Polygenic adaptation: a unifying framework to understand positive selection

Neda Barghi, Joachim Hermisson and Christian Schlötterer

Abstract | Most adaption processes have a polygenic genetic basis, but even with the recent explosive growth of genomic data we are still lacking a unified framework describing the dynamics of selected alleles. Building on recent theoretical and empirical work we introduce the concept of adaptive architecture, which extends the genetic architecture of an adaptive trait by factors influencing its adaptive potential and population genetic principles. Because adaptation can be typically achieved by many different combinations of adaptive alleles (redundancy), we describe how two characteristics — heterogeneity among loci and non-parallelism between replicated populations — are hallmarks for the characterization of polygenic adaptation in evolving populations. We discuss how this unified framework can be applied to natural and experimental populations.

Consider a population experiencing a novel selection pressure on a trait with a polygenic basis after an environmental change. The population responds by a rapid shift of the trait mean towards a new optimum. But how is this phenotypic change genetically encoded? Historically, two highly successful schools of evolutionary research have approached the question from different directions. Whereas quantitative genetics focuses on the phenotype, molecular population genetics relies on genomic signatures at selected and linked loci. Building on disjointed data sets, both schools developed diverging narratives of phenotypic adaptation: whereas quantitative genetics envisions adaptation via subtle allele frequency shifts at many loci, population genetics predicts independent selective sweeps, leading to phenotypic adaptation via single-locus mutations.

Recently, several studies sought to reconcile the two fields[1–4]. Which elements are needed in a joint framework of polygenic adaptation that accounts for a wide range of adaptive scenarios, and for different types of data? Which concepts and summary statistics are effective for characterizing these scenarios in theoretical and empirical work? Although much of the current discussion revolves around the predicted patterns of adaptation (sweeps versus shifts), the construction of a joint framework requires an analysis of the model assumptions that drive these patterns.

The molecular population genetic approach of adaptive evolution follows a reductionist paradigm. It works under the implicit assumption that selection on the phenotype translates into directional selection for single beneficial alleles towards a fixed target frequency[5,6] (usually fixation; FIG. 1a,b). Under this assumption, genotypic adaptation decouples from the phenotype and the adaptive process decomposes into a collection of largely independent events at single loci.

This reduction of phenotypic adaptation to individual loci allowed for a highly developed theory of selection footprints on linked neutral variation. The archetypical 'hard sweep' model[6] assumes constant selection on a new beneficial mutation, which rapidly increases from frequency 0 to 1. Consequently, the population genetic view of adaptation has often been characterized by large frequency changes and rapid fixation of unconditionally beneficial alleles[2,7]. However, although this mode leaves the clearest sweep signature, the modelling framework itself (the sweep model) readily allows for extensions.

For soft sweeps from standing genetic variation[8,9], later-beneficial alleles experience an initial phase of neutral or negative selection. Single-locus selection can also depend on spatial structure or allele frequency[10,11]. In this case, adaptive evolution may not drive beneficial alleles to fixation but lead to patterns of partial sweeps[12].

The general sweep model thus allows for allele frequency changes of any size, including small changes ('partial sweeps from standing genetic variation'). It also allows for weak selection and slow allele frequency changes — in which case, the model predicts that selection will only leave feeble footprints. The limitation of the approach is not the mode or magnitude of selection at single loci, but rather the assumption that concurrent allele frequency changes in the genomic background do not influence the single-locus selection response. In particular, the model implies that adaptation at each locus aims for the same target frequency across computational or experimental replicates. At the end of the adaptive phase, stochastic differences between replicates are only visible in the footprint on linked variation.

This reductionist approach also has important consequences for long-term evolution. If adaptation decomposes into single-locus events, the short-term response to positive selection during the adaptive phase determines the long-term patterns of adaptive divergence between populations or species. For the classical hard sweep model, in particular, adaptation simply proceeds by a series of single-locus substitutions. This concept of an adaptive walk is the basis of several influential approaches to describe the adaptive process over longer timescales[13–15].

Classical quantitative genetics is primarily concerned with phenotypes and does not aim at a detailed description of adaptation at the genotypic level. Indeed, it usually does not refer to genotype frequencies at all. The key insight by Fisher (1918)[16] is that the evidence of gradual phenotypic evolution collected by biometricians (for example, Galton[17]) is fully compatible with Mendelian genetics as long as sufficiently many genes contribute to the trait. In particular, the influential infinitesimal model[18–21] assumes an infinite number of loci as the genetic basis of the trait, each contributing an infinitely

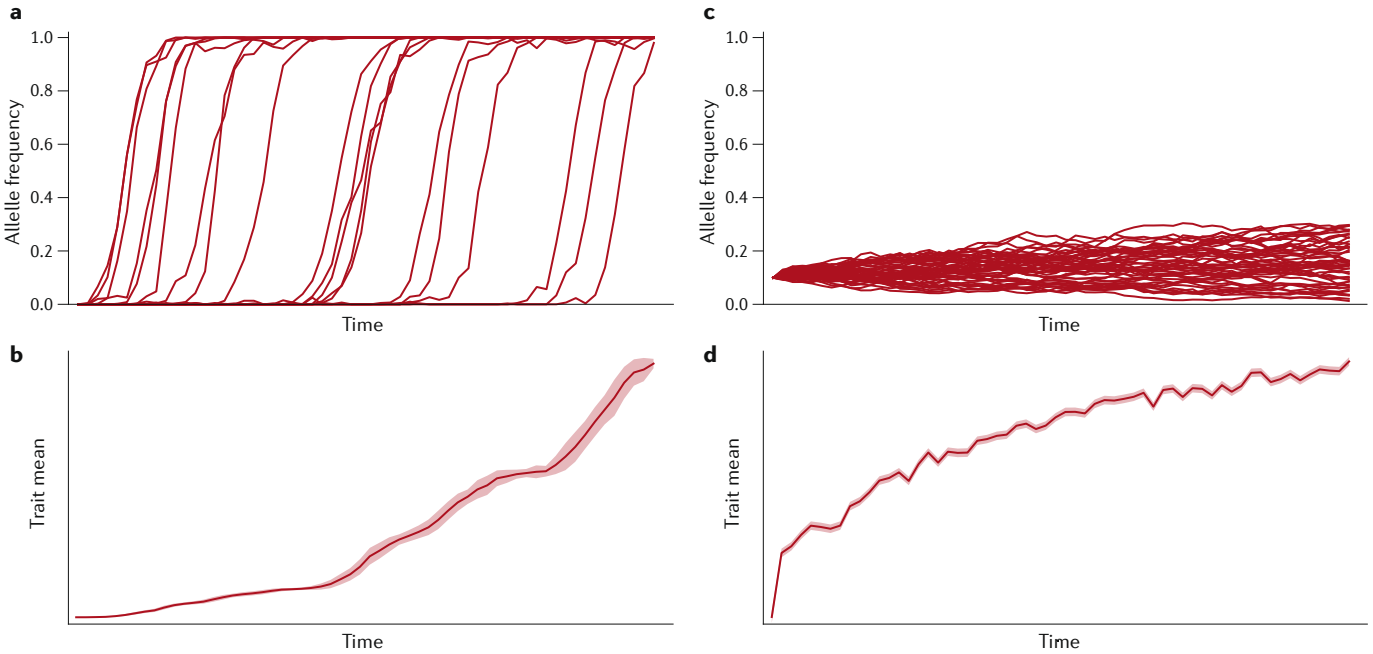


Fig. 1 | **Alternative adaptive paradigms: selective sweeps versus the infinitesimal model. a**,**b** | Selective sweeps. Beneficial mutations arise recurrently in a population and increase in frequency until fixation (part **a**). The consecutive fixation of beneficial alleles increases the fitness of the population (part **b**). **c**,**d** | The infinitesimal model. A large number of contributing alleles experience subtle frequency changes after a shift in the trait optimum (part **c**) that result in changes in the mean trait of the population (part **d**). For simulation parameters, see Supplementary information.

small amount. Consequently, the change in allele frequency due to selection on each single locus becomes vanishingly small[18]. Intuitively, phenotypic adaptation then occurs by subtle frequency shifts at many loci (FIG. 1c,d), leading to a view of genotypic adaptation in quantitative genetics that is as far away from the hard sweep model of molecular population genetics as can possibly be imagined.

However, similar to the sweep model, the quantitative genetic framework is not bound to a particular pattern (such as subtle shifts). Its predictions for phenotypic evolution, such as a smooth change in the trait mean under directional selection with a stable genetic variance, do not require the extreme genetic assumptions of the infinitesimal model. They hold, to a good approximation, for a much larger class of models[19–21] and allow, to some extent, for loci of larger effect and sweep-like changes in the allele frequency[19,22,23]. The crucial assumption is that the selection response is collective and dominated by many small contributions. It is not important whether any single locus contributes to the phenotype or not: there are no fixed target frequencies for alleles at single loci. Indeed, knowledge of the phenotype of an individual provides hardly any information about the underlying genotype[21].

This assumption also remains relevant over longer evolutionary timescales: as shifts in the trait optimum do not lead to fixation of alleles in the infinitesimal model, short-term adaptation cannot determine the patterns of adaptive divergence between populations or species. These long-term patterns, then, rather reflect evolutionary forces (such as genetic drift, migration or purifying selection) that act after the rapid adaptive phase and will typically differ strongly between independently evolving populations.

In this Perspectives article, we start by presenting the empirical evidence for polygenic adaptation and highlight the challenges and limitations of current approaches used. Next, we develop an integrated framework of polygenic adaptation by emphasizing redundancy, the salient feature of polygenic adaptation, which explains parallelism and heterogeneity. After discussing the mathematical foundations of polygenic adaptation, we illustrate how polygenic adaptation could be studied either in the laboratory, by experimental evolution or in natural systems that allow for replication.

Empirical evidence

One primary goal of molecular population genetics in the past decades has been the identification of selective sweeps[5]. Elaborate statistical tests have been developed to identify sweep signatures both for single loci and for genome scans[24,25], but even for some classic sweeps a polygenic adaptation model may fit as well (BOX 1). More recently, with plummeting sequencing costs and the availability of sequence data from huge population samples, the hitherto neglected genetic basis of quantitative traits once again became a topical study object[4]. The insight from many genome-wide association studies (GWAS) that most traits are highly polygenic in combination with the central role of the infinitesimal model in quantitative genetics culminated in the 'omnigenic' model[3]. This model suggests an important contribution of thousands of genes outside core pathways to heritability, thereby providing the abstract infinitesimal framework with a concrete mechanistic basis[3].

Based on the expectation of small effects for individual loci, the anticipated genomic signature of individual markers does not provide a strong enough signal to detect selection. For genome scans this is even more challenging, as the required multiple testing correction further reduces the sensitivity. Hence, new approaches to study the genetics of adaptation through polygenic traits were required.

A common feature of most empirical tests for polygenic adaptation is that the required statistical power is achieved by combining information from a priori defined sets of single-nucleotide polymorphisms (SNPs). Individual tests differ by the method used to extract a selection signal and how the SNP sets are defined.

Box 1 | **Sweep signatures: from oligogenic to polygenic adaptation**

Given that molecular population genetics has been aiming to detect the genomic signatures of selective sweeps, it is not surprising that several very convincing examples have been reported. For some of them, even the genotype–phenotype relationship has been experimentally confirmed. Nevertheless, as discussed in the main text, sweep-like selection signals are also possible with polygenic adaptation. Hence, it is important to question the underlying architectures, even for classic sweep signatures.

Oligogenic adaptation — the outcome of pleiotropic constraints?

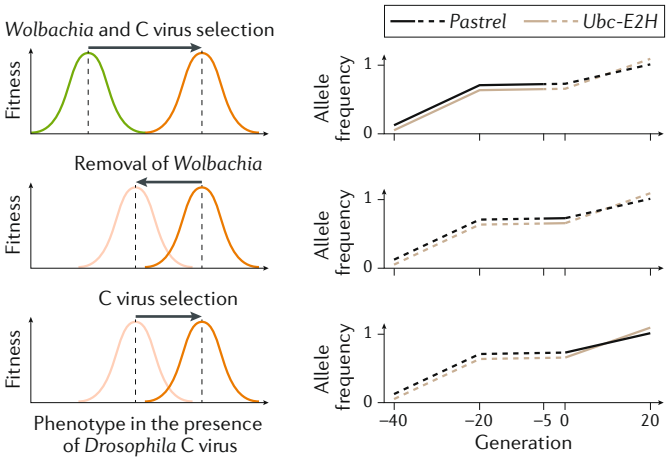
One very prominent example is coat colour in *Peromyscus* mice. In the Nebraskan deer mice, selection for the derived light coat colour left characteristic population signatures at the *Agouti* locus[122]. This was also experimentally confirmed by directly measuring selection in field experiments[123]. The coat colour of the closely related beach mouse has a different genetic basis in populations inhabiting the Atlantic and Gulf coasts. In the Gulf coast populations, *Mc1r* is the major contributor[124], but in Atlantic populations *Mc1r* is not involved in the colour polymorphism. With *Mc1r* and *Agouti* being the key players in many vertebrates[93], genetic redundancy is limited, although other genes in the pigmentation pathway could have been targeted as well. A similar situation is observed in *Arabidopsis thaliana*, where all naturally occurring glabrous plants (lacking trichomes) are caused by mutations in *GL1*, although many other genes also affect trichome development[125,126]. Both examples are a vivid demonstration of the difference between genetic architecture, where many genes could contribute to an adaptive phenotype, and adaptive architecture, where pleiotropy reduces redundancy at the genetic level.

Sweep signatures of rare large-effect alleles

The development of insecticide resistance is a very insightful example to demonstrate the complexity of distinguishing sweep and polygenic architectures. Genome scans identified a strong sweep signature around the gene *Cyp6g1* due to the insertion of a transposable element in the two closely related species *Drosophila simulans* and *Drosophila melanogaster*[94,95]. Overexpression of *Cyp6g1* fully rescues resistance to the insecticide DDT[127]. Nevertheless, selection for insecticide resistance in the laboratory using sublethal doses of DDT resulted in many resistance loci distributed across the entire genome[128]. This discrepancy has been explained by the *Cyp6g1* resistance allele being at too low a frequency in the founder population of the laboratory selection experiments[129]. Another interesting explanation for this discrepancy comes from studies of insecticide resistance in the Australian sheep blowfly, *Lucilia cuprina*[130]. Using a founder population, which was not mutation limited, independent resistance mutations mapping to the same position as a resistance allele from natural populations were obtained using high concentrations of the insecticide diazinon. Sublethal doses, however, resulted in a polygenic response. This pattern could be explained by the major effect allele being favoured for a distant trait optimum, but not for a close trait optimum. Here, alleles with smaller effect are favoured to prevent overshooting.

Very similar results have been obtained for herbicide resistance, where pronounced sweep signatures can be detected for target-site resistance[131]. Nevertheless, non-target-site-based resistance, which typically involves multiple loci, is the most important resistance mechanism worldwide for glyphosate and acetyl-CoA carboxylase inhibitors[132]. Similar to the Australian sheep blowfly, experiments with weak selection favoured non-target site-based resistance mechanisms[132].

It is important to note that sweep signatures do not preclude polygenic adaptation at other loci. Theoretical analyses even suggest that the contribution of large-effect alleles and small/intermediate-effect alleles have different temporal dynamics, with large-effect loci contributing more at an early phase of adaptation whereas alleles with smaller effects dominate at later stages[78]. A nice demonstration for this principle comes from the classic sweep signature associated with maize domestication. It is driven by the insertion of *Hopscotch*, a transposable element, in the regulatory region of the gene *teosinte branched 1* (REFS[133,134]). This transcription factor regulates a complex network of genes, which are involved in the architecture of the plant and inflorescence. As most of the genes in this network carry selection signatures from maize domestication[135], adaptation targeted the entire network, rather than only a single gene.



Incomplete sweeps — a hallmark of polygenic adaptation

Polygenic adaptation can be studied by changing trait optima, as is usually done in experimental evolution studies. An extension is to change the trait optimum after the population has reached the previous trait optimum — either in the same or opposite direction[136,137]. An interesting alternative experimental design was pursued in a *D. melanogaster* population, which was infected with *Wolbachia* and selected for resistance to the *Drosophila* C virus[138,139] (see the figure). This resistance was mediated by allele frequency changes at the genes *pastrel* and *Ubc-E2H*, which show strong sweep signatures[139]. After having reached the maximum resistance level, the allele frequency of these genes did not change any further, although the alleles had not reached fixation (see the figure, upper panels). This population was fixed for *Wolbachia*, an intracellular microorganism, which also contributes to C virus resistance. The evolved populations were treated with antibiotics to remove *Wolbachia*. In the absence of *Drosophila* C virus exposure, no allele frequency changes were noted in *Wolbachia*-free populations (see the figure, middle panels). Resuming *Drosophila* C virus exposure, the lack of *Wolbachia*-mediated resistance, which corresponds to moving the phenotype away from the trait optimum, resulted in a further frequency increase of both alleles. The frequency increase in the *pastrel* and *Ubc-E2H* genes restored the pre-antibiotic treatment resistance level (that is, the trait optimum was reached again (see the figure, lower panels))[138]. Hence, this example nicely demonstrates that stabilizing selection is operating even for resistance traits with few contributing factors.

Summary

The three examples of sweep-like signatures, which can be associated with adaptation to a trait optimum, may be a common phenomenon. We propose that polygenic adaptive architectures featuring redundancy and non-parallelism in the adaptive response may be the most likely scenario even for cases with apparent sweep signatures. Hence, polygenic adaptation should be routinely used to build hypotheses about the expected genomic response.

One early landmark approach identifies selection via the correlation of SNP frequencies in multiple populations with environmental variables. The covariance of allele frequencies across populations, which is created by the population history, is accounted for by a large panel of random SNPs[26]. Applying this method to humans[27] and *Arabidopsis thaliana*[28] resulted in a strong enrichment of non-synonymous SNPs among the variants with the strongest correlation. The other SNP categories,

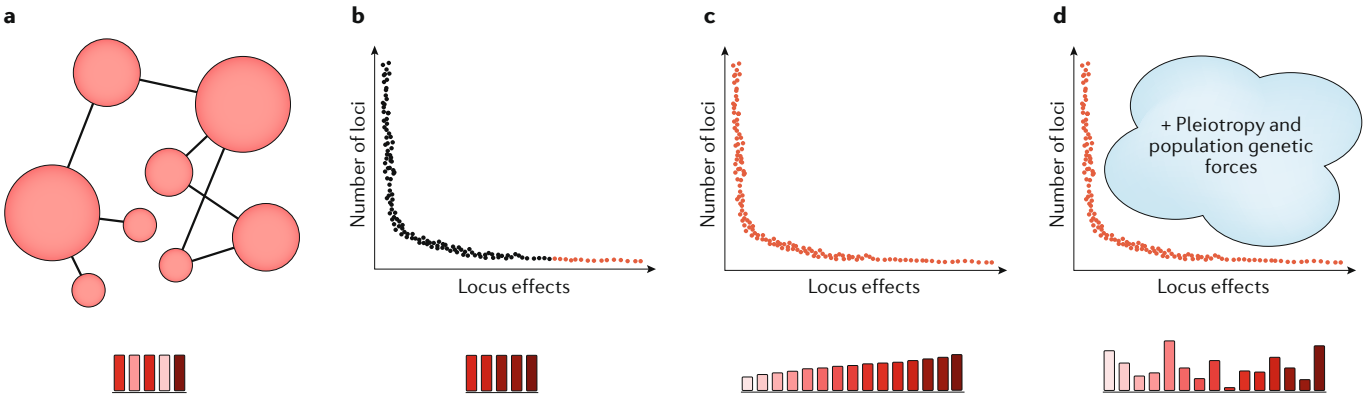


Fig. 2 | **Different approaches to characterize polygenic adaptation.** Upper panels indicate how candidate single-nucleotide polymorphisms (SNPs) are identified. Lower panels show the weighting of these SNPs: colour intensity corresponds to the locus identity and bar height depicts the locus weight. **a** | Candidate SNPs are obtained from predefined groups of genes using Gene Ontology terms, pathways or networks. Each candidate SNP has the same weight. **b** | Significant SNPs from genome-wide association studies constitute the group of candidate SNPs, which are all weighted equally. **c** | All SNPs are included as candidate SNPs, but weighted according to the variance explained (polygenic score). **d** | The SNPs in combination with the variance explained by them are an estimate of the genetic architecture. When these SNPs are reweighted by population genetic factors and pleiotropy, which determine how much a given SNP contributes to the new phenotype after a shift in the trait optimum, the adaptive architecture is described.

synonymous and intergenic, showed a much weaker signal. A more fine-grained classification of SNP sets relies on similar functional properties defined by Gene Ontology categories or pathways[29,30] (FIG. 2a). SNPs in genes of a given category can be treated jointly and significance is tested by comparison with a matched set of background markers. Although these approaches successfully provided evidence for polygenic adaptation, the lack of a direct link to phenotypes leaves some uncertainty about the nature of the selected traits.

GWAS, however, provide a powerful approach to link polygenic traits with their underlying genetic architecture. With phenotypes and genotypes from hundreds of thousands of individuals, human GWAS are particularly powerful[31,32]. The important step in going from association studies to the characterization of adaptive polygenic traits is the identification of a collective selection signature for contributing alleles[33,34]. Hence, another class of tests for polygenic adaptation involves grouping SNPs that show a significant association with the focal trait (FIG. 2b). If this group of SNPs differs for the selection summary statistic from background SNPs, this is considered evidence for polygenic adaptation. One example of this class of tests is the higher frequency of SNPs with significant GWAS effects for human height in northern Europeans than in southern European populations. This collective signature suggests that the size differences in human populations are caused by selection[35]. Similarly, elevated differentiation (measured by the fixation index, F_{ST}) of trait-associated alleles among populations without a specific spatial context has been used to detect polygenic adaptation. Contrasting human populations from three different continents detected selection for schizophrenia, waist-to-hip ratio and height[36]. Finally, the singleton density score, a selection signature that is particularly well suited for large sample sizes, has been found significantly elevated for height-increasing and pigmentation-associated GWAS variants[37]. The limitations of these approaches are discussed in the next section.

An alternative approach follows from the idea of genomic prediction[38], which was developed because GWAS markers with a significant effect explain only a small fraction of the total variation. Rather than focusing on a subset of (significant) SNPs, all of them are considered, regardless of whether they pass a nominal significance threshold (FIG. 2c). Unlike the other methods, the polygenic score does not treat each focal SNP equally, but weights each of them according to the explained phenotypic variance. Thus, polygenic scores account for a substantial fraction of the phenotypic variation that remains unexplained by significant markers. Selection signatures based on polygenic scores were detected in human populations for a diverse set of traits, including body size and skin pigmentation[39].

Limitations of current methods

For methods to study polygenic adaptation, increasing the statistical power by combining groups of SNPs is conceptually appealing, although this comes with several limitations and problems.

Are available GWAS data sets suited to studying adaptation? Powerful GWAS are currently limited to humans and species of agricultural importance, which directly reflects the type of phenotypes studied. Currently, studied traits are typically either of applied relevance or can be easily measured, which raises the question of their appropriateness for studying adaptive processes in natural populations[40]. Hence, more powerful studies on traits that are relevant to adaptation in the wild are needed, with the plant community spearheading this development[41,42].

The role of deleterious variants and pleiotropy. GWAS identifies variants that affect a trait irrespective of whether they are beneficial or deleterious. Evidence from human populations suggests that the influence of deleterious alleles on a broad range of phenotypes is pervasive[43,44]. In contrast to their substantial impact on GWAS, such deleterious alleles are not expected to contribute to polygenic adaptation. Including them in GWAS-based SNP sets could either dilute the signal or result in false-positive signals of polygenic adaption due to heterogeneous selection efficacy in different populations.

Many alleles contributing to a trait are pleiotropic[3] — they affect more than one phenotype. This is particularly relevant for the omnigenic model, which implies universal pleiotropy. The genomic response to selection on a polygenic trait is strongly affected by pleiotropy[34,45–50]. On the one hand, alleles are less likely to respond if they affect other phenotypes under stabilizing selection. Hence, even alleles

with a large effect size in GWAS may not respond during polygenic adaptation due to pleiotropy. On the other hand, selection on pleiotropically related traits may create spurious selection signals on the focal trait, even if this trait itself is not a selection target.

The challenge of population structure. Most GWAS target only a few populations; in human studies, these are typically populations with European ancestry. The implicit assumption of all polygenic selection signatures relying on candidate SNP sets from GWAS is that phenotypes can be predicted based on genotypes. Nevertheless, the linkage structure and allele frequencies differ between populations, and the extent to which the estimated effects can be transferred to genetically differentiated populations has not only been challenged for some traits in humans[32,51] but also in domestic animals[52,53] and plants[54]. It is not clear whether this translates only into reduced statistical power or could also lead to false positives in selection tests based on polygenic scores. However, a clear example for a false-positive selection signature based on polygenic scores comes from the reanalysis of human height using a different set of SNPs derived from a data set with less population stratification[55]. Two independent studies could not find a selection signature associated with body height with effects estimated from the less stratified population, but reproduced the original result with estimates from a structured population sample[56,57]. This result underlines the challenge to distinguish between demography and selection, in particular when many weak signals are combined.

An interesting approach to avoid the problem of population stratification combines GWAS-based effect sizes at one time point with genomic data before and after selection[58]. This approach provides the advantage that the estimated effect sizes are combined with realized allele frequency changes, and thus integrates selection response with effect sizes.

Single or multiple selection events? An implicit assumption of polygenic adaptation is that the observed pattern results from a single selection event on a single focal trait. However, the identification of the selected trait is challenging (BOX 2), and in many cases the observed signal may result from multiple independent selection events on a set of correlated traits, because the contributing loci are (for example) combined into the same Gene Ontology category. Similarly, it is conceivable that alleles of genes with different functions exhibit a similar

Box 2 | **The challenge of defining selected traits**

Trait hierarchy

Phenotypes are hierarchically organized (see the figure). High-level phenotypes, such as fitness and viability, are the outcome of many other underlying phenotypes. Any adaptation will result in higher fitness, but studying fitness alone will not provide much information about how and why the increase in fitness was achieved. This requires the identification of a lower-level phenotype, which needed to be altered by selection to increase fitness. The left and right panels in the figure represent individuals from two populations that have adapted by modifying the same intermediate-level phenotype, but through changes in different low-level phenotypes. In both individuals, the high-level phenotype (for example, fitness) has evolved in the same direction, but it is not possible to infer which of the underlying phenotypes has contributed to the change in high-level phenotype. Thus, the focus on high-level traits prevents further insights into the lower levels of the trait hierarchy that are the actual drivers behind this change. Therefore, we propose that traits of intermediate hierarchy are more likely to be informative about the underlying adaptive architecture.

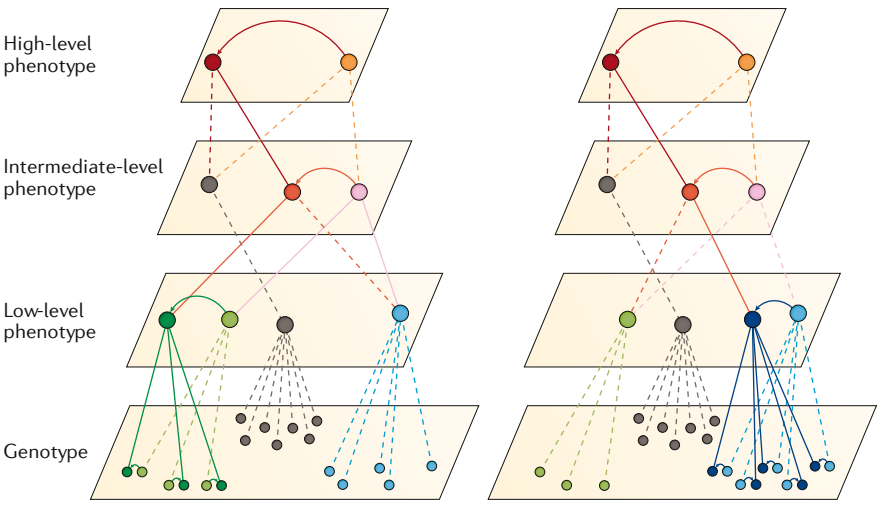
With genetic redundancy, many different genes could contribute to the selected phenotype, which limits the informativeness of a single gene for the adaptation process. Hence, finding the correct hierarchical level is one major challenge for studying adaptive phenotypes.

Example: altitude adaptation in humans. Populations living in high altitudes in Tibet and Ethiopia adapted by regulating red blood cell production[140]. The genomic analysis of Andean populations suggested a different target of selection. Here, genes associated with cardiovascular development and function showed the strongest selection signals[99]. Although all populations adapted to a similar environmental selective force, two different evolutionary solutions have been selected[141]. This raises the question of the adaptive phenotype. Whereas studies on Tibetan and Ethiopian populations suggest regulation of red blood cells to be the adaptive trait, analysis of Andeans points to cardiovascular phenotypes. It is apparent that the focus on either of the phenotypes as selection targets could result in incorrect conclusions, and another higher-level phenotype (for example, oxygen supply at high altitude) should be considered as the selection target, for which functional redundancy exists.

Pleiotropy

The genetic covariance between traits[142] makes it extremely challenging to distinguish between selected phenotypes and traits that covary either due to linkage disequilibrium (non-random association of single-nucleotide polymorphisms across loci) or pleiotropy (alleles that affect multiple phenotypes). As a consequence, research on adaptive phenotypes typically focuses on traits that are easy to score or for which an a priori hypothesis exists about their adaptive role.

Example: thermal budget hypothesis. A well-characterized pattern is the clinal variation in body pigmentation for many insects, in particular, *Drosophila* species[143]. According to the hypothesis, the more efficient absorption of sunlight by darker pigmentation provides a selective advantage in colder latitudes by providing thermal energy to insects. Although plausible, a detailed analysis of the relative importance of size and colour refuted the thermal budget hypothesis for small insects, such as *Drosophila* species[144]. On the other hand, an experimental evolution study in *Drosophila simulans* showed that temperature affects dopamine signalling, with increased dopamine synthesis in cold temperature regimes[145]. Because dopamine is an important precursor of the black melanin in fruit flies, it is conceivable that the naturally occurring cline for dopamine synthesis could explain the pigmentation cline.

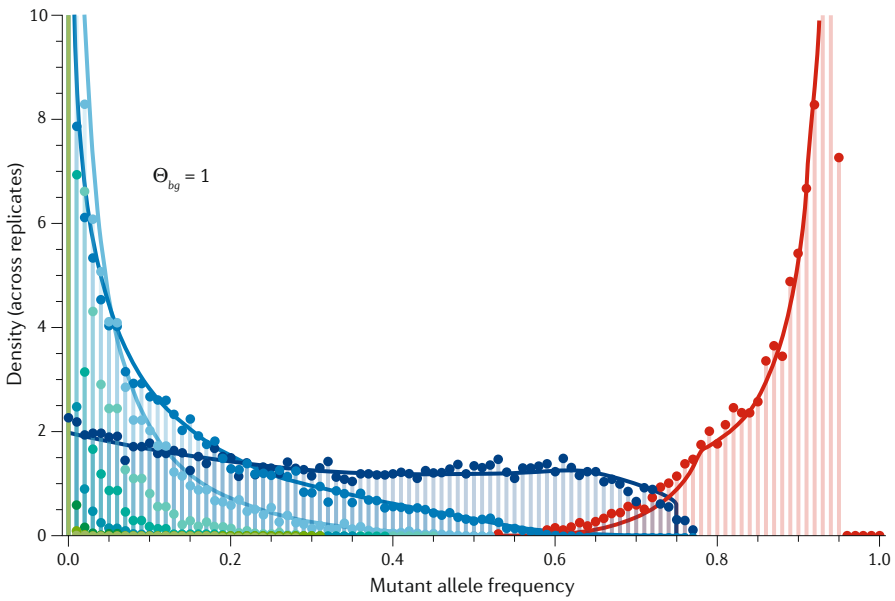


Box 3 | **Theory: the architecture of adaptation of a simple trait**

The simplest trait to feature genetic redundancy of its optimum has just two states, for example, a phenotype that is either resistant or non-resistant to a pathogen. Imagine that 'non-resistant' is the initial state of the population, and that the 'resistant' phenotype can evolve by a mutation at one of L different loci (for example, by shutting down a gene along a linear pathway that is exploited by the pathogen). There are thus L redundant ways to obtain the adaptive phenotype from the initial state, but mutation at multiple loci does not lead to further phenotypic changes. If (almost) all individuals in a population use the same mutation to produce the resistant phenotype, the genomic pattern is sweep-like. If most individuals use different alleles, adaptation occurs by polygenic shifts. In general, assume that adaptation occurs from mutation–selection–drift balance and we observe the population towards the end of the adaptive phase, when only a fraction, f_w, of individuals are non-resistant. Then, the probability that the frequency of mutant alleles at loci 1–L in a single evolutionary replicate given by $(x_1, \ldots, x_L)$ can be explicitly derived[77]

$$P[x_1, \ldots, x_L] = \frac{1}{B(\theta)} \prod_i x_i^{\theta-1} \left(\sum_i x_i\right)^{L\theta} \left(\sum_i \frac{f_w\, x_i}{1-x_i}\right)$$

where $\theta = 4N_e\mu$ is the population mutation parameter (N_e is the effective population size and μ the mutation rate per locus) and $B(\theta)$ is a normalization factor. $P[x_1, \ldots, x_L]$ defines the architecture of adaptation. One can show that, for this simple case, the shape of this architecture depends mostly on a single compound parameter $\theta_{bg} = (L-1)\theta$, combining the level of redundancy (measured by L) and the mutation parameter θ[77]. For $\theta_{bg} \ll 1$, adaptation is dominated by a single locus and we observe sweep-like signatures; for $\theta_{bg} \gg 1$, we obtain subtle frequency shifts at many loci; and at intermediate values $\theta_{bg} \approx 1$, we observe a heterogeneous polygenic adaptive response (see the figure). The figure shows the joint distribution of mutant allele frequencies when 95% of individuals have adapted ($f_w = 0.05$). The frequency distribution (across replicates) of the locus with the largest mutant frequency at this time is shown in red, the second largest in dark blue, and so forth. Although there is a major (sweep-like) frequency change at the first (red) locus, other alleles also increase in frequency, resulting in a mixed selection signal, which is typical for $\theta_{bg} = 1$. Lines are from the formula shown in this box, with simulation dots.



correlation with environmental variables. Hence, any a priori grouping of candidate loci (SNPs) may cause misleading results. Furthermore, it is possible that selection occurred at different time points, and thus the signature reflects different adaptation events. An extended method based on an admixture graph addresses this limitation and introduces a temporal component of adaptation. By estimating divergence and admixture events it is possible to infer selection strength on each branch, and thus to infer the extent to which populations experienced selection[59].

An integrated view

As discussed above, the deep rooting of current tests for polygenic adaptation in the infinitesimal model has several apparent limitations arising from the combination of the signature of multiple candidate loci. We propose an integrated framework for polygenic adaptation, which goes beyond the single-locus approach and combines key concepts of the infinitesimal model with a realistic genetic architecture and population genetic principles.

From genotype to phenotype: the role of genetic redundancy. The genetic architecture refers to "the pattern of genetic effects that build and control a given phenotypic character"[60]. It can be represented as a genotype–phenotype map, summarizing the effect sizes of all alleles affecting the focal trait, together with their interaction effects due to dominance or epistasis. Via collections of major and minor effect loci (and some epistatic relationships[61]), both quantitative trait locus (QTL) mapping and GWAS aim to provide this map.

One central property of the genetic architecture of polygenic traits is the general many-to-one relationship, that is, a single phenotype corresponds to a larger number of genotypes. These genotypes are thus redundant with respect to the phenotype they produce (BOX 3). Indeed, redundancy of the selection target is a generic property of any polygenic trait under stabilizing selection. In this case, the level of adaptive convergence is the phenotype rather than the genotype. This feature has two consequences for the patterns of molecular variation under polygenic adaptation. First, replicate populations that adapt to the same trait optimum are expected to show different adaptive responses on the genotypic level (FIG. 3a,b). Second, within each replicate, individuals can use different sets of alleles to produce the optimal phenotype. If many individuals in the adapted population use the same beneficial alleles, polygenic adaptation produces large frequency shifts and sweep-like patterns. Small shifts are observed if individuals mostly use different beneficial alleles.

With redundancy, adaptation is therefore necessarily a collective phenomenon across loci. This does not just mean that many loci contribute, but that adaptation at each of these loci can only be understood in a polygenic context. This is ignored in the classical sweep model. In stark contrast, the infinitesimal model assumes that conditioning on a trait value does not constrain the underlying allele frequencies at single genes at all[21]. The model thus assumes an architecture with infinite redundancy. Clearly, any integrative framework of polygenic adaptation needs to allow for trait architectures with quantifiable, intermediate

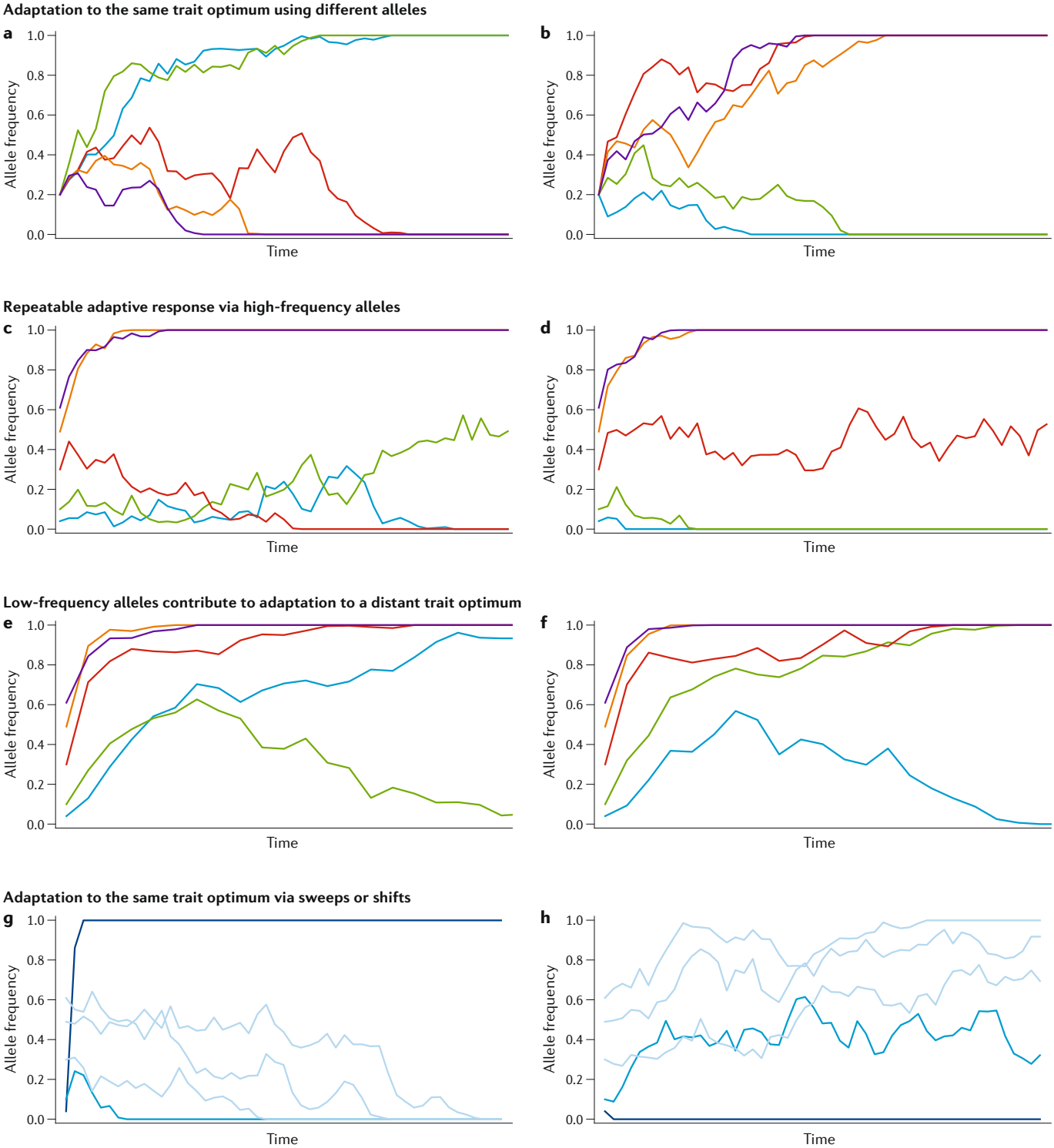


Fig. 3 | **Genetic redundancy and heterogeneity among loci are the main characteristics of polygenic adaptation.** Each line shows the trajectory of one allele contributing to the selected phenotype. **a**,**b** | Genetic redundancy results in non-parallel signatures of adaptation across replicate populations. Replicate populations (parts **a** and **b**) adapt to the same trait optimum using different alleles. **c**–**f** | Allele frequencies in the founder population (parts **c** and **d**) and distance to the trait optimum (parts **e** and **f**) are among the factors accounted for in the adaptive architecture of a trait. Heterogeneity of allele frequencies in replicate populations affects the adaptive response and parallelism across replicates, as alleles with higher starting frequency produce a more repeatable adaptive response (parts **c** and **d**). Founder populations with similar genetic architectures (parts **c**–**f**), when exposed to environments with different trait optima, will have different adaptive architectures. A close trait optimum (parts **c** and **d**) would require a frequency increase at few loci, whereas a further distant optimum (parts **e** and **f**) results in selection signatures at more loci. **g**,**h** | Adaptation to the new optimum can be achieved by a sweep-like signature of a large-effect locus (part **g**) or smaller frequency increases of multiple small-effect alleles (part **h**). The colour intensity in parts **g** and **h** reflect allelic effect sizes. For simulation parameters, see Supplementary information.

levels of redundancy. The importance of genetic redundancy for adaptation has been stressed before[49,62,63].

The architecture of adaptation: heterogeneity and parallelism. Although the genetic architecture specifies how the phenotype can be changed by mutation, it does not imply which alleles at which loci will be used during the course of adaptive evolution. For a trait optimum with redundant genetic basis, the genetic architecture shows how evolution can 'solve the problem' but does not quantify which solution is most likely. This is because the genetic architecture does not account for several key factors that are contingent on the selection scenario and that determine the contribution of individual alleles. These include, in particular, the starting frequencies of all potentially contributing alleles and the pleiotropic fitness constraints[49]. For example, large-effect alleles are less likely to be favoured when the new trait optimum is close or when the environment only leads to slow changes in the optimum; they are more likely to contribute after a rapid change to a distant new optimum[64–67]. Similarly, it is well known that large-effect alleles are disfavoured relative to small-effect alleles if negative pleiotropic effects on fitness are prevalent[13,45]. In a simple model, the effect size of the most-favoured allele declines proportionally to $\sim 1/\sqrt{n}$, where n is the number of pleiotropically affected traits[68].

To predict the pattern of polygenic adaptation, we therefore define the adaptive architecture of a trait to describe the adaptive potential for a given selection scenario. With redundancy, this architecture always takes the form of a probability distribution of allele frequencies that contribute to the adaptive response, both across replicates and across all loci in the genetic basis of the trait (a joint frequency distribution; BOX 3). Variance measures of this distribution provide crucial information about the adaptive scenario. For each locus, the variance across replicates measures the expected degree of parallelism (also called convergence or repeatability[49]) of adaptive evolution and informs us how representative a pattern from a single observation can be. For each replicate, the variance across loci reflects the heterogeneity of the adaptive response due to differences in the relative contribution of the genes. Reaching the new trait optimum, strongly contributing genes are characterized by high frequencies in all or most replicates, whereas weakly contributing genes are optional and do not appear in many individuals and/or replicates.

The variances (and all other parameters of the distribution) depend on the genetic architecture of the trait, but in addition also on the fitness constraints, the starting conditions and the population genetic forces that act during the adaptive phase (FIG. 2d). Whereas constant factors such as fitness constraints increase heterogeneity among loci but decrease differences among replicates (increased parallelism), stochastic forces due to mutation, migration and drift increase the heterogeneous response across both loci and replicates (reduced parallelism).

With this definition, the adaptive architecture follows the concept of the genetic architecture in that both describe so-called dispositional properties[69], that is, genetic architecture as the potential for phenotypic variation[60] and adaptive architecture as the potential for adaptation. Capturing adaptive potential is essential if we are interested in prediction (in which configuration polygenic adaptation is likely to occur) rather than in description (how it has occurred in a particular case). In contrast to the genetic architecture, the adaptive architecture does not just depend on mutation (the distribution of mutational effects) but also on all other population genetic forces — in particular, selection and drift — that act during the adaptive phase.

Modelling polygenic adaptation. An early analytical approach to characterize the footprint of polygenic adaptation focuses on the sweep signal (or lack thereof) at a single focal QTL in the presence of quantitative background variation[22]. Recently, the approach has been extended to include fluctuating selection[70]. As in quantitative genetics, background loci are not modelled explicitly, but are represented by a distribution that evolves with a constant variance. Explicit background loci are included in studies on the footprint at a focal QTL in two- to eight-locus models[71,72], but none of these studies considers polygenic patterns across multiple loci.

Several studies consider the polygenic signature that results from selection on an additive trait. Depending on initial levels of variation and the distance to the trait optimum, simulation studies find both sweep-like signatures at single loci and polygenic shifts[19,67,73,74]. Analytical approaches use a deterministic framework to predict how key aspects of the adaptive architecture depend on the distribution of locus effects if adaptation occurs from mutation–selection balance[23,75,76]. Whereas a deterministic approach does not allow for differences among replicates, the probabilistic nature of the adaptive architecture is revealed in the presence of genetic drift. Analytical results have been derived for a simple polygenic model with identical loci[77] (BOX 3) and in the limit of a highly polygenic trait[34,78].

Recombination and linkage constitute another source for heterogeneity of the adaptive architecture. Indeed, recent studies[79,80] of adaptive introgression using a version of the infinitesimal model that accounts for linkage[81] show that this necessary extension of the model already results in heterogeneous responses with large frequency changes of introgression haplotypes in some genomic regions[79,80].

The spatial population structure can be an important factor to shape the adaptive architecture, in particular for a trait with high genetic redundancy. If selection is homogeneous across a large spatial range, the adaptive architecture is expected to represent a mosaic of local replicates. This has been explored for a single adaptive step[82–85], but the models assume allelic exclusion and do not allow mutant alleles at different loci to co-segregate in the same region. Local adaptation of a polygenic trait leads to a pattern of multilocus clines[86]. After the initial adaptive phase, gene flow can result in a turnover of adaptive alleles and a long-term advantage of large-effect alleles, which are more robust to swamping[63].

Charting the adaptive architecture: replicates and time-series data. Replicates of adaptive evolution provide the most valuable source of information about the adaptive architecture. Clearly, replicates are essential to determine the degree of parallelism of adaption, which can be quantified by several indices[49,87,88]. Replicates are also required for the interpretation of the observed heterogeneity in the selection response across loci, to separate factors that are constant across replicates (for example, locus effects) from chance events (for example, recombination and genetic drift). Random allele frequency fluctuations during the early phase of the adaptation processes are a major factor determining to what extent a given allele will be contributing to the new trait optimum[89]. This effect is more pronounced in small to moderate-sized populations than in large natural populations. Hence, experiments with moderate population sizes provide a powerful approach to exploit redundancy and study the influence of various factors of the adaptive architecture. FIGURE 3c–f shows that alleles with a high starting

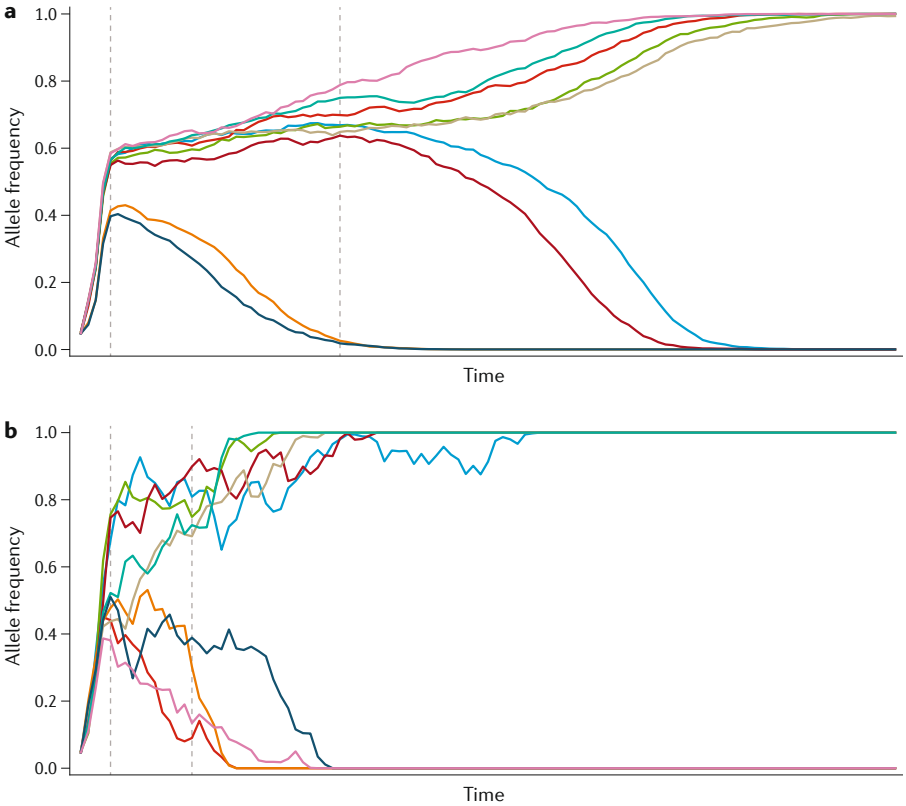


Fig. 4 | **Different stages of polygenic adaptation.** Phases are approximately demarcated by the vertical dotted lines. **a**,**b** | Phase 1: in this phase of directional selection, the trait mean adapts rapidly to match the new optimum. Phase 2: after the rapid adaptive phase, allele frequencies change under the weaker forces of genetic drift and selection against the segregation load, which produces suboptimal phenotypes. Phase 3: once drift has disturbed the allele frequency beyond a critical point, the allele fate (fixation or loss) is determined, and they are rapidly driven there by selection. As the sorting of alleles during phase 2 is mainly governed by the stochastic forces of genetic drift, phase 2 is longer in larger populations (part **a**, effective population size (N_e) = 36,000) than in small ones (part **b**, N_e = 450). The stochastic nature is also responsible for the transition from phase 2 to phase 3 differing among loci and precludes the definition of hard boundaries between these two phases. Hence, the dotted lines only roughly indicate the transition between phases 2 and 3. For simulation parameters, see Supplementary information.

frequency contribute more consistently to the optimal phenotype than low-frequency alleles. Although low-frequency alleles have variable contributions to optimal phenotypes (FIG. 3c,d), they make a more consistent contribution when the distance to the trait optimum is increased (FIG. 3e,f). Among replicates, extremely different selection responses can occur: either a sweep signature at a single locus (FIG. 3g) or small/moderate frequency shifts at several loci (FIG. 3h).

In addition to replicates, time-series data also provide important information about the adaptive scenario. Three different phases can be distinguished for a typical time series in polygenic adaptation. During the first phase, the contributing adaptive alleles increase in frequency. When the trait optimum is reached, the second phase starts with the contributing alleles being mainly affected by drift. In the third phase, the alleles are sorted until they reach fixation (FIG. 4). With phases 2 and 3 being specific to polygenic adaptation, time-series data are a powerful approach to uncover these dynamics.

Experimental study systems

Heterogeneity and parallelism (or non-parallelism) are two key features that characterize polygenic adaptation. Replicate populations, preferably with time-series data, are needed to study these features experimentally. In the following sections we discuss a range of study systems that have the potential to contribute to future studies of polygenic adaptation.

Natural populations. The ideal natural system to study polygenic adaptation would consist of multiple populations that independently evolve to the same environmental stressor. Genotypes and phenotypes of the founding populations would be known and the adaptation process on both levels could be followed in real time. Typically, however, studies of parallel evolution are retrospect and lack direct knowledge of the starting populations. In addition, most natural habitats are unique and true replicates rarely exist. Nevertheless, relaxing the criterion to similar, rather than identical, environments allows the identification of a few systems with potential to study polygenic adaptation in natural populations.

A parallel phenotypic response of different populations to similar environmental challenges is a classical hallmark of natural selection and has been documented across a wide range of species. In several cases, the genetic causes of parallel phenotypic evolution have also been documented[90–92]. Iconic examples (see also BOX 2) include the parallel evolution of crypsis in beach mice and several further species[93], the evolution of insecticide resistance in flies[94,95] and the loss of body armour in sticklebacks[96]. As case studies, these examples are valuable, in particular because they offer rather complete adaptive histories, from the ecological selection pressures down to the genotype. They typically also show a high level of parallelism on the genetic level and involve large changes in the allele frequency of key genes (that is, sweep-type architectures). However, their potential for the study of polygenic adaptation is still limited for two reasons. First, the level of replication is very low. Second, focus on the most striking cases of parallel evolution harbours a risk of bias and cannot be taken as representative of polygenic adaptation in general. Indeed, rapid evolution of extreme phenotypes may favour large-effect alleles in traits with oligogenic bases. In line with this expectation, adaptation occurs by loss of function of a key gene in several cases highlighted above.

These biases can be attenuated when we compare the evolution (parallel or not) across a broader range of traits in independent populations that are exposed to similar conditions. In nature, such conditions can be found in latitudinal[97] and altitudinal[98–102] clines, where similar phenotypic and, to some extent, also genetic changes are shared, sometimes even across continents. However, the level of replication is usually modest. More replication is offered in several cases of parallel colonization or range expansion. An example is the parallel river system of the Northern Range mountains in Trinidad. Waterfalls separate high and low-predation guppy populations

in different rivers, which has resulted in highly parallel phenotypic changes[103–105]. In sticklebacks, different contrasts — marine/freshwater, lake/stream and benthic/limnetic — have been studied in multiple rivers, making this a particularly powerful, highly replicated system[106–108]. The power to study heterogeneity of polygenic adaptation in natural replicate populations has been demonstrated in a study of benthic/limnetic adaptation[92]. Thirty-two phenotypes showed parallel evolution between two lakes, but about 50% of the underlying QTLs were not shared[92]. Although this is clearly less parallel than the famous plate armour trait, we are currently lacking a solid framework to decide how much parallelism is expected. Furthermore, as there is no direct access to the ancestral populations, and because local adaptation to unresolved ecological differences among lakes may reduce parallelism, it remains difficult to identify the causes of the observed heterogeneity in the adaptive response[109].

A barely exploited opportunity to study temporal dynamics is provided by sediment analyses, where each layer of sediment represents a different historic time. So far, sediment analysis has been largely employed to study presence/absence patterns[110], but the combination of steadily dropping sequencing costs with more refined enrichment methods provides the potential to study allele frequency changes across time. Some systems, such as *Daphnia*, can go even one step further. Diapausing eggs can be recovered from dated sediment layers and reactivated in the laboratory. The phenotypes of resulting parthenogenetic lineages can be studied in common garden experiments[111]. Hence, it is not only possible to follow allele frequency changes over time but also to monitor the associated phenotypic evolution.

Experimental evolution in the laboratory. Experimental introduction of populations is a classic experimental evolution setting in the wild, which has already provided interesting results[103–105,112] but typically suffers from low levels of replication and the challenge of environmental heterogeneity. Experimental evolution under controlled laboratory conditions overcomes these limitations and provides the opportunity to study adaptation for traits of interest without confounding factors, such as environmental heterogeneity, uncontrolled migration and variable population sizes. The major advantage of experimental evolution is the ability to replicate experiments at a scale that is rarely possible in the wild. Although most experimental evolution studies are on asexual microorganisms, we focus on outcrossing systems, as the combination of different alleles by recombination is essential for the multilocus nature of polygenic adaptation. Because of moderate population sizes and numbers of generations in typical experimental evolution studies with sexual eukaryotes, new mutations can be neglected[113]. This implies that adaptation needs to occur from standing genetic variation. Using freshly collected population samples as founders for experimental evolution allows one to build on natural frequency spectra for adaptive alleles and provides a natural setting to study

Glossary

Adaptive architecture
The measure of the probability that alleles contribute to adaptation. Adaptive architecture extends the genetic architecture by including further factors that influence the adaptive potential.

Adaptive introgression
New, favourable alleles are introduced into a population by migration.

Admixture graph
A representation of the divergence and admixture between populations.

Clines
Spatial patterns of allele frequency differences, which are maintained by a spatial selection gradient.

Common garden experiments
Experiments that, in order to control for the effects of the environment on phenotypes, measure the phenotypes of different genotypes in the same/similar environments.

Epistasis
Interaction between genes in a non-additive way.

Genetic architecture
Information about genes, with their associated effect sizes and patterns of pleiotropy, epistasis and dominance.

Genetic basis
The set of all loci contributing to a trait, but without reference to effect sizes or pleiotropy, epistasis or dominance.

Genetic drift
A stochastic process arising from the random sampling of gametes contributing to the next generation. In small populations, genetic drift can be strong and results in large, non-directional allele frequency changes.

Genome-wide association studies
(GWAS). A genetic technique that identifies statistically significant associations between phenotypes and underlying genetic variants. GWAS are particularly powerful, because they take advantage of recombination events that occurred historically in the focal population.

Infinitesimal model
The phenotype is determined by a very large (infinite) number of alleles, each with a very small effect, and by the environment.

Mutation–selection balance
An equilibrium situation for a population close to an adaptive optimum. The same number of new deleterious alleles are introduced into the population by mutation as are removed by purifying selection.

Non-synonymous SNPs
Single-nucleotide polymorphisms (SNPs) in protein-coding genes that result in an amino acid replacement.

Parallelism
(Also known as convergence or repeatability). Replicate populations reach the same trait values using the same set of alleles; non-parallelism is the possible consequence of redundancy. Parallelism has been also described for asexual microorganisms, where the same mutations are independently acquired in replicate populations.

Pleiotropy
A single gene affects multiple traits.

Polygenic traits
(Also known as complex traits). Quantitatively variable phenotypes that are affected by many contributing loci and the environment.

Purifying selection
Removal of deleterious alleles from a population.

Quantitative trait locus (QTL) mapping
A genetic mapping technique that relies on recombination events that occurred during the experiment.

Quantitative traits
Traits with a continuous distribution of phenotypes with a large number of contributing alleles.

Redundancy
Different combinations of alleles produce the same phenotypic value.

Selective sweeps
Classic selection signatures in molecular population genetics describing a pattern of reduced DNA polymorphism around the site of a recently fixed beneficial allele.

Singleton density score
A test statistic to detect selection based on the distance of singleton single-nucleotide polymorphisms nearest to the focal variant.

Soft sweeps
Different alleles at the same locus are favoured and contribute to adaptation. They can either be generated by recurrent mutations or they segregate in the population before the adaptive episode starts.

Stabilizing selection
Selection favours individuals with an intermediate trait value.

Standing genetic variation
Polymorphic sites segregating in a natural population.

Swamping
Beneficial alleles are driven to extinction by immigration of non-favoured alleles.

parallelism and heterogeneity. Combining experimental evolution with whole-genome sequencing of pooled individuals (Pool-Seq[114]) provides a powerful approach to study allele frequency dynamics during adaptation.

Most experimental evolution studies focus on the identification of parallel selection signatures across replicates[115–120]. Only recently was non-parallelism among replicates explicitly considered in a study of 10 replicate *Drosophila simulans* populations[88]. Computer simulations showed that the pattern of non-parallelism among replicates is not consistent with a selective sweep scenario of independent adaptation at different loci. The pattern of non-parallelism instead fits collective polygenic adaptation to a new trait optimum. Most selection targets started at low frequency in the founder population, but common targets could also be detected[88]. Low-frequency alleles were associated with stronger selection than common alleles[88]. The power of phenotypic and genetic time-series data was demonstrated by a study in yeast[121]. In contrast to other experimental evolution studies, which aimed for highly diverse founder populations, in this study the adaptive trajectories of only two genotypes was followed for almost 1000 generations in the presence of recombination. After an initial rapid frequency change, very little frequency change was observed, and the fitness variance among individuals was not strongly reduced during the plateauing phase. Supported by computer simulations, the authors concluded that, similar to the infinitesimal model, adaptation was facilitated by many loci with small effects. Further work is needed to reconcile this extreme polygenic architecture with results from more complex founder populations — in particular, the role of deleterious alleles needs clarification.

Outlook

Both experimental evolution and natural populations exhibiting parallel phenotypic evolution offer great potential to study non-parallelism and heterogeneity in the quest to understand the genetic basis of adaptation. We anticipate that the combination of time-series data with replicate populations will be a particularly powerful method to uncover the genetic architecture of adaptation. Of particular interest are specifically designed experimental evolution studies, both in the laboratory and, whenever possible, also in natural settings. Smaller population sizes increase genetic drift, resulting in more heterogeneity within populations and non-parallelism among populations, but with a sufficient level of replication it is possible to obtain reliable signatures that can be distinguished from neutral variation. Reducing the number of adaptive alleles by starting experiments with a small number of founders will result in stronger selection responses of individual alleles, providing the potential of further follow-up functional characterization. Once the adaptive architecture of a trait is well characterized by experimental evolution studies, it will be possible to expand the focus to natural populations to understand their dynamics in complex systems.

Neda Barghi[1], *Joachim Hermisson*[2]✉ *and Christian Schlötterer*[1]✉

[1]*Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria.*

[2]*Mathematics and BioSciences Group, Faculty of Mathematics and Max Perutz Labs, University of Vienna, Vienna, Austria.*

✉*e-mail: joachim.hermisson@univie.ac.at; christian.schloetterer@vetmeduni.ac.at*

https://doi.org/10.1038/s41576-020-0250-z

Published online: 29 June 2020

1. Pritchard, J. K. & Di Rienzo, A. Adaptation — not by sweeps alone. *Nat. Rev. Genet.* **11**, 665–667 (2010).
2. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010).
This key paper defines the concept of polygenic adaptation.
3. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
This landmark paper introduces the omnigenic model and links quantitative genetics with molecular biology.
4. Csillery, K., Rodriguez-Verdugo, A., Rellstab, C. & Guillaume, F. Detecting the genomic signal of polygenic adaptation and the role of epistasis in evolution. *Mol. Ecol.* **27**, 606–612 (2018).
5. Stephan, W. Selective sweeps. *Genetics* **211**, 5–13 (2019).
6. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
7. Jain, K. & Stephan, W. Modes of rapid polygenic adaptation. *Mol. Biol. Evol.* **34**, 3169–3175 (2017).
8. Hermisson, J. & Pennings, P. S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
9. Przeworski, M., Coop, G. & Wall, J. D. The signature of positive selection on standing genetic variation. *Evolution* **59**, 2312–2323 (2005).
10. Teshima, K. M. & Przeworski, M. Directional positive selection on an allele of arbitrary dominance. *Genetics* **172**, 713–718 (2006).
11. Ewing, G., Hermisson, J., Pfaffelhuber, P. & Rudolf, J. Selective sweeps for recessive alleles and for other modes of dominance. *J. Math. Biol.* **63**, 399–431 (2011).
12. Coop, G. & Ralph, P. Patterns of neutral diversity under general models of selective sweeps. *Genetics* **192**, 205–224 (2012).
13. Orr, H. A. The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–127 (2005).
14. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45 (1987).
15. Geritz, S. A. H., Kisdi, E., Meszena, G. & Metz, J. A. J. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57 (1998).
16. Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.* **52**, 339–433 (1918).
17. Galton, F. *Natural Inheritance* (MacMillan, 1889).
18. Bulmer, M. G. Effect of selection on genetic variability. *Am. Nat.* **105**, 201–211 (1971).
19. Barton, N. H. & Keightley, P. D. Understanding quantitative genetic variation. *Nat. Rev. Genet.* **3**, 11–21 (2002).
20. Turelli, M. Commentary: Fisher's infinitesimal model: a story for the ages. *Theor. Popul. Biol.* **118**, 46–49 (2017).
21. Barton, N. H., Etheridge, A. M. & Veber, A. The infinitesimal model: definition, derivation, and implications. *Theor. Popul. Biol.* **118**, 50–73 (2017).
This paper explains the infinitesimal model using modern statistical and population genetic terminology.
22. Chevin, L. M. & Hospital, F. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* **180**, 1645–1660 (2008).
This study is one of the first to investigate whether genomic sweep signatures are also possible for polygenic adaptation.
23. Jain, K. & Stephan, W. Rapid adaptation of a polygenic trait after a sudden environmental shift. *Genetics* **206**, 389–406 (2017).
This paper provides analytical approaches to describe cases of sweep signatures for large-effect alleles and small allele shifts for small-effect loci during polygenic adaptation.
24. Pavlidis, P. & Alachiotis, N. A survey of methods and tools to detect recent and strong positive selection. *J. Biol. Res.* **24**, 7 (2017).
25. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting natural selection in genomic data. *Annu. Rev. Genet.* **47**, 97–120 (2013).
26. Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J. K. Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411–1423 (2010).
27. Hancock, A. M. et al. Adaptations to climate-mediated selective pressures in humans. *PLoS Genet.* **7**, e1001375 (2011).
28. Hancock, A. M. et al. Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83–86 (2011).
29. Daub, J. T. et al. Evidence for polygenic adaptation to pathogens in the human genome. *Mol. Biol. Evol.* **30**, 1544–1558 (2013).
30. Gouy, A., Daub, J. T. & Excoffier, L. Detecting gene subnetworks under selection in biological pathways. *Nucleic Acids Res.* **45**, e149 (2017).
31. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
32. Robinson, M. R. et al. Population genetic differentiation of height and body mass index across Europe. *Nat. Genet.* **47**, 1357–1362 (2015).
33. Lohmueller, K. E., Mauney, M. M., Reich, D. & Braverman, J. M. Variants associated with common disease are not unusually differentiated in frequency across populations. *Am. J. Hum. Genet.* **78**, 130–136 (2006).
34. Sella, G. & Barton, N. H. Thinking about the evolution of complex traits in the era of genome-wide association studies. *Annu. Rev. Genom. Hum. Genet.* **20**, 461–493 (2019).
This comprehensive review discusses the connection between GWAS and polygenic adaptation.
35. Turchin, M. C. et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* **44**, 1015–1019 (2012).
36. Guo, J. et al. Global genetic differentiation of complex traits shaped by natural selection in humans. *Nat. Commun.* **9**, 1865 (2018).
37. Field, Y. et al. Detection of human adaptation during the past 2000 years. *Science* **354**, 760–764 (2016).
This paper introduces the singleton density score to detect polygenic adaptation.
38. Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
39. Berg, J. J. & Coop, G. A population genetic signal of polygenic adaptation. *PLoS Genet.* **10**, e1004412 (2014).
This paper presents a key method to detect polygenic adaptation from GWAS data.
40. Novembre, J. & Barton, N. H. Tread lightly interpreting polygenic tests of selection. *Genetics* **208**, 1351–1355 (2018).

41. Ferrero-Serrano, A. & Assmann, S. M. Phenotypic and genome-wide association with the local environment of *Arabidopsis*. *Nat. Ecol. Evol.* **3**, 274–285 (2019).
42. Yeaman, S. et al. Convergent local adaptation to climate in distantly related conifers. *Science* **353**, 1431–1433 (2016).
This paper is an empirical demonstration of convergent local adaptation between two distantly related tree species confronted with the same adaptive challenge.
43. Gazal, S. et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
44. Zeng, J. et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* **50**, 746–753 (2018).
45. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon Press, 1930).
46. Chevin, L. M., Martin, G. & Lenormand, T. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**, 3213–3231 (2010).
47. Hansen, T. F. & Houle, D. Measuring and comparing evolvability and constraint in multivariate characters. *J. Evolut. Biol.* **21**, 1201–1219 (2008).
48. Jones, A. G., Arnold, S. J. & Bürger, R. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**, 1747–1760 (2003).
49. Yeaman, S., Gerstein, A. C., Hodgins, K. A. & Whitlock, M. C. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS Genet.* **14**, e1007717 (2018).
This paper develops an index to quantify the amount of observed repeatability relative to theoretical expectations.
50. Simons, Y. B., Bullaughey, K., Hudson, R. R. & Sella, G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.* **16**, e2002985 (2018).
51. Carlson, C. S. et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol.* **11**, e1001661 (2013).
52. Erbe, M. et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy. Sci.* **95**, 4114–4129 (2012).
53. de Roos, A. P., Hayes, B. J. & Goddard, M. E. Reliability of genomic predictions across multiple populations. *Genetics* **183**, 1545–1553 (2009).
54. Guo, Z. et al. The impact of population structure on genomic prediction in stratified populations. *Theor. Appl. Genet.* **127**, 749–762 (2014).
55. Haworth, S. et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.* **10**, 333 (2019).
56. Berg, J. J. et al. Reduced signal for polygenic adaptation of height in UK Biobank. *eLife* **8**, e39725 (2019).
57. Sohail, M. et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* **8**, e39702 (2019).
58. Beissinger, T. et al. A simple test identifies selection on complex traits. *Genetics* **209**, 321–333 (2018).
59. Racimo, F., Berg, J. J. & Pickrell, J. K. Detecting polygenic adaptation in admixture graphs. *Genetics* **208**, 1565–1584 (2018).
60. Hansen, T. F. The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. S* **37**, 123–157 (2006).
61. Carlborg, O. & Haley, C. S. Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**, 618–625 (2004).
62. Goldstein, D. B. & Holsinger, K. E. Maintenance of polygenic variation in spatially structured populations: roles for local mating and genetic redundancy. *Evolution* **46**, 412–429 (1992).
63. Yeaman, S. Local adaptation by alleles of small effect. *Am. Nat.* **186**, S74–S89 (2015).
64. Kopp, M. & Hermisson, J. Adaptation of a quantitative trait to a moving optimum. *Genetics* **176**, 715–719 (2007).
65. Matuszewski, S., Hermisson, J. & Kopp, M. Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics* **200**, 1255–1274 (2015).
66. Jain, K. & Devi, A. Polygenic adaptation in changing environments. *EPL* **123**, 48002 (2018).
67. Stetter, M. G., Thornton, K. & Ross-Ibarra, J. Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. *PLoS Genet.* **14**, e1007794 (2018).
68. Orr, H. A. Adaptation and the cost of complexity. *Evolution* **54**, 13–20 (2000).
69. Wagner, G. P. & Altenberg, L. Perspective: complex adaptations and the evolution of evolvability. *Evolution* **50**, 967–976 (1996).
70. Chevin, L. M. Selective sweep at a QTL in a randomly fluctuating environment. *Genetics* **213**, 987–1005 (2019).
71. Pavlidis, P., Metzler, D. & Stephan, W. Selective sweeps in multilocus models of quantitative traits. *Genetics* **192**, 225–239 (2012).
72. Wollstein, A. & Stephan, W. Adaptive fixation in two-locus models of stabilizing selection and genetic drift. *Genetics* **198**, 685–697 (2014).
73. Franssen, S. U., Kofler, R. & Schlötterer, C. Uncovering the genetic signature of quantitative trait evolution with replicated time series data. *Heredity* **118**, 42–51 (2017).
This paper studies genomic signatures of polygenic adaptation in small, replicated experimental populations.
74. Thornton, K. R. Polygenic adaptation to an environmental shift: temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. *Genetics* (2019).
This study presents a comprehensive simulation of polygenic adaptation of a single additive trait.
75. de Vladar, H. P. & Barton, N. Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics* **197**, 749–767 (2014).
76. Jain, K. & Stephan, W. Response of polygenic traits under stabilizing selection and mutation when loci have unequal effects. *G3* **5**, 1065–1074 (2015).
77. Höllinger, I., Pennings, P. S. & Hermisson, J. Polygenic adaptation: from sweeps to subtle frequency shifts. *PLoS Genet.* **15**, e1008035 (2019).
This paper presents an analytical framework to describe the adaptive architecture of a polygenic trait.
78. Hayward, L. K. & Sella, G. Polygenic adaptation after a sudden change in environment. *bioRxiv* (2019).
This paper presents an analytical description of the adaptation process for a highly polygenic trait.
79. Sachdeva, H. & Barton, N. H. Replicability of introgression under linked, polygenic selection. *Genetics* **210**, 1411–1427 (2018).
80. Sachdeva, H. & Barton, N. H. Introgression of a block of genome under infinitesimal selection. *Genetics* **209**, 1279–1303 (2018).
This paper models adaptation for the infinitesimal model with linkage and finds that this can lead to large frequency changes.
81. Robertson, A. in *Proc. Int. Conf. Quant. Genet*. (eds, E. Pollak, O. Kempthorne & T. B. Bail) 307–322 (Iowa State Univ. Press, 1977).
82. Ralph, P. & Coop, G. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics* **186**, 647–668 (2010).
83. Ralph, P. L. & Coop, G. The role of standing variation in geographic convergent adaptation. *Am. Nat.* **186**, S5–S23 (2015).
84. Ralph, P. L. & Coop, G. Convergent evolution during local adaptation to patchy landscapes. *PLoS Genet.* **11**, e1005630 (2015).
85. Paulose, J., Hermisson, J. & Hallatschek, O. Spatial soft sweeps: patterns of adaptation in populations with long-range dispersal. *PLoS Genet.* **15**, e1007936 (2019).
86. Barton, N. H. Clines in polygenic traits. *Genet. Res.* **74**, 223–236 (1999).
87. Jaccard, P. Louis de distribution florale dans la zone alpine. *Bull. Soc. Vaud. Sci. Nat.* **38**, 69–130 (1902).
88. Barghi, N. et al. Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biol.* **17**, e3000128 (2019).
This paper provides empirical proof that polygenic adaptation can be detected in sufficiently well-replicated experimental evolution studies.
89. Nene, N. R., Dunham, A. S. & Illingworth, C. J. R. Inferring fitness effects from time-resolved sequence data with a delay-deterministic model. *Genetics* **209**, 255–264 (2018).
90. Martin, A. & Orgogozo, V. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**, 1235–1250 (2013).
91. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
92. Conte, G. L. et al. Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. *Genetics* **201**, 1189–1200 (2015).
This paper presents an innovative approach using QTL mapping to explore redundancy.
93. Harris, R. B. et al. The population genetics of crypsis in vertebrates: recent insights from mice, hares, and lizards. *Heredity* **124**, 1–14 (2019).
94. Catania, F. et al. World-wide survey of an Accord insertion and its association with DDT resistance in *Drosophila melanogaster*. *Mol. Ecol.* **13**, 2491–2504 (2004).
95. Schlenke, T. A. & Begun, D. J. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc. Natl Acad. Sci. USA* **101**, 1626–1631 (2004).
96. Colosimo, P. F. et al. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**, 1928–1933 (2005).
This classic study shows that the same allele is independently selected in many river populations of sticklebacks.
97. Adrion, J. R., Hahn, M. W. & Cooper, B. S. Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends Genet.* **31**, 434–444 (2015).
98. Yi, X. et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
99. Crawford, J. E. et al. Natural selection on genes related to cardiovascular health in high-altitude adapted andeans. *Am. J. Hum. Genet.* **101**, 752–767 (2017).
100. Bigham, A. et al. Identifying signatures of natural selection in tibetan and andean populations using dense genome scan data. *PLoS Genet.* **6**, e1001116 (2010).
101. Keller, I., Alexander, J. M., Holderegger, R. & Edwards, P. J. Widespread phenotypic and genetic divergence along altitudinal gradients in animals. *J. Evolut. Biol.* **26**, 2527–2543 (2013).
102. Halbritter, A. H. et al. Trait differentiation and adaptation of plants along elevation gradients. *J. Evolut. Biol.* **31**, 784–800 (2018).
103. Alexander, H. J., Taylor, J. S., Wu, S. S. & Breden, F. Parallel evolution and vicariance in the guppy (*Poecilia reticulata*) over multiple spatial and temporal scales. *Evolution* **60**, 2352–2369 (2006).
104. Hendry, A. P., Kelly, M. L., Kinnison, M. T. & Reznick, D. N. Parallel evolution of the sexes? Effects of predation and habitat features on the size and shape of wild guppies. *J. Evolut. Biol.* **19**, 741–754 (2006).
105. Reznick, D. N. & Bryga, H. Life-history evolution in guppies (*Poecilia Reticulata*): 1. phenotypic and genetic changes in an introduction experiment. *Evolution* **41**, 1370–1385 (1987).
106. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
107. Hohenlohe, P. A. et al. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* **6**, e1000862 (2010).
108. Jones, F. C. et al. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.* **22**, 83–90 (2012).
109. Stuart, Y. E. et al. Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat. Ecol. Evol.* **1**, 158 (2017).
110. Ficetola, G. F. et al. DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. *Sci. Adv.* **4**, eaar4292 (2018).
111. Hairston, N. G. et al. Lake ecosystems — rapid evolution revealed by dormant eggs. *Nature* **401**, 446–446 (1999).
112. Marques, D. A., Jones, F. C., Di Palma, F., Kingsley, D. M. & Reimchen, T. E. Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nat. Ecol. Evol.* **2**, 1128–1138 (2018).
113. Burke, M. K. et al. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* **467**, 587–590 (2010).
114. Schlötterer, C., Tobler, R., Kofler, R. & Nolte, V. Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* **15**, 749–763 (2014).
115. Kelly, J. K. & Hughes, K. A. Pervasive linked selection and intermediate-frequency alleles are implicated in an evolve-and-resequencing experiment of *Drosophila simulans*. *Genetics* **211**, 943–961 (2019).
116. Burke, M. K., Liti, G. & Long, A. D. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. *Mol. Biol. Evol.* **31**, 3228–3239 (2014).

117. Graves, J. L. Jr et al. Genomics of parallel experimental evolution in *Drosophila*. *Mol. Biol. Evol.* **34**, 831–842 (2017).
This paper presents an evolve and resequence study with a surprising level of parallel evolution on the genomic and phenotypic levels.
118. Mallard, F., Nolte, V., Tobler, R., Kapun, M. & Schlötterer, C. A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in *Drosophila*. *Genome Biol.* **19**, 119 (2018).
119. Sikkink, K. L., Reynolds, R. M., Ituarte, C. M., Cresko, W. A. & Phillips, P. C. Environmental and evolutionary drivers of the modular gene regulatory network underlying phenotypic plasticity for stress resistance in the nematode caenorhabditis remanei. *G3* **9**, 969–982 (2019).
120. Teotonio, H., Estes, S., Phillips, P. C. & Baer, C. F. Experimental evolution with *Caenorhabditis* nematodes. *Genetics* **206**, 691–716 (2017).
121. Kosheleva, K. & Desai, M. M. Recombination alters the dynamics of adaptation on standing variation in laboratory yeast populations. *Mol. Biol. Evol.* **35**, 180–201 (2018).
This study of experimental evolution with recombinants of two yeast strains suggests the presence of many contributing loci, which continue to segregate during the entire experiment.
122. Linnen, C. R., Kingsley, E. P., Jensen, J. D. & Hoekstra, H. E. On the origin and spread of an adaptive allele in deer mice. *Science* **325**, 1095–1098 (2009).
123. Barrett, R. D. H. et al. Linking a mutation to survival in wild mice. *Science* **363**, 499–504 (2019).
124. Steiner, C. C., Rompler, H., Boettger, L. M., Schoneberg, T. & Hoekstra, H. E. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Mol. Biol. Evol.* **26**, 35–45 (2009).
125. Hauser, M. T., Harr, B. & Schlötterer, C. Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene GLABROUS1. *Mol. Biol. Evol.* **18**, 1754–1763 (2001).
126. Hilscher, J., Schlötterer, C. & Hauser, M. T. A single amino acid replacement in ETC2 shapes trichome patterning in natural *Arabidopsis* populations. *Curr. Biol.* **19**, 1747–1751 (2009).
127. Daborn, P. J. et al. A single p450 allele associated with insecticide resistance in *Drosophila*. *Sci.* **297**, 2253–2256 (2002).
128. Crow, J. F. Genetics of insect resistance to chemicals. *Annu. Rev. Entomol.* **2**, 227–246 (1957).
129. Ffrench-Constant, R. H. The molecular genetics of insecticide resistance. *Genetics* **194**, 807–815 (2013).
130. McKenzie, J. A., Parker, A. G. & Yen, J. L. Polygenic and single gene responses to selection for resistance to diazinon in *Lucilia cuprina*. *Genetics* **130**, 613–620 (1992).
131. Kreiner, J. M. et al. Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. *Proc. Natl Acad. Sci. USA* **116**, 21076–21084 (2019).
132. Delye, C., Jasieniuk, M. & Le Corre, V. Deciphering the evolution of herbicide resistance in weeds. *Trends Genet.* **29**, 649–658 (2013).
133. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* **43**, 1160–1163 (2011).
134. Wang, R. L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
135. Studer, A. J., Wang, H. & Doebley, J. F. Selection during maize domestication targeted a gene network controlling plant and inflorescence architecture. *Genetics* **207**, 755–765 (2017).
136. Teotonio, H., Chelo, I. M., Bradic, M., Rose, M. R. & Long, A. D. Experimental evolution reveals natural selection on standing genetic variation. *Nat. Genet.* **41**, 251–257 (2009).
137. Teotonio, H. & Rose, M. R. Perspective: reverse evolution. *Evolution* **55**, 653–660 (2001).
138. Faria, V. G., Martins, N. E., Schlotterer, C. & Sucena, E. Readapting to DCV infection without *Wolbachia*: frequency changes of *Drosophila* antiviral alleles can replace endosymbiont protection. *Genome Biol. Evol.* **10**, 1783–1791 (2018).
139. Martins, N. E. et al. Host adaptation to viruses relies on few genes with different cross-resistance properties. *Proc. Natl Acad. Sci. USA* **111**, 5938–5943 (2014).
140. Huerta-Sanchez, E. et al. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol. Biol. Evol.* **30**, 1877–1888 (2013).
141. Beall, C. M. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc. Natl Acad. Sci. USA* **104**, 8655–8660 (2007).
142. Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226 (1983).
143. Rajpurohit, S., Parkash, R. & Ramniwas, S. Body melanization and its adaptive role in thermoregulation and tolerance against desiccating conditions in *Drosophila*. *Entomol. Res.* **38**, 49–60 (2008).
144. Willmer, P. G. & Unwin, D. M. Field analyses of insect heat budgets — reflectance, size and heating rates. *Oecologia* **50**, 250–255 (1981).
145. Jaksic, A. M. et al. Neuronal function and dopamine signaling evolve at high temperature in *Drosophila*. *Mol. Biol. Evol.* https://doi.org/10.1093/molbev/msaa116 (2020).

Acknowledgements
C.S. and J.H. are supported by various Austrian Science Fund (FWF) grants (W-1225-B20, P27630-B20, P29133-B29). Thanks to S. Allen, N. Barton, S.-K. Hsu, R. Kofler and T. MacKay for discussion and feedback on earlier versions of the manuscript.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Genetics thanks T. Mitchell Olds and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41576-020-0250-z.