ANNUAL REVIEWS

*Annual Review of Genomics and Human Genetics*

# Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies

Guy Sella[1,2,3] and Nicholas H. Barton[4]

[1]Department of Biological Sciences, Columbia University, New York, NY 10027, USA; email: gs2742@columbia.edu

[2]Department of Systems Biology, Columbia University, New York, NY 10032, USA

[3]Program for Mathematical Genomics, Columbia University, New York, NY 10032, USA

[4]Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria; email: nick.barton@ist.ac.at



Annu. Rev. Genom. Hum. Genet. 2019. 20:461–93

First published as a Review in Advance on July 5, 2019

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

https://doi.org/10.1146/annurev-genom-083115-022316



## Keywords

evolution, genome-wide association study, GWAS, quantitative genetics, complex traits, polygenic adaptation, genetic architecture

## Abstract

Many traits of interest are highly heritable and genetically complex, meaning that much of the variation they exhibit arises from differences at numerous loci in the genome. Complex traits and their evolution have been studied for more than a century, but only in the last decade have genome-wide association studies (GWASs) in humans begun to reveal their genetic basis. Here, we bring these threads of research together to ask how findings from GWASs can further our understanding of the processes that give rise to heritable variation in complex traits and of the genetic basis of complex trait evolution in response to changing selection pressures (i.e., of polygenic adaptation). Conversely, we ask how evolutionary thinking helps us to interpret findings from GWASs and informs related efforts of practical importance.

## 1. INTRODUCTION

Understanding the processes that generate differences in complex traits among individuals, populations, and species has been a central challenge to evolutionary biology since Darwin and Galton. Many, if not most, phenotypes of interest, including morphological, life history, and behavioral traits, as well as the risks of diseases, are genetically complex, in that heritable variation in these traits is due mostly to small contributions from many genetic loci (17, 102, 106, 155). Moreover, the heritable contribution typically accounts for a substantial proportion of the variation among individuals (119, 130, 178). Making sense of phenotypic differences among individuals therefore requires understanding the causes of genetic variation in complex traits. In turn, when selection pressures change, the response of the complex trait—the adaptive change—involves shifts in allele frequencies at many loci. Characterizing the polygenic response is therefore key to understanding the evolution of phenotypic differences among populations and species.

These topics have been studied for more than a century, but recent developments in the study of complex traits, notably in humans, make them especially timely. Until recently, empirical work on complex traits was limited largely to phenotypic observations, which allowed the estimation of, e.g., the genetic and environmental contributions to phenotypic variance, or the design of efficient breeding programs in agriculture (43, 106). Similarly, empirical studies motivated by evolutionary questions focused on phenotypic measurements of, e.g., the mutational input to phenotypic variance (58, 64, 72, 106, 115), the relationship between traits and components of fitness (94, 182), and the phenotypic response to selection (e.g., 11, 66, 185). Only over the past decade has it become feasible to systematically dissect the genetic basis of variation in complex traits (140), moving observations from a phenotypic (macroscopic) level to an allelic (microscopic) one. Notably, since 2007, genome-wide association studies (GWASs) in humans have identified many thousands of variants that are reproducibly associated with hundreds of complex traits, including susceptibility to a wide variety of diseases (177, 188). These studies have begun to reveal the genetic architecture of complex traits, particularly the numbers and genomic distributions of the variants that affect complex traits, and the distributions of their frequencies and effect sizes.

These new data hold the promise of substantial progress in our understanding of the evolution of complex traits. The genetic architecture of a trait is informative about the population genetic processes that gave rise to it; conversely, a better understanding of these processes can provide insight into practical questions, such as why GWASs have accounted for only a modest fraction of heritable variation [the missing-heritability problem (109)], why GWASs differ in success across traits, the design of future mapping studies, and prospects for predicting phenotypes from genomes. In addition, combining information about the effects of individual loci on traits with other kinds of data—notably, changes in allele frequency across space and time—should help us learn about polygenic adaptation.

With these objectives in mind, we review what is known about the processes that underlie variation in complex traits among individuals and populations. We have not aimed to be exhaustive, instead focusing on what we think is most relevant for realizing the opportunities for progress.

## 2. THE PREMISE

We begin by summarizing lines of evidence indicating that (*a*) variation in many traits is influenced by a large number of loci, (*b*) allelic effects on a trait are well approximated by a simple additive model, (*c*) alleles that affect one trait typically affect many others, and (*d*) complex traits are typically under stabilizing selection toward some optimum. Some of these assertions are better supported than others, but together they provide a plausible premise for beginning to think about

the evolution of complex traits. We rely on this premise later, when we consider the processes that shape heritable variation (Section 3) and polygenic adaptation (Section 4). For a summary of the notation and a glossary of key terms used throughout the article, see the **Supplemental Material**.

Supplemental Material >



## 2.1. Many Traits Are Highly Polygenic

The genetic basis of phenotypic variation spans a broad spectrum, from the Mendelian extreme, in which alleles at a single gene explain most heritable variance in a trait, to the highly polygenic extreme, in which thousands of alleles, distributed across the genome, contribute to heritable variance. Strong evidence suggests that many traits fall in the latter extreme.

This view is supported by more than a century of systematic artificial selection on a vast array of traits in plants and animals in agriculture and evolution experiments. The change in trait value in response to artificial selection is typically amazingly rapid, fairly steady, and without a limit in sight (11, 66, 185). Examples from agriculture abound, including a more than fourfold increase in the weight of eight-week-old chickens, with a near doubling of the proportion of breast meat, between 1957 and 2001 (59); a more than fourfold increase in grain oil content (from ∼5% to >20%) in 100 generations of selection on one strain of corn; and a threefold increase in grain protein content (from ∼10% to ∼30%) in another (117). There are also numerous examples from evolution experiments, notably in *Drosophila*, including an increase in flying speed in a wind tunnel from 2 cm $s^{-1}$ to 170 cm $s^{-1}$ (184), a more than tripling of the abdominal bristle number (200), and an ∼20% increase or decrease in body size (176), all within ∼100 generations. In these examples and many others, changes due to selection far exceed the range of variation in the original populations, within ∼40–100 generations.

It is hard to account for these rapid, sustained responses on the assumption that they arise from a few alleles of large effect. If such alleles had been present initially, they would have ascended to high frequencies shortly after the onset of directional selection. As they reached intermediate frequencies, the phenotypic variance and thus the rate of phenotypic change would have increased substantially (11, 204) (see Section 4.1). As alleles neared fixation, phenotypic change would have slowed down and eventually ground to a halt. Neither phenomenon is typically seen. Moreover, newly arising mutations cannot salvage this explanation: If only a few loci affect the trait, then the mutational target is small, and given that artificial selection is typically conducted on small populations (e.g., with effective sizes on the order of 100), the input of new mutations would be too low to account for the continuous and rapid changes that are observed.

By contrast, these observations do accord with genetic variation being spread thinly over many loci of small effect (68, 185). In fact, they are well approximated by the idealized infinitesimal model (9, 20, 46), in which genetic variation arises from infinitely many segregating alleles with infinitesimal effects. In this case, small changes to allele frequencies at numerous loci are expected to produce a rapid, steady, and sustained change in mean, with negligible effects on phenotypic variance. Although the small effective population sizes may entail a rapid loss of genetic variance, the large mutational input can sustain the response (64). This highly polygenic view is further supported by recent evolve-and-resequence experiments in *Drosophila*, in which significant changes in allele frequency are widely distributed across the genome (e.g., 6, 23, 170, 176). While rampant hitchhiking—in which alleles in linkage disequilibrium (LD) with causal loci are dragged to higher frequency—makes it difficult to estimate how many of the alleles are causal, their number is plausibly large (25, 81).

The high polygenicity and genome-wide distribution of variation of many traits are further supported by estimates of the heritable variation tagged by single-nucleotide polymorphisms (SNPs) in human GWASs (SNP heritability). For many traits, estimates of the heritability

contributed by chromosomes are proportional to their length (155), suggesting that causal variants are numerous and are distributed fairly uniformly across the genome. Nonetheless, traits vary in their polygenicity. For instance, Shi et al. (155) found that, for the most polygenic among 30 traits, schizophrenia and height, the top 1% of loci with the greatest contribution to (SNP) heritability collectively contribute only 4.2% and 6.5% of the total estimated (SNP) heritability, respectively; on the less polygenic end, for rheumatoid arthritis and lipid traits, the top 1% of loci account for 14–30% of heritability. Similarly, Loh et al. (102) estimated that more than 71% of 1-Mb windows in the genome contribute to heritability for schizophrenia, while similar lower bounds for hypertension and dyslipidemia were ∼60% and ∼30%, respectively. In some cases, the less polygenic among these traits have specific regions with large contributions, e.g., the HLA locus for rheumatoid arthritis, ulcerative colitis, and inflammatory bowel disease, or *FTO* for body mass index (BMI) and lipid traits, on an otherwise highly polygenic, fairly uniformly distributed background (155). While direct estimates of the number of causal variants contributing to heritability are still rare, Boyle et al. (17) estimated that ∼3.8% of the 1000 Genomes Project SNPs affect height, amounting to more than 100,000 common alleles [i.e., alleles with a minor allele frequency (MAF) of >1%].

Pritchard and colleagues (17, 100) have recently proposed a thought-provoking interpretation of these findings (also see 27). In their omnigenic model, the value of a given trait is determined by the expression level of a few core genes in relevant tissues. The expression levels of core genes, however, can be weakly affected by the expression level of any other gene in these same tissues, and thus by any variants that affect the expression of these other genes [i.e., expression quantitative trait loci (eQTLs)]. As an example, weak effects could arise through shared cellular machinery or by regulatory crosstalk within the RNA and protein network. Even though these *trans* effects may be minute, their sheer number (together with the fact that, due to selection, only weak allelic effects are expected to segregate at appreciable frequencies) leads them to account, in aggregate, for most of the heritable variance in the trait. This model is supported by several lines of evidence; for example, active chromatin in relevant tissues, rather than specific functional annotations, seems to be the best predictor of GWAS signals. Notably, it provides an explanation for why so many loci, distributed across the genome, contribute to the heritability of a given trait.

## 2.2. A Largely Additive Genotype-to-Phenotype Relationship

With numerous loci affecting a trait, how should we think about the relationship between an individual's genotypes at these loci and the person's trait value? In principle, quantitative genetics can describe any relationship between genotype, environment, and phenotype. The variance of a trait due to genetics ($V_G$) can be partitioned into additive ($V_A$), dominance ($V_D$), and a combined epistatic component ($V_I$), which itself can be partitioned into two-locus ($V_{AA}$, $V_{AD}$, and $V_{DD}$) and multilocus components ($V_{AAA}$, etc.); higher-order terms in this expansion are defined through the residuals of lower-order ones. Fisher introduced this expansion in his seminal 1918 paper (46), showing how in principle the components can be estimated from the phenotypic correlations among relatives (see 106).

Analyses based on the correlations among relatives and other lines of evidence (e.g., based on the response to artificial selection) in a variety of species suggest that the bulk of genetic variance in quantitative, complex traits is additive (see review and caveats in 67). This statement requires some clarification: It is formulated in terms of an analysis of variance, which is a descriptive statistical framework rather than a generative model. The processes that generate an individual's phenotype clearly involve highly complex and nonlinear interactions among many genes and external and internal environments. Nonetheless, the predominance of additive variance implies that

(on average) the deviation of an individual's trait value from the population mean, $z$, can be well approximated by a simple additive model:

$$z = g + e = \sum_{l=1}^{L} (a_l + a'_l) + e, \qquad 1.$$

where $g$ is the individual's breeding value, which is the additive genetic contribution from the $L$ segregating sites affecting the trait; $a_l$ and $a'_l$ are the effects of the parents' alleles at site $l$, scaled such that the population mean contribution at each site equals 0; and $e \sim N(0,\ V_E)$ is the additive environmental contribution. Importantly, $a_l$ and $a'_l$ reflect marginal allelic effects, averaged over the distribution of genetic backgrounds and environments in the population being considered. Assuming that segregating sites are biallelic, a given site with difference $a$ in the effect of alleles and MAF $x$ contributes $2a^2x(1-x)$ to genetic variance. The additive genetic variance, $V_A$, is a sum of these contributions; the total phenotypic variance is the sum of the additive genetic and environmental contributions, $V_P = V_A + V_E$; and the narrow-sense heritability, $h^2 = V_A/V_P$, is the proportion of phenotypic variance due to additive genetic effects.

From the perspective of phenotypic prediction, we would like to know how well the breeding value can be approximated by polygenic scores, constructed by adding up estimated allelic effects at subsets of loci associated with a trait. The narrow-sense heritability for traits is typically substantial (~0.1–0.9) (178), suggesting that polygenic scores should, in principle, account for a substantial proportion of the phenotypic variance. A current polygenic score for height, for example, for which $h^2 \approx 0.8$, relies on effect-size estimates for 20,000 SNPs and explains ~40% of the phenotypic variance in the self-described "white British" (96). In the near future, we should learn how well individual phenotypes can be approximated for a variety of traits across human populations.

The primacy of additive variance can be understood in several ways. Genetic variance can be modeled in terms of contributions arising from different combinations of alleles—from individual alleles, from the interactions of two alleles at a locus, from the interactions of two or more alleles at different loci, etc. Consequently, the additive variance can be expressed as a sum over terms proportional to second-order products of allele frequencies ($x_i x_j$), whereas $k$th-order variance components are sums of terms proportional to $2k$th-order products. Maki-Tanila & Hill (107) showed that, assuming minor alleles affecting quantitative traits tend to be rare (as are most neutral alleles, let alone selected ones), the additive component can contribute the bulk of genetic variance even when dominance and epistatic interactions are included. This argument does not assume that interactions have a negligible effect; rather, it states that even when their effect is substantial, most of it can be attributed to the marginal effects of individual alleles.

A distinct argument for the primacy of additive variance is that most segregating genetic variation arises from alleles with small effects on the trait, which are likely to translate into additive effects. It has long been known that mutations of small effect at a given locus tend to act additively, whereas those of large effect tend to be recessive (3, 16, 28, 56, 127, 191). Wright (191) explained this pattern in terms of the flux through a metabolic pathway subject to Michaelis–Menten-like enzyme kinetics. He argued that the flux would be only moderately reduced by halving the enzyme concentration, as occurs in loss-of-function heterozygotes, but is eliminated if the concentration drops to 0, as happens in loss-of-function homozygotes. In turn, the effect of mutations that only slightly reduce enzyme concentration is well approximated by the linear term of a Taylor expansion of flux with respect to concentration, around the normal (wild-type) expression level, such that the effect in homozygotes is approximately double that in heterozygotes.

This argument may extend to dominance and epistasis more generally. We can envision a multivariate Taylor expansion of a trait value with respect to some measure of gene activity for

each gene. The effect of sufficiently small perturbations in activity may be well approximated by the linear terms in the expansion, whereas larger perturbations amplify the importance of the higher-order, nonlinear effects (or disrupt smoothness altogether). This reasoning may help explain why additivity is insensitive to the particular choice of trait or to (smooth) transformations of the units in which it is measured. The argument remains incomplete, however, without an explanation of why small-effect mutations dominate genetic variation in complex traits. Possibilities include that most spontaneous mutations have small effects or that selection substantially reduces the frequencies and thus the contributions to variance of mutations with large, nonlinear effects.

Albeit speculative, these considerations are consistent with GWAS findings in humans, whereas evidence sometimes taken as providing arguments against additivity may not in fact apply to natural populations. For example, many studies have reported dominance and epistatic interactions in crosses between inbred lines and between individuals from diverged populations and species (e.g., 73). However, such crosses distort the allelic frequency spectrum (see 107) and bring together combinations of alleles that selection might prevent from cosegregating in the same population (akin to Dobzhansky–Muller incompatibilities in speciation) (48, 150, 156). Dominance and epistasis are also often observed for mutations of large effects, e.g., in knockout experiments (3, 153). The variants identified in human GWASs rarely if ever have such large effects, however; instead, their effect sizes are typically much smaller than the total phenotypic standard deviation (e.g., 101, 126, 152, 189). Extremely rare variants with large effects will tend to be missed by GWASs because of the reliance on genotyped common variants, but they are rarely found by other methods either (e.g., linkage studies and burden tests); even in traits for which they are more common (such as autism), their total contribution to phenotypic variance is small (74). Moreover, dominance effects in GWASs appear to be minor (e.g., 205). Similarly, considerable efforts to detect epistatic interactions in human GWASs have by and large come up empty-handed (e.g., 187, 203), possibly because they remain severely underpowered (187; but see 202). In summary, although the evidence is not airtight, it supports the notion that genetic variation in complex traits can be largely explained by additive effects.

From an evolutionary perspective, if we assume that higher-order genetic and environmental interactions are fairly constant, then the marginal additive effect of an allele should reflect its average effect on the trait during its sojourn (through time) in the population. Thus, in principle, the simple additive model (Equation 1) can be used to relate selection on traits to selection on individual alleles.

**2.2.1. A transformed-additive model for complex diseases.** While complex diseases are often described and studied as discrete traits, they are assumed to arise from an underlying continuous liability described by an additive model (but see 116). Falconer & Mackay (43) assumed that the disease state arises when liability exceeds a threshold, following Wright's (193) treatment of discrete complex traits. Risch (139) modeled the probability of developing a disease as a product of variants' odds ratios, which is equivalent to assuming that this probability is given by the logarithm of an underlying additive liability; this model is the premise for most disease mapping studies (e.g., linkage studies and case–control GWASs). Quite generally, we can think of the probability of developing a disease as a function of a set of continuous traits; assuming that this function is sufficiently smooth and allelic effect sizes are sufficiently small, the probability can be well approximated by a transformation of an underlying additive liability (but see 206). When additional constraints on model parameters are considered (e.g., that common diseases typically have prevalences of ∼0.1–2% and risk ratios for first-degree relatives of ∼4–10), the differences among most models become negligible, suggesting that they are practically interchangeable (161, 190).

To the best of our knowledge, there is not much evidence to indicate whether disease risk is well approximated by these models, in part because it is difficult to measure correlations among relatives in disease risk. Applications of polygenic scores to predict heritable disease risk may clarify this issue in years to come. For now, assuming transformed-additive models of complex-disease risk seems like a sensible starting point for thinking about the evolution of complex diseases.

## 2.3. Pleiotropy Is Pervasive

The high polygenicity of many traits almost inevitably implies extensive pleiotropy: If variation affecting a given trait spans a considerable portion of the functional genetic variation, then it is bound to overlap with variation affecting other traits (in addition to which causal variants in LD may also generate effective pleiotropy). Indeed, many of the variants identified in human GWASs are associated with more than one trait, and the extent of pleiotropy uncovered appears to be increasing rapidly with improvements in power and methodology (e.g., 4, 18, 34, 95, 128, 160, 164). The omnigenic model offers some intuition here, as an eQTL for one gene could affect expression of core genes for multiple traits in the same and in different tissues (17, 100).

The effects of variants on different traits can be correlated, e.g., due to shared regulatory pathways (100). Examples of genetic correlations among traits abound, both in classical quantitative genetics (145) and in GWASs. For example, Pickrell et al. (128) conducted a meta-analysis of GWASs to show that "variants that delay the age of menarche in women tend to delay the age of voice drop in men, decrease body mass index, increase adult height, and decrease risk of male pattern baldness" (p. 715). The same variants can also affect multiple traits without having correlated effects on these traits. We would expect this type of uncorrelated pleiotropy [sometimes referred to as type I (180)] to be abundant as well (100), but it is harder to identify, requiring knowledge of the genetic basis of variation. As we review below, these two kinds of pleiotropy have different evolutionary consequences.

Of course, we would be remiss if our consideration of pleiotropy did not note that traits are to some degree an arbitrary construction. In principle, a trait could be any feature of an organism that we can measure. With a potentially infinite number of measurable traits and with finite genetic variation, the genetic basis of variation in subsets of traits would inevitably overlap. However, more often than not the choice of traits is motivated by biological intuition, and many examples point to extensive pleiotropy among "biologically meaningful" traits. Considering selection on traits may allow us to better articulate what we mean by biologically meaningful.

## 2.4. Selection on Complex Traits

We can think of selected traits as fitness modules. In this view, fitness is a function of a set of selected traits, and altering the value of one while leaving others constant (within a functioning range) changes fitness. These sets of traits are organized in a nested hierarchy, ranging from fitness itself at the top, through viability and fecundity, down to the expression of every gene at every time and tissue. The value of a trait in this hierarchy derives from multiple traits below it; conversely, selection on a trait derives from its effects on the traits above it. Below, we consider selection on genetic variation affecting a given complex trait, which may or may not be a fitness module. Some of the models we describe posit direct effects of variation on fitness, while in others, selection derives from the effects of variation on intermediate fitness modules, which are themselves complex traits.

Selection on traits depends on an organism's ecology in the broad sense, bearing in mind that an organism's makeup influences the environmental attributes that affect it (98). In this review, we focus on relatively short evolutionary timescales, assuming that the traits that make up the fitness hierarchy remain constant. In a constant environment, we can envision a fitness landscape,

in which fitness is a function of the values of a set of selected traits. At steady state, we expect the population to be near some optimum in this landscape. We also expect the range of possible trait values around the optimum to be bounded; for example, fecundity cannot be infinite and may be constrained by a trade-off with viability. If fitness is maximized when the trait value is at the edge of its range, then directional selection pushes the trait against the boundary. If fitness is maximized when the trait takes an intermediate value, stabilizing selection pushes it toward its optimum.

There must be many quantitative traits, simple and complex, whose values need to be close to some intermediate optimum for the organism to function well. The morphology of a fly's wings is especially well studied: They have been precisely fine-tuned by selection, such that their many attributes—shape and weight (183), cellular density, and positioning of muscles and veins (1)—have specific combinations that optimize flight. Attaining this form entails many additional requirements, notably on gene expression and cellular patterning during development (1). Although any one of these attributes is affected by numerous mutations (115, 183, 186), the morphologies show remarkably little variation within populations and among closely related species (71, 183). More generally, the fossil record shows that morphological traits tend to remain fairly constant over long timescales (53), even though the morphology is highly unlikely to be neutral and changes rapidly when under artificial selection (5, 11, 68). This constancy of form and function indicates that stabilizing selection keeps many traits close to an optimum—and, indeed, this is essential for organisms to survive.

Other selected traits, especially at the very top and bottom of the fitness hierarchy, are clearly subject to directional selection. For instance, selection suppresses expression of genes in tissues where they are not needed, while at the highest level, fitness and its components are by definition under directional selection. It is less clear whether directional selection is common for intermediate, complex traits. As one example, the rate of somatic mutation is presumably selected to be as low as possible, as the incidences of cancers and other diseases increase as this rate increases, although it is likely bound from below by the costs of repair (162). More generally, while some complex diseases may reflect deviations of an underlying trait from an intermediate optimum (e.g., blood pressure or body mass index), others may arise in a setting in which directional selection simply acts to minimize liability. The literature on the evolution of complex traits assumes that there are many more complex selected traits under stabilizing selection than under directional selection, but, as we review next, this premise is hard to assess.

**2.4.1. Phenotypic measurements of selection.** In principle, the strength of directional and stabilizing (or disruptive) selection acting on quantitative traits can be estimated from the relationship between the trait values and fitness (94, 125). Directional selection on a trait in a single generation produces a change in its mean, known as the selection differential. For a single trait, the selection differential is the covariance of relative fitness, $w$, and the trait value, $z$—i.e., $S = \mathrm{Cov}(w, z)$. In natural populations, it is typically estimated by contrasting the means before and after some event that causes mortality, or by the covariance between fecundity and trait value, and thus reflects only a component of fitness. Because only heritable differences are passed down to the next generation, the selection response (i.e., the difference in mean across successive generations) is related to the selection differential by the breeder's equation (65, 104)—i.e., $\Delta\bar{z} = h^2 \cdot S$, where $h^2 = V_A/V_P$. The breeder's equation can be rewritten as $\Delta\bar{z} = V_A \cdot \beta$, where the selection gradient, $\beta = S/V_P = \mathrm{Cov}(w, z)/V_P$, is the regression coefficient of relative fitness on the trait value and measures the strength of directional selection on the trait. A similar logic applies to stabilizing (or disruptive) selection. With a single trait, the stabilizing selection differential is the quadratic regression coefficient of relative fitness on the trait, $\gamma = \mathrm{Cov}(w, (z - \bar{z})^2)/{V_P}^2$. Consider, for example, a trait that is normally distributed around an optimum with variance $V_P$ and

subject to Gaussian stabilizing selection, i.e., with fitness decreasing with distance from its optimum, $\delta z$, as $\exp(-\delta z^2/2V_S)$. If we plausibly assume that $V_S \gg V_P$ (see, e.g., 157), the stabilizing selection differential is $\gamma = -(V_S + V_P)^{-1} \cong -V_S{}^{-1}$, which is the strength of selection around the optimum. Stabilizing selection reduces (and disruptive selection increases) phenotypic variance, where the selection response in successive generations is $\Delta V_P = h^4 \cdot \mathrm{Cov}(w, (z - \bar{z})^2) = V_A{}^2 \cdot \gamma$.

When traits are correlated, which is often the case, estimates of the strength of selection using a single trait absorb indirect effects of selection on others. In principle, this problem can be overcome by treating multiple traits jointly in the multivariate framework introduced by Pearson (125) and extended by Lande & Arnold (94). Assuming multiple, normally distributed traits, the breeder's equation takes a vector form in the multidimensional trait space, $\Delta\bar{\vec{z}} = h^2 \cdot S = V_A \cdot \beta$, where $h^2 = V_A \cdot V_P{}^{-1}$ is the multivariate heritability matrix, and $V_A$ and $V_P$ are variance–covariance matrices. Similarly, the quadratic selection gradient and response take matrix forms. Importantly, the components of the selection gradients, e.g., $\beta = V_P{}^{-1} \cdot \mathrm{Cov}(w, \vec{z})$, are partial regression coefficients that measure the strength of selection on each trait, excluding indirect effects from other traits that are included in the analysis. Viewing the population on a multidimensional phenotype fitness landscape, the linear and quadratic selection gradients correspond to the gradient and curvature at the population's mean phenotype, respectively.

One of the earliest and most prominent applications of the multivariate method was to study selection on Galapagos finches, *Geospiza fortis* and *Geospiza scandens*, on the island of Daphne Major (reviewed in 54). During a drought in 1977, selection strongly favored birds who could crack large seeds. In subsequent years, changes in seed abundance reversed the direction of selection, and in both cases, beak morphology changed accordingly. The multivariate analysis showed that beak depth was the primary target of selection, and the multivariate breeder's equation predicted the changes in beak morphology (132). In this example, the method may have been successful because seed handling was a strong component of fitness and was influenced by only a few measurable morphological traits.

Other long-term studies of natural selection in the wild (84, chap. 30 of 182), are more difficult to interpret (32). In many studies, traits are estimated to be under appreciable directional selection and to have substantial additive genetic variance, yet trait means typically change little over long time spans (see, e.g., 53, 85). In addition, estimates of quadratic selection suggest that disruptive and stabilizing selection are equally common, which conflicts with the long-term stability of the traits' means and variances. In part, these discrepancies may reflect a focus on systems in which selection varies strongly in space (e.g., along geographic clines) or time (e.g., in response to severe disturbances), which might be unrepresentative of longer-term selection affecting heritable genetic variation.

These discrepancies may also reflect inherent difficulties in measuring selection, however. One problem is that, contrary to the premise of the breeder's equation, covariance between fitness and a trait might not reflect selection on the trait; rather, it could reflect a nonheritable environmental effect on both. For example, both fitness and the trait may be strongly affected by an individual's general condition (91), and variation in condition may be largely due to the environment. Another problem is that the multivariate approach relies on estimating variance–covariance matrices, which requires sample sizes that scale with the number of traits squared. Thus, in practice, most studies can afford to include only a few traits, and correlated traits that are excluded likely introduce substantial biases. This suggests that the prevalence of disruptive and stabilizing selection in meta-analyses largely reflects a combination of low statistical power and publication bias. Notably, only 16% of the reported estimates of quadratic selection terms are significant, and larger studies generally report weaker selection (84). In humans, larger sample sizes have recently become available; Sanjak et al. (147) recently applied the multivariate approach, using lifetime

reproductive success as a proxy for fitness in a sample size of more than 250,000 from the UK Biobank (168), orders of magnitude larger than previous studies. Nonetheless, they had to limit the number of traits included and apply an approximate multivariate regression. Estimates of selection gradients, accounting for multivariate correlations, were substantially reduced compared with previous findings; moreover, they found a large excess of stabilizing relative to disruptive selection.

Such large data sets in humans, with extensive phenotypic, environmental, and genomic information, may offer a way forward. In particular, genomic information makes it possible to estimate covariances between fitness components and breeding values (rather than phenotypic values) of a trait. The selection response is proportional to such covariances: The change in mean is $\Delta\bar{z} = \mathrm{Cov}(w, (g - \bar{g}))$ and in variance is $\Delta V_P = \mathrm{Cov}(w, (g - \bar{g})^2)$, where $g$ is the trait's breeding value, and $\bar{g}$ is its mean (131, 143). Importantly, because these equations rely on heritable variation, they should not be affected by environmental effects or predicated on selection acting directly on the trait being considered. This approach has been applied to identify selected changes in individual traits, using both pedigrees in species in the wild (e.g., 32) and genomic data in humans (e.g., 12, 88, 118, 147). It can also be applied to multiple traits and, accounting for their heritability matrix, can provide estimates of the strength of selection on individual traits. In practice, even with large data sets, statistical power is limiting. However, it may be possible to apply dimensionality reduction (e.g., principal component analysis) to reduce the number of traits included, while limiting errors due to those that are not.

**2.4.2. Limits on the number of selected traits.** While it is plausible that many complex traits are under stabilizing selection, their number and the strength of selection on them remain unknown. Interestingly, theory suggests that both are strongly constrained (7). Consider, for example, $n$ uncorrelated traits that are normally distributed around the optimal $n$-dimensional phenotype, each with variance $V_P$, where each trait is subject to Gaussian stabilizing selection of strength $V_S^{-1}$. If we plausibly assume that variation in individual traits has a minor effect on mean fitness, i.e., $V_S \gg V_P$ (see, e.g., 157), then variation in all traits will reduce fitness by $\sim\exp(-n(V_P/2V_S))$ relative to the optimal phenotype. If organisms below a given absolute fitness cannot function, this genetic load would impose strong constraints on the number of independently selected traits. However, this constraint would be weaker if most selection is soft, e.g., based on competition between individuals rather than on differences in absolute fitness, or if there are negative epistatic interactions in effects on fitness (87).

A more robust (though weaker) constraint is set by the actual variance in relative fitness among individuals. The standard deviation in relative fitness increases with the number of traits as $\sim\exp(n(V_P/2V_S)^2)$. While the standard deviation in fitness is difficult to measure (but see 30), it is unlikely to exceed ~1/4 and is bound from above by offspring numbers (7). Assuming the traditional estimate of $V_P/V_S \sim 1/20$ (174) would suggest an unreasonably low number of ~400 traits (although $V_P/V_S$ is likely smaller typically, and thus the number of traits could be somewhat larger). Such constraints on the strength and extent of stabilizing selection definitely merit further investigation.

## 3. THE PROCESSES THAT SHAPE GENETIC VARIATION IN COMPLEX TRAITS

With these preliminary considerations in mind, we turn to one of our central questions: How is the ubiquitous heritable variation in complex traits shaped and maintained? This question has been studied for more than a century, largely through mathematical models (50, 182). Until recently,

the focus was on explaining how substantial heritable variation in numerous traits is maintained, guided by phenotypic observations. More recently, emphasis has shifted toward explaining observations about the genetic architectures of complex traits emerging from human GWASs. We briefly review the earlier work, then move to what models tell us about genetic architecture and how they can be related to GWAS discoveries.

### 3.1. Maintenance of Variation

Quantitative genetic variation could be maintained in many ways, reflecting the many processes influencing genetic diversity. These processes can be broadly divided into mutation–selection–drift balance and balancing selection (31, 52). Balancing selection takes many forms, including heterozygote advantage, frequency-dependent advantage of rare alleles, and temporally and spatially varying selection pressures. Nevertheless, there is little evidence indicating that it plays a major role in maintaining genetic variation in general and variation in complex traits in particular (but see 30, 81), although this might be due partly to the difficulties in identifying its subtler forms [especially fluctuating selection pressures (51)].

One argument against the ubiquity of balancing selection is that if it maintained alleles at substantial frequencies and these alleles had substantial effects on complex traits, they would have been among the first to be identified in human GWASs (see Section 3.3). We are unaware of such examples, with one notable exception. The special role of the major histocompatibility complex (MHC, or HLA in humans) loci in adaptive immunity (171) likely renders them subject to persistent balancing selection (167). Specifically, arms races between the host immune system and pathogens likely favor rare alleles at these loci (167), helping to explain why the MHC is the most genetically variable region in the genomes of humans and other species (40, 61). Additionally, since MHC alleles affect general health throughout life, through both the immunity they confer and their pleiotropic effects (e.g., on the risk of autoimmune disorders), they likely affect the values of many complex traits, plausibly explaining why the MHC region is highly enriched for GWAS hits across many traits (97, 172).

One might also view migration–selection balance as a form of balancing selection, one that may help explain why major loci (i.e., loci that contribute substantially to genetic variance) are found for specific traits. Theory suggests that when different populations are subject to opposing selection pressures yet exchange migrants, major alleles may evolve that are beneficial in a given environment (e.g., via loss-of-function mutations, inversions, or the agglomeration of multiple alleles into a compound supergene), thereby reducing the deleterious migration load (86, 198, 199). Such processes may explain why GWASs with relatively small samples of *Arabidopsis thaliana* collected over a diverse ecological range have identified major loci for traits such as flowering time, which show high genetic differentiation along seasonal clines (99). They may also explain findings of major loci, with fairly old alleles, for skin pigmentation in humans, also discovered in a GWAS with a small sample size collected over a wide range; in this case, selection is thought to vary with UV radiation levels across the world and in particular across Africa (36). Similar processes may explain why marine stickleback populations carry extremely old, rare alleles, which were central to repeated adaptive radiations into freshwater environments (77); such loci might have facilitated other adaptive radiations as well (199). However, theory suggests that the conditions in which migration–selection balance produces major loci are quite restrictive; notably, mutations affecting the selected trait should be clustered in low-recombination regions (86, 197–199). In humans, the scarcity of major loci found in GWASs suggests that such mechanisms affected the architecture of few traits. More generally, while various kinds of balancing selection doubtless contribute to the maintenance of genetic variation, they depend on the idiosyncrasies of a trait and so seem unlikely to explain ubiquitous heritability.

By contrast, mutation–selection balance, in which mutations continually replenish genetic variation while stabilizing and/or directional selection (alongside genetic drift) eliminates it, is plausibly universal (182). The conditions for mutation–selection balance affecting many traits—i.e., a substantial mutational input and selection on genetic variation affecting the traits—are largely established by the evidence reviewed above.

The pervasiveness of mutation–selection balance is further supported by experimental work on a variety of traits and taxa (182), in which the input of variation per generation due to mutation, $V_M$, is measured either by the accumulation of mutations in a genetically homogeneous population (see 106) or by relaxing selection in a heterogeneous population (e.g., 115). Estimates of the ratio of mutational and environmental variances, $V_M/V_E$, known as the mutational heritability, are typically ~0.0006–0.006 (72, 106, 115). To interpret these values, note that the mutational input of variance per generation $V_M \cong 2U \cdot E(a^2)$, where $U$ is the rate of mutations per gamete affecting the trait, and that the average mutational effect size squared, $E(a^2)$, is likely smaller than $0.1 \cdot V_P$, based, e.g., on the dearth of large-effect alleles in the response to artificial selection (see Section 2.1). Assuming such an upper bound and accounting for heritability, the estimates of $V_M/V_E$ suggest remarkably high mutation rates of ~0.006–0.06 per gamete per generation affecting a variety of traits. Assuming a typical mutation rate of $\sim 10^{-8}$ per base pair per generation for multicellular eukaryotes, these values suggest mutational target sizes of ~0.15–1.5 Mb, echoing the evidence for high polygenicity based on the response to artificial selection and human GWASs.

The same experiments were used to estimate the ratio of the genetic and mutational variances, $V_A/V_M$. This ratio, also referred to as persistence time, is a measure of the number of generations required for mutation to replenish quantitative genetic variation. If most genetic variation is deleterious, the persistence time should be an average of the inverse selection coefficient; the specific averaging depends on the relation between effect sizes and selection coefficients (see below). If most variation is effectively neutral, the persistence time is much longer, on the order of twice the effective population size, $2N_e$. Estimates of the persistence time are typically on the order of 100 generations (72, 115), much lower than $N_e$ for any species, suggesting that the bulk of quantitative genetic variation is selected against (as opposed to being neutral or under balancing selection) and lending strong support to the dominant role of mutation–selection balance in maintaining quantitative genetic variation.

For all these reasons, understanding how mutation–selection balance shapes quantitative genetic variation has been a major theoretical endeavor over the past four decades (76). One of the early fault lines, the so-called Gaussian versus house of cards debate (174), centered on whether the distribution of allelic effect sizes should be modeled as Gaussian (83, 92) or rather reflects an (unknown) distribution for newly arising mutations (141, 174, 194). One might expect an approximate Gaussian distribution if segregating variation at a locus consisted of (largely) nonrecombining haplotypes carrying many mutations affecting the trait; this would occur if the per-site mutation rate, $u$, far exceeds the recombination rate, $r$. However, we now know that typically $u/r \sim 1$ or less for all mutations, and for mutations affecting a given trait, it is likely to be at least an order of magnitude smaller (e.g., in humans, only ~10% of mutations are under selection; see, e.g., 137). For this reason and others, the Gaussian assumption has been largely abandoned, and LD is now thought to play a minor role in mutation–selection balance (174). Instead, as we review next, models of mutation–selection balance over the past three decades disagree primarily in their assumptions about the relationship between allelic effects on a trait and on fitness.

### 3.2. Mutation–Selection Balance Models of Genetic Architecture

Mutation–selection balance models can be thought of as follows: Mutations affecting the trait arise at a rate $U$, which depends on the number of contributing loci or its target size, and have

some distribution of selection coefficients, $s$. The mode and strength of selection on these mutations, genetic drift, and demographic processes together determine allele frequencies; alongside the mutation rate, they also determine the number of segregating sites. We assume that segregating sites are biallelic; this infinite-sites assumption is a good one for most species, including humans (as $\theta = 4N_e u \ll 1$ per site). We further assume a diploid panmictic population with constant size $N$, but later consider the effects of nonequilibrium demography.

The relationship between a mutation's effect on the trait and on fitness is then modeled using one of two approaches. One traces back to models of mutation–selection balance at a single locus (31, 52, 57, 195): Mutations are assumed to be unconditionally deleterious and semidominant, and the conditional probability distribution of allelic effect sizes, $a$, given their selection coefficient, $P(a|s)$, takes a given form by assumption. The other approach is generative, in the sense that the mode and strength of selection on alleles and $P(a|s)$ arise from modeling the selection acting on traits. The distribution of selection coefficients and $P(a|s)$ determine the joint distribution of frequencies, $x$, and effect sizes of alleles influencing the trait. In particular, they determine the distribution of allelic contributions to genetic variance, $v = 2a^2x(1-x)$, e.g., whether most of the heritability arises from strongly selected, extremely rare alleles or more common, weakly selected ones.

Because the distribution of selection coefficients is generally unknown, we focus primarily on what the assumptions made about the conditional distribution $P(a|s)$ imply about the relative contribution of mutations to variance, given their selection coefficient. At one extreme, the trait in question does not affect fitness, but the alleles that influence the trait have uncorrelated, unconditionally deleterious effects on fitness. Barton (7) and Kondrashov & Turelli (87) showed that these assumptions result in apparent stabilizing selection, because individuals with trait values farther from the population mean tend to carry a greater burden of deleterious mutations and thus have lower fitness. Pritchard (133) considered the genetic architecture of complex-disease risk under these assumptions and showed that, under these conditions, the largest per-mutation contribution to genetic variance would arise from mutations that are effectively neutral and therefore ascend to higher frequencies than other mutations, but happen to have large effects on the trait.

At the other extreme, selection on alleles affecting a trait stems entirely from the trait itself, typically assumed to be under stabilizing selection (e.g., 79, 141, 174, 194). The distribution of phenotypes and allelic dynamics under this model turn out to be relatively simple. When the number of contributing loci is large enough for the trait to be highly polygenic, phenotypes are approximately normally distributed, because they arise from a sum of (approximately) independent contributions from many segregating sites. The phenotypic mean is held tightly at the optimum, because selection becomes stronger with its distance from the optimum and any displacement is quickly adjusted for by minor changes to allele frequencies at many loci. Consequently, individual alleles barely affect the phenotypic mean, and selection acts primarily to reduce their contribution to phenotypic variance. For a biallelic site, this implies that (*a*) selection on the site does not depend on other loci, (*b*) selection takes an underdominant form, acting against the rarer allele, and (*c*) the strength of selection scales with the squared effect size ($s = a^2/V_S$). The allele frequency dynamics are fully characterized by the first two moments of frequency change per generation (41):

$$E(\Delta x) \cong -\frac{a^2}{2V_S}x(1-x)\left(\frac{1}{2}-x\right) \text{ and } V(\Delta x) \cong \frac{x(1-x)}{2N}. \qquad 2.$$

The effect of selection on the contribution to genetic variance follows (79, 157) (**Supplemental Figure 1*a***). In this case, increasing the effect size of a mutation has countervailing effects on its expected contribution to variance: It increases the difference in trait value among individuals with

Supplemental Material >

and without the mutation, but it also strengthens the selection that acts to reduce the frequency of the mutation. Consequently, and in contrast to the purely pleiotropic model, the greatest expected variance arises from mutations under moderate and strong selection (roughly $2Ns > 3$), which contribute approximately $v_S = 2V_S/N$ per unit mutational input.

In practice, we would expect most traits to fall somewhere between the two extreme assumptions about pleiotropic effects. As reviewed above, there are compelling reasons to believe that quantitative genetic variation is highly pleiotropic. Yet contrary to the purely pleiotropic model, we expect alleles with large effects on a given trait (whether selected or not) to tend to have large effects on other traits that are under selection and thus on fitness [e.g., because they cause greater perturbations to gene expression, affecting many traits (17, 100)]. Motivated by these considerations, Eyre-Walker (42) and Caballero et al. (24) considered models in which the correlation between allelic effects on a trait and those on fitness can vary between the purely pleiotropic and direct selection extremes. Interestingly, when plausible, intermediate correlation strengths are assumed, the predictions of these models diverge. In the Eyre-Walker model, mutations with large effect sizes and small selection coefficients are too rare to contribute substantially to genetic variance, and therefore strongly selected mutations with large effect sizes have the greatest contribution to variance. By contrast, in the Caballero et al. model (first presented in 80), the distribution of effect sizes conditional on a given selection coefficient can have thicker tails, and thus weakly selected mutations with larger effect sizes make the greatest contribution. In both cases, the distributions were chosen largely for mathematical convenience. Their diverging predictions suggest that we need to understand more about the statistical relationship between selection coefficients and effect sizes to know what to expect.

Simons et al. (157) approached this problem by deriving (rather than postulating) this relationship, in a model with stabilizing selection on multiple traits. They focused on the architecture of one of these traits (say, the first), where the total number of traits, $n$, quantifies the degree of pleiotropy. With multiple traits, the selection coefficient of an allele is proportional to the sum of its squared effects over all selected traits—i.e., $s = a^2/V_S = \sum_{i=1}^{n} a_i^2/V_S$, where $a_i$ (its effect on trait $i$) is measured in appropriate units. In other words, mutations with a given selection coefficient lie on a hypersphere with a corresponding radius in the $n$-dimensional trait space. Thus, if mutations are equally likely to point in any direction in trait space (i.e., are isotropic), the distribution $P(a_1|s)$ derives from the projection of mutations on the hypersphere on the first dimension (**Supplemental Figure 1*b***). In reality, mutations tend to have larger effects on some traits than on others, and their effects on different traits tend to be correlated (110). Simons et al. (157) showed that when mutational effects are described by a variance–covariance matrix, the distribution $P(a_1|s)$ can be approximated by assuming that mutation is isotropic in a model with an effective rather than an actual number of traits, $n_e$; if, e.g., effect sizes on the focal and other traits are highly correlated, then $n_e$ is less than $n$. Importantly, when genetic variation is highly pleiotropic (e.g., $n_e \geq 10$), the distribution $P(a_1|s)$ is well approximated by the normal distribution $a_1 \sim N(0, (V_S/n_e) \cdot s)$. The genetic architecture of a trait follows from this distribution and the first two moments of change in allele frequency, which take the same form as Equation 2.

The effect of selection on the relative contribution to genetic variance does not depend on the number of traits, and takes the single-trait form (**Supplemental Figure 1*a***) where the expected contribution under strong selection generalizes to $v_S = 2V_S/n_eN$. Thus, similar to the case with a single trait, more weakly selected alleles make smaller contributions to genetic variance. By contrast, the distribution of variance among loci with a given selection coefficient is affected by the number of traits. When variation is highly pleiotropic, as seems plausible, and selection is sufficiently strong (roughly $2Ns > 5$), this distribution approaches a limit, which depends only

Supplemental Material >

on $v_S$ and is thus insensitive to the selection coefficient and degree of pleiotropy (**Supplemental Figure 1*c***).

Supplemental Material >

**3.2.1. Possible extensions.** We can envision several important generalizations to this kind of geometric model. One would be to relax the assumption that the (effective) number of traits affected by mutations is fixed. Notably, given that different genes affect distinct numbers of traits and that genes that affect more traits are likely to evolve under stronger selective constraint, we would expect the number of traits to vary across mutations and covary with the strength of selection. Another generalization would be to relax the assumption that all selected traits are subject to stabilizing selection. To the best of our knowledge, there are currently no models in which the architecture of a complex trait derives from directional selection on its value (e.g., a model of liability), let alone models that include traits under both directional and stabilizing selection. One might also envision mutational effects on a focal trait to partially reflect contributions that are not selected. Such effects can be incorporated by adding an underlying neutral trait. The mean value of such a trait would slowly drift (44, 105), and its variance would be affected by apparent stabilizing selection (7, 87). Lastly, one might consider the effects of selected traits that are not highly polygenic, e.g., when intermediate fitness modules do not fully capture the selection acting on quantitative genetic variation (see Section 2.4). Arguably, models that incorporate such extensions would capture the major factors that are thought to affect genetic variation in complex traits under mutation–selection balance.

**3.2.2. The effects of nonequilibrium demography.** Thus far, we have considered how the genetic architecture of complex traits is shaped at steady state, notably under constant selection and population size. We consider the response to changing selection pressures below (Section 4.1) and argue that, despite its evolutionary importance, it likely has little effect on genetic architecture. By contrast, nonequilibrium demographic changes, particularly ubiquitous changes in population size and admixture from other populations, can dramatically affect genetic architecture. The effects of gene flow definitely merit investigation but are not straightforward and have barely been studied (but see 150, 156). We therefore focus on the better-understood effects of changes in population size (103, 158, 159).

Changes in population size profoundly affect both the numbers and frequencies of segregating variants, and these effects are modulated by selection, because more strongly selected mutations spend less time segregating in the population than do more weakly selected ones. Consider, for example, the case of Europeans, who have been inferred to have gone through an extended bottleneck after leaving Africa and, very recently, undergone explosive population growth (35, 149, 181). Strongly deleterious genetic variation is sufficiently young for much of it to have arisen after the bottleneck or even since the onset of growth, when the population size, $N$, was large. This implies that the number of strongly selected mutations that entered the population was high, as the expected number per generation is $2NU$, but that their initial frequencies were low, i.e., $1/2N$. The total, strongly selected variance is approximately unaffected, because the two effects cancel out, but the distribution of the variance among sites is profoundly affected, with many more sites segregating but at proportionally lower MAFs, and thus with lower per-site contributions to genetic variance (159). Weakly deleterious variation, by contrast, would have largely arisen before or during the out-of-Africa bottleneck, which implies a lower input of mutations and stronger genetic drift that accelerated the loss of most mutations and boosted the frequency of those remaining. We would therefore expect fewer weakly selected segregating sites, but with greater MAFs and per-site contributions to variance (159). More generally, we expect the total contribution of strongly or weakly selected mutations to variance to be fairly insensitive to changes in population

size (and instead depend primarily on their mutational input). By contrast, the number of segregating sites and the distribution of their contributions to variance should be markedly influenced, with strongly selected variation more affected by the more recent population sizes than weakly selected variation.

### 3.3. Relating Evolutionary Models to Findings from Genome-Wide Association Studies

In the early days of GWASs, evolutionary models were used primarily to inquire whether we should expect studies of modest size to succeed at mapping quantitative genetic variation (140), with much attention revolving around the common disease–common variant hypothesis (26, 133, 138). More recently, the questions have been recast in terms of explaining the apparent failure of large GWASs to map much of the heritable variance in many traits, the so-called missing-heritability conundrum (109). While many factors have been argued to contribute to the missing heritability, estimates of SNP heritability, based on all GWAS SNPs rather than only those that are genome-wide significant (GWS), indicate that the bulk of heritable variance derives from loci that current GWAS are underpowered to detect (70, 154, 196).

To a first approximation, a causal biallelic locus will be GWS if its contribution to variance, $2a^2x(1-x)$, exceeds a threshold determined by the study's power, $v^*$. This threshold is proportional to the variance of the background noise for identifying individual loci given the sample size, $m$, i.e., $v^* \propto V_P/m$. Current, large-scale GWASs rely on genotyping and imputation, which, in effect, introduces an additional threshold on MAF (which depends primarily on the size of the imputation reference panel). When GWS associations are represented in terms of their estimated frequencies and effect sizes, they are often distributed tightly above the variance threshold (e.g., **Supplemental Figure 2**). The question about missing heritability can be recast as asking where the remaining loci reside on such plots. For example, are they mostly strongly selected loci with relatively large effects, which evade identification because their minor alleles are so rare, or are they relatively weakly selected loci with relatively small effect sizes, which evade detection because of their small contributions to variance (i.e., due to limitations in sample size)? Fitting evolutionary models to GWAS findings can help to answer these questions.

Supplemental Material >

#### 3.3.1. Fitting models of architecture to findings from genome-wide association studies.

Agarwala et al. (2) took a pioneering step in this direction, asking about the evolutionary parameters that are consistent with the empirical findings for type 2 diabetes. They used forward population genetic simulations to generate samples of genomes, incorporating both neutral and selected mutations, where the latter were assumed to arise from a distribution inferred for nonsynonymous mutations. They ascribed liabilities to these genomes under a range of models that vary both in their mutational target size for the disease (which determined how many of the selected sites were picked to be causal) and in the coupling between selection and effect size, assuming Eyre-Walker's model (which determined how effect sizes were ascribed to these causal sites). The environmental contribution and liability threshold were chosen to match the heritability and prevalence of type 2 diabetes. To determine whether models were consistent with observations, they performed GWASs on their simulated data sets and compared the numbers of GWS associations with the one observed for type 2 diabetes. They also considered summaries of epidemiological and linkage studies, but their inclusion did little to narrow down the range of possible models. While they were able to rule out the pleiotropic and direct selection extremes of Eyre-Walker's model, they were left with a wide range of possible genetic architectures.

More recently, Mancuso et al. (108) applied a similar approach to the study of prostate cancer in men of African ancestry. To this end, they relied on targeted sequencing at 63 loci found to affect

disease risk in a larger GWAS and estimated that rare variants at these loci (with MAFs of 0.1–1%) account for ∼12% of the heritability in risk (on the liability scale) compared with ∼17% for common variants. The contribution of rare variants far exceeds the neutral expectation, indicating that variation affecting disease risk is subject to purifying selection. Mancuso et al. (108) then used their heritability estimate as a summary statistic to infer the coupling between selection and effect size, again assuming Eyre-Walker's model (42), and as in the study by Agarwala et al. (2), they were left with a wide range of possible architectures. These results suggest that using richer summaries of genetic architecture may be required.

Simons et al. (157) took a step in this direction by asking whether the distribution of variances among the loci identified in GWASs for height and BMI in Europeans accords with their theoretical predictions (see Section 3.2). They predicted that, assuming the loci are highly pleiotropic and under moderate to strong selection (because otherwise they contribute much less to variance), the distribution of variances among them is well approximated by a closed form that depends on a single parameter, $v_S$ (**Supplemental Figure 1*c***). They used the observed distribution for the ∼700 GWS associations for height and ∼80 for BMI to estimate this parameter and found that the theoretical distribution provided a good fit for either trait (**Supplemental Figure 3*a***), whereas models with low pleiotropy did not. They then predicted the numbers of moderately and strongly selected loci and corresponding variances that would be captured with future increases in study sizes by extrapolating their fitted distributions to corresponding variance thresholds (**Supplemental Figure 3*b***). Similarly, they estimated that moderately and strongly selected mutations affecting height have a target size of ∼5 Mb and account for ∼50% of the heritable variance, whereas for BMI these values are ∼1 Mb and only ∼15%, respectively. The lower proportion of variance for BMI than for height may explain why the GWAS for BMI succeeded in mapping substantially less heritable variance (∼3–5% versus ∼20%) despite its larger size (∼340,000 versus ∼250,000).

Supplemental Material >

The Simons et al. (157) inferences should be considered preliminary, however, as they do not account for the effects of historical changes in population size. Simulations that incorporate changes in the population sizes of Europeans suggest that only moderately selected loci ($s \sim 10^{-3}$) would be identified by these GWASs (157), because the contribution to variance per segregating, strongly selected locus has been reduced by population growth after the out-of-Africa bottleneck (see Section 3.2). Thus, their estimates should be attributed only to moderately selected loci, and the heritability that they do not account for could be due to loci under stronger and/or weaker selection. Interestingly, the same results suggest that the reliance on genotyping rather than resequencing had little effect on the loci identified, because moderately selected loci that exceed the variance threshold imposed by current sample sizes also exceed the MAF threshold imposed by genotyping.

**3.3.2. Potential extensions.** The first steps toward evolutionary inferences based on association studies highlight some outstanding challenges. Assuming that heritable variation in complex traits is predominantly shaped by mutation–selection balance, the key evolutionary parameters include the rate of mutations affecting the trait (i.e., its target size), the distribution of selection coefficients of these mutations, and the extent of coupling between selection coefficients and their effects on the focal trait. If genetic variation in most complex traits affects many selected traits, then this coupling may follow some universal form, with few parameters, akin to the effective number of traits in the study by Simons et al. (157). Given this relationship and the distribution of selection coefficients, the mutation rate should be easy to infer, e.g., from the number of loci that are GWS. Despite recent progress in fine mapping (148) and in accounting for LD in GWASs more generally (e.g., 19), uncertainty about which loci are causal—and hence about their numbers, MAFs, and effect sizes—remains a major limitation. Setting this problem aside, the main

challenge will be to infer the distribution of selection coefficients; doing so would likely require relying on richer summaries from GWASs than those used to date. One idea would be to consider the joint distribution of frequencies and effect sizes of GWS associations (e.g., rather than only their distribution of variances), also accounting for demographic effects. As we noted, however, current GWS loci likely reflect a fairly narrow range of selection effects; further progress therefore depends on moving beyond these GWS associations.

Such a step might converge with the development of methods for estimating SNP heritability. As noted, such methods have already been instrumental in learning about missing heritability (e.g., 196) and polygenicity (e.g., 102, 155). Yet most approaches rely on strong assumptions about genetic architecture and are sensitive to varying these assumptions (70, 165). Notably, earlier work assumed that the contributions of SNPs to genetic variance are normally distributed and do not depend on their frequency, i.e., that $E(a^2|x) \propto [x(1-x)]^{-1}$. More recent work assumed the more flexible $\alpha$ model, in which $E(a^2|x) \propto [x(1-x)]^{\alpha}$ for some $\alpha$, which is estimated from data. Estimates of $\alpha$ are generally negative, suggesting, as we would expect, that mutations with larger effect sizes are more strongly selected against (e.g., 151, 165). Nonetheless, there is little reason to think that selection produces genetic architectures that are well approximated by the $\alpha$ model. A principled approach would be to parameterize the relationship between frequency and effect size in terms of the distribution of selection coefficients and the relationship between selection coefficients and effect sizes, and to directly estimate these parameters. If feasible, such an approach could be highly informative about selection and produce more reliable heritability estimates. An alternative approach may be to build on the recent method of Pasaniuc and colleagues (70), who relied on the LD matrix among pairs of SNPs to estimate SNP heritability without making strong assumptions about genetic architecture. Extensions of this method may allow for the distribution of heritable variance (or even the joint distribution of frequencies and effect sizes) to be estimated across SNPs genome-wide, which in turn could be used to estimate key evolutionary parameters.

Such evolutionary inferences can inform answers to question of practical import, notably about the merits of different study designs. We already described how they could be used to predict the gains from increased study sizes and the utility of moving from genotyping to resequencing. Similar reasoning can also help to quantify the benefits of increasing the size of imputation reference panels in genotyping-based studies. Additionally, knowing more about evolutionary parameters could inform phenotypic prediction, e.g., by providing bounds on how precise we can expect it to be both within and across ancestries (under specific assumptions).

More fundamentally, these inferences will help to elucidate the evolutionary processes that underlie phenotypic variation in complex traits. If the first decade of GWASs is any indication, we should expect increases in the scale and quality of studies in the near future to greatly increase our knowledge about the genetic architecture of a wide variety of traits. Applying evolutionary inferences to these data will teach us about the mutational target sizes and about direct and pleiotropic selection shaping genetic variation in a wide variety of traits. Testing the fit of alternative models will also enlighten us about the effects of other factors, such as how recent changes in environment may have impacted the genetic architecture (38). In turn, comparing inferences for different traits may help explain differences in their genetic architecture and resolve long-standing puzzles about phenotypic variation, such as the greater heritability in morphological compared with life history traits (178).

## 4. POLYGENIC ADAPTATION

Given that many selected traits are highly polygenic (see Sections 2.1 and 2.4), the adaptive response to changing selective pressures must often involve shifts in such traits, accomplished

through changes to allele frequencies at the many segregating loci that affect them. We would therefore expect polygenic adaptation in complex traits to be ubiquitous. This view traces back to the dawn of population and quantitative genetics (47, 192) and has been supported by more than a century of studies of the response to artificial selection on many traits in plants and animals, and more recently by GWASs revealing the highly polygenic basis of variation in many traits in humans (see Section 2.1).

The quest for the genetic basis of adaptation in humans has also lent support to this view. When genome-wide polymorphism data sets first became available, this search was largely predicated on the monogenic model of a hard selective sweep (78, 114), in which adaptation proceeds by the fixation of a new or initially rare beneficial mutation of large effect (e.g., 120, 146, 179). Subsequent analyses echoed studies of artificial selection in other species, indicating that hard sweeps were rare, at least over the past ∼500,000 years of human evolution (33, 63). Yet it seems plausible that humans adapted in myriad ways during this time period, and they definitely experienced substantial changes in selection pressures, notably during more recent expansions across the globe. These considerations, bolstered by the aforementioned evidence, have largely shifted the search for the genetic basis of human adaptation toward polygenic adaptation (134, 135).

## 4.1. Models of Polygenic Adaptation

Theoretical work on polygenic adaptation has focused primarily on two scenarios. The first is motivated by the response to sustained artificial selection, which is modeled as either truncation selection (142) or stabilizing selection, with the optimal phenotype moving at a constant rate in a given direction (e.g., 22, 29, 89, 90, 112, 113). In natural populations, however, quantitative traits are unlikely to be subject to long-term continuous change in one direction. Instead, they are likely subject to long-term stabilizing selection, with intermittent shifts of the optimum in different directions (see Section 2.4). Other studies therefore assume a scenario in which a sudden change of environment induces an instantaneous shift in the optimum of a trait under stabilizing selection (37, 60, 75, 93, 169). While one can envision more elaborate scenarios—e.g., in which the optimum and/or strength of stabilizing selection vary frequently—this simple scenario provides a sensible starting point for thinking about polygenic adaptation in nature, and is our focus here. We further assume that the effective population size is appreciable (unlike in many cases of artificial selection) and set aside complications arising from LD and pleiotropy.

Lande (93) considered the phenotypic response to selection under this scenario in the infinitesimal limit, in which genetic variation arises from infinitely many segregating loci with infinitesimal effect sizes. In this limit, the new optimum is approached at a rate that is proportional to both the additive genetic variance in the trait, $V_A$, and the strength of stabilizing selection, $V_S^{-1}$. The distance from the new optimum $t$ generations after an instantaneous shift of $\Delta$ at time 0 is thus $D(t) = \Delta \cdot \exp(-V_A t/V_S)$. As an illustration, with $V_A/V_S \approx 1/20$, half the distance to the new optimum is traversed in ∼14 generations. Importantly, this is much faster than when adaptation arises from one or a few beneficial alleles, even if the beneficial alleles were segregating prior to the onset of selection—let alone if new mutations were required (see, e.g., 33). Polygenic adaptation can be so rapid because it requires only tiny changes to allele frequencies at the numerous loci contributing to genetic variation. When the number of segregating loci and their effect sizes are finite, the phenotypic response to selection can be more complicated than in Lande's approximation, because it involves changes to higher moments of the phenotypic distribution (21, 60, 69, 175). Nonetheless, so long as the trait is highly polygenic, it remains the case that small increases in the frequencies of numerous segregating alleles that affect the trait in the same direction allow populations to rapidly adapt to large shifts in optimum, even of many phenotypic standard deviations.

The genetic basis of the adaptive response is largely shaped by the genetic architecture of the trait prior to the shift in optimum (37, 60, 75). To see why, consider the expected change in frequency per generation of an allele with effect size $a$ and frequency $x$, at various times after the shift (21, 37, 60, 75, 194), which is given by

$$E(\Delta x) \cong \frac{1}{2V_S} D(t) a x(1-x) - \frac{1}{4V_S} a^2 x(1-x)\left(\frac{1}{2} - x\right) \qquad 3.$$

[assuming that $|a| \ll \sqrt{V_S}$ and $\Delta < \sqrt{V_S}$ (74)]. The first term on the right-hand side corresponds to directional selection: It pushes alleles with effects that are aligned with the shift to higher frequencies, but its strength weakens as the distance to the new optimum, $D$, decreases. The second term corresponds to stabilizing selection: It acts to reduce the frequency of minor alleles regardless of the direction of their effect, and dominates when $D$ is small. In particular, it shapes the genetic architecture at mutation–selection–drift balance, prior to the shift in optimum (when $D = 0$; see Equation 2). As reviewed in Section 3.2, the expected contribution to variance is then greatest for sites with intermediate and large effects, where it is approximately constant at $E(2a^2x(1-x)) \cong v_S$, implying that for such sites, on average, $x(1-x)$ declines roughly as $1/a^2$.

Immediately after the shift in optimum, the change in allele frequency due to directional selection is proportional to $ax(1-x)$ and thus is greater for alleles with intermediate effect sizes than for alleles with large effects (60). In turn, the contribution of an allele to phenotypic change is proportional to $a^2x(1-x)$ and is therefore fairly insensitive to its effect sizes (so long as it is sufficiently large) (60). Shortly afterward, once the population mean has largely caught up with the new optimum (and $D$ is small), stabilizing selection dominates the allelic dynamics. Because large-effect alleles started at low frequencies and are not likely to have neared a frequency of 1/2 by this time, they are highly unlikely to further increase in frequency, let alone fix, with stabilizing selection now acting against them. Alleles with intermediate effects that are aligned with the shift reach higher frequencies by this time and are therefore more likely to increase in frequency at this second stage, eventually leading to an excess fixation of intermediate-effect, aligned alleles. As a result, and counterintuitively, over longer timescales, the contribution of intermediate-effect alleles to phenotypic change supplants the contribution of alleles with large effect (60). Eventually, the contribution of all transient frequency changes is replaced by fixations at a small subset of loci, and the equilibrium architecture is restored around the new optimum.

Such considerations clarify that the response to changes in selection pressures is governed by the initial genetic architecture. However, we have yet to understand precisely how the change in the phenotypic distribution across individuals depends on genetic architecture and to better characterize the allelic trajectories that underlie the response. The trajectories are particularly important because they determine how polygenic adaptation will affect patterns of linked, neutral genetic variation (8) and how we might go about identifying and interpreting population genetic footprints of polygenic adaptation. We would also like to understand how the response to selection on a quantitative trait is affected by pleiotropy, LD among selected loci, and demographic parameters, notably the effective population size.

Nonetheless, our current understanding provides a basis for a few useful, educated guesses. One is that polygenic adaptation likely has minimal effects on the genetic architecture of a trait, because at most loci it causes only tiny changes to allele frequencies and thus only weakly perturbs the distribution of allele frequencies (although we cannot rule out larger perturbations if, e.g., the environment fluctuates on particular timescales). If this reasoning is correct, then even in the presence of polygenic adaptation, the genetic architecture of complex traits will be shaped predominantly by long-term stabilizing (and sometimes directional) selection and will be well

approximated by existing models (Section 3.2). Another plausible guess is that polygenic adaptation arises predominantly from standing variation and primarily from variants that were segregating at relatively high MAFs, which would suggest modest effects on levels of neutral genetic variation at linked sites. Lastly, if polygenic adaptation causes, on average, only tiny changes to allele frequencies per locus, its effect at any individual locus will easily be masked by the effects of genetic drift and/or selection on other traits affected by the locus or by those in LD with it. Therefore, identifying polygenic adaptation is likely to require pooling the evidence for changes in frequencies across many loci affecting a given trait.

## 4.2. Identifying Polygenic Adaptation in Humans

All the methods recently proposed for identifying polygenic adaptation in humans are based on combining signals of changes in allele frequency across many loci that affect a given trait and testing whether these changes tend to affect the trait in a given direction. GWASs facilitate such tests by identifying many loci that are associated with variation in a given trait and estimating allelic effects at these loci. The signal of coordinated frequency changes is usually tested against an empirical null model, generated by matching key attributes of the test set to random sets of putatively neutral sites.

In theory, a statistically significant result provides strong evidence that directional selection acted on genetic variation affecting the trait, due to selection on that trait or on genetically correlated traits. Importantly, such a finding could reflect a corresponding change in the mean trait value in response to a shift in optimal phenotype and/or a response to a countervailing shift in the environmental contribution to the phenotype (e.g., due to a change of diet) without a change in trait value.

In practice, the reliance on subtle signals aggregated over many loci identified in GWASs renders tests extremely sensitive to systematic biases (10, 13, 14, 121, 163). As an illustration, imagine a GWAS for height based on a sample taken from two genetically differentiated populations, where the individuals sampled from one are, on average, taller than those from the other, due to environmental rather than genetic differences between them. Without appropriate controls for population structure, the effect-size estimate of a given SNP would reflect a sum of a true genetic effect, arising from the causal loci that it tags, and an environmental contribution, which arises because the allele-frequency difference between populations at the SNP absorbs some of the environmental difference in heights among the individuals sampled from them (144). Consequently, alleles with higher frequencies in the tall population will tend to be identified as increasing height, and SNPs that are highly differentiated between populations will tend to have larger effect-size estimates and smaller $p$ values.

Both of the existing approaches for identifying polygenic adaptation are sensitive to the problems highlighted by our example. The first set of methods relies on frequency differences of trait-increasing alleles among extant populations (13, 144, 173). In our example, higher frequencies of height-increasing alleles would be found in the taller population, and the frequency differences would be greater at SNPs with larger estimated effect sizes or lower $p$ values. The second set instead leverages genealogical footprints of past increases in the frequency of trait-increasing (or trait-decreasing) alleles in a single population (39, 45, 166). In our example, there would be evidence for past increases in the frequency of height-increasing alleles in the tall population, and again, the evidence would be stronger for SNPs with larger estimated effect size or lower $p$ values. While all current GWASs employ controls for population structure, it is currently not known how well these correct for biases in tests of polygenic adaptation: They may obscure true signals or create artifactual ones (or both). Recent work suggests that much of the existing evidence for polygenic adaptation is driven by subtle biases in GWAS estimates (see below). Nonetheless,

assuming that systematic biases in GWASs will be overcome, the methods that have been used are likely to be valuable; we therefore briefly review them.

As noted, the first approach relies on allele-frequency differences among extant populations [or, by extension, in archaic ones (111)]. In the first such study in humans, Turchin et al. (173) considered allele frequencies at SNPs significantly associated with a trait in two populations. They applied a sign test to assess whether the number of SNPs in which trait-increasing alleles had a higher frequency in one population was significantly different from that expected by chance (also see 49, 123). They also examined whether the average frequency of trait-increasing alleles in one population was significantly greater than in the other, compared with the expectation under neutrality. To that end, they constructed an empirical null distribution using appropriately matched sets of putatively neutral SNPs (i.e., for average frequency of each SNP in the test set). Berg & Coop (13) generalized this approach to more than two populations. Rather than a simple average over frequencies, they considered a trait's mean polygenic score in each population, $\sum_i \hat{a}_i \hat{x}_i$, in which the estimated frequency of an allele at SNP $i$, $\hat{x}_i$, is weighted by its estimated effect, $\hat{a}_i$. Polygenic scores should predict the genetic contribution to a trait more accurately than the number of trait-increasing alleles (see Section 2.2), suggesting that, in principle, they should provide greater power. Berg & Coop (13) tested for overdispersion and outliers of mean polygenic scores among populations relative to empirical null distributions constructed using sets of putatively neutral SNPs. Importantly, in addition to matching test and control SNPs for average frequency, they also accounted for the effects of shared ancestry among populations under neutrality (also see 124, 144).

The second set of methods relies on genealogical signatures of past increases in the frequency of trait-increasing (or trait-decreasing) alleles in a single population. Field et al. (45) introduced the first method in this vein, relying on large genomic samples from a single population. They reasoned that a recent and rapid change in allele frequency driven by selection would result in shorter terminal branches in the genealogy of the favored allele, and therefore that the haplotypes flanking the beneficial alleles should carry fewer mutations that are singletons in the sample (for related ideas used to study purifying selection, see, e.g., 82). Focusing on a biallelic SNP, they relied on the distribution of distances to the nearest singleton across individuals carrying each genotype in order to estimate the (log) ratio of the mean tip-branch lengths corresponding to the two alleles; they then standardized these estimates within bins of derived allele frequencies to define the singleton density score (SDS). Large deviations of SDS from an empirically constructed null distribution should be well powered to detect the action of selection (notably since SDS utilizes much more information than extant frequencies), with larger sample sizes affording sensitivity to more recent selection (because of shorter tip lengths under the neutral null model). For complex traits, they defined a signed version of SDS (tSDS), such that selection favoring trait-increasing alleles corresponds to positive values. Given this setup, if selection acted to increase the trait value, the tSDS of causal SNPs and of SNPs that tag them will tend to be positive, more so for SNPs with lower GWAS $p$ values.

Two recent methods take the same general approach but rely on explicit inferences of the genealogies of SNPs associated with a trait rather than summaries of tip-branch lengths. This approach has great potential, because the set of underlying genealogies provides complete information about the evolutionary history of a sample. Notably, Edge & Coop (39) relied on estimated genealogical trees to approximate the time course of allele frequencies at SNPs associated with a trait and used them to approximate the time course of the polygenic score. They tested for polygenic adaptation in a given time frame by comparing the change in polygenic score with an empirically constructed null distribution. In principle, the test should be well powered farther back in time than one based on SDS, with power increasing with sample size. Yet the reliability of the estimated time course inevitably degrades substantially farther back in time, because the number

of lineages remaining in the genealogy of the sample decreases, and thus so does power. In practice, the performance of this method strongly depends on the quality of the genealogical inference.

Most recently, Speidel et al. (166) introduced a promising and computationally efficient method for reconstructing genealogical trees in large samples, alongside a new test for polygenic adaptation. Given the estimated tree for a SNP, they considered the number of lineages at the time of the most recent common ancestor of the focal allele and tested whether the branching rate that led to the current frequency of the allele is significantly greater than that of the other lineages present when the allele first arose. Significance levels were derived based on the symmetry of branching rates expected under neutrality and thus depend only on the topology of the inferred tree. For selection on complex traits, they tested whether derived trait-increasing (or trait-decreasing) alleles at GWS SNPs showed stronger evidence of directional selection relative to randomly sampled control alleles of the same frequency. The reliance on topology lends robustness to errors in other features of the inferred trees, at the cost of averaging the signal of adaptation over time. Speidel et al. (166) further argued that relying only on the direction of effects of alleles at GWS SNPs, which are much less sensitive to biases in GWASs (14, 163), lends some robustness to these biases, although their test could still be affected by biases in identifying the set of GWS SNPs.

More generally, all existing methods are affected by systematic biases in GWASs, such as those illustrated by the height example described above. Notably, two recent studies indicate that the reported evidence for polygenic adaptation of height in Europe, arguably the clearest example until now, was largely if not entirely driven by such biases (14, 163). Specifically, the signatures of adaptation that were based on the Genetic Investigation of Anthropometric Traits (GIANT) meta-analysis GWAS do not replicate or are strongly attenuated based on a GWAS in the UK Biobank, a newer, more homogeneous data set that allows better control for population structure. Moreover, an earlier replication of the signature that relied on a sibling based-GWAS, which is largely immune to effects of population structure, turned out to have resulted from a bug. Additionally, Berg et al. (14) showed that LD score regression (19), a popular method used to control for the effects of population structure in GWASs, is not as robust as had been hoped, notably in the presence of effects of selection at linked loci. Taken together, these analyses strongly suggest that much of the reported evidence for polygenic adaptation in height, and plausibly in other traits, was driven by subtle, systematic biases in GWASs. It remains unclear, however, to what extent more recent results, relying on GWASs in the UK Biobank and more powerful methods (39, 166), are still affected.

**4.2.1. Future directions.** How will we be able to tell? Ultimately, this is a question about the relative magnitude of signals and biases. In that regard, modeling should help us develop quantitative expectations for the magnitude of shifts in allele frequencies and polygenic scores given plausible shifts in, e.g., optimal trait values. Modeling may also provide a handle on the magnitude of biases in tests of polygenic adaptation and allow for better understanding of the trade-offs between signals and biases offered by different choices, e.g., about the use of different subsets of SNPs, or corrections for population structure, or sign tests versus polygenic scores. Other important factors that warrant consideration include the loss of power and possibly the bias caused by relying on effect-size estimates in an extant population to infer selection in the past, notably the past of other populations (e.g., due to gene–environment interaction effects and changes in LD structure), the potential effects of asymmetric power to detect trait-increasing versus trait-decreasing minor alleles in case–control GWASs (154), and the suitability of a neutral null model versus one that assumes stabilizing selection. Nevertheless, modeling will necessarily be limited by our imagination in devising appropriate scenarios.

In that regard and more generally, better data sets and improved methodology will be crucial. Indeed, it was the failure to replicate signature of polygenic adaptation using large and relatively homogeneous UK Biobank samples that highlighted the sensitivity to these biases. Once family-based GWASs, which are largely immune to effects of population structure, become sufficiently large to facilitate well-powered estimates of effect sizes, they will offer a promising way forward (see, e.g., 201). More generally, while the problems posed by biases in GWASs seem quite daunting at present, the pace at which powerful new methods and better data sets are emerging allows for cautious optimism.

Assuming these problems are overcome and polygenic adaptation in multiple traits and populations is identified, what is next? One challenge will be to place adaptation events in a given trait in the context of human evolutionary history, namely within the complex web and chronology of population splits and admixture events and introgression from archaic humans (e.g., 55, 62, 129). Some methodology to address this challenge has already been developed, primarily using extant allele frequencies (15, 39, 111, 136). Another challenge will be to home in on targets of selection—e.g., to infer whether signals of adaptation for multiple traits in a given population reflect selection on each of the traits or rather selection on fewer traits that are genetically correlated with the others. Berg et al. (15) took an important step in this direction by extending the Pearson–Lande–Arnold framework (see Section 2.4) to handle changes in polygenic scores over many generations. A related challenge will be to learn what might have driven past selection events. With this goal in mind, Berg et al. (15) tested for correlations between polygenic scores in extant populations and ecological variables, controlling for the relatedness among populations. Beyond the inherent interest in selection pressures that led to polygenic adaptation, successfully relating inferences of adaptation with their putative drivers would serve to enhance our confidence in such inferences.

## 5. DISCUSSION

Understanding the evolutionary processes that produce heritable variation in complex traits is essential to interpreting the results of GWASs. GWASs measure the distribution of heritability in the genome, a distribution that reflects the outcome of numerous mutational perturbations to the phenotype, each of which is typically too small to measure in experimental assays. Thus, GWASs offers unique insight into biological function. Importantly, however, the relationship between function and heritability is not straightforward: The mutations detected in GWASs are those that have been sifted through the sieve of evolution. Specifically, the relationship between the effect of an allele on a trait and its frequency, which together determine the contribution of a locus to heritability, is shaped in nontrivial ways by selection and demography. Understanding how evolutionary processes shape the distribution of heritability in the genome is therefore essential to making sense of GWASs.

In turn, GWASs may well end up providing the kind of data that evolutionary biologists were missing in order to elucidate the processes that maintain heritable variation in complex traits and that underlie their adaptation. We now think that ubiquitous heritable variation in complex traits is maintained primarily by a balance between mutation and selection. In principle, the key features of this balance can be inferred from the genetic architecture revealed by GWASs of these traits. We also think that polygenic adaptation via complex traits should be ubiquitous and a major mode of adaptation. In principle, we can learn about polygenic adaptation in recent human evolution by combining GWAS data with population genetic analyses. We have tried to articulate concrete challenges and avenues toward achieving these goals throughout the review; doubtless there is much more to say.

Some of the most exciting prospects for progress involve connecting the evolution of complex traits with other aspects of biology. At the molecular end, we could envision relating the processes that shape heritable variation with the cellular processes that translate this variation into phenotypic differences. Notably, Pritchard and colleagues (17, 100) recently proposed simple models relating phenotypic variation in complex traits to allelic effects on gene expression. We also have evolutionary models that relate mutational effects on multiple, selected complex traits with the genetic architecture of a given trait (157). By combining these approaches, we can ask, e.g., how *cis*-acting mutational effects on gene expression translate into contributions to trait heritability. Doing so may help to explain why the heritability of many complex traits is widely distributed across the genome, rather than being concentrated around specific genes and pathways that are important to those traits (e.g., 17, 102, 155). It may also help to interpret GWAS results in terms of function—the initial goal of these efforts. For example, it may turn out that low-frequency, large-effect associations are more indicative of genes that directly affect a trait, whereas common, smaller-effect ones are indicative of more general attributes of gene regulatory networks (17, 100, 122). Thus, we believe that the integration of evolutionary models and GWAS data may also turn out to be key to learning about biology from GWASs.

At the organismal end, we would like to understand how high-dimensional phenotypic variation is encoded in genomes and molded by selection. GWASs in humans reveal that heritable variation in many traits is extremely polygenic (e.g., 17, 102, 155) and highly pleiotropic (e.g., 128), while other evidence suggests that many traits must be held fairly constant for the organism to function. In turn, theory suggests that there are inherent limitations on the number of traits that can be independently selected (7). These considerations raise theoretical questions about how selection can shape heritable variation in many traits simultaneously. They also suggest that extensions of approaches by Pearson (125), Lande & Arnold (94), Robertson (143), and Price (131) and their application to large data sets, which include genomic and high-dimensional phenotypic information, may allow us to estimate an effective number of selected traits and understand how selection acts on myriad traits simultaneously and effectively (see, e.g., 147).

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

We thank Peter Andolfatto, Jeremy Berg, Graham Coop, Laura Hayward, Thomas Lenormand, Tobias Lenz, Gil McVean, Jonathan Pritchard, Molly Przeworski, Lluis Quintana-Murci, and Yuval Simons for helpful discussions, and Jeremy Berg, Vince Buffalo, Aravinda Chakravarti, Graham Coop, Arbel Harpak, Laura Hayward, Jonathan Pritchard, Matt Osmond, Yuval Simons, Monty Slatkin, and especially Molly Przeworski for comments on the manuscript. G.S. was funded by National Institutes of Health grant GM115889.

## LITERATURE CITED

1. Abouchar L, Petkova MD, Steinhardt CR, Gregor T. 2014. Fly wing vein patterns have spatial reproducibility of a single cell. *J. R. Soc. Interface* 11:20140443
2. Agarwala V, Flannick J, Sunyaev S, GoT2D Consort., Altshuler D. 2013. Evaluating empirical bounds on complex disease genetic architecture. *Nat. Genet.* 45:1418–27

3. Agrawal AF, Whitlock MC. 2011. Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* 187:553–66
4. Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, et al. 2013. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLOS Genet*. 9:e1003455
5. Backer LC, McNeel SV, Barber T, Kirkpatrick B, Williams C, et al. 2010. Recreational exposure to microcystins during algal blooms in two California lakes. *Toxicon* 55:909–21
6. Barghi N, Tobler R, Nolte V, Jaksic AM, Mallard F, et al. 2019. Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biol*. 17:e3000128
**7. Barton NH. 1990. Pleiotropic models of quantitative variation. *Genetics* 124:773–82**
8. Barton NH. 2000. Genetic hitchhiking. *Philos. Trans. R. Soc. B* 355:1553–62
9. Barton NH, Etheridge AM, Veber A. 2017. The infinitesimal model: definition, derivation, and implications. *Theor. Popul. Biol.* 118:50–73
10. Barton NH, Hermisson J, Nordborg M. 2019. Population genetics: Why structure matters. *eLife* 8:e45380
11. Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. *Nat. Rev. Genet.* 3:11–21
12. Beauchamp JP. 2016. Genetic evidence for natural selection in humans in the contemporary United States. *PNAS* 113:7774–79
13. Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. *PLOS Genet*. 10:e1004412
**14. Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, et al. 2019. Reduced signal for polygenic adaptation of height in UK Biobank. *eLife* 8:e39725**
15. Berg JJ, Zhang X, Coop G. 2017. Polygenic adaptation has impacted multiple anthropometric traits. bioRxiv 167551. **https://doi.org/10.1101/167551**
16. Bourguet D. 1999. The evolution of dominance. *Heredity* 83:1–4
**17. Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169:1177–86**
18. Bulik-Sullivan BK, Finucane HK, Anttila V, Gusev A, Day FR, et al. 2015. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* 47:1236–41
19. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, et al. 2015. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47:291–95
20. Bulmer MG. 1985. *The Mathematical Theory of Quantitative Genetics*. Oxford, UK: Oxford Univ. Press
21. Burger R. 1991. Moments, cumulants, and polygenic dynamics. *J. Math. Biol.* 30:199–213
22. Burger R. 1999. Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* 153:1055–69
23. Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. 2010. Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature* 467:587–90
24. Caballero A, Tenesa A, Keightley PD. 2015. The nature of genetic variation for complex traits revealed by GWAS and regional heritability mapping analyses. *Genetics* 201:1601–13
25. Castro JPL, Yancoskie MN, Marchini M, Belohlavy S, Hiramatsu L, et al. 2019. An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. *eLife* 8:e42014
26. Chakravarti A. 1999. Population genetics—making sense out of sequence. *Nat. Genet.* 21:56–60
27. Chakravarti A, Turner TN. 2016. Revealing rate-limiting steps in complex disease biology: the crucial importance of studying rare, extreme-phenotype families. *BioEssays* 38:578–86
28. Charlesworth B. 1979. Evidence against Fisher's theory of dominance. *Nature* 278:848–49
29. Charlesworth B. 1993. Directional selection and the evolution of sex and recombination. *Genet. Res.* 61:205–24
30. Charlesworth B. 2015. Causes of natural variation in fitness: evidence from studies of *Drosophila* populations. *PNAS* 112:1662–69
31. Charlesworth B, Charlesworth D. 2010. *Elements of Evolutionary Genetics*. Greenwood Village, CO: Roberts & Co.
32. Charmantier A, Garant D, Kruuk LEB. 2014. *Quantitative Genetics in the Wild*. Oxford, UK: Oxford Univ. Press

**7. Reviews models where the focal trait is neutral, so that variation is shaped by selection on the pleiotropic side effects of alleles.**

**14. Shows that evidence for selection on polygenic traits (e.g., Refs. 13 and 45) mostly vanishes when using the more homogeneous UK Biobank data (see also Ref. 163).**

**17. Argues that complex traits are influenced by essentially all expressed loci, via weak interactions that propagate through the regulatory network.**

33. Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, et al. 2009. The role of geography in human adaptation. *PLOS Genet*. 5:e1000500
34. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, et al. 2011. Pervasive sharing of genetic effects in autoimmune disease. *PLOS Genet*. 7:e1002254
35. Coventry A, Bull-Otterson LM, Liu X, Clark AG, Maxwell TJ, et al. 2010. Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat. Commun.* 1:131
36. Crawford NG, Kelly DE, Hansen MEB, Beltrame MH, Fan S, et al. 2017. Loci associated with skin pigmentation identified in African populations. *Science* 358:867–87
37. de Vladar HP, Barton NH. 2014. Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics* 197:749–67
38. Di Rienzo A, Hudson RR. 2005. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet*. 21:596–601
39. Edge MD, Coop G. 2019. Reconstructing the history of polygenic scores using coalescent trees. *Genetics* 211:235–62
40. Ejsmond MJ, Radwan J. 2015. Red Queen processes drive positive selection on major histocompatibility complex (MHC) genes. *PLOS Comput. Biol.* 11:e1004627
41. Ewens WJ. 2004. *Mathematical Population Genetics*. New York: Springer
42. Eyre-Walker A. 2010. Evolution in health and medicine Sackler colloquium: genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *PNAS* 107(Suppl. 1):1752–56
43. Falconer DS, Mackay TFC. 1996. *Introduction to Quantitative Genetics*. Harlow, UK: Longman
44. Felsenstein J. 1988. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19:445–71
45. Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, et al. 2016. Detection of human adaptation during the past 2000 years. *Science* 354:760–64
**46. Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52:399–433**
47. Fisher RA. 1930. *The Genetical Theory of Natural Selection*. Oxford, UK: Clarendon
48. Fraïsse C, Gunnarsson PA, Roze D, Bierne N, Welch JJ. 2016. The genetics of speciation: insights from Fisher's geometric model. *Evolution* 70:1450–64
49. Fraser HB. 2013. Gene expression drives local adaptation in humans. *Genome Res.* 23:1089–96
**50. Galton F. 1877. Typical laws of heredity. *Nature* 15:492–95, 512–14, 532–33**
51. Gillespie JH. 1991. *The Causes of Molecular Evolution*. Oxford, UK: Oxford Univ. Press
52. Gillespie JH. 2004. *Population Genetics: A Concise Guide*. Baltimore, MD: Johns Hopkins Univ. Press
53. Gingerich PD. 1983. Rates of evolution: effects of time and temporal scaling. *Science* 222:159–61
54. Grant PR, Grant BR. 2014. *40 Years of Evolution: Darwin's Finches on Daphne Major Island*. Princeton, NJ: Princeton Univ. Press
55. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–22
56. Greenberg R, Crow JF. 1960. A comparison of the effect of lethal and detrimental chromosomes from Drosophila populations. *Genetics* 45:1153–68
**57. Haldane JBS. 1927. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proc. Camb. Philos. Soc.* 23:838–44**
58. Halligan DL, Keightley PD. 2009. Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* 40:151–72
59. Havenstein G, Ferket P, Qureshi M. 2003. Growth, livability, and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poult. Sci.* 82:1500–8
60. Hayward L, Sella G. 2019. Polygenic adaptation after a sudden change in environment. Manuscript in preparation
61. Hedrick PW, Whittam TS, Parham P. 1991. Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. *PNAS* 88:5897–901
62. Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, et al. 2014. A genetic atlas of human admixture history. *Science* 343:747–51

**46. Reconciled Mendelian genetics with biometry and introduced analysis of trait variance into its components (environmental, additive genetic, dominance, etc.).**

**50. First statement of what became known as the infinitesimal model, where the breeding values of offspring are normally distributed around the mean of the parents.**

**57. Showed how deleterious alleles can be maintained by a balance between mutation and selection.**

63. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, et al. 2011. Classic selective sweeps were rare in recent human evolution. *Science* 331:920–24

**64. Showed that mutation contributes substantially to genetic variance, consistent with a highly polygenic architecture.**

**64. Hill WG. 1982. Rates of change in quantitative traits from fixation of new mutations. *PNAS* 79:142–45**

65. Hill WG. 2014. Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. *Genetics* 196:1–16
66. Hill WG. 2016. Is continued genetic improvement of livestock sustainable? *Genetics* 202:877–81
67. Hill WG, Goddard ME, Visscher PM. 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLOS Genet*. 4:e1000008
68. Hill WG, Kirkpatrick M. 2010. What animal breeding has taught us about evolution. *Annu. Rev. Ecol. Evol. Syst.* 41:1–19
69. Hill WG, Rasbash J. 1986. Models of long term artificial selection in finite population. *Genet. Res.* 48:41–50
70. Hou K, Burch KS, Majumdar A, Shi A, Mancuso N, et al. 2019. Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture. bioRxiv 526855. **https://doi.org/10.1101/526855**
71. Houle D, Bolstad GH, van der Linde K, Hansen TF. 2017. Mutation predicts 40 million years of fly wing evolution. *Nature* 548:447–50
72. Houle D, Morikawa B, Lynch M. 1996. Comparing mutational variabilities. *Genetics* 143:1467–83
73. Huang W, Richards S, Carbone MA, Zhu D, Anholt RRH, et al. 2012. Epistasis dominates the genetic architecture of Drosophila quantitative traits. *PNAS* 109:15553–59
74. Huguet G, Ey E, Bourgeron T. 2013. The genetic landscapes of autism spectrum disorders. *Annu. Rev. Genom. Hum. Genet.* 14:191–213
75. Jain K, Stephan W. 2017. Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics* 206:389–406
76. Johnson T, Barton NH. 2005. Theoretical models of selection and mutation on quantitative traits. *Philos. Trans. R. Soc. B* 360:1411–25
77. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61
78. Kaplan NL, Hudson RR, Langley CH. 1989. The "hitchhiking effect" revisited. *Genetics* 123:887–99
79. Keightley PD, Hill WG. 1988. Quantitative genetic variability maintained by mutation-stabilizing selection balance in finite populations. *Genet. Res.* 52:33–43
80. Keightley PD, Hill WG. 1990. Variation maintained in quantitative traits with mutation–selection balance: pleiotropic side-effects on fitness traits. *Proc. R. Soc. B* 242:19900110
81. Kelly JK, Hughes KA. 2019. An examination of the evolve-and-resequence method using *Drosophila simulani*. *Genetics* 211:943–61
82. Kiezun A, Pulit SL, Francioli LC, van Dijk F, Swertz M, et al. 2013. Deleterious alleles in the human genome are on average younger than neutral alleles of the same frequency. *PLOS Genet*. 9:e1003301
83. Kimura M. 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. *PNAS* 54:731–36

**84. Reviews the many estimates of selection in nature made using the Pearson–Lande–Arnold method.**

**84. Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, et al. 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* 157:245–61**

85. Kinnison MT, Hendry AP. 2001. The pace of modern life II: from rates of contemporary microevolution to pattern and process. *Genetica* 112–113:145–64
86. Kirkpatrick M, Barton NH. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* 173:419–34
87. Kondrashov AS, Turelli M. 1992. Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. *Genetics* 132:603–18
88. Kong A, Frigge ML, Thorleifsson G, Stefansson H, Young AI, et al. 2017. Selection against variants in the genome associated with educational attainment. *PNAS* 114:E727–32
89. Kopp M, Hermisson J. 2007. Adaptation of a quantitative trait to a moving optimum. *Genetics* 176:715–19

90. Kopp M, Hermisson J. 2009. The genetic basis of phenotypic adaptation I: fixation of beneficial mutations in the moving optimum model. *Genetics* 182:233–49
91. Kruuk EB, Slate J, Pemberton JM, Brotherstone S, Guinness F, Clutton-Brock T. 2002. Antler size in red deer: heritability and selection but no evolution. *Evolution* 56:1683–95
92. Lande R. 1975. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* 26:221–35
93. Lande R. 1976. Natural-selection and random genetic drift in phenotypic evolution. *Evolution* 30:314–34
**94. Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–26**
95. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. 2012. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 28:2540–42
96. Lello L, Avery SG, Tellier L, Vazquez AI, de Los Campos G, Hsu SDH. 2018. Accurate genomic prediction of human height. *Genetics* 210:477–97
97. Lenz TL, Spirin V, Jordan DM, Sunyaev SR. 2016. Excess of deleterious mutations around HLA genes reveals evolutionary cost of balancing selection. *Mol. Biol. Evol.* 33:2555–64
98. Lewontin RC. 2000. *The Triple Helix: Gene, Organism, and Environment*. Cambridge, MA: Harvard Univ. Press
99. Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO. 2010. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *PNAS* 107:21199–204
100. Liu X, Li YI, Pritchard JK. 2019. *Trans* effects on gene expression can drive omnigenic inheritance. *Cell* 177:1022–34.e6
101. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, et al. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518:197–206
102. Loh PR, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, et al. 2015. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* 47:1385–92
103. Lohmueller KE. 2014. The impact of population demography and selection on the genetic architecture of complex traits. *PLOS Genet.* 10:e1004379
104. Lush JL. 1937. *Animal Breeding Plans*. Ames, IA: Coll. Press
105. Lynch M, Hill WG. 1986. Phenotypic evolution by neutral mutation. *Evolution* 40:915–35
106. Lynch M, Walsh B. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer
107. Maki-Tanila A, Hill WG. 2014. Influence of gene interaction on complex trait variation with multilocus models. *Genetics* 198:355–67
108. Mancuso N, Rohland N, Rand KA, Tandon A, Allen A, et al. 2016. The contribution of rare variation to prostate cancer heritability. *Nat. Genet.* 48:30–35
109. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–53
110. 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
111. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528:499–503
112. Matuszewski S, Hermisson J, Kopp M. 2014. Fisher's geometric model with a moving optimum. *Evolution* 68:2571–88
113. Matuszewski S, Hermisson J, Kopp M. 2015. Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics* 200:1255–74
114. Maynard Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* 23:23–35
115. McGuigan K, Aguirre JD, Blows MW. 2015. Simultaneous estimation of additive and mutational genetic variance in an outbred population of *Drosophila serrata*. *Genetics* 201:1239–51
116. Mitchell KJ. 2012. What is complex about complex disorders? *Genome Biol.* 13:237
117. Moose SP, Dudley JW, Rocheford TR. 2004. Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends Plant Sci.* 9:358–64

**94. Revived and extended Pearson's methods for estimating selection on multiple traits.**

118. Mostafavi H, Berisa T, Day FR, Perry JRB, Przeworski M, Pickrell JK. 2017. Identifying genetic variants that affect viability in large cohorts. *PLOS Biol.* 15:e2002458
119. Mousseau TA, Roff DA. 1987. Natural selection and the heritability of fitness components. *Heredity* 59(Part 2):181–97
120. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. *Genome Res.* 15:1566–75
121. Novembre J, Barton NH. 2018. Tread lightly interpreting polygenic tests of selection. *Genetics* 208:1351–55
122. O'Connor LJ, Schoech AP, Hormozdiari F, Gazal S, Patterson N, Price AL. 2018. Polygenicity of complex traits is explained by negative selection. bioRxiv 420497. **https://doi.org/10.1101/420497**
123. Orr HA. 1998. Testing natural selection versus genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* 149:2099–104
124. Ovaskainen O, Karhunen M, Zheng C, Arias JM, Merila J. 2011. A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics* 189:621–32
125. Pearson K. 1903. I. Mathematical contributions to the theory of evolution.—XI. On the influence of natural selection on the variability and correlation of organs. *Philos. Trans. R. Soc. A* 200:1–66
126. Perry JRB, Day F, Elks CE, Sulem P, Thompson DJ, et al. 2014. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* 514:92–97
127. Phadnis N, Fry JD. 2005. Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics* 171:385–92
128. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. 2016. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* 48:709–17
129. Pickrell JK, Reich D. 2014. Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet.* 30:377–89
130. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, et al. 2015. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 47:702–9
131. Price GR. 1970. Selection and covariance. *Nature* 227:520–21
132. Price TD, Grant PR, Gibbs HL, Boag PT. 1984. Recurrent patterns of natural selection in a population of Darwin's finches. *Nature* 309:787–89
133. Pritchard JK. 2001. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* 69:124–37
134. Pritchard JK, Di Rienzo A. 2010. Adaptation – not by sweeps alone. *Nat. Rev. Genet.* 11:665–67
135. Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* 20:R208–15
136. Racimo F, Berg JJ, Pickrell JK. 2018. Detecting polygenic adaptation in admixture graphs. *Genetics* 208:1565–84
137. Rands CM, Meader S, Ponting CP, Lunter G. 2014. 8.2% of the human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLOS Genet.* 10:e1004525
138. Reich DE, Lander ES. 2001. On the allelic spectrum of human disease. *Trends Genet.* 17:502–10
139. Risch N. 1990. Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.* 46:222–28
140. Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* 273:1516–17
141. Robertson A. 1956. The effect of selection against extreme deviants based on deviation or on homozygosis. *J. Genet.* 54:236–48
142. Robertson A. 1960. A theory of limits in artificial selection. *Proc. R. Soc. B* 153:234–49
143. Robertson A. 1966. A mathematical model of the culling process in dairy cattle. *Anim. Sci.* 8:95–108
144. Robinson MR, Hemani G, Medina-Gomez C, Mezzavilla M, Esko T, et al. 2015. Population genetic differentiation of height and body mass index across Europe. *Nat. Genet.* 47:1357–62
145. Roff DA. 1996. The evolution of genetic correlations: an analysis of patterns. *Evolution* 50:1392–403
146. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–37

147. Sanjak JS, Sidorenko J, Robinson MR, Thornton KR, Visscher PM. 2018. Evidence of directional and stabilizing selection in contemporary humans. *PNAS* 115:151–56
148. Schaid DJ, Chen W, Larson NB. 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* 19:491–504
149. Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* 46:919–25
150. Schiffman JS, Ralph PL. 2018. System drift and speciation. bioRxiv 231209. **https://doi.org/10.1101/231209**
151. Schoech AP, Jordan DM, Loh P-R, Gazal S, O'Connor LJ, et al. 2019. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun.* 10:790
152. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, et al. 2012. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* 44:991–1005
153. Segre D, Deluna A, Church GM, Kishony R. 2005. Modular epistasis in yeast metabolism. *Nat. Genet.* 37:77–83
154. Sham PC, Purcell SM. 2014. Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* 15:335–46
155. Shi H, Kichaev G, Pasaniuc B. 2016. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* 99:139–53
156. Simon A, Bierne N, Welch JJ. 2018. Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evol. Lett.* 2:472–98
**157. Simons YB, Bullaughey K, Hudson RR, Sella G. 2018. A population genetic interpretation of GWAS findings for human quantitative traits. *PLOS Biol.* 16:e2002985**
158. Simons YB, Stella G. 2016. The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. *Curr. Opin. Genet. Dev.* 41:150–58
159. Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* 46:220–24
160. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, et al. 2011. Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.* 89:607–18
161. Slatkin M. 2008. Exchangeable models of complex inherited diseases. *Genetics* 179:2253–61
162. Sniegowski PD, Gerrish PJ, Johnson T, Shaver A. 2000. The evolution of mutation rates: separating causes from consequences. *BioEssays* 22:1057–66
**163. Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, et al. 2019. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* 8:e39702**
164. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. 2013. Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* 14:483–95
165. Speed D, Hemani G, Johnson MR, Balding DJ. 2012. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* 91:1011–21
166. Speidel L, Forest M, Shi S, Myers S. 2019. A method for genome-wide genealogy estimation for thousands of samples. bioRxiv 550558. **https://doi.org/10.1101/550558**
167. Spurgin LG, Richardson DS. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc. R. Soc. B* 277:979–88
168. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, et al. 2015. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Med.* 12:e1001779
169. Thornton KR. 2018. Polygenic adaptation to an environmental shift: temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. bioRxiv 505750. **https://doi.org/10.1101/505750**
170. Tobler R, Franssen SU, Kofler R, Orozco-terWengel P, Nolte V, et al. 2014. Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Mol. Biol. Evol.* 31:364–75

**157. Models stabilizing selection on multiple traits to show that the distribution of the contributions of a SNP to trait variance takes a characteristic form.**

**163. Shows that evidence for selection on polygenic traits (e.g., Refs. 13 and 45) mostly vanishes when using the more homogeneous UK Biobank data (see also Ref. 14).**

171. Trowsdale J. 2011. The MHC, disease and selection. *Immunol. Lett.* 137:1–8
172. Trowsdale J, Knight JC. 2013. Major histocompatibility complex genomics and human disease. *Annu. Rev. Genom. Hum. Genet.* 14:301–23
173. Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, et al. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* 44:1015–19
**174. Turelli M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor. Popul. Biol.* 25:138–93**
175. Turelli M, Barton NH. 1990. Dynamics of polygenic characters under selection. *Theor. Popul. Biol.* 38:1–57
176. Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLOS Genet*. 7:e1001336
177. Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90:7–24
178. Visscher PM, Hill WG, Wray NR. 2008. Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.* 9:255–66
179. Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLOS Biol.* 4:e72
180. Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* 12:204–13
181. Wall JD, Przeworski M. 2000. When did the human population size start increasing? *Genetics* 155:1865–74
182. Walsh B, Lynch M. 2018. *Evolution and Selection of Quantitative Traits*. Oxford, UK: Oxford Univ. Press
183. Weber KE. 1990. Selection on wing allometry in *Drosophila melanogaster*. *Genetics* 126:975–89
184. Weber KE. 1996. Large genetic change at small fitness cost in large populations of *Drosophila melanogaster* selected for wind tunnel flight: rethinking fitness surfaces. *Genetics* 144:205–13
185. Weber KE, Diggins LT. 1990. Increased selection response in larger populations. II. Selection for ethanol vapor resistance in *Drosophila melanogaster* at two population sizes. *Genetics* 125:585–97
186. Weber KE, Eisman R, Morey L, Patty A, Sparks J, et al. 1999. An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. *Genetics* 153:773–86
187. Wei WH, Hemani G, Haley CS. 2014. Detecting epistasis in human complex traits. *Nat. Rev. Genet.* 15:722–33
188. Wellcome Trust Case Control Consort. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–78
189. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* 46:1173–86
190. Wray NR, Goddard ME. 2010. Multi-locus models of genetic risk of disease. *Genome Med.* 2:10
191. Wright S. 1929. Fisher's theory of dominance. *Am. Nat.* 63:274–79
192. Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159
193. Wright S. 1931. On the genetics of number of digits of the guinea pig. *Anat. Rec.* 51:115
194. Wright S. 1935. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *J. Genet.* 30:243–56
195. Wright S. 1937. The distribution of gene frequencies in populations. *PNAS* 23:307–20
196. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–69
197. Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *PNAS* 110:E1743–51
198. Yeaman S, Otto SP. 2011. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* 65:2123–29
199. Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* 65:1897–911

**174. Reviews evidence that variation is maintained by a mutation–selection balance and argues (contra Ref. 93) that rare alleles are responsible.**

200. Yoo BH. 1980. Long-term selection for a quantitative character in large replicate populations of *Drosophila melanogaster*: 1. Response to selection. *Genet. Res.* 35:1–17
201. Young AI, Frigge ML, Gudbjartsson DF, Thorleifsson G, Bjornsdottir G, et al. 2018. Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.* 50:1304–10
202. Young AI, Wauthier FL, Donnelly P. 2018. Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nat. Genet.* 50:1608–14
203. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, et al. 2013. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLOS Genet*. 9:e1003520
204. Zhang XS, Hill WG. 2005. Predictions of patterns of response to artificial selection in lines derived from natural populations. *Genetics* 169:411–25
205. Zhu Z, Bakshi A, Vinkhuyzen AA, Hemani G, Lee SH, et al. 2015. Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* 96:377–85
206. Zuk O, Hechter E, Sunyaev SR, Lander ES. 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *PNAS* 109:1193–98

Annual Review of Genomics and Human Genetics

Volume 20, 2019

# Contents

**Errata**

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at http://www.annualreviews.org/errata/genom