MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 1429–1448 doi: 10.1111/mec.13559

INVITED REVIEWS AND SYNTHESES

The genetic consequences of selection in natural populations

TIMOTHY J. THURMAN*† and ROWAN D. H. BARRETT*

*Redpath Museum and Department of Biology, McGill University, Sherbrooke Street West, Montreal, Quebec, Canada H3A 1B1, †Smithsonian Tropical Research Institute, Panamá, Panamá

Abstract

The selection coefficient, *s*, quantifies the strength of selection acting on a genetic variant. Despite this parameter's central importance to population genetic models, until recently we have known relatively little about the value of *s* in natural populations. With the development of molecular genetic techniques in the late 20th century and the sequencing technologies that followed, biologists are now able to identify genetic variants and directly relate them to organismal fitness. We reviewed the literature for published estimates of natural selection acting at the genetic level and found over 3000 estimates of selection coefficients from 79 studies. Selection coefficients were roughly exponentially distributed, suggesting that the impact of selection at the genetic level is generally weak but can occasionally be quite strong. We used both nonparametric statistics and formal random-effects meta-analysis to determine how selection varies across biological and methodological categories. Selection was stronger when measured over shorter timescales, with the mean magnitude of *s* greatest for studies that measured selection within a single generation. Our analyses found conflicting trends when considering how selection varies with the genetic scale (e.g., SNPs or haplotypes) at which it is measured, suggesting a need for further research. Besides these quantitative conclusions, we highlight key issues in the calculation, interpretation, and reporting of selection coefficients and provide recommendations for future research.

Keywords: evolution, genetics, meta-analysis, natural populations, natural selection, selection coefficient

Received 12 September 2015; revision received 21 December 2015; accepted 27 January 2016

Introduction

Since the publication of Lande and Arnold's landmark methods for calculating selection on quantitative phenotypic traits (Lande & Arnold 1983), the study of selection in natural populations has exploded. Hundreds of studies have generated thousands of estimates of selection on phenotypic traits, and the last 15 years have seen a number of influential reviews and meta-analyses of this data on phenotypic selection. These studies have improved our understanding of the strength and form of phenotypic selection in natural populations (Hoekstra *et al.* 2001; Kingsolver *et al.* 2001; Hereford *et al.* 2004), demonstrated its role in creating phenotypic diversity (Rieseberg *et al.* 2002), and shown how selection varies through time and space (Siepielski *et al.* 2009; Kingsolver & Diamond 2011; Siepielski *et al.* 2011; Morrissey & Hadfield 2012; Siepielski *et al.* 2013).

Of course, biologists have long recognized that natural selection must be transmitted to the genetic level for adaptive evolutionary change to occur. Population genetic models explicitly account for natural selection's role in changing allele frequencies with the parameter *s*, the selection coefficient (Hartl & Clark 2007). Although *s* can have slightly different meanings in different models (Box 1), it generally describes the relative fitness advantage or disadvantage of an allele at a genetic locus. The genetic selection coefficient is thus similar to phenotypic selection gradients and differentials and quantifies the magnitude of natural selection acting on genetic variants. Compared to measures of phenotypic

Correspondence: Timothy J. Thurman, Fax: 514 398 3185; E-mail: timothy.thurman@mail.mcgill.ca

Box 1. The meaning(s) of *s*

The selection coefficient, *s*, can have slightly different meanings in different evolutionary models. In most models *s* represents the difference in mean relative fitness between a reference genotype and another genotype. By definition, the reference genotype has a relative fitness of one. The choice of reference genotype, however, leads to subtle differences in the properties of *s*. First, consider directional selection at a locus with two alleles, A and a, with allele A having higher fitness. Researchers studying mutation have tended to focus on selection against new, deleterious alleles. The homozygote of the most-fit allele is used as the reference genotype, such that $s = 1 - w_{aa}$ and $w_{aa} = 1\text{-}s$. In this case, *s* quantifies the strength of selection against the deleterious allele and has a range from 0 to 1. Studies of adaptation, however, typically focus on selection in favour of beneficial alleles and thus set the homozygote of the less-fit allele as the reference: $s = w_{AA} - 1$, or $w_{AA} = 1 + s$. Here, *s* measures the strength of selection acting in favour of the beneficial allele and has a range from 0 to infinity, as genotypes can have a >100% fitness advantage, at least in theory. It should be noted that under this scenario s_{for_AA} is not equal in magnitude to $s_{against\_aa}$ (see supplementary methods). When the magnitude of *s* is small the difference between s_{for_AA} and $s_{against\_aa}$ will be small, but as the strength of selection increases the difference grows. When $s_{against\_aa}$ equals 1 (a lethal allele), s_{for_AA} equals infinity.

So far this model has ignored dominance, which has important implications for the calculation of *s*. In population genetic models of directional selection, dominance is most often accounted for with the dominance coefficient, *h*. In the single locus, two-allele model described above, the fitness of each genotype would be $w_{AA} = 1$, $w_{Aa} = 1 - hs$, $w_{aa} = 1 - s$. When $h = 0$, A is completely dominant and $w_{AA} = w_{Aa}$. When $h = 1$, A is completely recessive and $w_{aa} = w_{Aa}$. Although the definition of *s* remains the same, the calculated value of *s* could change significantly depending on the assumed or known level of dominance and the method used to estimate selection. While methods that estimate *s* by directly measuring fitness differences between homozygotes are robust to changes in *h*, methods that track changes in allele frequency or that measure fitness in heterozygotes are sensitive to assumptions about dominance. The dominance coefficient was rarely empirically estimated in the studies included in our database. Most studies assumed additive fitness effects ($h = 0.5$) or calculated multiple possible *s* values under difference assumptions of dominance.

In the case of over- or underdominance, slightly different genetic models are used. The heterozygote is defined as the reference, and selection coefficients for or against each homozygote are calculated. Selection may be assumed to be symmetric such that *s* for each homozygote is equal, but other models allow *s* to vary, and might use s_1 and s_2 or *s* and *t* to denote the two selection coefficients.

In the simple, one-locus models described above, *s* quantifies the direct fitness effects of the genetic variant. In real organisms, of course, allelic variants do not occur in isolation. Each generation, the fate of an allele is determined by both the direct effects of that locus on fitness and by the indirect effects of selection operating on other sites that are in linkage disequilibrium (LD) with the focal locus (Smith & Haigh 1974; Charlesworth *et al.* 1993). The situation is analogous to correlated selection on phenotypic traits (Lande & Arnold 1983). At the phenotypic level, biologists can use multiple regression to distinguish between direct selection and total (direct and correlated) selection on a specific phenotype (selection gradients and differentials, respectively; Lande & Arnold 1983; Brodie *et al.* 1995). At the genetic level, isolating the direct effects of an individual locus on fitness is quite difficult (Barrett & Hoekstra 2011). Accounting for the effects of linked sites requires either (i) sufficient recombination to break apart associations with other alleles, (ii) complex, multigeneration crosses such as near-isogenic lines, (iii) replicate populations subject to the same experimental treatment, (iv) sufficient sample sizes and genetic variation such that selected alleles are present in multiple genetic backgrounds, or (v) transgenics. In most other cases, genetic selection coefficients should be interpreted as being analogous to phenotypic selection differentials, not gradients.

selection, however, we know relatively little about the values of *s* in natural populations of organisms. The questions that have been considered in reviews of phenotypic selection remain unanswered at the genetic level: How strong is selection at the genetic level? Is selection most often directional, overdominant, or frequency dependent? What is the distribution of selection coefficients in natural populations, and does that distribution change according to the temporal or genetic scale at which selection is measured? These are important questions in evolutionary biology, but it is only recently that biologists have had sufficient genetic data to address them empirically.

Theoretical models have examined these issues, but their results are difficult to apply to natural populations for a variety of reasons. The first difficulty is the

conceptual division between theories of positive selection and theories of negative selection. The designation of selection as positive or negative is determined by the choice of the allele used as the reference for calculating relative fitnesses and is thus somewhat arbitrary (Box 1). Nevertheless, theoretical models often consider only one mode of selection, and this difference in focus can lead to different results. For example, theoretical models of the fitness effects of new mutations find that beneficial mutations fixed during adaptation are likely exponentially distributed, while the distribution of deleterious mutations can be complex and multimodal (see reviews by Orr 2005b; Eyre-Walker & Keightley 2007; Rockman 2012). Although it is easy to delimit positive and negative selection in theoretical models, drawing this distinction is more difficult in natural systems, where there is considerable debate about whether most populations are at fitness optima (and thus likely to experience mostly negative selection) or maladapted (and thus allowing an opportunity for positive selection). Reciprocal transplant experiments find frequent but not ubiquitous local adaptation (reviewed in Kawecki & Ebert 2004; Leimu & Fischer 2008; Hereford 2009) and published estimates of phenotypic selection indicate that mean trait values for the majority of traits are within two standard deviations of the fitness optimum (Estes & Arnold 2007). Whether these patterns indicate widespread adaptation or maladaptation is open to interpretation, so it is difficult to know *a priori* which set of theory to apply (Hendry & Gonzalez 2008).

Second, within the broad fields of positive and negative selection, theoretical predictions vary greatly based on the assumptions and parameters of specific models. Consider, for example, theories of adaptation that predict the distribution of fitness factors fixed during an adaptive bout (Orr 2005a,b). Models that assume a stationary fitness optimum (e.g., Orr 1998, 2003; Kryazhimskiy *et al.* 2009) predict a different distribution of selection coefficients than models with a moving optimum (e.g., Collins *et al.* 2007; Kopp & Hermisson 2007, 2009a,b). Other factors that can influence this distribution include correlations between traits (Martin & Lenormand 2006), migration between populations (Yeaman & Whitlock 2011), the use of novel versus standing genetic variation (Hermisson & Pennings 2005; Barrett & Schluter 2008), the distance to the fitness optimum (Barrett *et al.* 2006; Seetharaman & Jain 2013), and the number of fitness optima (Martin & Lenormand 2015). Once again, applying this theory requires knowledge of parameters (e.g., amount of migration between locally adapted populations, current level of (mal)adaptation in the population, movement of fitness optima) that can be difficult to estimate for natural populations. Finally, theoretical models usually examine selection at a scale that can be difficult for empiricists to access in natural populations. Most of the models mentioned above consider the fitness effects of single point mutations. Often, biologists must measure selection on different alleles of a gene or QTL; selection acting on these larger genomic intervals might have different properties from selection on SNPs.

In summary, this array of theory is informative but difficult to apply. Until recently, obtaining the data necessary to address these questions empirically was challenging. Although population geneticists have inferred selection at the genetic level by observing changes in Mendelian phenotypes for many years (Appendix S2), direct estimation of selection on genetic variation was only made possible by the revolution in molecular genetic techniques that occurred in the 1970s and 1980s. These methods, and the next-generation sequencing technologies that followed, have allowed researchers to detect natural selection at the genetic level using a variety of methods. We briefly discuss these methods below; see Linnen & Hoekstra (2009) and Hohenlohe *et al.* (2010) for more thorough reviews.

Many observational approaches to quantifying selection rely on measuring changes in allele frequency, which can be detected directly with molecular genetic techniques. Allele frequency changes can occur over time (e.g., an increase in frequency over multiple generations, Nsanzabana *et al.* 2010), across an environmental gradient (e.g., a decrease in frequency across a gradient of insecticide application, Lenormand *et al.* 1999), or between contrasting environments (e.g., frequency differences between two locally adapted populations, Hoekstra *et al.* 2004). Another important observational approach is the detection of selection from features of DNA sequence variation. These features can include (but are not limited to) haplotype structure (e.g., Quesada *et al.* 2003), patterns of linkage disequilibrium (LD) around a selected locus (e.g., Ohashi *et al.* 2011), and reduction in variation around a selected locus (e.g., Orengo & Aguade 2007). The limitation of these approaches is that they alone cannot explicitly determine the process that led to the observed patterns of allele frequency or nucleotide variation (Barrett & Hoekstra 2011). Thus, observational approaches often use population genetic modelling, simulations, and statistical analysis to rule out the possibility that only genetic drift or other neutral forces (e.g., demographic changes) could have produced the observed pattern (Excoffier *et al.* 2009; Li *et al.* 2012; Vitti *et al.* 2013). Nevertheless, all estimates of selection likely contain some imprecision due to drift. This problem of determining causality can sometimes be mitigated with experimental approaches. By tracking changes in allele frequency during experimentally controlled selection in the field, researchers can accurately measure selection and identify the agent imposing it (Linnen & Hoekstra 2009).

Over the past few decades, biologists have made use of all of these techniques, and others (e.g., Robinson *et al.* 2012), to quantify selection acting at the genetic level in natural populations. In this study, we gathered those estimates of selection to address a number of key questions. Given the difficulty of applying population genetic theory to predict the distribution of selection coefficients, we first plotted this distribution to see how it differs between biological and methodological categories. Next, we used nonparametric statistics and generalized linear mixed models to quantify how the magnitude of selection varies across temporal and genetic scales. Meta-analyses of phenotypic selection and evolutionary rates indicate that strong phenotypic selection is rarely maintained for long and that long-term estimates of selection or rates of evolutionary change tend to be weaker than short-term estimates (Gingerich 1983; Hoekstra *et al.* 2001; Kinnison & Hendry 2001; Siepielski *et al.* 2009). We predicted that this inverse relationship between strength of selection and temporal scale would be true of selection at the genetic level as well. We hypothesized that selection would also vary based on the genetic unit at which it was measured. Specifically, we assumed that strength of selection on a locus would be proportional to the amount of phenotypic variance that the genetic unit can explain. We reasoned that, with some exceptions, larger genetic units (e.g., allelic variants of a gene or QTL) would tend to have larger phenotypic effects than SNPs. Thus, we predicted that the strength of selection would increase with genetic scale. Finally, we highlight a number of important issues regarding the calculation and interpretation of selection coefficients and make recommendations for further research that will improve our understanding of this important evolutionary parameter.

Materials and methods

Literature search

To assemble our database, we searched for journal articles reporting selection coefficients in a number of ways. First, we searched the Web of Science database system using three different search terms: 'selection coefficient*', 'genotyp* selection' and 'adapt* gene'. We excluded books and search results from scientific fields outside of ecology and evolution (see supplementary methods). Second, we searched the preliminary literature database of Siepielski *et al.* 2013, a meta-analysis of spatial variation in phenotypic selection, for journal articles that were excluded from their analysis for studying genotypes instead of phenotypes. Third, we searched the weekly tables of contents from a number of journals that focus on evolutionary biology and genetics (see supplementary methods). Finally, while determining which studies met our inclusion criteria, we noted references to papers that might have reported selection coefficients and added those to our database. In total, we examined approximately 2200 papers for estimates of natural selection at the genetic level.

Inclusion criteria

To be included in our quantitative analysis, studies needed to satisfy three criteria. First, the study had to report a selection coefficient on a genetic unit (s). Estimates of s that were equal to zero were not included, as they did not detect selection acting on a locus (see supplementary analysis). Selection coefficients can have different meanings under different population genetic models, but in most cases they quantify the difference in mean relative fitness between the most- and least-fit homozygotes (Box 1). We excluded a small number of studies that reported selection coefficients that did not follow this model and thus had different properties from the rest of the calculated estimates. We also analysed directional selection separately from over- and underdominance. Selection coefficients scaled by effective population size (e.g., γ or δ) were excluded, as were estimates of s that reported a range of possible values without specifying a median or point estimate. A number of studies reported relative fitnesses for genotypes without explicitly calculating a selection coefficient. In those cases we used the relative fitnesses to calculate selection against the less-fit homozygote ($s = 1\text{-}w_{aa}$).

Second, selection coefficients needed to be calculated for a specific genetic unit. For our analysis, we categorized these units as either 'SNP', which includes point mutations and single nucleotide polymorphisms, or 'haplotype', which includes all larger genetic units (e.g., insertions or deletions of more than one base pair, allelic variants of genes, allozymes, microsatellite loci, and quantitative trait loci). A number of studies used DNA sequence data to estimate the distribution of selection coefficients or average strength of selection acting on a set of genetic loci or type of mutation but did not calculate locus-specific selection coefficients. For example, Turchin *et al.* (2012) estimated the average strength of selection on ~1400 individual SNPs associated with increased height in Europeans, but did not report estimates of s for each SNP. These average selection coefficients were excluded from our analysis.

Finally, studies needed to measure selection operating in natural populations. Thus, we excluded measures of selection in laboratory populations or in domesticated plants and animals. Estimates of selection in humans were included, as were estimates of selection from

experimentally manipulated natural populations or organisms introduced into suitable habitat.

For each study that satisfied these criteria, we recorded the absolute value of *s*, whether selection was positive or negative, any measures of error, the statistical significance of the coefficient, the number of generations over which selection was measured, the genetic unit at which selection was measured, the method used to calculate the selection coefficient, whether the estimate of selection came from observation or a manipulative experiment, and other information (Table S2, Supporting information). We modified this raw database in three ways to prepare it for quantitative analysis. First, to avoid pseudoreplication, we removed estimates of selection that were calculated from the same data as other selection coefficients. In most cases this occurred when one study reported alternative estimates of selection for the same genetic unit under different evolutionary parameters (e.g., generation times or degrees of dominance). When the authors deemed one set of parameters most biologically plausible, we included the selection coefficient from that model. Otherwise, we flipped a coin to randomly select one selection coefficient to include. Some studies calculated selection coefficients from data previously reported in other studies. If the original study also reported selection coefficients, we included whichever study reported more selection coefficients. If the studies reported equal numbers of selection coefficients, we included the original study.

Second, some studies reported selection coefficients from the same data at different temporal scales or for different fitness components. In such cases, we included the selection estimates from the shorter timescales or more subdivided fitness components in our analysis, as including only the overall component might obscure relevant selection and result in pseudoreplication. For example, Bérénos and colleagues calculated selection coefficients based on selection for survival, reproductive success, and overall lifetime fitness (Bérénos *et al.* 2015). We included the measures of selection on survival and reproductive success, but did not include the lifetime fitness selection coefficients in our quantitative analysis.

Finally, we standardized all estimates of selection as the magnitude of selection against the disfavoured allele. Because positive selection for beneficial alleles and negative selection against deleterious alleles are calculated under slightly different models, they are not directly comparable (see Box 1). Fortunately, a given estimate of positive selection on an allele can be easily converted into the estimate of negative selection against the corresponding disfavoured allele, assuming a diallelic system. Hereafter, references to the distribution of selection coefficients or the mean magnitude of selection coefficients refer to these converted estimates.

Quantitative analysis

Selection on phenotypes can be measured using standardized, regression-based methods that allow straightforward comparison in a meta-analysis (Kingsolver *et al.* 2012; Morrissey & Hadfield 2012; Siepielski *et al.* 2013). Selection at the genetic level, on the other hand, can be measured with many different methods, and this diversity complicates formal meta-analysis. We therefore analysed the database using a variety of statistical techniques. All analysis was performed in R version 3.0.1 (R Core Team 2015).

First, we followed the example of early syntheses of phenotypic selection coefficients by plotting the distribution of selection coefficients, observing how this distribution differs between biological and methodological categories, and using nonparametric statistics to evaluate differences in the mean magnitude of selection between categories (Endler 1986; Kingsolver *et al.* 2001). Because two studies accounted for over 90% of selection estimates (see results), we performed all nonparametric analysis on both the full dataset and the subset of estimates excluding these two studies, hereafter referred to as the reduced dataset.

Some studies reported multiple selection coefficients, and failing to correct for autocorrelation within studies could influence our conclusions (Gurevitch & Hedges 1999). To account for this, we implemented generalized linear mixed models (GLMMs) in a Bayesian framework using the R package MCMCglmm (Hadfield 2010). We included study as a random factor and used the exponential distribution to model our response variable, the selection coefficient. For the fixed effects, we specified independent normal distributions with mean = 0 and large variance (10^9). For the random effects, we used parameter expansion, which results in scaled *F* priors, to improve convergence. We used flat inverse-Wishart priors for the residual variance (a full specification of the models and priors, including the function calls in MCMCglmm, can be found in the supplementary methods). We first modelled the distribution of selection coefficients without any predictor variables to see how accounting for autocorrelation within studies influenced our results. We then ran separate models specifying the direction of selection, type of study, time period of selection and genetic unit as explanatory variables to understand whether the strength of selection differed between these categories.

Measurement error can have a significant effect on the conclusions drawn from meta-analyses of selection (Morrissey & Hadfield 2012). Unfortunately, relatively few studies reported measures of error around their estimates of selection, and those that did often used different methods to calculate error bounds. For this reason, we were unable to account for measurement error

in our analysis of all reported selection coefficients. To gain some understanding of how measurement error might influence our results, we performed three GLMMs on the subset of our data for which standard errors were reported or could be calculated and compared their estimates of the mean selection coefficient. We used the same normal priors for the fixed effects, but did not use parameter expansion and instead used flat inverse-Wishart priors for both the random effects and residual variance. The first model included study as a random factor, the second incorporated measurement error and the third incorporated both terms.

Results

Database results

Of the more than 2200 studies we examined, only 79 (~3.5%) met all the inclusion criteria. After accounting for pseudoreplication and multiple temporal scales within a study, the database contained 3416 directional selection coefficients and 70 instances of heterozygote advantage. Most of the directional selection coefficients came from two studies. Anderson *et al.* 2014 reported 2793 estimates of selection, and Gompert *et al.* 2014 contained 300 selection coefficients (see Box 2). All of the methodological and biological categories were well represented (see Table 1). Of the 79 studies, 15 reported selection coefficients for overdominant selection (see Box 3).

Distributions and nonparametric analysis

Overall, directional selection coefficients were roughly exponentially distributed (coefficient of variation = 1.05, CV = 1 for exponential distributions). Estimates of the strength of selection ranged from extremely weak ($s = 9.9 \times 10^{-5}$) to extremely strong (maximum $s = 1$ for two lethal mutations, otherwise maximum $s = 0.891$) (Fig. 1a). The mean selection coefficient of the full dataset was 0.135 (95% CI: 0.131–0.140, determined by 10 000 bootstrap replicates), while the mean of the reduced dataset was significantly smaller at 0.093 (95% CI: 0.078–0.110; Wilcoxon rank sum test, W = 697656, $P = 3.45 \times 10^{-15}$). The distribution of the reduced dataset was also roughly exponential (Fig. 1b, dark gray bars).

In the full dataset, there was a significant difference in mean strength of selection across categories of statistical significance (Kruskal–Wallis rank sum test, $\chi^2 = 325$, d.f. = 2, $P = 2.2 \times 10^{-16}$), with significant estimates of selection being much greater than estimates that were not significant or of unknown significance (Figs 2a and 3a, Table 2). In the reduced dataset, there was no difference among statistical categories (Kruskal–Wallis rank sum test, $\chi^2 = 1.79$, d.f. = 2, $P = 0.4$; Fig. 3a, Table 2). Estimates of negative selection had larger mean selection coefficients than estimates of positive selection in both the full and reduced datasets (Fig. 3b, Table 2). The mean strength of selection was greater for manipulative experiments than for observational estimates in both the full and reduced datasets (Table 2).

The distribution of selection coefficients varied based on the time period over which selection was measured (Figs 2b and 3c, Table 2). When studies did not report the number of generations over which selection was measured, we searched the literature for estimates of generation time for the studied organism and used these to coarsely estimate the number of generations over which selection was measured. We grouped estimates of selection into four categories: selection within a generation, short-term selection operating over <200 generations, long-term selection operating over 200 or more generations, and estimates for which the time period was unclear or unspecified. The mean magnitude of s was significantly different across categories (full dataset: Kruskal–Wallis rank sum test: $\chi^2 = 122$, d.f. = 3, $P = 2.2 \times 10^{-16}$; reduced dataset: Kruskal–Wallis rank

Table 1 Summary of database and directional selection coefficients. Numbers in parenthesis indicate the number of selection coefficients in the reduced dataset

Full dataset		Directional selection							
Studies	79	Taxon Type		Unit of Selection		Time Period		Statistical significance	
Taxa	30	Vertebrates	202	SNP	2160 (131)	Within generation	3131 (38)	Significant	398 (106)
Total # s	3556	Invertebrates	350 (50)	Haplotype	1256 (192)	Short term	125	Not significant	2822 (21)
Positive	1596 (224)	Plants	2844 (51)			Long term	141	Not reported	196
Negative	1820 (99)	Microbes	20			Unspecified	19		
Overdominant	140*								

*Estimates of overdominant selection report two selection coefficients per locus, one for the selective advantage over each of the two homozygotes.

Box 2. Field Studies of Selection: Anderson *et al.* 2014 and Gompert *et al.* 2014

Two studies reported a large portion of the selection coefficients in our database. Both studies tracked changes in allele frequency on hundreds to thousands of loci in large-scale field experiments, and there was no *a priori* understanding of whether these markers would influence fitness. This is in contrast to many of the other papers in our database, and in principle, such field studies could give a more unbiased view into how selection operates across the genome. However, details of the experimental design and analytical procedures for these studies can also influence the selection coefficients they report, so it is useful to discuss each paper in more detail.

Anderson *et al.* 2014

Anderson and colleagues used multiyear field transplants to study local adaptation and fitness trade-offs in *Boechera stricta*, a perennial mustard native to the Rocky Mountains. Anderson *et al.* crossed plants from two potentially locally adapted populations in Colorado and Montana to create 172 F_6 recombinant inbred lines (RILs), and genotyped each RIL at 62 microsatellite loci and 102 SNPs. They planted two cohorts containing replicates of each RIL and parental line into two common gardens near the source populations, and tracked each cohort for multiple years, measuring survival, flowering success, and fecundity for each individual. From this individual-level data on fitness components, they calculated relative fitnesses for the different genotypes at each locus and used permutation to estimate selection coefficients and significance thresholds for each genotyped locus (Anderson *et al.* 2013, 2014). This permutation procedure does not calculate error bounds, so the precision of each estimate is unknown. They calculated s at both experimental sites for multiple within-generation episodes of selection and multiyear selection coefficients based on lifetime flowering probability and fruit production. For our quantitative analysis, we included all within-generation estimates of selection, but not the lifetime selection coefficients (see main text). We also used the more conservative genomewide threshold when classifying estimates of s as significant or insignificant. Thus, most estimates of s were insignificant, and this might tend to increase the mean of the significant category. However, Anderson *et al.* calculated and reported a selection coefficient for every genetic marker at every instance of selection, regardless of strength or significance. There is therefore no within-study publication bias, and Anderson *et al.* present an objective report of the impact of selection in their experiments. Their study is also unusual in that it calculates selection coefficients for each locus in two locations across multiple time periods, providing some insight into how selection at the genetic level can vary through space and time.

Gompert *et al.* 2014

Gompert and colleagues studied two ecotypes of *Timema cristinae* stick insects that are differentially adapted to live on the host plants *Adenostoma fasciculatum* and *Ceanothus spinosus*. Visual predation by birds drives phenotypic divergence in *T. cristinae*: insects with a white dorsal stripe are cryptic on *Adenostoma* and conspicuous on *Ceanothus*, while the opposite is true of unstriped morphs (Sandoval 1994; Nosil 2004; Nosil & Crespi 2006). Gompert *et al.* collected 500 total *T. cristinae* from a mostly *Adenostoma*-adapted population that receives some gene flow from other populations with different host plants (Nosil *et al.* 2012). They cut off a portion of leg from each individual for tissue sampling and transplanted groups of insects onto individual *Adenostoma* and *Ceanothus* plants in experimental blocks at a nearby site. After 8 days they resampled the experimental plants and recaptured 140 insects, from which they took a postselection tissue sample. Using a genotype-by-sequencing approach, they determined the pre- and postselection allele frequencies of almost 200 000 SNPs. They developed statistical models to identify loci that showed parallel changes in allele frequency across experimental blocks that were unlikely to occur due to drift alone and then used MCMC to calculate the mean selection coefficient and 95% credible intervals for these loci. Thus, for quantifying selection, Gompert *et al.* take a different approach from Anderson *et al.* Although they have the data, in principle, to calculate selection coefficients for all loci, they calculate selection coefficients only for loci that demonstrated large, parallel allele frequency changes across experimental blocks. Weak selection is unlikely to drive such changes, and the Gompert *et al.* method is thus biased against the quantification of weak selection. Indeed, the distribution of selection coefficients reported in Gompert *et al.* is quite different from the distribution of both Anderson *et al.* 2014 and all other estimates of s (Fig. 1b).

Box 3. Heterozygote Advantage

Overdominant selection was rarely detected, with only 140 estimates of s from 15 studies (70 instances of heterozygote advantage, two selection coefficients per instance). With so few estimates, it is difficult to draw firm conclusions about the strength of overdominant selection, especially because most estimates were insignificant or did not report statistical significance (Fig. B3). Overall, selection ranged from very weak ($s = 0.0003$) to very strong ($s = 1$ for homozygote lethal alleles). The distribution of overdominant selection coefficients was significantly different from that of directional selection coefficients (Kolmogorov–Smirnov test, $D = 0.34$, $P = 1.14 \times 10^{-14}$) and was more uniformly distributed, although weak estimates of selection were still most common. Multiple studies reported selection coefficients for HLA loci in humans or MHC loci in other vertebrates. These immune system genes are classic examples of heterozygote advantage (Hedrick 2012). Heterozygote advantage was also detected at a number of allozyme loci in various species of plants, although determining phenotypic effects and agents of selection on these loci is more difficult. The prevalence of heterozygote advantage and its importance for the maintenance of genetic variation has long been a topic of debate (Lewontin & Hubby 1966; Garrigan & Hedrick 2003; Mitchell-Olds *et al.* 2007; Hedrick 2012; Fijarczyk & Babik 2015). There are few cases in which heterozygote advantage has been suggested in natural populations (Hedrick 2012), and, as we find in this study, even fewer cases in which selection has been quantified. This may be due to the inherent difficulties in detecting heterozygote advantage (Garrigan & Hedrick 2003). For example, genome scans can be used to detect a signature of balancing selection in nucleotide polymorphism data, which may be indicative of heterozygote advantage. However, other processes can also lead to a signature of balancing selection, including spatial or temporal variation in selection and frequency-dependent selection (Fijarczyk & Babik 2015). Distinguishing between these possibilities is often not possible using DNA sequence data alone (Hedrick 2012). Alternatively, heterozygote advantage may be rarely detected because it is, in fact, rare. Resolving the debate over whether heterozygote advantage is truly rare or simply difficult to detect is beyond the scope of our study.

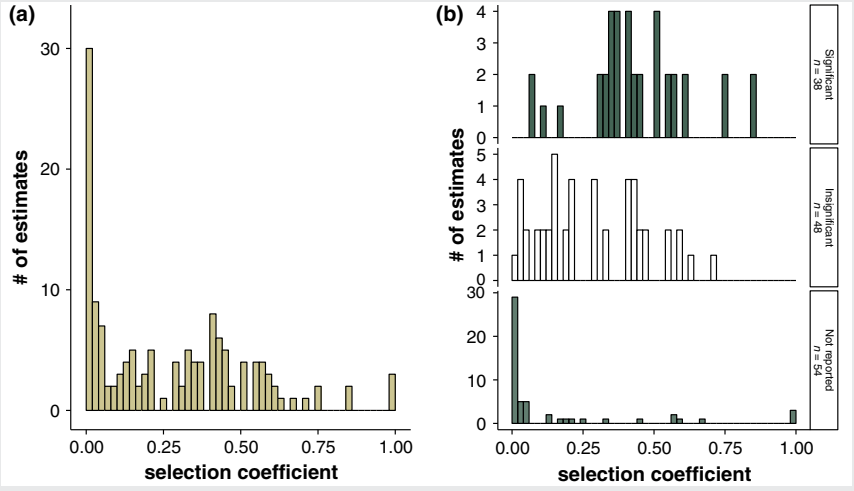

Fig. B3 The distribution of overdominant selection coefficients. (a) The distribution of all overdominant selection coefficients. (b) The distribution of overdominant selection coefficients across different categories of statistical significance.

sum test: $\chi^2 = 48$, d.f. = 3, $P = 2.1 \times 10^{-10}$; Fig. 3b). In both datasets, the mean strength of selection decreased as the timescale over which selection was measured increased. The distribution of selection coefficients also varied with the genetic scale at which selection was measured (Figs 2c and 3d). In the full dataset, the mean strength of selection was greater for SNPs than for haplotypes, although this difference was marginally nonsignificant. In the reduced dataset, however, selection was significantly stronger on haplotypes (Table 2).

GLMM results

The results of our GLMMs were qualitatively similar to the results obtained using nonparametric statistics. First, we modelled the mean selection coefficient of the full dataset while accounting for autocorrelation within

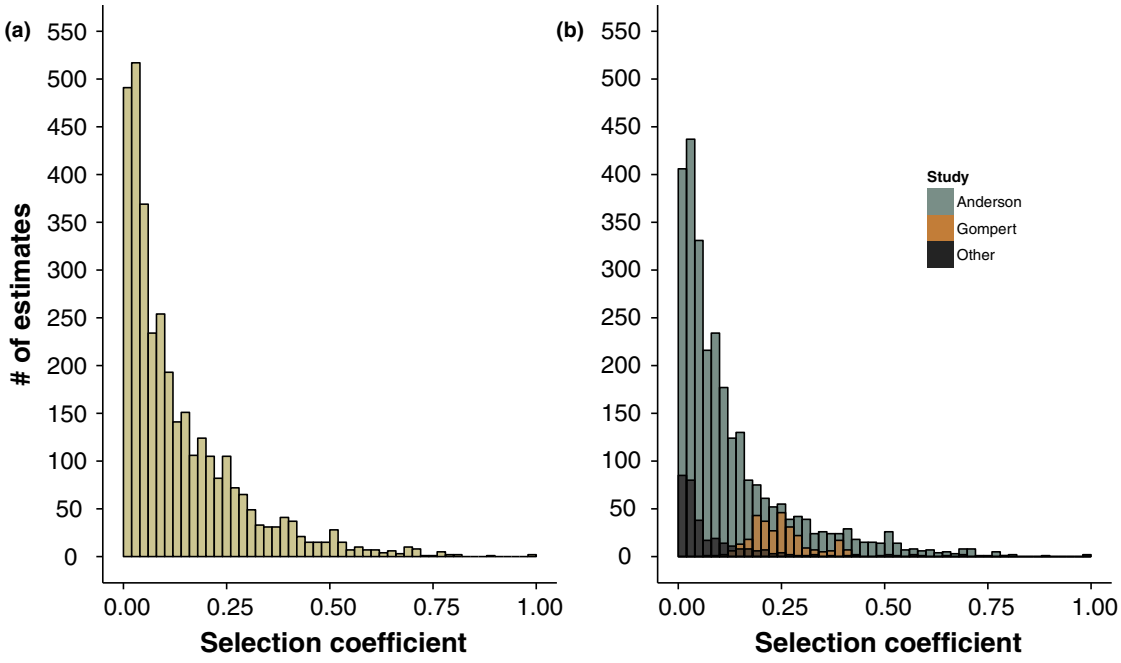

Fig. 1 The distribution of directional selection coefficients, *s*. (a) The distribution of directional selection coefficients included in the quantitative analysis. All selection coefficients are represented as selection against the less-fit allele. (b) Directional selection coefficients, coloured by the study in which they were reported. Anderson *et al*. 2014; in light blue, reported 2793 selection coefficients. Gompert *et al*. 2014, in orange, contained 300 selection coefficients. All other studies, in dark gray, contained 323 estimates of selection.

studies by including study as a random factor. This GLMM estimated a mean overall selection coefficient of 0.095 (posterior mode, 95% HPD interval: 0.066–0.124). These confidence intervals do not overlap with those of the uncorrected mean selection coefficient of the full dataset (0.135, 95% CI: 0.131–0.140). However, the GLMM estimate is very similar to the mean of the reduced dataset (0.093, 95% CI: 0.078–0.110), albeit with less precision. The GLMMs that incorporated predictor variables found results similar to the nonparametric analyses, but with weaker estimates of the strength of selection and wider confidence intervals, such that differences between categories were not always statistically significant (see Table 3 for posterior modes and 95% HPD interval estimates for all models). Negative selection was slightly stronger but not significantly different from positive selection. Selection estimates from experimental studies were nearly equal to estimates from observational studies, in contrast to the nonparametric results. Selection over long timescales was significantly weaker than both selection over short timescales and selection within a generation. The GLMM that included genetic scale as a predictor estimated that selection was stronger on haplotypes than on SNPs, although this difference was not significant.

The GLMMs we performed to evaluate the effects of measurement error indicated that autocorrelation had a much greater effect on our dataset than imprecise estimation of selection coefficients (Fig. 4, Table 3). Compared to the uncorrected mean *s* estimated by bootstrapping, the GLMMs that incorporated measurement error had smaller estimates of mean *s* and wider confidence intervals, as might be expected. However, incorporating measurement error had much less effect than accounting for autocorrelation within a study, which greatly reduced the estimate of mean *s*. This analysis could only be performed on the approximately 10% of estimates for which we could calculate standard errors, so generalizing these results to the full dataset requires caution. However, these models indicate that the results of the GLMM on the full dataset, which accounts for autocorrelation, are probably robust to measurement error.

Discussion

In this study, we report the results of the first meta-analysis of published estimates of selection coefficients in natural populations. Our search through the literature has uncovered a dynamic and growing field, with researchers using a wide variety of methodological and analytical techniques to understand how genetic variation influences fitness across a diverse set of taxa. Together, these estimates allow us to take the first steps towards answering fundamental questions about how

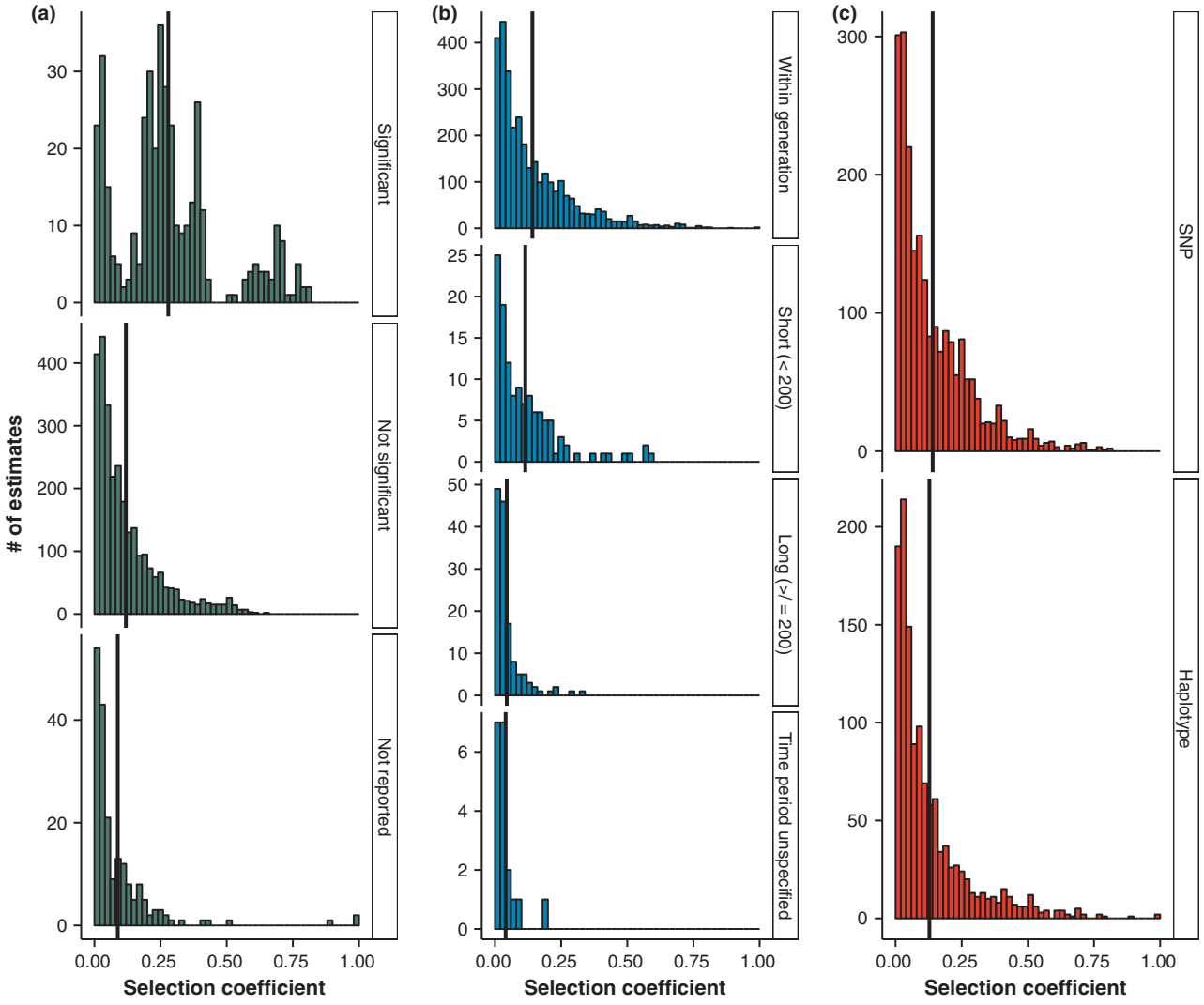

Fig. 2 The distribution of directional selection coefficients in the full dataset for different biological and methodological categories. Coefficients were categorized by (a) statistical significance, (b) time period over which selection was measured and (c) genetic scale at which selection was measured. The vertical line in each histogram marks the uncorrected mean of selection coefficients in that category.

natural selection operates at the genetic level. The vast majority of selection coefficients reported were for directional selection, with heterozygote advantage rarely detected (Box 3). We found that directional selection coefficients were roughly exponentially distributed, a pattern similar to estimates of selection on phenotypes. Although most estimates of s were small, some studies detected very strong selection ($s > 0.5$), especially on short timescales. Selection varied as predicted with temporal scale, as selection measured over long time periods was significantly weaker than selection measured over shorter periods. Selection also varied with the size of the genetic unit at which it was measured, although our different analyses found conflicting trends.

Before discussing these conclusions in more detail, it is important to note some limitations of this dataset. As with most meta-analyses, our study likely contains some biases, a number of which could tend to inflate estimates of selection. First, researchers may have chosen to study genetic loci that have an *a priori* expectation of being under strong selection ('research bias', Gurevitch & Hedges 1999). For example, a number of candidate gene studies examined insecticide resistance alleles (e.g., Lenormand *et al.* 1998) or drug resistance alleles (e.g., Roper *et al.* 2003), which are expected to be under strong selection. Even studies that started without *a priori* candidates and took a genomewide approach to detecting selection (e.g., Anderson *et al.* 2014; Gompert *et al.* 2014) studied populations that could be expected to be under strong selection for local adaptation. Similarly, there may be publication bias against reporting insignificant or weak estimates of

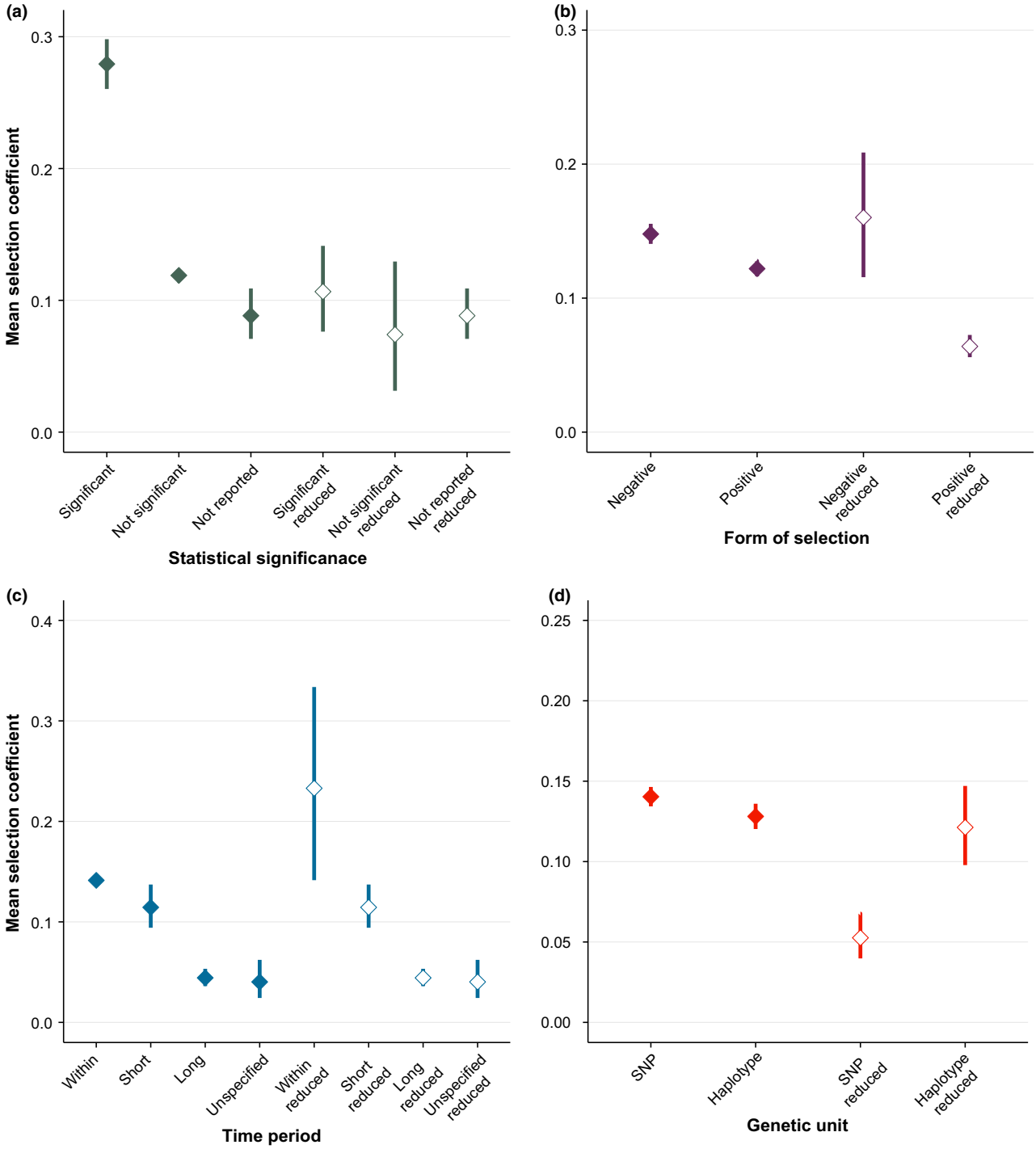

Fig. 3 Summary of mean selection coefficients across different biological and methodological categories. Diamonds and error bars represent the mean and 95% confidence intervals based on 10 000 bootstrap replicates. Unfilled diamonds represent the reduced dataset and filled diamonds represent the full dataset. Selection coefficients were categorized by (a) statistical significance, (b) form of selection, (c) timescale and (d) genetic scale. **N.B.** that estimates of selection for beneficial alleles were converted into selection against the less favoured allele. The means and confidence intervals presented here are from those standardized estimates.

selection (the well-known 'file drawer problem', Rosenthal 1979). In our dataset there appears to be some bias against weak estimates of selection (see supplementary analysis and Figs S2–S5, Supporting information), but there was clearly bias against statistically insignificant estimates. Nearly all insignificant estimates of selection

Table 2 Mean *s* and 95% confidence intervals (determined by 10 000 bootstrap replicates) of various methodological and biological categories, for both the (A) full dataset and (B) reduced dataset.

	(A) Full dataset		(B) Reduced dataset	
	Mean	95% CI	Mean	95% CI
Overall mean selection coefficient	0.135	0.131–0.140	0.093	0.078–0.110
Statistical significance				
Significant	0.279	0.260–0.298	0.106	0.076–0.141
Not significant	0.118	0.114–0.123	0.074	0.031–0.129
Not reported	0.088	0.070–0.108	0.088	0.070–0.108
Form of selection				
Positive selection	0.121	0.116–0.127	0.063	0.055–0.072
Negative selection	0.147	0.140–0.155	0.160	0.115–0.208
Type of study				
Experimental	0.140	0.135–0.144	0.122	0.084–0.167
Observational	0.090	0.074–0.108	0.090	0.073–0.108
Time period				
Within generation	0.141	0.136–0.146	0.232	0.141–0.333
Short term (<200 gens.)	0.114	0.094–0.137	0.114	0.094–0.137
Long term (≥200 gens.)	0.044	0.036–0.053	0.044	0.036–0.053
Not specified	0.040	0.024–0.062	0.040	0.024–0.062
Genetic unit				
Haplotypes	0.128	0.120–0.135	0.121	0.097–0.147
SNPs	0.140	0.134–0.146	0.052	0.039–0.067

Table 3 Results of the generalized linear mixed models. Estimates are the posterior mode and lower and upper bounds of the 95% highest posterior density interval. (A) Results of GLMMs performed on the full dataset. Bolded text shows the form of the fixed effect model specification, and normal text shows each fixed factor within that analysis. All models incorporated study ID as a random factor. (B) Results of GLMMs performed on the subset of data for which standard errors could be calculated. Bolded text shows the form of the fixed effect model specification, and normal text lists the random factors included in the three models: study ID, measurement error or both.

	Results of generalized linear mixed models	
	Posterior mode	95% HPD interval
(A) Full dataset		
Selection coefficient ~ 1	0.095	0.066–0.124
Selection coefficient ~ form of selection		
Positive selection	0.086	0.065–0.117
Negative selection	0.096	0.074–0.133
Selection coefficient ~ type of study		
Experimental	0.097	0.057–0.163
Observational	0.095	0.067–0.124
Selection coefficient ~ time period		
Within generation	0.201	0.123–0.351
Short term (<200 gens.)	0.111	0.077–0.174
Long term (≥200 gens.)	0.036	0.023–0.065
Not specified	0.032	0.017–0.079
Selection coefficient ~ genetic unit		
Haplotypes	0.093	0.067–0.124
SNPs	0.086	0.065–0.123
(B) subset with standard errors		
Selection coefficient ~ 1		
Study ID	0.141	0.055–0.231
Measurement error	0.186	0.174–0.202
Study ID + measurement error	0.113	0.050–0.217

came from the Anderson *et al.* 2014 and Gompert *et al.* 2014 studies. In the reduced dataset, there were only 21 insignificant selection coefficients, compared to 106 significant estimates and 196 with unreported statistical significance. Insignificant selection was thus rarely reported outside of the context of genomewide studies of selection, in which many insignificant estimates are expected. Perhaps this is not surprising, given the preeminence of neutral theory and the desire to avoid adaptationism (Gould & Lewontin 1979; Nielsen 2009; Barrett & Hoekstra 2011). However, we agree with other authors who have urged researchers to think of selection coefficients as continuous variables and not to overemphasize categorical distinctions between 'significant' and 'insignificant' selection coefficients (Gompert 2016). Failing to report selection coefficients because they are insignificant puts too much emphasis on *P*-values, too little on effect sizes and confidence intervals, and leads to publication bias (Halsey *et al.* 2015).

Finally, the full database of selection coefficients is largely made up of estimates from two studies that combined large-scale field experiments with genomewide sampling to generate hundreds of estimates of selection (Anderson *et al.* 2014; Gompert *et al.* 2014; see Box 2). Experimental evolution studies have important advantages over other methods of detecting selection, as researchers can track evolution in real time and control or mitigate some of the demographic and ecological factors that complicate the detection and quantification of selection. However, these methods also have limitations, especially for detecting weak selection (see Box 2). While more studies of this type will surely follow, for now they complicate the analysis of this dataset. We have sought to account for this with a variety of statistical techniques, but the accumulation of more estimates of selection at the genetic level will ensure that future meta-analyses of natural selection at the genetic level are not unduly influenced by a few studies. Despite these limitations, this dataset is our best source of information for both preliminary conclusions about selection at the genetic level and for informing future research.

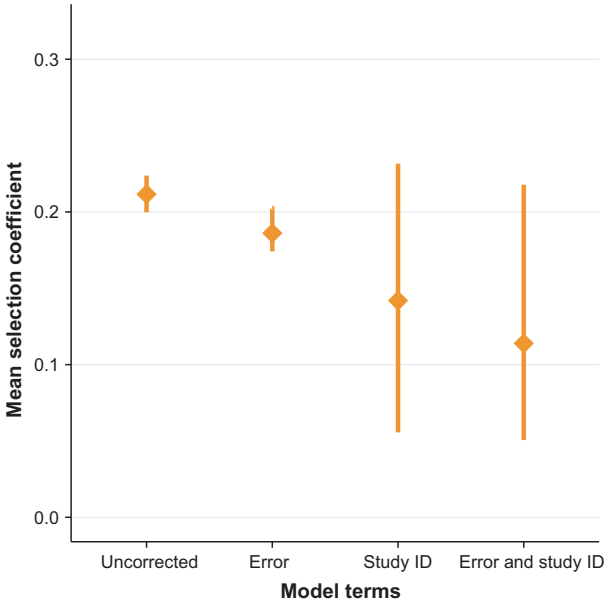

Fig. 4 Effect of accounting for autocorrelation and measurement error, in the subset of data for which standard errors were reported or could be calculated. The uncorrected estimate shows the mean and 95% confidence interval of the selection coefficient, based on 10 000 bootstrap replicates. The other estimates are the posterior mode of the estimate of mean s from the three GLMMs that incorporated measurement error (error), study as a random factor (study ID), or both (error and study ID). Error bars show the upper and lower bounds of the 95% highest posterior density interval.

Quantitative results

Our analysis found a number of important quantitative results. First, selection coefficients could be quite large. The uncorrected mean and median of the full dataset were 0.135 and 0.082, respectively, and there were 112 estimates of selection coefficients >0.5. Selection at the genetic level is often assumed to be rather weak. For example, some studies in this database that used simulations to quantify selection considered coefficients only within a narrow range (e.g., 0–0.1 in Ohashi *et al.* 2004; 0–0.03 in Gerbault *et al.* 2009). While many published estimates of selection coefficients are indeed small, our results show that researchers cannot discount the possibility of large selection coefficients for genetic variants, especially over short timescales. Of course, whether a given coefficient represents 'significant' or 'strong' selection is a matter of perspective. All alleles are affected by genetic drift, and where to draw the line between 'selected' and 'neutral' alleles is a matter of debate. Multiple definitions have been proposed, and most rely on an understanding of the effective population size (N_e), recognizing that selection will be less efficient in smaller populations (Nei *et al.* 2010). When estimates of effective population size are unavailable, as in most of the studies in our database, Nei suggests a threshold of approximately $|s| = 0.001$ for vertebrates (Nei 2005). Under this relaxed definition of neutrality, nearly all (3411 of 3416) of the selection coefficients in our database are not neutral.

Second, the exponential distribution of s is very similar to the distribution of phenotypic selection coefficients reported in other studies (Hoekstra *et al.* 2001; Kingsolver *et al.* 2001). This is not necessarily expected, as genetic selection coefficients are fundamentally different from phenotypic selection differentials and gradients. While selection coefficients against a disfavoured allele range from 0 to 1 (see Box 1), selection differentials and gradients are calculated via linear regression and their range is thus unrestricted in theory, although in practice the absolute values of most estimates fall between 0 and 1 (Kingsolver *et al.* 2001). There is no clear theoretical expectation that both phenotypic and genetic selection coefficients should be exponentially distributed. Kingsolver *et al.* (2001) note that, if most organisms are well adapted to their environments, phenotypic directional selection should be normally distributed around a mean of 0. Indeed, more recent meta-analyses of phenotypic selection have used a folded-normal distribution to model the absolute values of selection gradients (Hereford *et al.* 2004; Kingsolver *et al.* 2012; Morrissey & Hadfield 2012). Multiple genetic models of adaptation predict that the fitness effects of adaptive mutations during a single adaptive walk will be exponentially distributed (Orr 2005a). However, the assumptions that underlie those predictions (i.e., a single population adapting to a relatively close, stationary fitness peak solely through the fixation of new mutations) do not apply to our broad dataset, and other models make different predictions about how selection coefficients might be distributed (e.g., Kopp & Hermisson 2009b, who model adaptation to a moving fitness optimum and predict a unimodal distribution with mutations of intermediate effect dominating).

Instead of referring to disparate phenotype- or genotype-level theories, another way to explain the similarity in these distributions could come from understanding how selection at these levels is linked. Selection does not act directly at the genetic level; rather it acts on phenotypes and is then transmitted to the genetic level based on the genetic architecture of the trait(s) under selection. Assuming that the phenotypic effects of an allele are proportional to its fitness effects, it may be possible to work downward from the empirical, roughly exponential distribution of phenotypic selection coefficients to derive an expected distribution for genetic selection coefficients. To do so properly would require some understanding of the number and phenotypic effect sizes of the loci underlying the

selected phenotypic traits, as well as the degree of pleiotropy. The general genetic architecture of traits subject to selection is a topic of much debate (Rockman 2012; Lee *et al.* 2014). The two opposing views could be characterized as 'exponential-like' (i.e., traits are controlled by one or a few loci with large phenotypic effects and many loci with small phenotypic effects) and 'infinitesimal' (i.e., traits are controlled by hundreds to thousands of loci of extremely small effect). Interestingly, the observed exponential form of selection coefficients acting on phenotypes may be transmitted to the genetic level to produce an L-shaped (exponential-like) distribution of selection on alleles, regardless of whether the allelic effects on a phenotype are drawn from an exponential or a uniform distribution, assuming that the strength of selection acting on a trait does not influence its genetic architecture and there is no pleiotropy (see Appendix S1, Supporting information, co-authored with S. Otto, for theory). Although some genotype–phenotype maps transmit the exponential-like distribution of phenotypic selection unchanged to the genetic level, not all maps will do so. It remains an open theoretical question to determine which genotype–phenotype maps are most consistent with our observations.

The impacts of natural selection at the genetic level varied across a number of biological and methodological categories. Statistically significant estimates of selection tended to be stronger than insignificant ones, which is unsurprising given that stronger selection is easier to distinguish from drift than weak selection. The mean value of selection coefficients that did not have estimates of error or significance was similar to the mean of insignificant selection coefficients (Fig. 3), which may suggest that many of these estimates are statistically insignificant. Or course, the statistical significance of an estimate of selection is dependent on the power of the procedures used to estimate it. Unfortunately, analysing the power of each study in our database was not feasible. Statistical significance will only be indicative of the biological relevance of a variant's fitness effect with sufficient power: underpowered studies may be unable to distinguish selection from drift. Conversely, significant estimates should not be misunderstood to mean that only selection is driving allele frequency change. All alleles in finite populations are influenced by drift; significant estimates of s simply indicate that drift alone could not cause the observed change. Again, we emphasize that selection coefficients are continuous variables; it is preferable to interpret their statistical and biological significance by considering their confidence intervals, not their P-values alone. And, absent knowledge of experimental power for each study, we cannot distinguish estimates of s that are insignificant due to neutrality from those that are insignificant due to insufficient power. Thus, we caution against overinterpreting the differences we observe across statistical categories.

Estimates of negative selection were of greater magnitude than estimates of positive selection in both the full and reduced datasets, although this difference was not significant in the GLMM. Selection coefficients for both forms of selection were roughly exponentially distributed (Fig. S1, Supporting information). In some sense, comparing the magnitude and distributions of these categories is not biologically informative, as the designation of selection as positive or negative is relative (Box 1). The difference in magnitude between these categories perhaps suggests research bias, with researchers who focus on negative selection choosing to study populations that experience slightly stronger selection. It is reasonable to expect that experimental manipulations may be associated with selection that is stronger than selection that is simply observed. While the nonparametric statistics indicated that this was the case, the estimates of mean s for experimental and observational studies were nearly equal in the GLMM. The vast majority of estimates of selection from experiments came from the Anderson *et al.* 2014 and Gompert *et al.* 2014 studies, only eight other studies contributed 31 total estimates, so there is little statistical power for firm conclusions.

Natural selection on shorter timescales tended to be stronger than selection on longer timescales, as we predicted. This was true in both the full and reduced datasets, and the GLMM corroborated the trend, although differences between some categories were insignificant. The absolute differences in magnitude between categories were fairly small: mean s for long-term estimates was 0.044 in both datasets, while mean s within a generation was ~ 3× greater (0.141) in the full dataset and ~5× greater (0.232) in the reduced dataset. This overall trend may be partially due to the fact that studies over shorter time periods, and especially within generations, are often unable to distinguish between direct and indirect selection on a locus, which could lead to larger estimates of s (see Box 1, Box 2). However, the patterns we see in the strength of genetic selection coefficients are consistent with those observed in measures of evolutionary rates and phenotypic selection. Short-term rates of phenotypic change are often orders of magnitude greater than long-term rates (Gingerich 1983), phenotypic selection on viability is stronger when measured over shorter time periods (Hoekstra *et al.* 2001), and long-term rates of phenotypic evolution are often slower than one would expect when extrapolating from short-term estimates of phenotypic selection (Kinnison & Hendry 2001). This tendency for evolutionary rates, phenotypic selection coefficients and

genetic selection coefficients to be of smaller magnitude when measured over longer periods of time is likely to be partially a mathematical artefact of averaging that is inherent to all measures that compare differences between initial and final states (Gingerich 1983). Such measures must assume that the rate of change (in the case of s, change in allele frequencies) is constant between measured instances and will thus average out the instantaneous rates into a less extreme long-term rate. However, the effects of averaging almost certainly reflect biological reality. Meta-analysis of phenotypic selection shows that selection may fluctuate through time such that short-term estimates of selection are not indicative of long-term trends (Siepielski *et al.* 2009; but see Morrissey & Hadfield 2012).

This effect is illustrated in the few studies that examined selection on the same locus or loci through time. For example, Barrett *et al.* (2008) found opposing patterns of strong selection at different life stages on an allele for reduced armour plating in threespine sticklebacks, such that the lifetime s was much weaker than the per-life-stage estimates of selection coefficients. Anderson *et al.* also found negative correlations between selection coefficients across some (but not all) episodes of selection, indicative of fitness trade-offs within a generation (see table 4 in Anderson *et al.* 2014 and our supplementary analysis). Those trade-offs did not necessarily lead to estimates of weak lifetime selection. For example, plants in Montana experienced trade-offs between flowering/fruiting and overwinter survival. However, selection on survival was relatively weak and selection on both fruiting and flowering was quite strong, leading to large estimates of lifetime s.

Perhaps the best example of how temporal variation can affect the magnitude of selection estimates comes from a study on drug resistance alleles in the malaria parasite, *Plasmodium falciparum* (Taylor *et al.* 2012). The authors calculated both annual selection coefficients and overall selection coefficients on mutations at individual codons across a nine-year study. Annual selection coefficients varied in magnitude and direction and were often statistically insignificant. The selection coefficients calculated across all nine years, however, were smaller in magnitude, statistically significant, and favoured resistance alleles. This study could not be included in our quantitative analyses, as the regression-based selection coefficients they calculated were not comparable to the other estimates in our dataset. However, it clearly demonstrates that long-term patterns of selection are the result of fluctuating moment-by-moment forces of selection.

The magnitude of s also varied based on the genetic unit at which selection was measured, but interpreting those trends is more complicated. We predicted that selection would be stronger on haplotypes than on SNPs, as allelic variants for haplotypes should, in general, have larger phenotypic effects than allelic variants of SNPs. In both the reduced dataset and the GLMM this prediction was supported, although the difference in mean s between these categories was not significant in the GLMM. In the reduced dataset, the difference in mean s was large (0.052 for SNPs, 0.121 for haplotypes). In the GLMM, the difference was much smaller (posterior mode of $s = 0.086$ for SNPs, $s = 0.093$ for haplotypes). In the full dataset, however, selection was stronger on SNPs, although this difference was marginally not significant. Some of this inconsistency across analyses is due to the outsized effects of the Anderson *et al.* 2014 and Gompert *et al.* 2014 studies. Gompert *et al.* 2014; in particular, measured 300 instances of selection on SNPs, and their methods were biased towards the detection of very strong selection (Box 2). This may have inflated the mean s for SNPs in the full dataset. While the GLMM accounts for autocorrelation within studies, the confidence intervals around the estimate of mean s for both categories are quite broad, indicating little statistical support for either interpretation. There are also other possible hypotheses for how selection might vary with genetic scale. For example, larger genetic units could contain multiple loci with contrasting fitness effects, so that they experience weaker selection that is the average effect of the individual loci contained within them. In that case haplotypes would tend to experience weaker selection than SNPs, as we see in the full dataset. Given the conflicting trends among the different datasets and methods of analysis, it seems that we need further data before we can determine whether, how, and why the strength of selection varies with genetic scale.

Recommendations for future research

In addition to our observations about the distribution and variation of selection coefficients, our review of the literature uncovered a number of important issues to consider when studying natural selection at the genetic level. First, consider that the acceptance rate for inclusion in our dataset was extremely low (~3.5%). Of course, this low rate is partly due to our inclusion criteria, as we excluded some studies that quantified selection in ways that were incompatible with our analysis. Another possible reason might have been our Web of Science search terms. They seemed to be simultaneously too broad (our search results included many studies on agricultural plants, purely theoretical models, and phenotypic selection coefficients) and ineffective at locating studies (we found almost as many studies that reported selection coefficients by searching references as we did in our Web of Science searches).

While these reasons are certainly part of the explanation, we suspect that the discrepancy between the number of plausible studies and the number of studies that report estimates of *s* exposes a larger issue: natural selection is frequently invoked or detected, but very rarely quantified, even in studies which contain raw data from which selection coefficients could be calculated. Of course, not all biologists are interested in quantifying natural selection, and it is understandable that many researchers do not take this final step. However, we hope that our analysis has shown that quantifying selection can lead us toward answers for important questions in evolutionary biology. We therefore encourage researchers to endeavour, not only to detect selection, but also to quantify its strength.

What, then, are the best practices for calculating, reporting and interpreting selection coefficients? Methods for calculating selection coefficients will depend on the type of data available to a researcher. An extensive review of methods is beyond the scope of this work; for specifics, we direct readers to previous reviews of methods for the detection and quantification of selection (Linnen & Hoekstra 2009; Hohenlohe *et al.* 2010; Vitti *et al.* 2013), to the examples cited in our introduction, and to the papers within our literature database (see supplemental material). We also note that new methods are frequently being developed, especially methods which estimate selection coefficients based on sequence data (Charlesworth & Wright 2004; Slatkin 2008; Messer & Neher 2012; Chen & Slatkin 2013; Vitalis *et al.* 2014; Foll *et al.* 2015). Whatever method is used, researchers should take careful note of the mathematical model used to calculate *s* so that its biological meaning is clear. Of particular importance is understanding whether models calculate positive selection or negative selection, as these quantities are not directly comparable without a conversion (see Box 1). Further, researchers should calculate statistical significance, ideally from some form of confidence interval, and be cognizant of the specific statistical issues particular to their data (e.g., considerations of multiple testing, linkage between sites, and/or population structure). When feasible, researchers should also seek to calculate or determine other parameters that will aid in the interpretation of selection coefficients. These include experimental power (to establish the minimum *s* that could be reliably detected), source of genetic variation (i.e., standing genetic variation or new mutations), effective population size, and the ancestral allelic state of the locus under selection.

At minimum, researchers should clearly report (i) the model used to calculate *s*, (ii) some form of confidence interval for the estimate of *s*, and (iii) the data necessary to understand the time period over which selection was measured (ideally in generations). Researchers should report both significant and insignificant estimates to reduce publication bias. As genomic data become more available, the question of whether to calculate selection coefficients on all loci versus only those that show evidence of selection will become more important. This decision will depend, at least in part, on the interests and computational resources of the researcher. When calculating selection for all loci is not feasible, we recommend researchers follow the example of Gompert *et al.* (2014) by clearly stating the selection criteria for quantified loci.

We also strongly recommend that researchers report estimates of effective population size, which aids in interpreting the strength of selection. Information about the source of genetic variation and levels of linkage disequilibrium in the population tested is also valuable, as levels of LD can determine the extent to which researchers can partition genetic selection as direct or indirect. This complication arises in the application of one-locus models of selection, which assume that *s* represents direct selection, to natural populations in which allelic variants are also influenced by selection on linked loci and *s* should properly be interpreted as quantifying both direct and indirect selection. Models used to study genomewide selection often have more parameters (genetic variants) than statistical replicates (individuals), inhibiting the ability to measure direct selection (Gompert *et al.* 2014). Linkage disequilibrium, epistasis and pleiotropy can all complicate the simple goal of measuring the direct fitness effects of an allele and muddle the distinction between direct and indirect selection (Barton & Servedio 2015). Further theoretical work to address these issues will be especially welcome. Nevertheless we note that, in many cases, quantification of direct selection is not necessarily the ultimate goal. Understanding direct selection is crucial for elucidating the genetic and phenotypic mechanisms that drive adaptive evolutionary change. However, the total amount of selection (both direct and indirect) that impacts a locus is what drives allele frequency change each generation, and understanding it is more important for predicting the trajectory of evolution.

Conclusions

Our analysis has taken important first steps towards improving our understanding of the impacts of selection at the genetic level. Where should researchers direct their attention with future studies of selection at the genetic level? Keeping in mind our methodological guidelines above, simply accumulating more estimates of selection will be extremely useful. Our conclusions are necessarily limited by the data that have been published so far, and the practice of estimating genetic

selection coefficients is still rather young. More estimates of selection from a wider variety of taxa are needed for a fuller understanding of how natural selection shapes genetic variation. Fortunately, technological advances in collecting and analysing genetic data make it possible to quantify selection without requiring *a priori* knowledge of selection, and to do so in the context of manipulative field experiments. And, as with phenotypic selection, it will be informative to consider how selection coefficients vary with space, time, and across sexes and life-history stages. Such studies will give insight into fundamental questions about local adaptation, developmental trade-offs, and sexual conflict. We expect that, in the coming years, the number and scope of studies that quantify selection at the genetic level will rapidly increase. With larger datasets, future researchers will be able to more conclusively answer the questions we have begun to consider here.

Acknowledgements

We thank J. Anderson, W. Eanes, P. Gerbault, Z. Gompert, J. Prunier and S. Taylor for providing data, K. Gotanda for access to the preliminary database of Siepielski *et al*. 2013, and J. Hadfield and S. Diamond for answering questions about MCMCglmm. S. Otto, P. Nosil and W.O. McMillan provided helpful comments on an earlier version of this manuscript. S. Otto also co-authored Appendix S1 (Supporting information). We thank L. Bernatchez, A. Eyre-Walker and two anonymous reviewers for helpful comments that improved this manuscript. This work was supported by a STRI-McGill-NEO fellowship to T.J.T. and funding from a Canada Research Chair and a Natural Science and Engineering Research Council Discovery grant to R.D.H.B.

References

Anderson JT, Lee C-R, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, **22**, 699–708.

Anderson JT, Lee C-R, Mitchell-Olds T (2014) Strong selection genome-wide enhances fitness trade-offs across environments and episodes of selection. *Evolution*, **68**, 16–31.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767–780.

Barrett R, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, **23**, 38–44.

Barrett RDH, M'Gonigle LK, Otto SP (2006) The distribution of beneficial mutant effects under strong selection. *Genetics*, **174**, 2071–2079.

Barrett RDH, Rogers SM, Schluter D (2008) Natural selection on a major armor gene in threespine stickleback. *Science*, **322**, 255–257.

Barton NH, Servedio MR (2015) The interpretation of selection coefficients. *Evolution*, **69**, 1101–1112.

Bérénos C, Ellis PA, Pilkington JG *et al*. (2015) Heterogeneity of genetic architecture of body size traits in a free-living population. *Molecular Ecology*, **24**, 1810–1830.

Brodie ED III, Moore AJ, Janzen FJ (1995) Visualizing and quantifying natural selection. *Trends in Ecology and Evolution*, **10**, 313–318.

Charlesworth B, Wright SI (2004) The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics*, **168**, 1071–1076.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

Chen H, Slatkin M (2013) Inferring selection intensity and allele age from multilocus haplotype structure. *G3*, **3**, 1429–1442.

Collins S, De Meaux J, Acquisti C (2007) Adaptive walks toward a moving optimum. *Genetics*, **176**, 1089–1099.

Cook LM (2003) The rise and fall of the *carbonaria* form of the peppered moth. *The Quarterly Review of Biology*, **78**, 399–417.

Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton, New Jersey.

Estes S, Arnold SJ (2007) Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *American Naturalist*, **169**, 227–244.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, **8**, 610–618.

Fijarczyk A, Babik W (2015) Detecting balancing selection in genomes: limits and prospects. *Molecular Ecology*, **24**, 3529–3545.

Fisher RA, Ford EB (1947) The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity*, **1**, 143–174.

Foll M, Shim H, Jensen JD (2015) WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources*, **15**, 87–98.

Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution*, **57**, 1707–1722.

Gerbault P, Moret C, Currat M, Sanchez-Mazas A (2009) Impact of selection and demography on the diffusion of lactase persistence. *PLoS ONE*, **4**, e6369.

Gingerich PD (1983) Rates of evolution: effects of time and temporal scaling. *Science*, **222**, 159–161.

Gompert Z (2016) Bayesian inference of selection in a heterogeneous environment from genetic time-series data. *Molecular Ecology*, **25**, 121–134.

Gompert Z, Comeault AA, Farkas TE *et al*. (2014) Experimental evidence for ecological selection on genome variation in the wild. *Ecology Letters*, **17**, 369–379.

Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society B*, **205**, 581–598.

Gurevitch J, Hedges LV (1999) Statistical issues in ecological meta-analyses. *Ecology*, **80**, 1142–1149.

Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R Package. *Journal of Statistical Software*, **33**, 1–22.

Halsey LG, Curran-Everett D, Vowler SL, Drummond GB (2015) The fickle *P* value generates irreproducible results. *Nature Methods*, **12**, 179–185.

Hartl DL, Clark AG (2007) *Principles of Population Genetics*. Sinaeur Associates, Sunderland, Massachusetts.

Hedrick PW (2012) What is the evidence for heterozygote advantage selection? *Trends in Ecology and Evolution*, **27**, 698–704.

Hendry AP, Gonzalez A (2008) Whither adaptation? *Biology and Philosophy*, **23**, 673–699.

Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *The American Naturalist*, **173**, 579–588.

Hereford J, Hansen TF, Houle D (2004) Comparing strengths of directional selection: how strong is strong? *Evolution*, **58**, 2133–2143.

Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, **169**, 2335–2352.

Hoekstra HE, Hoekstra JM, Berrigan D *et al.* (2001) Strength and tempo of directional selection in the wild. *Proceedings of the National Academy of Sciences*, **98**, 9157–9160.

Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, **58**, 1329–1341.

Hohenlohe PA, Phillips PC, Cresko WA (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International Journal of Plant Sciences*, **171**, 1059–1071.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Kettlewell H (1958) A survey of the frequencies of *Biston betularia* (L.)(Lep.) and its melanic forms in Great Britain. *Heredity*, **12**, 51–72.

Kingsolver JG, Diamond SE (2011) Phenotypic selection in natural populations: what limits directional selection? *The American Naturalist*, **177**, 346–357.

Kingsolver JG, Hoekstra HE, Hoekstra JM *et al.* (2001) The strength of phenotypic selection in natural populations. *American Naturalist*, **157**, 245–261.

Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, **26**, 1101–1118.

Kinnison MT, Hendry AP (2001) The pace of modern life II: from rates of contemporary microevolution to pattern and process. *Genetica*, **112–113**, 145–164.

Kopp M, Hermisson J (2007) Adaptation of a quantitative trait to a moving optimum. *Genetics*, **176**, 715–719.

Kopp M, Hermisson J (2009a) The genetic basis of phenotypic adaptation I: fixation of beneficial mutations in the moving optimum model. *Genetics*, **182**, 233–249.

Kopp M, Hermisson J (2009b) The genetic basis of phenotypic adaptation II: the distribution of adaptive substitutions in the moving optimum model. *Genetics*, **183**, 1453–1476.

Kryazhimskiy S, Tkacik G, Plotkin JB (2009) The dynamics of adaptation on correlated fitness landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18638–18643.

Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.

Lee YW, Gould BA, Stinchcombe JR (2014) Identifying the genes underlying quantitative traits: a rationale for the QTN programme. *AoB PLANTS*, **6**, plu004.

Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PLoS One*, **3**, e4010.

Lenormand T, Guillemaud T, Bourguet D, Raymond M (1998) Evaluating gene flow using selected markers: a case study. *Genetics*, **149**, 1383–1392.

Lenormand T, Bourguet D, Guillemaud T, Raymond M (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature*, **400**, 861–864.

Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, **54**, 595–609.

Li J, Li H, Jakobsson M *et al.* (2012) Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Molecular Ecology*, **21**, 28–44.

Linnen CR, Hoekstra HE (2009) Measuring natural selection on genotypes and phenotypes in the wild. *Cold Spring Harbor Symposia on Quantitative Biology*, **74**, 155–168.

Martin G, Lenormand T (2006) A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution*, **60**, 893–907.

Martin G, Lenormand T (2015) The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. *Evolution*, **69**, 1433–1447.

Mathieson I, Mcvean G (2013) Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics*, **193**, 973–984.

Messer PW, Neher RA (2012) Estimating the strength of selective sweeps from deep population diversity data. *Genetics*, **191**, 593–605.

Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, **8**, 845–856.

Morrissey MB, Hadfield JD (2012) Directional selection in temporally replicated studies is remarkably consistent. *Evolution*, **66**, 435–442.

Nei M (2005) Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution*, **22**, 2318–2342.

Nei M, Suzuki Y, Nozawa M (2010) The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics*, **11**, 265–289.

Nielsen R (2009) Adaptionism—30 years after Gould and Lewontin. *Evolution*, **63**, 2487–2490.

Nosil P (2004) Reproductive isolation caused by visual predation on migrants between divergent environments. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 1521–1528.

Nosil P, Crespi BJ (2006) Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences*, **103**, 9090–9095.

Nosil P, Gompert Z, Farkas TE *et al.* (2012) Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 5058–5065.

Nsanzabana C, Hastings IM, Marfurt J *et al.* (2010) Quantifying the evolution and impact of antimalarial drug resistance: drug use, spread of resistance, and drug failure over a 12-year period in Papua New Guinea. *The Journal of Infectious Diseases*, **201**, 435–443.

Ohashi J, Naka I, Patarapotikul J *et al.* (2004) Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *The American Journal of Human Genetics*, **74**, 1198–1208.

Ohashi J, Naka I, Tsuchiya N (2011) The impact of natural selection on an *ABCC11* SNP determining earwax type. *Molecular Biology and Evolution*, **28**, 849–857.

Orengo DJ, Aguade M (2007) Genome scans of variation and adaptive change: extended analysis of a candidate locus close to the *phantom* gene region in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **24**, 1122–1129.

Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, **52**, 935–949.

Orr HA (2003) The distribution of fitness effects among beneficial mutations. *Genetics*, **163**, 1519–1526.

Orr HA (2005a) Theories of adaptation: what they do and don't say. *Genetica*, **123**, 3–13.

Orr HA (2005b) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–127.

Owen DF, Clarke CA (1993) The medionigra polymorphism in the moth, *Panaxia dominula* (Lepidoptera: Arctiidae): a critical re-assessment. *Oikos*, **67**, 393–402.

Quesada H, Ramírez UEM, Rozas J, Aguadé M (2003) Large-scale adaptive hitchhiking upon high recombination in *Drosophila simulans*. *Genetics*, **165**, 895–900.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rieseberg LH, Widmer A, Arntz AM, Burke JM (2002) Directional selection is the primary cause of phenotypic diversification. *Proceedings of the National Academy of Sciences*, **99**, 12242–12245.

Robinson SJ, Samuel MD, Johnson CJ, Adams M, McKenzie DI (2012) Emerging prion disease drives host selection in a wildlife population. *Ecological Applications*, **22**, 1050–1059.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, **66**, 1–17.

Roper C, Pearce R, Bredenkamp B *et al.* (2003) Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet*, **361**, 1174–1181.

Rosenthal R (1979) The file drawer problem and tolerance for null results. *Psychological Bulletin*, **86**, 638–641.

Rosser N, Dasmahapatra KK, Mallet J (2014) Stable *Heliconius* butterfly hybrid zones are correlated with a local rainfall peak at the edge of the Amazon basin. *Evolution*, **68**, 3470–3483.

Sandoval CP (1994) The effects of the relative geographic scales of gene flow and selection on morph frequencies in the walking-stick *Timema cristinae*. *Evolution*, **48**, 1866–1879.

Seetharaman S, Jain K (2013) Adaptive walks and distribution of beneficial fitness effects. *Evolution*, **68**, 965–975.

Siepielski AM, DiBattista JD, Carlson SM (2009) It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, **12**, 1261–1276.

Siepielski AM, DiBattista JD, Evans JA, Carlson SM (2011) Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1572–1580.

Siepielski AM, Gotanda KM, Morrissey MB *et al.* (2013) The spatial patterns of directional phenotypic selection (P Thrall, Ed,). *Ecology Letters*, **16**, 1382–1392.

Slatkin M (2008) A Bayesian method for jointly estimating allele age and selection intensity. *Genetics Research*, **90**, 129–137.

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**, 23–35.

Taylor SM, Antonia A, Feng G *et al.* (2012) Adaptive evolution and fixation of drug-resistant *Plasmodium falciparum* genotypes in pregnancy-associated malaria: 9-year results from the QuEERPAM study. *Infection, Genetics and Evolution*, **12**, 282–290.

Turchin MC, Chiang CW, Palmer CD *et al.* (2012) Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Publishing Group*, **44**, 1015–1019.

Vitalis R, Gautier M, Dawson KJ, Beaumont MA (2014) Detecting and measuring selection from gene frequency data. *Genetics*, **196**, 799–817.

Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting natural selection in genomic data. *Annual Review of Genetics*, **47**, 97–120.

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration–selection balance. *Evolution*, **65**, 1897–1911.

R.D.H.B. and T.J.T. designed the research. T.J.T. performed the literature review and analysis. T.J.T. and R.D.H.B. wrote the manuscript.

Data accessibility

The literature databases, databases of selection coefficients, input data files, analysis code and code to construct figures are available as supplementary material, or upon request from the authors. See the supplemental text for a description of files.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 The distribution of selection coefficients in the full dataset, categorized by direction of selection (positive or negative).

Fig. S2 Analysis of publication bias: Figure 1 of the main text, with selection coefficients binned at 0.01 intervals.

Fig. S3 Analysis of publication bias: the relationship between sample size and estimated selection coefficient, *s*, in the full and reduced datasets.

Fig. S4 Analysis of publication bias: the relationship between sample size and precision of estimation of *s*, for all data which reported error bounds (top) and excluding studies with N > 5000 (bottom).

Fig. S5 Analysis of publication bias: the relationship between the precision of the estimate of *s* (size of error bounds) and the magnitude of *s*, for all selection coefficients with error bounds.

Table S1 Description of the directional selection database.

Table S2 Summary of results of a sensitivity analysis to determine if our conclusions are robust to the inclusion of null estimates of selection ($s = 0$).

Appendix S1 Genetic architecture and the distribution of phenotypic and genetic selection coefficients.

Appendix S2 Selection on Mendelian phenotypes.