Brown, M. Thomas, D. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559–575.
Rowe, S. J., A. Rowlatt, G. Davies, S. E. Harris, D. J. Porteous *et al.*, 2013 Complex variation in measures of general intelligence and cognitive change. PLoS One 8: e81189.
Simons, Y. B., M. C. Turchin, J. K. Pritchard, and G. Sella, 2014 The deleterious mutation load is insensitive to recent population history. Nat. Genet. 46: 220–224.
Spencer, C., Z. Su, P. Donnelly, and J. Marchini, 2009 Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 5: e1000477.
Thornton, K., A. Foran, and A. Long, 2013 Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. PLoS Genet. 9: e1003258.
Torgerson, D. G., A. R. Boyko, R. D. Hernandez, A. Indap, X. Hu *et al.*, 2009 Evolutionary processes acting on candidate *cis*-regulatory regions in humans inferred from patterns of polymorphism and divergence. PLoS Genet. 5: e1000592.
Uemoto, Y., R. Pong-Wong, P. Navarro, V. Vitart, C. Hayward *et al.*, 2013 The power of regional heritability analysis for rare and common variant detection: simulations and application to eye biometrical traits. Front. Genet. 4: 232.
UK10K Consortium, 2015 The UK10K project identifies rare variants in health and disease. Nature 526: 82–90.
Visscher, P., 2008 Sizing up human height variation. Nat. Genet. 40: 489–490.
Visscher, P., M. Brown, M. McCarthy, and J. Yang, 2012 Five years of GWAS discovery. Am. J. Hum. Genet. 90: 7–24.
Wang, K., S. Dickson, C. Stolle, I. Krantz, D. Goldstein *et al.*, 2010 Interpretation of association signals and identification of causal variants from genome-wide association studies. Am. J. Hum. Genet. 86: 730–742.
Wood, A., T. Esko, J. Yang, S. Vedantam, T. Pers *et al.*, 2014 Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46: 1173–1186.
Wray, N., S. Purcell, and P. Visscher, 2011 Synthetic associations created by rare variants do not explain most GWAS results. PLoS Biol. 9: e1000579.
Wu, M., S. Lee, T. Cai, Y. Li, M. Boehnke *et al.*, 2011 Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. 89: 82–93.
Yang, J., B. Benyamin, B. McEvoy, S. Gordon, A. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569.
Yang, J., S. Lee, M. Goddard, and P. Visscher, 2011 GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88: 76–82.
Yang, J., A. Bakshi, Z. Zhu, G. Hemani, A. E. Vinkhuyzen *et al.*, 2015 Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. 47: 1114–1120.
Zaitlen, N., P. Kraft, N. Patterson, B. Pasaniuc, G. Bhatia *et al.*, 2013 Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet. 9: e1003520.
Zuk, O., E. Hechter, S. Sunyaev, and E. Lander, 2012 The mystery of missing heritability: genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA 109: 1193–1198.

Communicating editor: D. J. de Koning

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177220/-/DC1

The Nature of Genetic Variation for Complex Traits Revealed by GWAS and Regional Heritability Mapping Analyses

Armando Caballero, Albert Tenesa, and Peter D. Keightley

DOI: 10.1534/genetics.115.177220

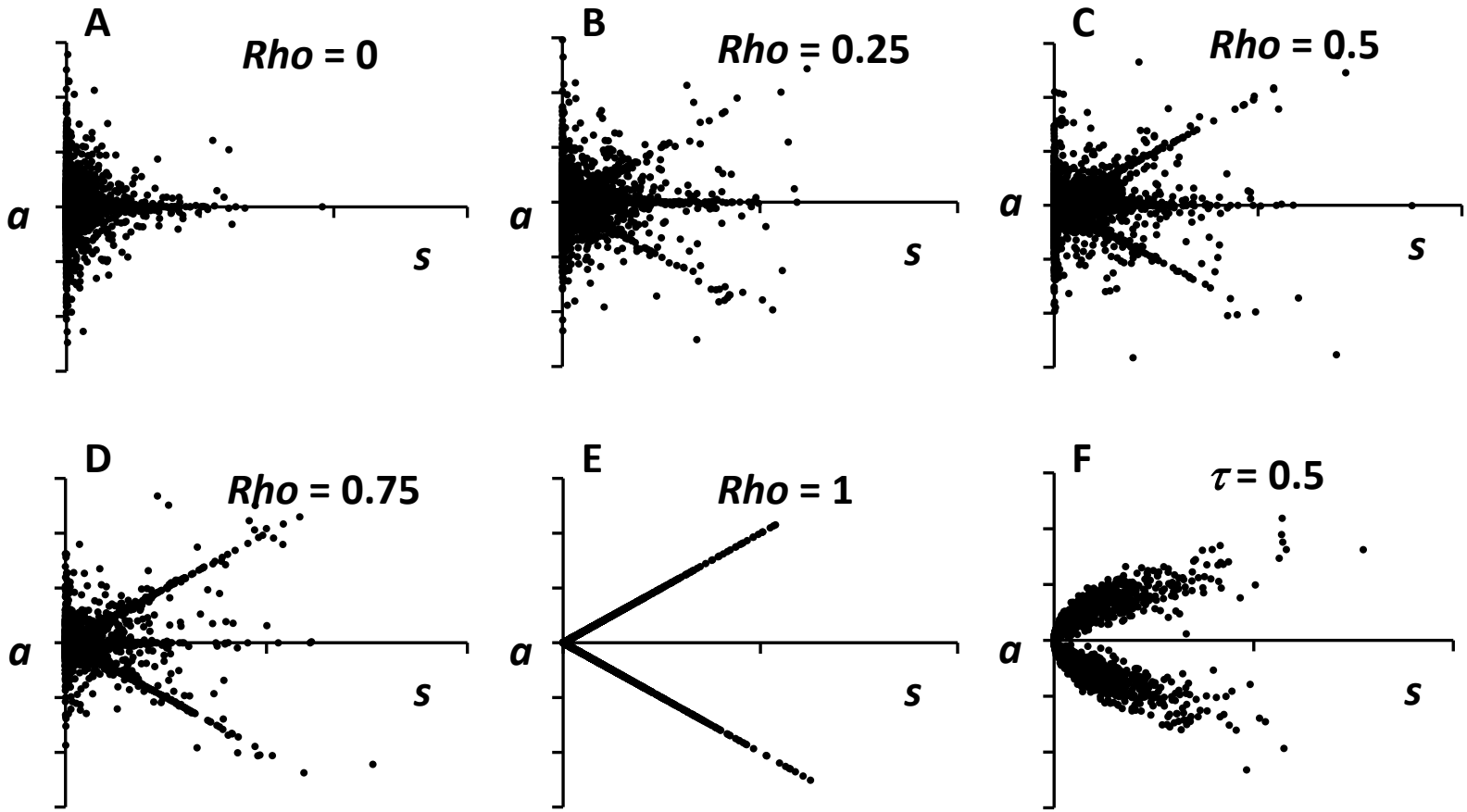

Figure S1. Joint distribution of effects on the quantitative trait (*a*) and fitness (*s*) for different values of the correlation between trait and fitness effects (*rho*). Graphs (A-E) are obtained with a bivariate gamma distribution. Graph (F) illustrates the pleiotropic model proposed by Eyre-Walker (2010), where mutation trait effects (*a*) are related to fitness *s* effects by the expression $a = \delta s^{\tau} (1 + \varepsilon) C$. In this case, $\delta = 1$, $\tau = 0.5$, $C = 3$, and ε is a normal deviate with mean zero and standard deviation 0.25.

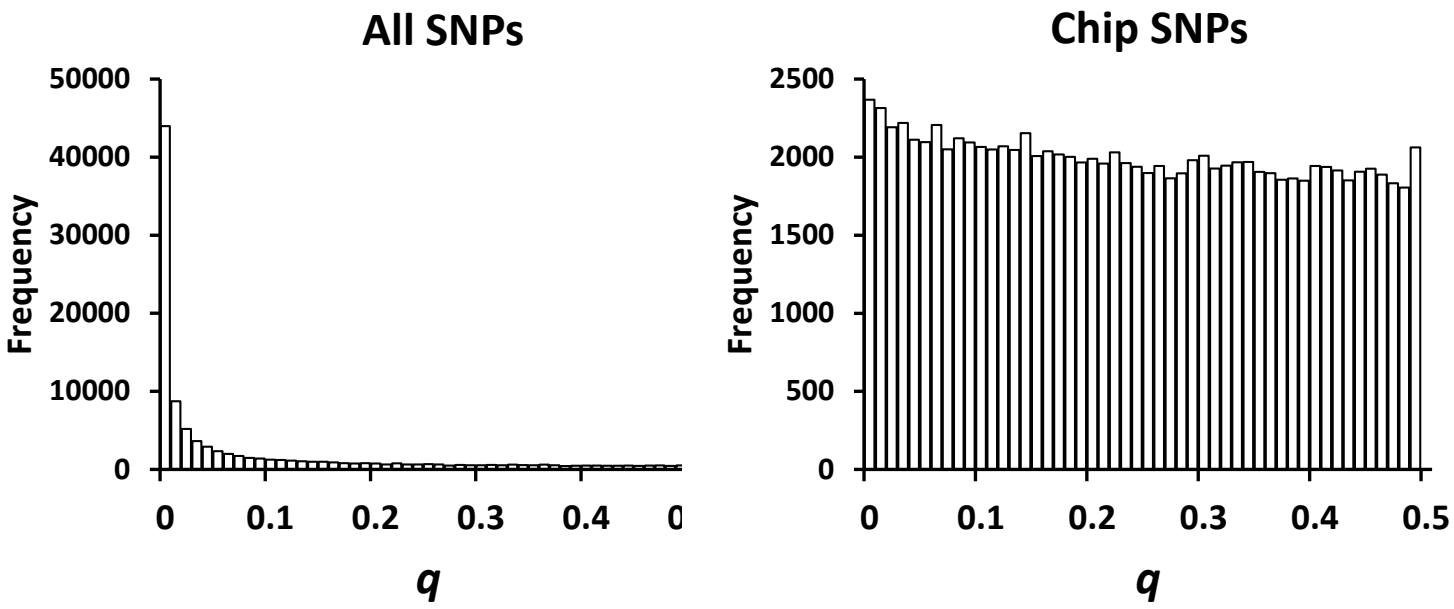

Figure S2. Distribution of SNP minor allele frequencies (MAF) in the genome. (A) SNPs simulated. (B) SNPs analysed in the Chip panel.

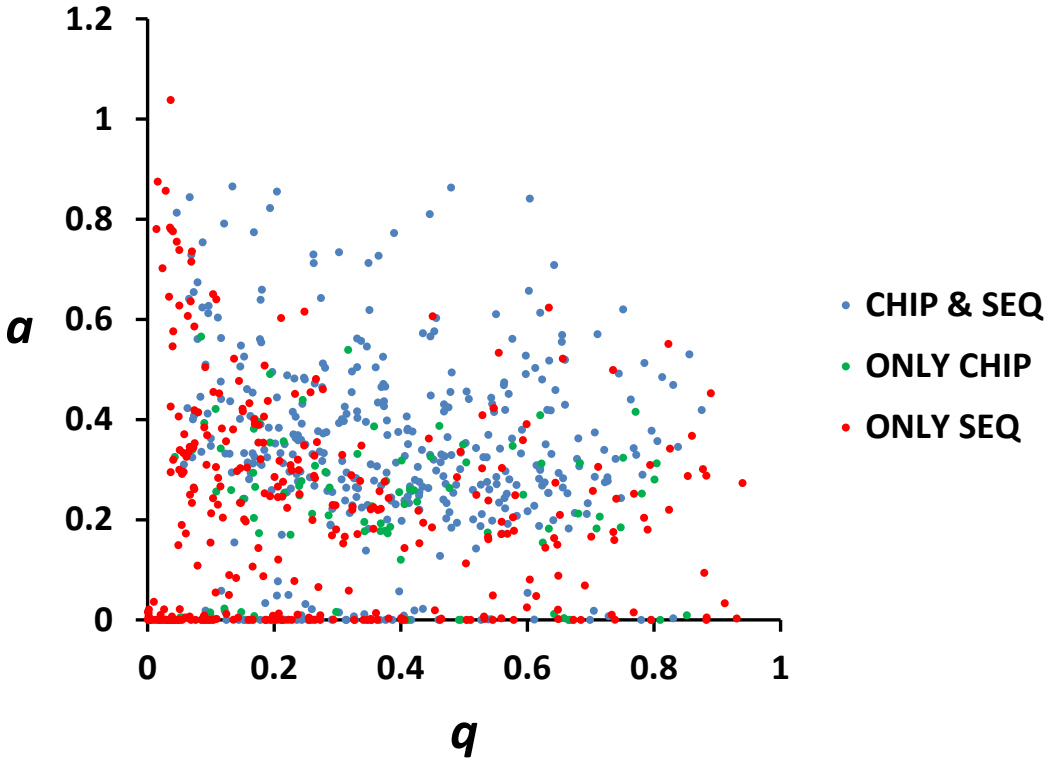

Figure S3. Joint distribution of effects on the trait (*a*; in phenotypic standard deviations) and frequencies (*q*) of all detected QTLs, separating those detected by the Chip analysis and Sequence analysis (blue), those detected only in the Chip analysis (green) and those detected only in the Sequence analysis (red). Results combining all genomes and values of *rho*.

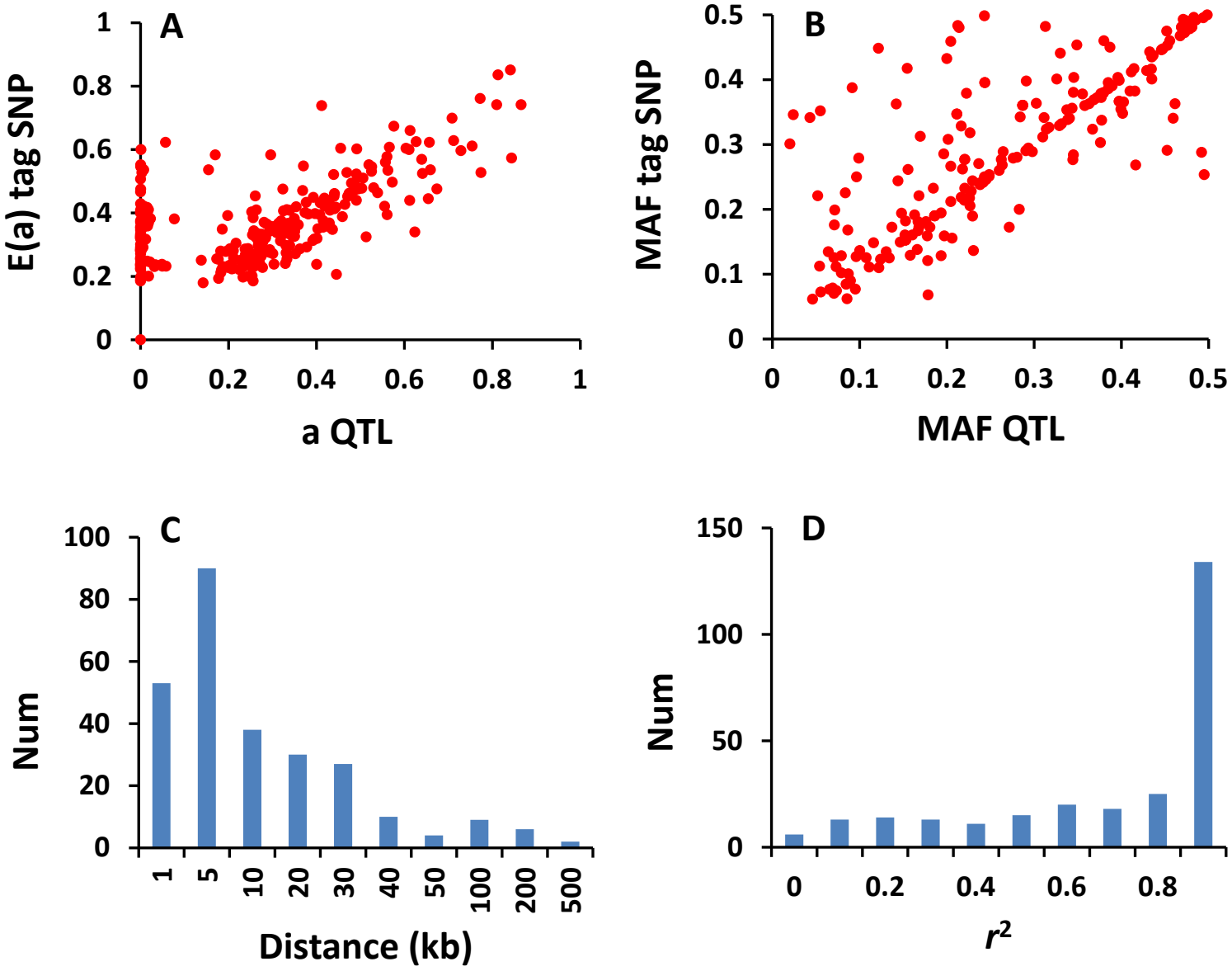
A
E(a) tag SNP
a QTL
0
0.2
0.4
0.6
0.8
1
B
MAF tag SNP
MAF QTL
0
0.1
0.2
0.3
0.4
0.5
C
Num
Distance (kb)
0
20
40
60
80
100
1
5
10
20
30
40
50
100
200
500
D
Num
r^2
0
50
100
150
0.2
0.4
0.6
0.8

Figure S4. Relationship between the most associated tag neutral SNPs and their corresponding causal QTLs in the Chip analysis (results combining all genomes and values of *rho*). (A) Relationship between estimated effects on the trait for the tag SNPs [E(*a*)] against the corresponding true effects (*a*) in the causal QTLs (effects shown in phenotypic standard deviations). (B) Relationship between the minor allele frequencies (MAF) of the tag SNPs and their corresponding causal QTLs. (C) Distribution of distances (in kilobases) between the tag SNPs and their corresponding causal QTLs. (D) Distribution of linkage disequilibrium (r^2) between the tag SNPs and their corresponding causal QTLs.

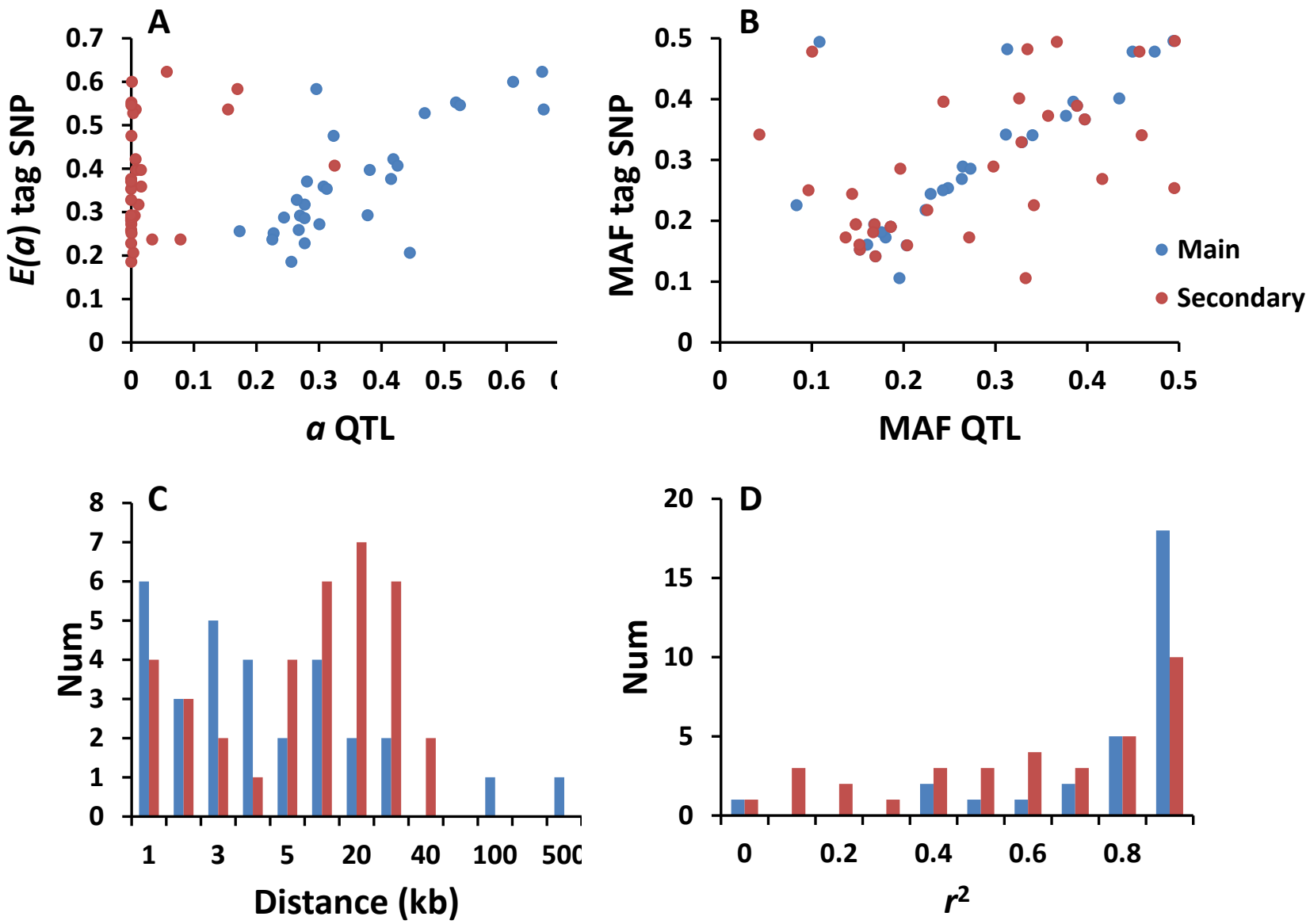
A
E(a) tag SNP
a QTL
B
MAF tag SNP
MAF QTL
Main
Secondary
C
Num
Distance (kb)
D
Num
r2

Figure S5. Relationship between the most associated tag neutral SNPs and their corresponding causal QTLs in the Chip analysis (results combining all genomes and values of *rho*) in those cases where the significance of a given significant SNP is due to two or more QTLs. The main QTL (that with the largest effect on the trait) is shown in blue, whereas the secondary QTLs are shown in red. (A) Relationship between estimated effects on the trait for the tag SNPs [E(*a*)] against the corresponding true effects (*a*) in the causal QTLs (effects shown in phenotypic standard deviations). (B) Relationship between the minor allele frequencies of the tag SNPs and their corresponding causal QTLs. (C) Distribution of distances (in kilobases) between the tag SNPs and their corresponding causal QTLs. (D) Distribution of linkage disequilibrium (r^2) between the tag SNPs and their corresponding causal QTLs.

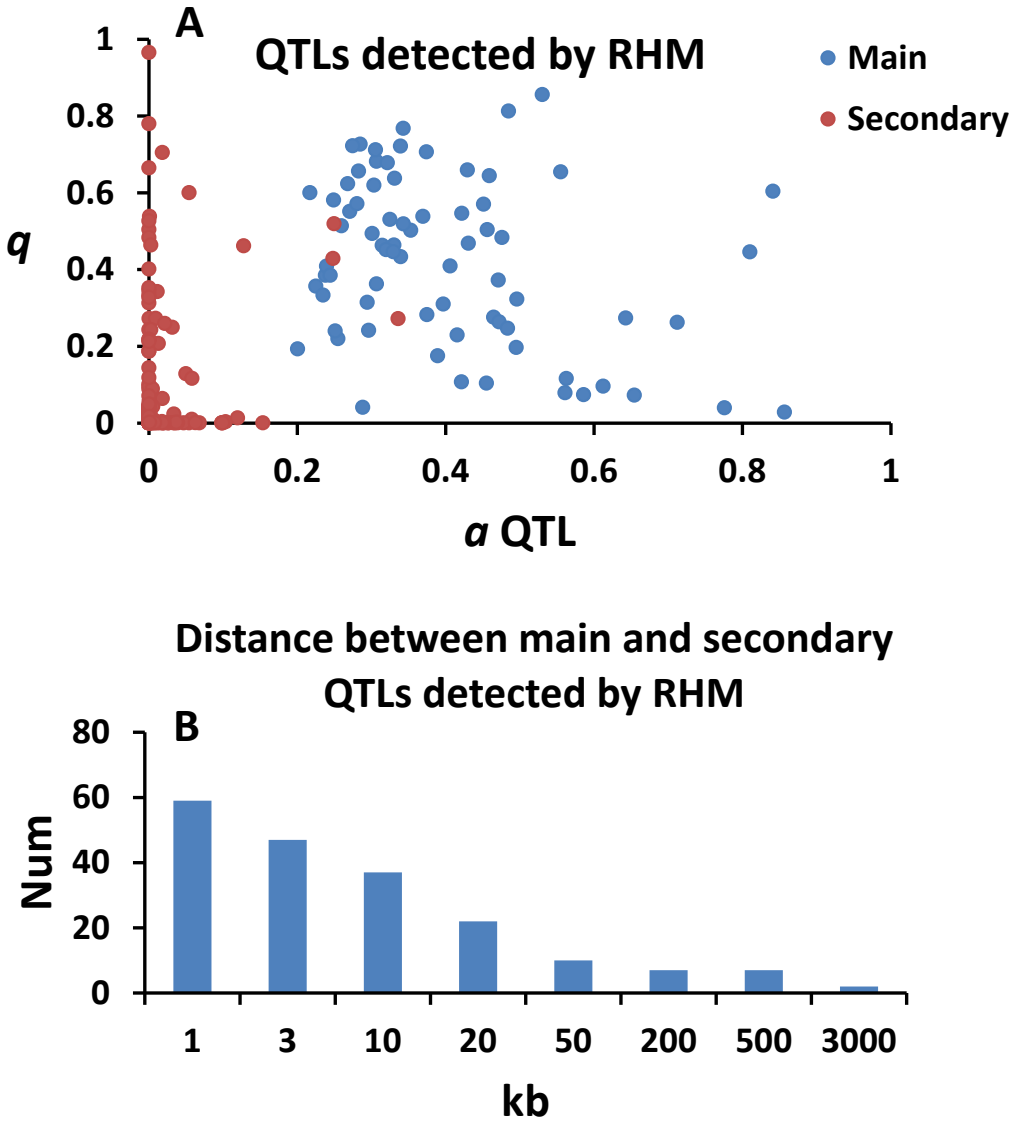

Figure S6. (A) Distribution of absolute effects (*a*; in phenotypic standard deviations) and frequencies (*q*) for QTLs detected by RHM method, distinguishing between the QTLs with the largest effect in a given group of significant regions (Main; in blue) and the remaining detected QTLs (Secondary; in red). (B) Distribution of distances (in kilobases) between the many and secondary QTLs. There were in total 105 main QTLs and 309 corresponding secondary QTLs. Thus, there were about 3 QTLs of small effect discovered because of the signal of a large effect one.

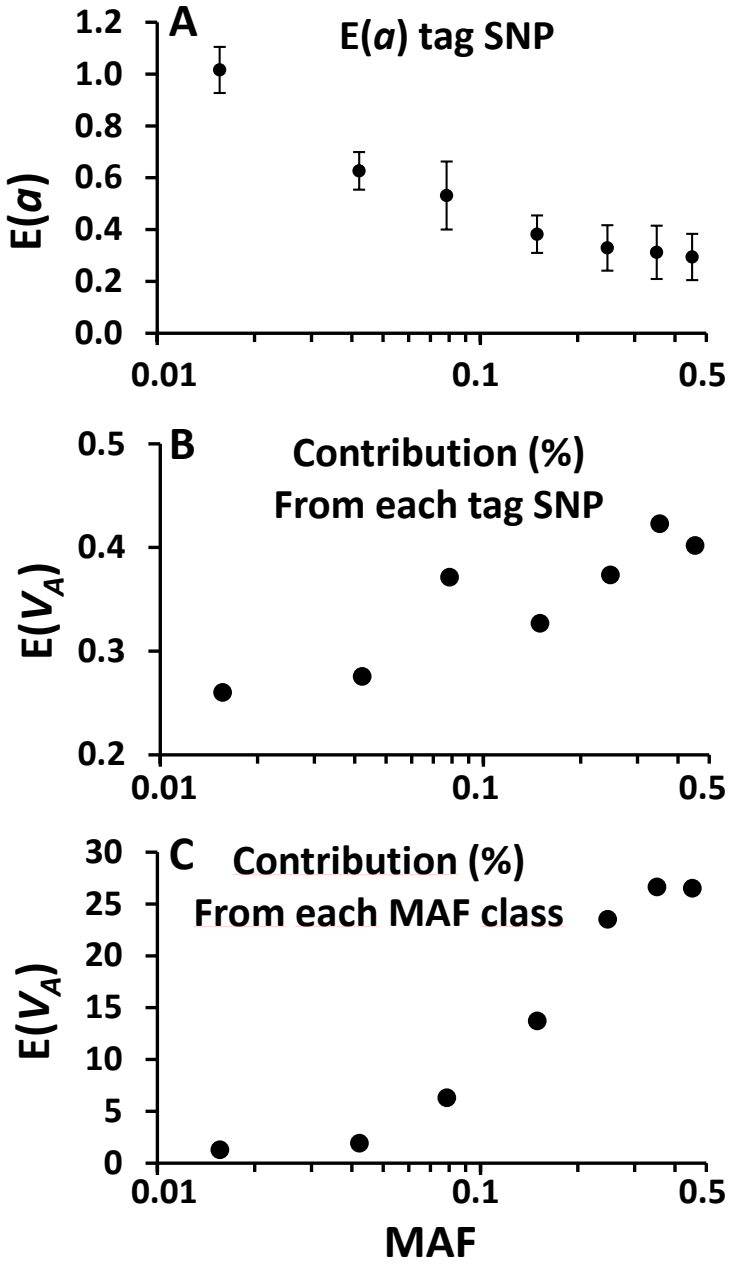
A
E(a) tag SNP
E(a)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.01
0.1
0.5
B
Contribution (%)
From each tag SNP
E(V_A)
0.5
0.4
0.3
0.2
0.01
0.1
0.5
C
Contribution (%)
From each MAF class
E(V_A)
30
25
20
15
10
5
0
0.01
0.1
0.5
MAF

Figure S7. Allelic spectrum showing the simulated estimated homozygous effects (E(*a*)) and contribution to additive genetic variance (E(V_A)) of significant SNPs, against their minor allelic frequency (MAF), for the scenario *rho* = 0.5. The 263 SNPs are arbitrarily classified in 7 MAF classes (ranges 0.01-0.03, 0.03-0.05, 0.05-0.1, 0.1-0.2, 0.2-0.3, 0.3-0.4 and 0.4-0.5) which correspond to the 7 dots in each of the panels. Panel A gives the mean homozygous effect (E(*a*)) of the SNPs in the class, bars denoting one standard deviation at each side of the mean. This panel clearly shows that significant SNPs of larger estimated effect have lower frequencies. Panel B shows the contribution to the predicted additive variance (E(V_A)) from each single SNP in each class, as a percentage of the total predicted additive variance from all SNPs. Panel C shows the contribution to the predicted additive variance (E(V_A)) from all SNPs in each class, again as a percentage of the total predicted additive variance from all SNPs. Panels B and C show that SNPs with lower frequencies contribute proportionately less to E(V_A) than common SNPs.

Table S1. Average numbers, effects and frequencies (standard errors in parenthesis) of significant SNPs and QTLs detected per genome in the Chip and Sequence analysis for each value of the correlation between quantitative trait effects and fitness effects (*rho*). Mean effects are given in phenotypic standard deviations (psd) for the trait.

rho	**0**		**0.25**		**0.5**		**0.75**	
	Chip	**Seq**	**Chip**	**Seq**	**Chip**	**Seq**	**Chip**	**Seq**
Number of significant SNPs	21.0	60.0	50.0	128.0	61.0	188.80	48.50	145.70
	(3.45)	8.87	(5.97)	(10.62)	(5.21)	(22.88)	(7.10)	(15.27)
Absolute estimated effect (psd)	0.18	0.16	0.17	0.12	0.16	0.11	0.18	0.16
	(0.02)	(0.04)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Frequency	0.37	0.38	0.40	0.39	0.39	0.37	0.34	0.32
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Number of QTLs detected	11.00	16.05	20.84	30.50	30.00	43.83	20.00	29.84
	(0.83)	(2.36)	(0.85)	(1.52)	(0.95)	(1.68)	(1.81)	(2.12)
Effect (psd)	0.25	0.24	0.28	0.23	0.27	0.24	0.35	0.33
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Frequency	0.39	0.39	0.41	0.34	0.39	0.35	0.35	0.28
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)

Table S2. Averaged statistics regarding the relationship between causal QTLs and tag SNPs in the Chip and Sequence analysis for each value of the correlation between quantitative trait effects and fitness effects (*rho*). First row: average number of significant SNPs associated to a given causal QTL. Second row: average number of QTLs responsible for the significance of a given tag SNP. Third row: median distances (in kb) between the tag SNPs and their corresponding causal QTLs. Fourth row: average linkage disequilibrium (r^2) between the tag SNPs and their corresponding causal QTLs.

rho	**0**		**0.25**		**0.5**		**0.75**	
	Chip	**Seq**	**Chip**	**Seq**	**Chip**	**Seq**	**Chip**	**Seq**
Num. significant SNPs per causal QTL	2.67	5.76	3.90	6.17	3.19	7.00	3.32	6.89
Num QTLs causal of a significant SNP	1.26	1.41	1.21	1.29	1.25	1.38	1.19	1.28
Median distance (kb)	4.66	9.36	4.79	53.76	4.48	17.20	2.96	59.53
Linkage disequilibrium (r^2)	0.73	0.50	0.76	0.36	0.74	0.24	0.77	0.33

Table S3. Number of regions and QTLs detected by RHM using different significance cut-off values (α). The columns give the total number of significant regions, the number of those carrying QTLs (and its %), the number of QTLs detected and the proportion of the additive genetic variance (V_A) explained by them, and finally, the total number of QTLs detected which are coincident with QTLs detected by GWAS with cut-off probability of 10^{-8}. The results are the sum for all simulations under scenario with *rho* = 0.5. The lower part of the table shows the corresponding results when, for the same simulations which give the results of the upper part of the table, the phenotypes of individuals are randomly permuted.

Probability(α)	Number of significant regions	Number of significant regions with QTLs	Number of detected QTLs	% VA explained	Number of QTLs coincident with GWAS
10^{-4}	3932	872 (22%)	877	45.9	200
10^{-5}	2215	598 (27%)	606	40.8	186
10^{-6}	1512	450 (30%)	460	37.0	175
10^{-7}	1098	360 (33%)	363	34.0	153
10^{-8}	838	290 (35%)	298	31.1	135
10^{-9}	656	246 (37%)	261	29.2	121
10^{-10}	510	196 (38%)	220	26.8	110
10^{-11}	400	154 (38%)	173	23.1	88
10^{-12}	326	117 (36%)	118	19.8	64
Permutation of phenotypes					
10^{-4}	417	17 (5%)	18	0.0	0
10^{-5}	42	4 (9%)	4	0.0	0
10^{-6}	3	0 (0%)	0	0.0	0
10^{-7}	1	0 (0%)	0	0.0	0
10^{-8}	0	0 (0%)	0	0.0	0
10^{-9}	0	0 (0%)	0	0.0	0
10^{-10}	0	0 (0%)	0	0.0	0
10^{-11}	0	0 (0%)	0	0.0	0
10^{-12}	0	0 (0%)	0	0.0	0

Table S4. Example of results by the method of Regional Heritability Mapping (RHM). In this simulation run for 10 Mb of sequence (the scenario refers to *rho* = 0.5), 18 regions were significant with probability $P < 10^{-6}$. The SNP code starting and ending the regions are shown in the first two columns. The estimated additive genetic variance (V_A) for the region is shown in the third column. The four QTLs detected in the regions, along with their effects and frequencies are shown in the hand-right columns. One of the QTLs had a large homozygous effect (0.4 phenotypic standard deviations, psd) while the other three had very small effects. The explanation suggested is that the three minor QTLs are mostly detected because of the signal of the major one, which is likely to be responsible for the significance of the 18 nearby genome regions.

Region			Detected QTLs		
Starting SNP code	Ending SNP code	Estimated V_A in the region	Code	Effect (a) in psd	Frequency
11841	11860	21.96			
11861	11880	9.95			
11871	11890	15.23			
11881	11900	13.55			
11891	11910	11.10			
11901	11920	10.82			
11911	11930	13.02	11923	-0.40	0.31
11921	11940	16.82			
11931	11950	13.53			
11941	11960	14.41	11960	2×10^{-4}	0.026
11951	11970	19.03			
11961	11980	17.78	11977	7×10^{-5}	0.0005
11971	11990	12.60	11978	3×10^{-6}	0.007
11981	12000	23.94			
11991	12010	20.74			
12001	12020	15.30			
12051	12070	16.36			
12061	12080	17.44			

File S1: Simulation procedure

The protocol involved the simulation of a complex trait correlated with fitness in a large population subject to mutation, drift, selection, and recombination. The genotypic and phenotypic data were analyzed by GWAS using information on typical SNP chips or full sequence data, and Regional Heritability Mapping (RHM) of short genome windows. The whole procedure was carried out by shell scripts which involve the following codes and programs available on request:

(1) A perl script (**run-slim-bivariate.pl**) gets the organization of the genome in a given piece of chromosome (introns, exons, etc.) through the program **Tabix** (Li *et al.* 2011), and runs the software **SLiM** (Messer 2013) which performs individual-based forward simulations of sequences in a population of size *N*, considering mutation, drift and selection. The code of SLiM was modified to provide mutations with pleiotropic effects on a quantitative trait and fitness with a given correlation *rho* between 0 and 1. The bivariate gamma deviates were generated using the GVTR algorithm of Schmeiser and Lal (1982). Tabix software delivers an input file (*ensembl.cds.bed.gz*) containing the CDS coordinates for the fragment of the genome considered. The parameters assumed for the run of SLiM are the population size, the number of generations, the mutation rate, the migration rate, the population size assumed in the last generation, the sample size, and the correlation assumed between fitness and trait effects. The values of the parameters are given in the main text. For the case of *rho* = 0, a value of 0.01 was used instead, as the bivariate routine does not work with an exact value of zero. Each run of SLiM includes 10 Mb of sequence for 10000 generations. For each value of *rho*, 1800 runs were carried out.

(2) A C program (**SNPSLIMBVoutGENinput.c**) opens the SLiM output file (*slimout*) and gets the positions, fitness effects, trait effects and frequencies of the mutations in the sampled population, as well as the individual genotypes for all SNPs. The program makes the data files necessary for the GWAS analyses. Genotypic values of individuals for fitness and the quantitative trait are obtained using a multiplicative fitness model in the first case and an additive model in the second. Environmental deviations are added to the genotypic values for the quantitative trait to obtain phenotypic values. The files *posfile*, *genofile* and *phenofile*, including the genome positions of the mutations, the genotypes of SNPs and the individual phenotypic values, respectively, are produced to be analysed by GWAS. Files *plinkposfile* and *plinkphenofile* (*qt.phe*) are ready to be inputs for the software PLINK (Purcell *et al.* 2007). Files with the lists of all SNPs

(*list_allsnps*), and with those SNPs to be used in a chip analysis (*list_chipsnps*), sampled so that a uniform distribution of frequencies is assumed, are also obtained. The former is used when the Sequence analysis is performed whereas the second is used for the Chip analysis. Genotypic variances for the quantitative trait are calculated and kept in the file *genotypicvariance*.

(3) For the Chip analysis, the software **PLINK** (Purcell *et al.* 2007) is run to prune out SNPs with high linkage disequilibrium ($r^2 > 0.90$) using sliding windows of 1000 SNPs with overlap of 500. The resulting file with the SNPs to be analyses is called *plink.prune.in*. In the scenarios with Sequence analysis this step is skipped, as all SNPs are considered in the analysis.

(4) An R script (*test.R3NEW* for the Chip analysis and *test.R3NEW_RESEQ*) are aimed at performing GWAS using the software **GenABEL** (Aulchenko et al. 2007). For the Chip analysis, filtering of SNPs for MAF (< 0.05), deviate from Hardy-Weinberg equilibrium (HWE) (false discovery rate, FDR < 0.2) and exclusion of individuals with unusually high heterozygosity (FDR < 1%) are made with the function *check.marker*. Cleaned SNPs which have not been pruned out because of high linkage disequilibrium in the previous step are considered for analysis. These filterings are not made for the Sequence analysis. Regression analysis for individual SNPs is carried out with the function *qtscore*. A file (*ProbEffects*) is obtained with probabilities values for the SNPs and their estimated effects.

(5) A C program (**SNPSLIMBVvariance.c**) then gathers results from SLiM and GenABEL to obtain data on additive and dominance variances as well as a number of calculations and files with the SNPs. Significant SNPs are pinpointed according to a cut-off probability value which is an input in the program. Files with the distributions of frequencies and effects are produced for all SNPs, cleaned SNPs, selected SNPs (QTLs), chip SNPs and significant SNPs. Files with lists (*list_clsnps*, *list_selsnps*, *list_sigsnps*) and information (*LDselSNP*, *LDsigSNP*, *SNPselTable*, *SNPsigTable*, *SNPsigselTable*, *SNPallTable*) from selected, significant, and all SNPs are produced. Values of additive variance from QTLs, and estimated additive variance from estimated effects of significant SNPs are obtained. A file (*outfile_VA*) is produced with the summary of the results.

(6) An R script (*testCOV.pruebaNEW* for the Chip analysis and *testCOV.pruebaNEW_RESEQ*) are aimed at finding the causal QTLs responsible for the significance of each SNP. Every QTL from the list of selected SNPs is fitted into the

regression model as a covariate and GWAS is performed again for each of them. The loss of significance of a previously significant SNP implies that the considered QTL is involved in that significance. The list of causal QTLs is held in the file *SNPcovTable*, and the linkage disequilibrium between the tagged SNPs and the causal QTLs in the file *SNPLDTable*.

(7) For the Chip analysis, the **REACTA** software (Cebamanos *et al.* 2014), a version of **GCTA** (Yang *et al.* 2011), is used to obtained unconstrained estimates of the additive genetic variance using the whole chip with the command: ./reacta --bfile data --make-grm --pheno qt.phe --reml --reml-no-constrain --out va_allsnpsun.

(8) A C program (**SNPSLIMBVprocess.c**) takes the information of the causal QTLs and the corresponding tagged SNPs to make a number of calculations: average number of SNPs tagged to each QTL, average number of QTLs responsible for the significance of a given SNP, average distance and linkage disequilibrium between causal QTLs and tagged SNPs, the tagged SNPs most associated with each QTL, etc. which are kept in the files: *rep_avedisld_file*, *rep_causal_file*, *rep_sigpersel_file*, *list_assocsnps* and *REP_SUMMARY_FILE*.

(9) The **REACTA** analysis of sliding windows of 20 SNPs with overlap of 10 SNPs is made with the command: reacta --bfile data --make-grm --pheno qt.phe --reml --multi-region 20 10 --out va_regional. Two C programs (**SNPreacta1.c** and **SNPreacta2.c**) read the results of the analysis and extract the significant windows and the QTLs included in them with a given threshold probability value, keeping them in the files *outSIGreacta* and *outQTLreacta*.

(10) Permutation of phenotypic values for GWAS and REACTA analyses is carried out with the programs **PermutationPHE.c**, **PermutationPLINKPHE.c** and **PermutationPER.c**.

References

Cebamanos, L., A. Gray, I. Stewart, and A. Tenesa, 2014 Regional heritability advanced complex trait analysis for GPU and traditional parallel architectures. Bioinformatics 30: 1177-1179.

Li, H. 2011 Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics 27: 718-719.

Messer, P. 2013 SLiM: Simulating Evolution with Selection and Linkage. Genetics 194: 1037-1039.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. Ferreira, D. Bender, J. Maller, P. Sklar, P. de Bakker, M. Daly, M., *et al.*, 2007 PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559-575.

Schmeiser, B. W., and R. Lal, 1982 Bivariate gamma random vectors. Oper. Res. 30: 355–374.

Yang, J., S. Lee, M. Goddard, and P. Visscher, 2011 GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88: 76-82.