Published in final edited form as:
Nat Rev Genet. 2012 December ; 13(12): 867–877. doi:10.1038/nrg3322.

Adaptive evolution: evaluating empirical support for theoretical predictions

Carrie F. Olson-Manning, **Maggie R. Wagner**, and **Thomas Mitchell-Olds***
Department of Biology, Box 90338, Program in Genetics and Genomics, Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708

Abstract

Adaptive evolution is shaped by the interaction of population genetics, natural selection and underlying network and biochemical constraints. Variation created by mutation, the raw material for evolutionary change, is translated into phenotypes by flux through metabolic pathways and by the topography and dynamics of molecular networks. Finally, the retention of genetic variation and the efficacy of selection depend on population genetics and demographic history. Emergent high-throughput experimental methods and sequencing technologies allow us to gather more evidence and to move beyond the theory in different systems and populations. Here we review the extent to which recent evidence supports long-established theoretical principles of adaptation.

Introduction

The emergence of adaptive alleles has fascinated evolutionary biologists for decades, inspiring the development of three broad, unifying concepts that underlie adaptive evolution. First, an allele's probability of fixation depends on the **effective population size** (*Ne*) of the population in which it arises. Large populations experience more mutations than small ones and thus harbor more potentially adaptive alleles. Furthermore, advantageous segregating alleles are less likely to be lost by random genetic drift in large populations and are more likely eventually to become fixed.

Second, an adaptive allele's fate depends on its frequency when positive selection begins to act. For *de novo* mutations, initial frequencies (p_0) are very low, and alleles are likely to be lost by chance, even when they are favored by selection. By contrast, neutral or mildly deleterious alleles may drift to intermediate frequency before becoming advantageous. Such intermediate frequency alleles are less likely to be eliminated by drift and more likely to be fixed by selection.

Third, the magnitude of an allele's beneficial effect determines how efficiently selection can act to increase its frequency. An allele's **selection coefficient** is positively related to its level of control over advantageous traits but is negatively related to possible deleterious pleiotropic effects. These emergent effects of a gene are partially underlain by molecular and biochemical phenomena such as epistatic interactions, linkage to other alleles and gene network position.

Together, *Ne*, p_0 and selection coefficients provide a foundation for understanding the fate of adaptive alleles (Fig. 1), but modern evolutionary biology requires more detailed and nuanced descriptions of these principles. In some systems, our relatively advanced

*tmo1@duke.edu.

The authors declare that they have no competing financial interests.

understanding of biochemistry and molecular biology allows us to explore in depth the functional mechanisms that determine the phenotypic and fitness effects of a mutation. In addition, the emergence of adaptive alleles depends on evolutionary as well as cellular constraints.

Advances in empirical methods have made possible the study and interpretation of genomic signals of adaptive evolution in different populations and systems. Here we review recent work that highlights the interplay of factors that constrain the emergence and maintenance of adaptive alleles. Moving beyond the theory of adaptive evolution, we discuss how adaptations can arise, what biochemical and physiological constraints they encounter, the nature of the **genetic architecture** and how the population context of alleles may determine their fate.

Some important disclaimers are required. Natural selection is not the only force that changes allele frequencies; other evolutionary forces include genetic drift, mutation, migration and biased gene conversion[1]. These non-selective processes influence many complex features of organisms. For instance, the assembly of genetic networks through gene duplication and subfunctionalization may be substantially influenced by neutral forces[2]. Also, most of the examples presented here have not explicitly tested the effect of genes or mutations on fitness (but see Ref. 3 for a discussion of this topic). We have chosen examples of phenotypes that plausibly increase fitness in nature, although in most cases the agents and mechanisms of selection were not investigated. We also highlight examples with experimentally imposed selection that avoid this pitfall but that have limitations of their own (reviewed in Ref. 4). In addition, we also consider cases in which factors such as deleterious standing variation, genetic drift, mutation pressure and migration influence the fate of alleles.

The impact of population size

Population size and adaptation: the theory

The probability that any adaptive allele will eventually become fixed or lost depends on the demography of the population in which it arises. One indicator of demographic influences is *N*e (Ref. 5), which can be influenced by factors such as sex ratio, variation in offspring number, inbreeding, mode of inheritance, age structure, changes in population size, spatial structure and genetic structure (reviewed in Ref. 6). In humans and other species with expanding populations, very large samples (consisting of >10,000 individuals) are needed to identify the rare, young polymorphisms that reflect current *N*e values and that may contribute to disease or adaptation from new mutations[7–9]. Smaller studies (involving hundreds of individuals) sample older, more common polymorphisms, reflecting historical *N*e values. Such standing variation may contribute to soft sweeps.

Populations with large *N*e values are likely to fix adaptive alleles regardless of how strongly they are favored (Fig. 1). They also encounter many mutations each generation6 and therefore have shorter waiting times for adaptive alleles to arise. Small populations, by contrast, sample far fewer mutations per generation and lose many favorable alleles to drift, even those with large selective advantages.

Evidence for population size effects

Despite abundant information on human genetics, we have only begun to understand how our recent evolutionary history influences the emergence of adaptive alleles. Consistent with our small historical *N*e of ~10,000 (Ref. 10), there are few clear examples of new alleles that have spread rapidly to fixation in human populations (reviewed in Ref. 11). Instead, recent population expansion has increased human *N*e (to ~1.1 million in Europe)[9], resulting in many rare variants that contribute to complex trait variation and disease[7,12].

Comparison of species with different *N*e values finds that orthologous proteins are less constrained in species with lower *N*e (Ref. 13). *Arabidopsis thaliana* also has a small historical *N*e and harbours excess rare and potentially deleterious mutations[14]. By contrast, *Drosophila melanogaster* has a large *N*e and appears to gain adaptive alleles frequently from new mutations[15]. The efficacy of selection was measured on three species of mice with similar ecologies but very different *N*e values[16]. As expected, the species with the highest historical *N*e had the most adaptive evolution and the most purifying selection[16]. Similar effects of *N*e have been found in two subspecies of rabbits[17]. However, in two species of *Drosophila*, higher *N*e values did not correlate with estimated adaptive evolution[18], emphasizing that other factors also have an impact on adaptive evolutionary change.

Where do adaptive alleles come from?

Predictions of origins and initial frequency

An allele favoured by natural selection may be either a new mutation[19] or a pre-existing, segregating genetic variant. If adaptation primarily depends on new mutations, then adaptive change may be limited by the time needed for new mutations to arise and by their stochastic loss. New recessive mutations are primarily found in heterozygotes and are therefore invisible to natural selection until they drift to a higher frequency, when recessive homozygotes become more frequent. In sexual diploids, this bias against the establishment of beneficial recessive mutations is known as Haldane's sieve[20]. This bias is less important in inbreeding species, in which recessive homozygotes occur more frequently[21]. However, experimental studies to date have not found consistent patterns in the degree of dominance for new mutations (reviewed in Ref. 22).

Alternatively, a pre-existing neutral allele can drift to intermediate frequency and can later become advantageous when selection pressures change owing to environmental change or to colonization of a new habitat. Alleles from standing variation have several advantages over new mutations. First, populations do not have to wait for new mutations to arise because the alleles are already present. Second, the allele may be at a higher frequency, decreasing the probability of loss by drift (Fig. 1). This is especially true for beneficial recessive mutations, as dominance has little influence on the fixation of alleles from standing variation, but it is very important for new mutations or very low frequency alleles[23]. Such pre-existing alleles may have already recombined onto several genetic backgrounds, which also 'hitchhike' to a higher frequency with the favored allele[19] in a process known as a soft sweep (Box 1).

Box 1

Inferences about adaptive evolution from sequence data: Caveats

Adaptive alleles may be derived from new mutations or from pre-existing standing variation. In a classic (hard) selective sweep, new mutations appear on a single haplotype that quickly goes to fixation. Therefore, when the genome is scanned after a hard sweep, a pattern of reduced nucleotide variation may be found near the adaptive mutation. By contrast, in soft sweeps, favored variants occur on multiple genetic backgrounds (haplotypes), and the variation among haplotypes obscures the reduction in polymorphism around favored variants[111], making soft sweeps difficult to detect. Such soft sweeps can occur if old, standing genetic variation has recombined onto several genetic backgrounds before it becomes advantageous or if independent advantageous mutations occur in parallel. Finally, signatures of selective sweeps may be obscured by population structure, and variation may resemble selection on standing variation[112]. Indeed, identifying soft selective sweeps has been one of the most challenging problems in molecular population genetics. However, a recent analysis of changes in allele

frequency at height-associated polymorphisms finds clear evidence for selection on pre-existing polygenic variants — an analytical approach that should be broadly applicable[38].

Several population characteristics influence the probability of a hard versus a soft sweep. Soft sweeps are more likely when populations are large or when mutation rates are high[113], whereas small populations or low mutation rates favorr hard sweeps from a single, new mutation. In addition, soft sweeps are more likely in widely distributed species with low migration rates[27], facilitating parallel sweeps in different parts of a species range.

Epistasis and rejecting neutrality

Many tests for selection use evidence of fixation of multiple adaptive mutations. For example, the McDonald–Kreitman test[114] examines polymorphism and interspecific divergence at multiple sites to detect natural selection. However, epistasis among sites in a protein may cause false rejection of the null hypothesis of neutral evolution[115]. Instead, a locus with an extreme McDonald–Kreitman statistic might actually reflect a neutral permissive mutation that changed protein functional constraints and that enabled subsequent neutral mutations to occur. Thus, given the demonstrated importance of epistasis within proteins[84,85], inferences of adaptive evolution made from the McDonald–Kreitman tests must be made with caution.

	Polymorphism	
	Old	New
Single origin	soft	hard
Parallel mutations	soft	soft

Adaptive change does not necessarily require fixation of advantageous alleles. Instead, modest changes in allele frequencies at many loci can increase population fitness[11]. The raw material for polygenic adaptation may come from standing variation, and slight changes in gene frequencies support the fine-tuning of a trait to a new or changing environment[24].

Does theory predict whether adaptive alleles are more likely to arise from new mutations, fixation of standing variation or modest changes at many loci? The answer lies in the effect size of the mutation on fitness[25], which is influenced by gene network position, linkage to other loci and epistatic interactions, as discussed in the section below.

Evidence for adaptive alleles arising from new mutations

Because large populations sample more new mutations and can efficiently select for weakly favored alleles, they should respond to natural selection more readily than small populations. In *D. melanogaster*, which has a very high *N*e, several adaptations have been attributed to new mutations[15,26]. In one striking example, convergent insecticide resistance arose in isolated *D. melanogaster* populations several times in the past 50 years[15]. Many insecticides target a neuronal signalling enzyme encoded by the *Acetylcholine esterase* (*Ace*) locus, and four sites in *Ace* have mutated to cause insecticide resistance in different local genetic backgrounds. In non-selective environments, the derived mutations are deleterious, but they confer a strong advantage in the presence of insecticide. The haplotype structure and deleterious effects under non-selective conditions rule out the possibility that these mutations arose from standing genetic variation (but see Box 1). This example demonstrates local hard sweeps for insecticide resistance, and a soft sweep with parallel changes at the species scale, as predicted by theory[27]. These results suggest that under strong selection, *D. melanogaster* can quickly sample and fix new adaptive alleles in response to selective

pressure. These results are consistent with genome-wide studies of polymorphisms[28] that find lower variation around amino acid substitutions than around synonymous substitutions, suggesting that selective sweeps are common in *D. melanogaster*.

Evidence for adaptive alleles arising from pre-existing variants

Adaptations arising from standing variation may be at moderate frequency when they become advantageous. Many well-characterized adaptive alleles arose from standing variation[29–31]. For instance, light coloration in beach mice has a single origin that has probably been selected from standing variation in melanocortin 1 receptor (*Mc1r*) in mainland populations[32]. Similarly, an allele for reduced plate armor in sticklebacks segregates in the source ocean population and has gone to fixation independently in several freshwater colonizations[30,33]. Several other examples (reviewed in Ref. 11) have shown that standing variation facilitates fast adaptive responses to strong selection.

Although the fixation of alleles is clearly important to adaptation, it is not required for large changes in phenotype[11], which may result from slight changes in allele frequency at many loci. Such changes are usually difficult to detect, yet several directed evolution studies in *D. melanogaster*[34,35] support this model of adaptation by polygenic allele frequency changes. Human populations that span a wide range of geographic locations and climates also appear to have evolved in this manner[36–38].

Because soft sweeps have been difficult to detect in natural populations (Box 1), empirical studies have so far struggled to infer their importance relative to hard sweeps. Patterns of nucleotide polymorphism support a role for hard sweeps in wild *D. melanogaster* populations[26], but a long-term selection experiment for accelerated development suggested soft or incomplete selective sweeps[34]. Few examples of hard sweeps have been found in genome-wide studies of species with a small historical *N*e (namely, humans and *Arabidopsis thaliana*[14,39]), although soft sweeps at multiple loci influencing human height have been inferred in European populations[38]. Thus, some combination of soft sweeps and poly-genic allele frequency changes may influence adaptive variation of complex traits.

Adaptation by introgression from other species

Introgression from closely related species appears to be an important but underappreciated source of new adaptive alleles. For example, dark coloration arose through artificial selection in domestic dogs by means of a mutation in *Mc1r*. This allele introgressed into wolves and rose in frequency in the boreal forest but not in tundra habitats, where the ancestral grey color is putatively advantageous[40]. Introgression between domesticated species and their wild relatives is an intriguing source of new alleles, because humans can maintain breeding populations of domesticated species that carry mutations that would be disfavored in wild populations. The exchange of color pattern mimicry genes by congeneric *Heliconius* butterflies[41] shows that natural selection can facilitate adaptive evolution by introgression. Similarly, a human leukocyte antigen (HLA) class I allele originating in archaic humans introgressed into modern Asian populations[42]. This allele spread and later introgressed back into African populations, suggesting that an important contributor to immune defense in modern humans arose through admixture.

Strength of selection

Theoretical models of effect size

Genes vary in their level of control over phenotypes, and the evolutionary fate of new mutations depends on their effect size on selectively important traits. Two contrasting models represent extremes along a continuum of effect sizes. Under the polygenic or

infinitesimal model, trait variation is controlled by many loci with small effects[24]. As the number of loci controlling a phenotype increases, their individual effects tend to decrease, and the average strength of selection on any locus is reduced[5]. Consequently, most beneficial mutations have a high probability of being lost, even in large populations, because their selective coefficients are small[5].

Alternatively, some adaptive alleles can be classified as **major genes**[43]. Because advantageous major alleles are theoretically much rarer than small-effect alleles[44] but are also less susceptible to loss by genetic drift[45], intermediate-size mutations are likely to be over-represented during adaptive evolution. Under some circumstances, this may result in an exponential distribution of effect sizes, with a few major and many minor genes contributing to adaptation[43,46]. Major-effect mutations include many unconditionally deleterious alleles (for example, Mendelian diseases) as well as variants with environment-specific fitness (for example, see Ref. 29), which might become adaptive when environments change.

Observations of effect size

Genomic analyses reveal a broad diversity of genetic architectures, including many small-effect genes, major genes and mixtures of polygenic and Mendelian determinants of trait variation. Human height exemplifies many of these patterns. At least 180 loci control variation within and among populations and together explain ~10% of phenotypic variation in height. Meta-analysis[47] finds that some small-effect genes controlling human height also have pleiotropic influences on other traits, especially physiological components related to type 2 diabetes risk. For height and many other traits, intermediate frequency alleles have small phenotypic effects, whereas rare alleles may have larger effects on traits[7,9,12] (Fig. 2).

Large studies of some traits support both the infinitesimal and major gene models. In *D. melanogaster*, strong artificial selection has resulted in large changes in SNP frequencies with the involvement of mostly small-effect loci[34,35], although some field experiments have found evidence of large-effect quantitative trait loci (QTLs)[48]. In general, most individuals express polygenic variation, whereas phenotypically extreme individuals may reflect non-additive gene action or rare, large-effect alleles[49]. These extremes may involve additional loci and pathways contributing to human disease and can be very valuable for breeding domesticated species[50]. Surprisingly, genes that contribute to both Mendelian and complex diseases in humans[51] show different molecular and network properties compared to loci controlling only Mendelian or only complex diseases; specifically, genes that control both complex and Mendelian diseases are longer, more highly transcribed, more tissue-specific in their expression and are embedded in more complex protein networks. The discovery of patterns in these genes is an important step towards predictive approaches to complex trait variation.

Networks and effect size: theory

The position of a gene product in its biochemical or regulatory network influences its effect size. Theory predicts that the first enzyme in a metabolic pathway will have more influence on the total output of a pathway than downstream enzymes[52,53] (but see Ref. 54). The same principle should apply to perception signal transduction pathways, in which the genes that directly interact with environmental stimuli, pathogen effectors or pheromones are upstream of all other genes that modulate responses to those exogenous signals. Similarly, a highly connected transcription factor (Fig. 3, nodes A and E) that integrates or directs expression of many genes is predicted to have stronger phenotypic effects than a peripheral regulator that influences only a few relevant genes[55] (Fig. 3, node F). Finally, even a gene with low **connectivity** can have a large effect if it has high **centrality**[56] (Fig. 3, node D).

Pleiotropic effects may change the net advantage of an allele. Which loci tend to affect multiple phenotypes? Crosstalk between biochemical pathways results from the partitioning of precursors and intermediates, such as when an enzyme's product is a substrate for multiple downstream enzymes[57]. Similarly, transcription factors with high connectivity to multiple gene networks affect more traits than those with only a few target genes[55](Fig. 3, nodes A and B).

Because natural selection works more efficiently on loci that exert large effects on phenotypes, the evolution of such loci is more likely to be constrained by purifying selection[58]. Most new mutations will quickly be purged from these loci, which will therefore accumulate little genetic diversity. By contrast, other loci may tolerate more mutational changes and may accumulate genetic variants that could become adaptive after a change in environment or genetic background[59]. The trade-off between strong positive selection and purifying selection has been demonstrated by simulated evolution of both biochemical[53] and regulatory networks[55](Box 2).

Box 2

Comparing biochemical and regulatory networks

How different are the evolutionary consequences of position in metabolic versus regulatory networks? The two network types reflect very different biological processes: biochemical pathways consist of a series of enzymes modifying chemical substrates, whereas regulatory networks may involve phosphorylation changes or soluble proteins binding to DNA. Nevertheless, they share analogous features that have important consequences for adaptive evolution.

For example, the phenotype produced by a linear metabolic pathway is the concentration of final chemical product, and for a linear regulatory cascade it is the concentration of final gene product. Genes included in more than one such pathway can influence multiple phenotypes. If a biochemical intermediate is generated faster than the next enzyme can process it, it will remain in solution or be diverted to compatible enzymes in other pathways, thus altering non-target phenotypes and creating possibly toxic intermediates[57]. Similarly, if the concentration of a transcription factor is higher than can simultaneously bind to a target gene, it may bind to alternative targets[116]. Finally, in much the same way as enzymes can evolve different substrate affinities and catalytic rates, transcription factors have evolvable affinities for DNA sequence motifs. In both cases, the interaction between two genes can be affected by mutations in either locus.

Recent simulations of regulatory and metabolic pathway evolution have identified similar patterns. A regulatory network adapting to ecological challenges *in silico* quickly developed a hierarchical structure dominated by a few global regulators[55]. Notably, these patterns were not observed in parallel simulations of regulatory network evolution under neutral conditions[55]. The analogous simulation of metabolic pathway evolution found that upstream enzymes tend to have the most control over phenotype and also be the target of natural selection[53]. In both cases, large-effect genes in upstream parts of the protein networks evolved under strong positive selection and then became constrained by strong purifying selection, after which further evolution of the system proceeded only through neutral or nearly neutral mutations in peripheral genes of small effect.

In conclusion, metabolic and regulatory networks share many of the properties that determine the effect size and pleiotropy of the networked genes. Enzymes and transcription factors with high connectivity — that is, receiving input from many other genes and/or transmitting output to many other genes — are likely to be highly

pleiotropic. In general, genes with high centrality — that is, genes at bottlenecks or branch points — are particularly important components of protein networks[56,57].

Networks and effect size: observations

When a molecular network is well-characterized, the evolutionary histories of genes occupying different positions can be compared. This approach was used to investigate selection on human innate immunity against bacterial pathogens[60]. The authors used resequencing data from 132 innate immune system genes to test the effects of network position on evolvability. Their results conform well to theoretical predictions: genes with a high connectivity and/or centrality harbor significantly less nucleotide diversity than genes on the periphery of the network[60]. The genes that are most likely to show signatures of positive or balancing selection are pathogen recognition receptors (the most upstream part of the network) and the downstream functional modules that carry out the immune response. By contrast, the most strongly conserved genes integrate inputs from the numerous pathogen recognition receptors and coordinate the activity of the downstream functional modules[60]. These highly connected, highly centralized global regulators are presumably so essential to the immune response that slightly deleterious mutations are quickly eliminated.

Similar studies of the insulin and target of rapamycin (TOR) pathways, which are instrumental in growth regulation and are highly conserved among eukaryotes[61], provide an interesting contrast to theoretical expectations and to the immunity case study[60]. Centrally located and highly connected genes in the human insulin and TOR pathways show increased positive selection[62]; this is the opposite pattern from the human immunity network, in which central genes are under stronger purifying selection and the peripheral genes show positive selection[60]. Furthermore, contrary to theoretical predictions, downstream genes in the insulin and TOR pathways are more constrained than upstream genes in *Caenorhabditis* spp.[63], vertebrates[64] and *Drosophila* spp.[65]. Several hypotheses are invoked to explain these unexpected patterns of selection on insulin and TOR network genes. The tendency of downstream genes to be more constrained than upstream genes was first attributed to a positive correlation between network position and intensity of pleiotropy[65]. Alvarez-Ponce *et al.*[64] revised this hypothesis to suggest that the pleiotropic effects of downstream genes simply have greater fitness consequences than those of upstream genes. Jovelin and Philips[63] concluded that differential expression level is the best explanati on for stronger purifying selection on downstream insulin and TOR genes. Usually, however, upstream genes show evidence of stronger selection, whether it be purifying[53,60] or positive selection[66–68]. Notably, selection on genes that interact directly with the environment has been implicated in the evolution of olfactory perception in mammals[69] and light perception in cichlid fishes[70]. Because some examples support predictions from network theory but others consistently defy them, it is clear that our understanding of adaptive gene network evolution is incomplete.

Regulatory variation: circumventing pleiotropic constraint

Deleterious pleiotropic effects may be circumvented by mutations in regulatory elements that allow flexibility in the amount, timing and location of gene expression[71]. For instance, a missense mutation could simultaneously eliminate all functions of a protein, which could have serious pleiotropic consequences. By contrast, a regulatory mutation might alter gene expression only in specific tissues, developmental stages or environmental conditions. The duplication or deletion of regulatory elements could incrementally increase or decrease concentration of a gene product. Thus, the modular nature of gene regulatory sequence may allow a gene to evolve with less pleiotropic constraint[71].

In practice, adaptive molecular evolution may involve coding mutations in key proteins: for example, tetrodotoxin resistance in garter snakes[72]. Regulatory variants, however, have recently been implicated in the evolution of adaptive traits, such as sex pheromones in moths[73] and resistance to Lassa fever in humans[74]. Notably, mutations in tissue-specific *cis*-regulatory elements allowed the loss of seed shattering in rice and *Brassica* spp.[75] and pelvises in sticklebacks[76]; in both cases, full knockout of the target gene had lethal or clearly deleterious pleiotropic effects.

Epistasis and adaptation

Predicting the effects of epistasis

To understand how genes influence traits under selection, we must consider how their genetic context influences their selection coefficients. Epistatic interactions are commonly visualized on a theoretical surface known as an **adaptive landscape**[77]. The shape of an adaptive landscape heavily depends on the extent and nature of **epistasis** (Fig. 4).

In the absence of epistasis, adaptive alleles are equally advantageous regardless of genetic background (Fig. 4a). In this scenario, populations can readily evolve towards a nearby adaptive peak. However, if epistasis is common, the benefit of an adaptive allele depends on pre-existing and subsequent mutations, potentially preventing a population from climbing the nearest fitness peak[78]. Even the type of epistasis may influence the progress towards a local peak. **Sign epistasis**[79] (Fig. 4c) creates a rough fitness landscape in which the chronological order of adaptive substitutions is important, because many mutational pathways contain unfit steps. Evolution to a fitness optimum may be impossible in some genetic backgrounds because all single mutational steps towards the optimum are inaccessible owing to deleterious epistatic interactions (Fig. 4c). However, with **negative epistasis** (Fig. 4b), a series of adaptive substitutions can take place in almost any order: the first substitution causes the greatest fitness gain and subsequent changes have diminishing returns. Most theoretical and experimental research in this area has focused on haploids owing to the additional complications that arise with diploidy[80].

Evidence for the effects of epistasis

When epistatic interactions occur among mutations, the order in which multiple adaptive changes arise may make a large difference. It was noted that the rate of adaptation slows as populations approach an optimum: large-effect changes occur first, and minor changes occur last[81]. This observation could be caused by one of two principles. In one scenario, the first allele to become fixed does so because it has the highest selective advantage regardless of background, and major alleles generally reach fixation more quickly. Alternatively, the selective advantage of any beneficial allele is context-dependent, such that any early-arising allele has larger beneficial effects than any subsequent mutation.

Two independent experimental evolution studies find that epistasis between adaptive alleles shows diminishing returns from late-occurring mutations. In one study, the endogenous *Methylobacterium extorquens* methanol metabolism pathway was replaced with an engineered pathway and selected for use of methanol as the sole carbon source[82]. Fitness measurements of all combinations of the five beneficial mutations that arose during adaptation showed that four of the five mutations exhibit decelerating fitness gains during adaptation (Fig. 4f). The other study investigated the beneficial mutations that arose during a long-term *Escherichia coli* evolution experiment on glucose-limited media[83]. Again, when all combinations of beneficial mutations were tested, all mutations increased fitness, but smaller selective increments were seen in higher fitness backgrounds. In both of these scenarios, the order of the mutations appears to be irrelevant, as the first mutation gives a

large increase in fitness, the second gives a lesser gain, and so on. Thus, in these cases, all mutational trajectories are accessible. The magnitude of the fitness gain caused by an adaptive allele heavily depends on the preceding mutations, but its sign is always positive.

In contrast to the smooth landscape seen for multi-locus adaptation, it was found the landscape of within-protein adaptive alleles is very rough[84]. Only 18 of 120 possible trajectories of beneficial mutations in the *E. coli TEM beta-lactamase* gene were favored by natural selection (Fig. 4f). This pattern also holds for the fitness landscape of isopropylmalate dehydrogenase (IMDH)[84], in which only 29% of possible mutational trajectories are selectively accessible. Within-protein epistasis is strong, and substantial sign epistasis makes the chronological order of adaptive alleles important. As shown by Fig. 4f, each of the five mutations increases the selective advantage, but their positive effect strongly depends on the background in which they arose. Given the clear importance of sign epistasis in proteins[85], neutral, permissive mutations probably occur frequently[86], constantly changing selective coefficients of nearby mutations (Box 1).

Among the well-characterized examples of natural adaptive alleles, epistasis appears to be extremely important. In beach mice, the interaction between MC1R and agouti controls coat colour. The agouti protein determines the expression of MC1R and is a stronger antagonist in light coloured mice, in which MC1R function is impaired[29]. Within-protein epistasis is also responsible for widespread drug resistance in the H1N1 virus[87]. Although they do not provide the same level of mechanistic detail as experimental evolution, these natural examples indicate that epistasis has an important role in natural variation, at least in the large-effect genes that are most easily detected in nature.

Recombination and adaptation

Predictions of the effects of recombination

Another phenomenon that affects the emergence of adaptive alleles is recombination, which can separate adaptive alleles from linked maladaptive ones and thus can facilitate their fixation[88] (Fig. 5c). Tightly linked deleterious sites reduce the net selective coefficients of adaptive alleles. Additionally, alleles can spread among populations more easily when they are in regions of high recombination. Thus, theory predicts that adaptive alleles are more likely to arise in areas of high recombination (reviewed in Ref. 89).

Alternatively, recombination may decrease fitness by separating co-adapted alleles[90] (Fig. 5a, b). For example, when populations adapt to different niches, alleles that are advantageous in one environment may be deleterious in another. Decreased recombination between these sets of co-adapted genes reduces the production of offspring that are maladapted to both environments (Fig. 5a). Thus, chromosomal inversions containing multiple co-adapted alleles increase fitness because they suppress recombination between divergent types (Fig. 5b), resulting in fewer unfit recombinant offspring[91]. In some circumstances, therefore, reduced recombination may be advantageous.

Evidence of the effects of recombination

Generally, recombination is expected to increase the efficacy of selection for adaptive alleles (Fig. 5c). For instance, proteins in regions of low recombination suffer from a larger load of deleterious mutations and fewer beneficial mutations compared to those in regions of high recombination[92]. Additional evidence comes from the positive correlation between recombination rate and nucleotide diversity[93,94], which could be generated by **background selection** or hitchhiking near adaptive alleles. In humans, divergence among populations is negatively correlated with recombination rate[95], suggesting that alleles in areas of high recombination can spread more easily among populations because they are less constrained

by linked deleterious mutations. Although this is only an indirect link between recombination and the emergence of adaptive alleles (and may be confounded by background selection), it is clear that recombination can increase the efficiency of natural selection[90].

When multiple loci control complex phenotypes, however, recombination may break up groups of co-adapted alleles. Therefore, a non-recombining inversion can function as a single locus, and the collective effects of multiple loci may jointly be favored[96]. Two recent studies[97,98] document cases in which sympatric but ecologically distinct ecotypes maintain their respective sets of adaptive alleles on inversions (Fig. 5d, e). These examples highlight the importance of inversions for maintaining specific trait combinations in locally adapted, sympatric species. However, several questions remain. How common are adaptive inversions in nature? How do inversions gain adaptive alleles? Does an inversion event prevent the gain of new genes in the future, or can new adaptive alleles be 'captured' by an inversion after it has been established?

Models of migration and adaptation

Migration influences the emergence of advantageous alleles in several ways. Gene flow between populations that are locally adapted to divergent environments may increase genetic variability within habitats[99], some of which may lend itself to adaptation. Indeed, recent theory suggests that high gene flow among locally adapted populations encourages the emergence of fairly large-effect QTLs[100] that may consist of several linked, small-effect loci. However, if the migration from one habitat overwhelms the other, migration from the source could prevent adaptation to marginal habitats[101]. Finally, gene duplicates may be selectively favored when nearby populations experience maladaptive gene flow[102] because duplicated local alleles may be able to mask maladapted immigrant alleles. In particular, tandem duplications can increase in frequency when populations experience immigration of locally deleterious alleles, suggesting that such migration–selection balance would favour increased copy number variation, which is common in many species (for example, see Ref. 103).

Migration and adaptation in practice

The prediction that migration–selection balance causes higher levels of genetic variation in environmentally heterogeneous areas is supported by a large study in lodgepole pines, in which regional climatic heterogeneity predicts complex trait variation in common gardens[104]. This positive correlation between environmental and genetic variation suggests that the immigration of deleterious alleles from divergent environments increases local genetic diversity. However, a recent experiment with *D. melanogaster* in heterogeneous laboratory environments found no evidence that environmental heterogeneity increases genetic variation[105]. Several recent studies of gene flow among plant populations support the prediction that immigration of maladapted alleles may limit the capacity of populations to adapt to local environmental conditions[106,107]. Future progress in understanding the role of migration–selection balance and whether environmental heterogeneity can maintain genetic variation will benefit from experimental studies in natural populations[105], especially for known QTLs experiencing natural selection.

Conclusions and perspective

The integration of evolutionary theory with modern genetics and genomics has yielded deep insights into the emergence of adaptive alleles. Although a full understanding of evolution must also consider non-selective forces[3], a focus on the process of adaptation is especially

valuable in medicine, agriculture and conservation biology, in which the functionality of adaptive changes has direct importance for humans and our environment.

Recent empirical evidence supports some major theoretical predictions but not others. For instance, the implication of both small-effect and large-effect QTLs in genetic adaptation suggests that both the infinitesimal and Mendelian models may be correct[49,51]. Similarly, the importance of epistasis[82,83,87] and recombination[97,98] in adaptive evolution is well-supported, and examples of adaptive alleles originating both from new mutations[15,26] and standing genetic variation[29–32] have been documented.

However, conflicting results prevent resolution of long-standing theoretical debates about the roles of hard and soft selective sweeps[14,26,39], and the effect of N_e on the efficiency of selection[18]. It is possible that these contradictions are simply due to inherent difficulties in detecting sequence signatures of adaptive evolution (Box 1); the availability of more sequences and improved analytical methods may eventually lead to more conclusive results. Additionally, although some studies validate predictions from molecular network theory[60,66], other examples consistently defy expectations[62–65], indicating that the effects of pathway and network position on adaptive value are more complex than previously thought. More detailed studies of the functional effects of gene product interactions may shed light on these confusing patterns.

This Review has addressed adaptive evolution and how population genetic factors influence the emergence of alleles. Thus, we have not discussed the selective environment itself, which is obviously a major factor shaping the course of adaptive evolution. For example, biotic agents of selection such as mate choice, predators or pathogens are more likely to elicit responses through large-effect QTLs[68] and strong balancing selection than are abiotic selection pressures[108,109]. Furthermore, adaptation in complex environments will probably differ from adaptation in simple environments, such as those in laboratory experiments — not only because selection pressures vary with the environment but also because gene effects often interact with environmental stimuli. Alleles that are beneficial in one habitat often have different fitness effects in other environments, and this has important consequences for population variation and evolutionary change.

Our empirical understanding of adaptive evolution has greatly progressed in the past decade, yet many challenges remain. Pathway and network analyses of complex traits[110] offer great potential to elucidate the causes of phenotypic variation and evolutionary change. Apart from agents of selection and gene-by-environment interactions (which are best investigated using a combination of functional genomics, evolutionary genetics and environmental context), many questions about adaptation at the molecular level could be resolved by additional work in the systems reviewed here. Experimental evolution is a particularly promising method for isolating individual factors affecting adaptation and for explicitly testing the interplay among such factors, yet we do not know whether the inferences drawn from such studies can apply to natural systems. Studies of human complex traits and disease have catalyzed a revolution in omics technologies that benefits studies in many organisms, promising additional opportunities to test evolutionary hypotheses.

Acknowledgments

We thank J. T. Anderson, D. Lowry, P. Magwene, A. Schmid and three anonymous reviewers for helpful discussion and comments. We also acknowledge the authors of the many excellent studies that could not be included in this review due to space limitations. This work was supported by the National Institutes of Health (award R01 GM086496 to TM-O; Training Grant 5T32GM007754-32 to Duke University Program in Genetics and Genomics) and the National Science Foundation (award EF-0723447 to TM-O; dissertation grant 1110445 to CO-M).

GLOSSARY

Selection coefficient	The proportional change in fitness due to a mutation
Pleiotropy	The control of multiple traits by a single locus
Effective population size (N_e)	The size of an ideal population that would experience genetic drift at the same rate as the actual population
Genetic architecture	Also called trait architecture, this is a summary of the allelic effects and frequencies of the genes underlying a given phenotypic trait and their emergent properties, such as epistasis, pleiotropy and modularity
Introgression	The introduction of a gene into a population or species by crossing with a different population or species
Major genes	Phenotypic differences between alleles at a locus can range from large phenotypic effects (major genes) to small effects (minor genes). A major gene at intermediate frequency will control a fairly large proportion of trait variation but would have little effect on variation when rare
Connectivity	A gene's degree of connectivity is the number of direct links to or from other genes in its network
Centrality	Sometimes referred to as 'between-ness', a gene's degree of centrality is the number of shortest paths between pairs of other genes that must pass through it
Epistasis	Interaction between genes. The dependence of the effect of a mutation on other sites in the same gene or in other loci, resulting in non-additive effects on phenotype
Adaptive landscape	A conceptual surface that describes the fitness of all possible genotypic combinations of an organism
Sign epistasis	Epistatic interactions for which the phenotypic effect of a mutation has a different sign in different genetic backgrounds
Negative epistasis	Epistatic interactions that lessen the magnitude of trait changes, such that combined effects of beneficial mutations are less than their individual effects
Mullerian mimicry	When two or more poisonous species mimic each other's warning signals for shared predators
Background selection	The purging of non-deleterious alleles that are closely linked to deleterious sites

References

1. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genom Hum Genet. 2009; 10:285–311.
2. Lynch M. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA. 2007; 104:8597–8604. [PubMed: 17494740]
3. Barrett RDH, Hoekstra HE. Molecular spandrels: tests of adaptation at the genetic level. Nature Rev Genet. 2011; 12:767–780. This Review emphasizes that genetic architecture and pleiotropy can

confound inferences of adaptive evolution, hence conclusive evidence that a locus is adaptive must demonstrate functional connections between genotype, phenotype and fitness. [PubMed: 22005986]

4. Kawecki TJ, et al. Experimental evolution. Trends Ecol Evol. 2012; 27:547–560. [PubMed: 22819306]
5. Crow, JF.; Kimura, M. Burgess. 1970. An Introduction to Population Genetics Theory.
6. Charlesworth B. Fundamental conceptsx in genetics: effective population size and patterns of molecular evolution and variation. Nature Rev Genet. 2009; 10:195–205. [PubMed: 19204717]
7. Nelson MR, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 2012; 337:100–104. [PubMed: 22604722]
8. Tennessen JA, Akey JM. Parallel adaptive divergence among geographically diverse human populations. PLoS Genet. 2011; 7:e1002127. [PubMed: 21698142]
9. Keinan A, Clark AG. Recent explosive human population growth has resulted in an excess of rare genetic variants. Science. 2012; 336:740–743. This theoretical analysis of expanding human populations predicts an excess of young, rare alleles that may contribute to human disease. Detection of such alleles requires very large population samples, such as the recent empirical studies reported in references 7 and 8, which correspond well to these models. [PubMed: 22582263]
10. Yu N, Jensen-Seaman MI, Chemnick L, Ryder O, Li WH. Nucleotide diversity in gorillas. Genetics. 2004; 166:1375–1383. [PubMed: 15082556]
11. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr Biol. 2010; 20:R208–R215. [PubMed: 20178769]
12. Tennessen JA, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012; 337:64–69. [PubMed: 22604720]
13. Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. Quantifying the slightly deleterious mutation model of molecular evolution. Mol Biol Evol. 2002; 19:2142–2149. [PubMed: 12446806]
14. Cao J, et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nature Genet. 2011; 43:956–963. [PubMed: 21874002]
15. Karasov T, Messer PW, Petrov DA. Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. PLoS Genet. 2010; 6:e1000924. [PubMed: 20585551]
16. Phifer-Rixey M, et al. Adaptive evolution and effective population size in wild house mice. Mol Biol Evol. 2012; 29:2949–2955. [PubMed: 22490822]
17. Carneiro M, et al. Evidence for widespread positive and purifying selection across the European rabbit (*Oryctolagus cuniculus*) genome. Mol Biol Evol. 2012; 29:1837–1849. [PubMed: 22319161]
18. Andolfatto P, Wong KM, Bachtrog D. Effective population size and the efficacy of selection on the X chromosomes of two closely related *Drosophila* species. Genome Biol Evol. 2011; 3:114–128. [PubMed: 21173424]
19. Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974; 23:23–35. [PubMed: 4407212]
20. Haldane JBS. A mathematical theory of natural and artificial selection. Part II the influence of paritial self-fertilisation, inbreeding, assortative mating, and selective fertilisation on the composition of Mendelian populations, and on natural selection. Biol Rev. 1924; 1:158–163.
21. Charlesworth B. Evolutionary rates in partially self-fertilizing species. Am Nat. 1992; 140:126–148. [PubMed: 19426068]
22. Halligan DL, Keightley PD. Spontaneous mutation accumulation studies in evolutionary genetics. Annu Rev Ecol Evol Syst. 2009; 40:151–172.
23. Teshima KM, Przeworski M. Directional positive selection on an allele of arbitrary dominance. Genetics. 2006; 172:713–718. [PubMed: 16219788]
24. Lynch, M.; Walsh, B. Genetics and Analysis of Quantitative Traits. Sinauer Press; 1998.
25. Hermisson J, Pennings PS. Soft sweeps molecular population genetics of adaptation from standing genetic variation. Genetics. 2005; 169:2335–2352. [PubMed: 15716498]
26. Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the *Drosophila* genome? PLoS Genet. 2009; 5:e1000495. [PubMed: 19503600]

27. Ralph P. Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics. 2010; 186:647–668. [PubMed: 20660645]
28. Sattath S, Elyashiv E, Kolodny O, Rinott Y, Sella G. Pervasive adaptive protein evolution apparent in diversity patterns around amino acid substitutions in *Drosophila simulans*. PLoS Genet. 2011; 7:e1001302. [PubMed: 21347283]
29. Steiner CC, Weber JN, Hoekstra HE. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol. 2007; 5:e219. [PubMed: 17696646]
30. Colosimo PF, et al. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science. 2005; 307:1928–1933. [PubMed: 15790847]
31. Jones FC, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012; 484:55–61. [PubMed: 22481358]
32. Domingues VS, Poh YP, Peterson BK. Evidence of adaptation from ancestral variation in young populations of beach mice. Evolution. 2012; 66:3209–3223. [PubMed: 23025610]
33. Barrett RDH, Schluter D. Adaptation from standing genetic variation. Trends Ecol Evol. 2008; 23:38–44. [PubMed: 18006185]
34. Burke MK, et al. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature. 2010; 467:587–590. [PubMed: 20844486]
35. Teotónio H, Chelo IM, Bradic M, Rose MR, Long AD. Experimental evolution reveals natural selection on standing genetic variation. Nature Genet. 2009; 41:251–257. [PubMed: 19136954]
36. Tennessen J, Akey J. Parallel adaptive divergence among geographically diverse human populations. PLoS Genet. 2011; 7:e1002127. [PubMed: 21698142]
37. Hancock A, et al. Adaptations to climate-mediated selective pressures in humans. PLoS Genet. 2011; 7:e1001375. [PubMed: 21533023]
38. Turchin MC, et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nature Genet. 2012; 44:1015–1019. This landmark paper reports the development of methods to test for soft selective sweeps at many loci and shows substantial natural selection on standing variation at hundreds of genes influencing human height. [PubMed: 22902787]
39. Hernandez RD, et al. Classic selective sweeps were rare in recent human evolution. Science. 2011; 331:920–924. [PubMed: 21330547]
40. Anderson TM, et al. Molecular and evolutionary history of melanism in North American grey wolves. Science. 2009; 323:1339–1343. [PubMed: 19197024]
41. Dasmahapatra KK, et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 2012; 487:94–98. [PubMed: 22722851]
42. Abi-Rached L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011; 334:89–94. [PubMed: 21868630]
43. Orr H. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution. 1998; 52:935–949.
44. Fisher, R. The Genetical Theory of Natural Selection. Oxford Univ. Press; 1930.
45. Kimura M. On the probability of fixation of mutant genes in a population. Genetics. 1962; 47:713–719. [PubMed: 14456043]
46. Rockman MV. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution. 2012; 66:1–17. [PubMed: 22220860]
47. Lanktree MB, et al. Meta-analysis of dense genecentric association studies reveals common and uncommon variants associated with height. Am J Hum Genet. 2011; 88:6–18. [PubMed: 21194676]
48. Loeschcke V, Kristensen TN, Norry FM. Consistent effects of a major QTL for thermal resistance in field-released *Drosophila melanogaster*. J Insect Physiol. 2011; 57:1227–1231. [PubMed: 21708160]
49. Chan Y, et al. Common variants show predicted polygenic effects on height in the tails of the distribution, except in extremely short individuals. PLoS Genet. 2011; 7:e1002439. [PubMed: 22242009]

50. Xu K, et al. *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature. 2006; 442:705–708. [PubMed: 16900200]
51. Jin W, Qin P, Lou H, Jin L, Xu S. A systematic characterization of genes underlying both complex and Mendelian diseases. Hum Mol Genet. 2012; 21:1611–1624. This analysis finds an important overlap between Mendelian and complex disease genes and shows that genes contributing to both disease categories show substantial functional differences from other loci. [PubMed: 22186022]
52. Kacser H. The molecular basis of dominance. Genetics. 1981; 97:639–666. [PubMed: 7297851]
53. Wright KM, Rausher MD. The evolution of control and distribution of adaptive mutations in a metabolic pathway. Genetics. 2010; 184:483–502. [PubMed: 19966064]
54. Tarutani Y, et al. *Trans*-acting small RNA determines dominance relationships in *Brassica* self-incompatibility. Nature. 2010; 466:983–986. [PubMed: 20725042]
55. Jenkins DJ, Stekel DJ. *De novo* evolution of complex, global and hierarchical gene regulatory mechanisms. J Mol Evol. 2010; 71:128–140. [PubMed: 20680619]
56. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comp Biol. 2007; 3:e59.
57. Yamada T, Bork P. Evolution of biomolecular networks — lessons from metabolic and protein interactions. Nature Rev Mol Cell Biol. 2009; 10:791–803. [PubMed: 19851337]
58. Ohta T. Slightly deleterious mutant substitutions in evolution. Nature. 1973; 246:96–98. [PubMed: 4585855]
59. Stoltzfus A. On the possibility of constructive neutral evolution. J Mol Evol. 1999; 49:169–181. [PubMed: 10441669]
60. Casals F, et al. Genetic adaptation of the antibacterial human innate immunity network. BMC Evol Biol. 2011; 11:202. [PubMed: 21745391]
61. van Dam TJ, Zwartkruis FJ, Bos JL, Snel B. Evolution of the TOR pathway. J Mol Evol. 2011; 73:209–220. [PubMed: 22057117]
62. Luisi P, et al. Network-level and population genetics analysis of the insulin/TOR signal transduction pathway across human populations. Mol Biol Evol. 2012; 29:1379–1392. [PubMed: 22135191]
63. Jovelin R, Phillips PC. Expression level drives the pattern of selective constraints along the insulin/TOR signal transduction pathway in *Caenorhabditis*. Genome Biol Evol. 2011; 3:715–722. [PubMed: 21849326]
64. Alvarez-Ponce D, Aguadé M, Rozas J. Comparative genomics of the vertebrate insulin/TOR signal transduction pathway: a network-level analysis of selective pressures. Genome Biol Evol. 2011; 3:87–101. [PubMed: 21149867]
65. Alvarez-Ponce D, Aguadé M, Rozas J. Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 *Drosophila* genomes. Genome Res. 2009; 19:234–242. [PubMed: 19141596]
66. Olson-Manning CF, Lee C-R, Rausher MD, Mitchell-Olds T. Evolution of flux control in the glucosinolate pathway in *Arabidopsis thaliana*. Mol Biol Evol. Aug 25.2012 10.1093/molbev/mss204.
67. Flowers JM, et al. Adaptive evolution of metabolic pathways in *Drosophila*. Mol Biol Evol. 2007; 24:1347–1354. [PubMed: 17379620]
68. Prasad KVSK, et al. A gain-of-function polymorphism controlling complex traits and fitness in nature. Science. 2012; 337:1081–1084. [PubMed: 22936775]
69. Gilad Y, Bustamante CD, Lancet D, Pääbo S. Natural selection on the olfactory receptor gene family in humans and chimpanzees. Am J Hum Genet. 2003; 73:489–501. [PubMed: 12908129]
70. Spady TC, et al. Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. Mol Biol Evol. 2005; 22:1412–1422. [PubMed: 15772376]
71. Wray GA. The evolutionary significance of *cis*-regulatory mutations. Nature Rev Genet. 2007; 8:206–216. [PubMed: 17304246]

72. Feldman CR, Brodie ED, Pfrender ME. Genetic architecture of a feeding adaptation: garter snake (*Thamnophis*) resistance to tetrodotoxin bearing prey. Proc Biol Sci. 2010; 277:3317–3325. [PubMed: 20522513]
73. Albre J, et al. Sex pheromone evolution is associated with differential regulation of the same desaturase gene in two genera of leafroller moths. PLoS Genet. 2012; 8:e1002489. [PubMed: 22291612]
74. Andersen K, et al. Genome-wide scans provide evidence for positive selection of genes implicated in Lassa fever. Phil Trans R Soc B. 2012; 367:868–877. [PubMed: 22312054]
75. Arnaud N, Lawrenson T, Ostergaard L. The same regulatory point mutation changed seed-dispersal structures in evolution and domestication. Curr Biol. 2011; 21:1215–1219. [PubMed: 21737279]
76. Chan YF, et al. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science. 2010; 327:302–305. This study demonstrated that the parallel adaptive evolution of pelvis loss in stickleback fish occurred via repeated mutations in a tissue-specific regulatory element, which eliminated protein function in pelvic tissue while avoiding lethal pleiotropic effects of protein knockout in other tissues. [PubMed: 20007865]
77. Wright S. The roles of mutation, inbreeding, cross-breeding and selection in evolution. Proc 6th Int Congress Genet. 1932; 1:356–366.
78. Povolotskaya IS, Kondrashov FA. Sequence space and the ongoing expansion of the protein universe. Nature. 2010; 465:922–926. [PubMed: 20485343]
79. Weinreich DM, Watson RA, Chao L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution. 2005; 59:1165–1174. [PubMed: 16050094]
80. Taylor CF, Higgs PG. A population genetics model for multiple quantitative traits exhibiting pleiotropy and epistasis. J Theor Biol. 2000; 203:419–437. [PubMed: 10736218]
81. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-term experimental evolution in *Escherichia coli* 1. Adaptation and divergence during 2,000 generations. Am Nat. 1991; 138:1315–1341.
82. Chou HH, Chiu HC, Delaney NF, Segre D, Marx CJ. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science. 2011; 332:1190–1192. [PubMed: 21636771]
83. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. Negative epistasis between beneficial mutations in an evolving bacterial population. Science. 2011; 332:1193–1196. [PubMed: 21636772]
84. Weinreich DM, Delaney NF, DePristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 2006; 312:111–114. [PubMed: 16601193]
85. Lunzer M, Golding GB, Dean AM. Pervasive cryptic epistasis in molecular evolution. PLoS Genet. 2010; 6:e1001162. [PubMed: 20975933]
86. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. Crystal structure of an ancient protein: evolution by conformational epistasis. Science. 2007; 317:1544–1548. [PubMed: 17702911]
87. Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB. Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet. 2011; 7:e1001301. [PubMed: 21390205]
88. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. Genet Res. 1966; 8:269–294. [PubMed: 5980116]
89. Charlesworth B. The effects of deleterious mutations on evolution at linked sites. Genetics. 2012; 190:5–22. [PubMed: 22219506]
90. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics. 2006; 173:419–434. [PubMed: 16204214]
91. Dobzhansky, TG. Genetics of the Evolutionary Process. Columbia Univ. Press; 1970.
92. Presgraves DC. Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr Biol. 2005; 15:1651–1656. [PubMed: 16169487]
93. Cutter AD, Choi JY. Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*. Genome Res. 2010; 20:1103–1111. [PubMed: 20508143]
94. Cutter AD, Moses AM. Polymorphism, divergence, and the role of recombination in *Saccharomyces cerevisiae* genome evolution. Mol Biol Evol. 2011; 28:1745–1754. [PubMed: 21199893]

95. Keinan A, Reich D. Human population differentiation is strongly correlated with local recombination rate. PLoS Genet. 2010; 6:e1000886. [PubMed: 20361044]
96. Hoffmann AA, Rieseberg LH. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst. 2008; 39:21–42. [PubMed: 20419035]
97. Joron M, et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature. 2011; 477:203–206. [PubMed: 21841803]
98. Lowry DB, Willis JH. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. 2010; 8:e1000500. [PubMed: 20927411]
99. Spichtig M, Kawecki T. The maintenance (or not) of polygenic variation by soft selection in heterogeneous environments. Am Nat. 2004; 164:70–84. [PubMed: 15266372]
100. Yeaman S, Whitlock M. The genetic architecture of adaptation under migration-selection balance. Evolution. 2011; 65:1897–1911. [PubMed: 21729046]
101. Kawecki T. Adaptation to marginal habitats. Annu Rev Ecol Evol Syst. 2008; 39:321–342.
102. Yanchukov A, Proulx S. Invasion of gene duplication through masking for maladaptive gene flow. Evolution. 2012; 66:1543–1555. [PubMed: 22519789]
103. Zhou J, Lemos B, Dopman EB, Hartl DL. Copy-number variation: the balance between gene dosage and expression in *Drosophila melanogaster*. Genome Biol Evol. 2011; 3:1014–1024. [PubMed: 21979154]
104. Yeaman S, Jarvis A. Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. Proc R Soc B. 2006; 273:1587–1593.
105. Yeaman S, Chen Y, Whitlock MC. No effect of environmental heterogeneity on the maintenance of genetic variation in wing shape in *Drosophila melanogaster*. Evolution. 2010; 64:3398–3408. [PubMed: 20624178]
106. Anderson JT, Geber MA. Demographic source-sink dynamics restrict local adaptation in Elliott's blueberry (*Vaccinium elliottii*). Evolution. 2010; 64:370–384. [PubMed: 19703223]
107. Paul J, Sheth S, Angert A. Quantifying the impact of gene flow on phenotype-environment mismatch: a demonstration with the scarlet monkeyflower *Mimulus cardinalis*. Am Nat. 2011; 178:S62–S79. [PubMed: 21956093]
108. Gan X, et al. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature. 2011; 477:419–423. [PubMed: 21874022]
109. Louthan AM, Kay KM. Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. BMC Evol Biol. 2011; 11:60. [PubMed: 21385379]
110. Califano A, Butte AJ, Friend S, Ideker T, Schadt E. Leveraging models of cell regulation & GWAS data in integrative network-based association studies. Nature Genet. 2012; 44:841–847. [PubMed: 22836096]
111. Nielsen R, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 2005; 3:976–985.
112. Wall JD, Andolfatto P, Przeworski M. Testing models of selection and demography in *Drosophila simulans*. Genetics. 2002; 162:203–216. [PubMed: 12242234]
113. Pennings PS, Hermisson J. Soft sweeps II-Molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol. 2006; 23:1076–1084. [PubMed: 16520336]
114. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature. 1991; 351:652–654. [PubMed: 1904993]
115. Fay JC. Weighing the evidence for adaptation at the molecular level. Trends Genet. 2011; 27:343–349. [PubMed: 21775012]
116. Segal E, Widom J. From DNA sequence to transcriptional behaviour: a quantitative approach. Nature Rev Genet. 2009; 10:443–456. [PubMed: 19506578]
117. Cutter AD. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. Genetics. 2005; 172:171–184. [PubMed: 16272415]
118. Shapiro JA, et al. Adaptive genic evolution in the *Drosophila* genomes. Proc Natl Acad Sci USA. 2007; 104:2271–2276. [PubMed: 17284599]

119. Charlesworth J, Eyre-Walker A. The rate of adaptive evolution in enteric bacteria. Mol Biol Evol. 2006; 23:1348–1356. [PubMed: 16621913]

Biographies

Carrie Olson-Manning received her B.S. in genetics and evolution at the University of Minnesota, USA. Her interest in the genetic and biochemical basis of evolution began while working as an undergraduate in the laboratory of Tony Dean. To pursue her interest in natural variation, Carrie is working towards a Ph.D. with Thomas Mitchell-Olds at Duke University, USA. Her dissertation research focuses on the genetic basis of a plant chemical defense against herbivory and the evolution of biochemical pathways.

Maggie R. Wagner graduated from the University of Michigan, Ann Arbor, USA with a B.S. in plant biology. She became fascinated with the intersection of ecology, microevolution, and phenotypic diversity during a research internship at Harvard Forest, Massachusetts, USA and several years working at the University of Michigan Herbarium. Her Ph.D. research with Thomas Mitchell-Olds at Duke University, North Carolina, USA examines the evolution and genetic basis of reaction norms and genotype-environment interactions in plants under spatially variable herbivory pressure.

Thomas Mitchell-Olds received his Ph.D. from the University of Wisconsin-Madison, USA, and spent one year as an NIH postdoc in human genetics. He is a former Director of the Max-Planck Institute of Chemical Ecology in Jena, Germany. At Duke University he is Professor in the Department of Biology and Institute for Genome Sciences and Policy, and is active in the graduate programs in Biology, Genetics and Genomics, and in Computational Biology and Bioinformatics. His research focuses on the functional basis and evolutionary causes of genetic variation in populations.

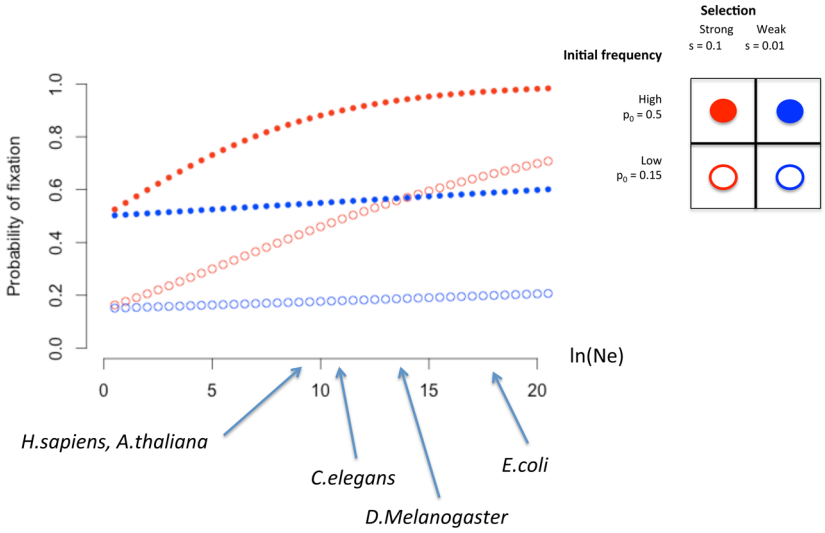

Figure 1. Determinants of fixation of adaptive alleles

The interaction of the selective coefficient (*s*), effective population size (*N*e) and initial frequency (*p*0) determines the probability of fixation of an adaptive allele. The *x* axis is log-transformed for clarity. *Homo sapiens*[10], *Arabidopsis thaliana*[14] and *Caenorhabditis elegans*[117] all have small historical *N*e values (of ~10,000), and the probability of fixing a strongly adaptive allele is much lower than for populations with a high *N*e (such as *Drosophila melanogaster*[118] and *Escherichia coli*[119], which have *N*e values of 1.1 million and 25 million, respectively). Higher initial frequency increases the chances of fixation of adaptive alleles[45].

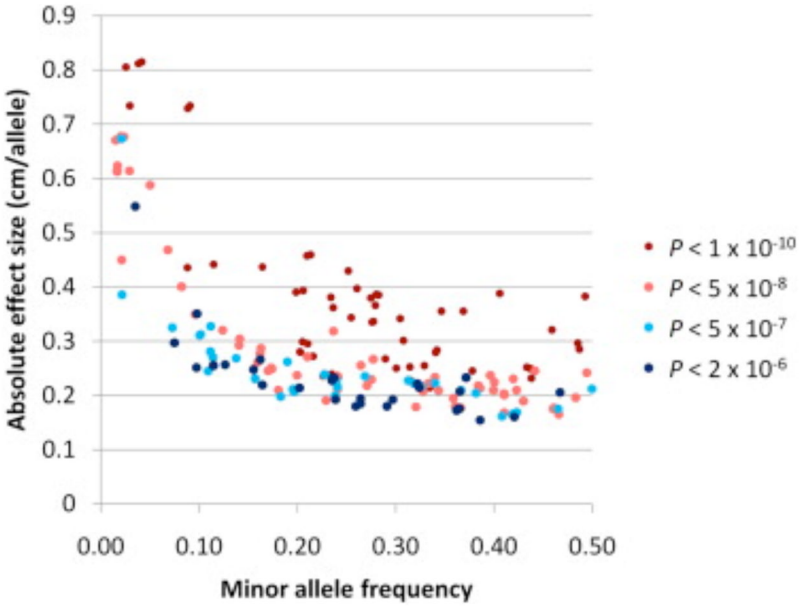

Figure 2. Genetic architecture of human height

Findings from a meta-analysis illustrate the population frequency of the rarer allele on the horizontal axis versus the effect size of that allele on the vertical axis. Colors indicate levels of statistical significance for the association of alleles with the trait. Results show that large effect alleles are fairly rare for this heritable, selectively important trait[47]. Figure reproduced, with permission, from Ref. 47 © (2011) Cell Press.

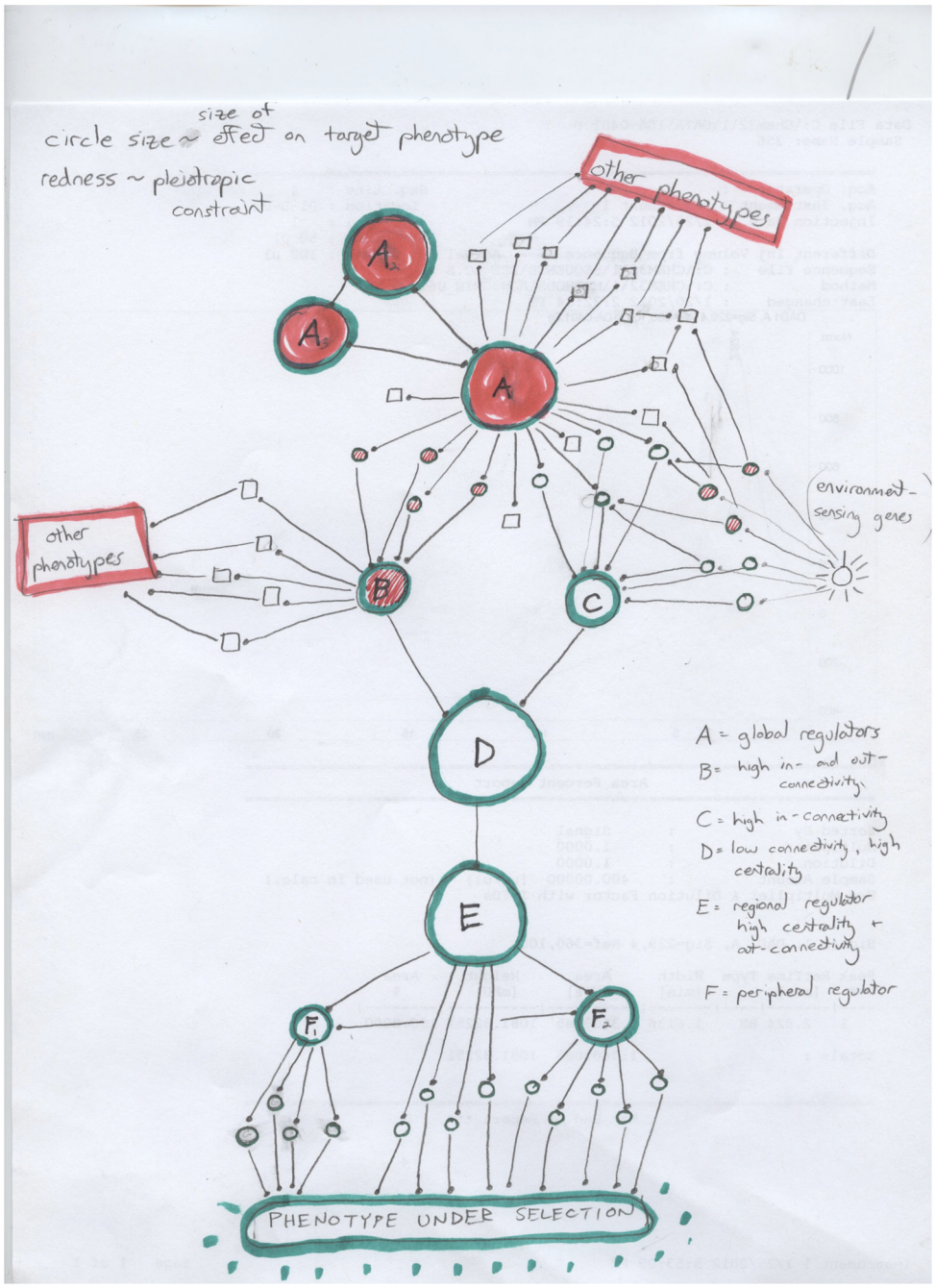

Figure 3. Importance of network effects for adaptation

A gene's position in a regulatory network influences its effects on a target phenotype and on other traits. Circular node sizes are proportional to the gene's effect on the selected phenotype; the intensity of red coloration is proportional to effects on other traits (where no color indicates no effect on other traits). Square nodes have no effect on the target phenotype owing to the directionality of the network, but they may influence other phenotypes. Small black circles indicate the directions of the interactions in the network; modes of interaction are not specified. Genes encoding upstream proteins (A) often have large effects because they control many downstream genes influencing a trait, although pleiotropy mediated by other connections may weaken net selection on these 'hub' genes.

Compared with protein A, the gene encoding protein B has fewer pleiotropic connections but also less control over the target phenotype. Protein C has lower pleiotropic constraint than protein B and integrates more upstream signals — including environmental inputs — resulting in higher evolvability. Proteins at central or 'bottleneck' positions (D) often have large effects, even if they have few direct connections with other proteins. Proteins with both high centrality and high connectivity (E) may be influenced by large-effect, adaptive alleles. Genes encoding downstream, peripheral proteins (F1 and F2) may have small to moderate effects and are more likely to accumulate neutral or nearly neutral variants than relatively upstream genes of large effect (D and E), which may evolve under positive and then purifying selection.

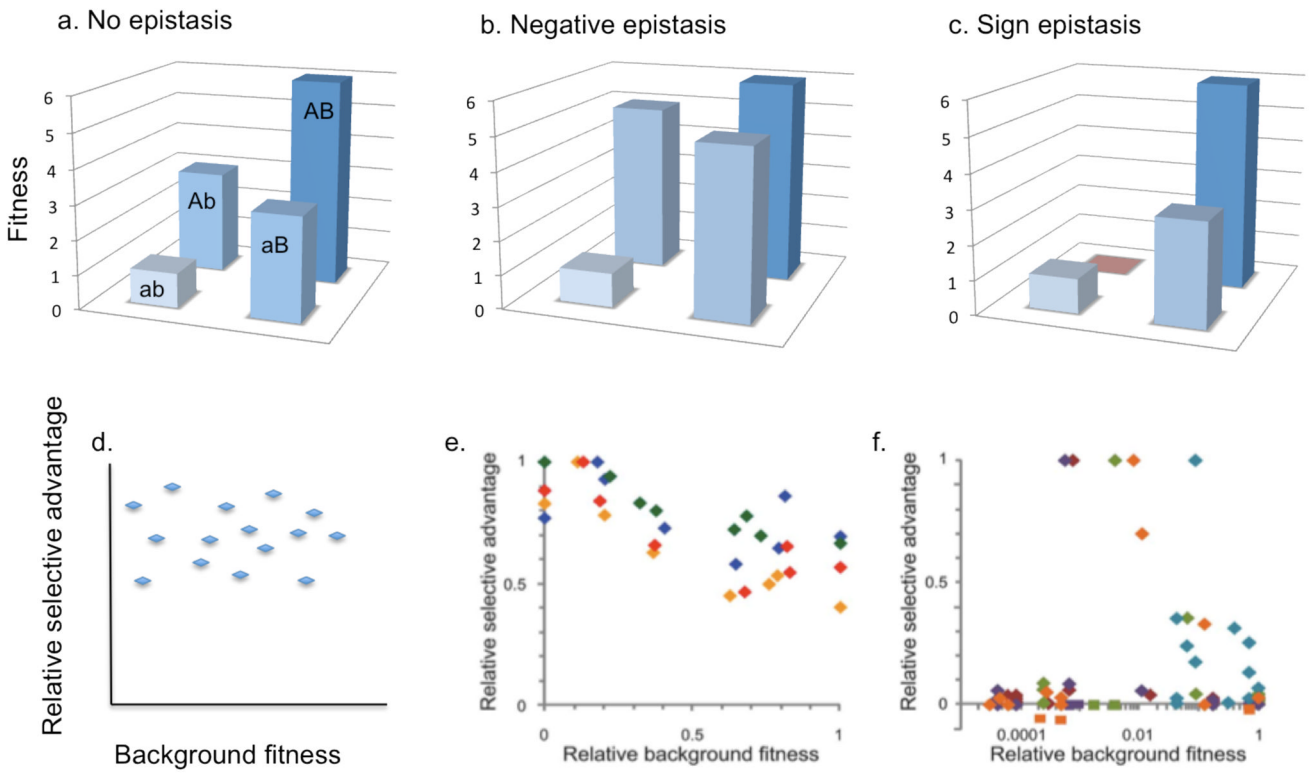

Figure 4. Hypothetical mutational trajectories

Panels **a**, **b** and **c** illustrate possible transitions from a low-fitness progenitor (*ab*) to a high-fitness derived state (*AB*). Panels **d**, **e** and **f** illustrate mutational trajectories under different epistaticconditions, as predicted in Refs. 67,69. **a** | When epistasis is absent, a change from *a* to *A* or from *b* to *B* has identical fitness consequences, regardless of the other locus. **b** | With negative epistasis, transition from *ab* to either *aB* or to *Ab* gives an identical fitness increase. However, the subsequent mutation to *AB* further increases fitness by a small amount. Thus, the fitness effect of *a* to *A* and of *b* to *B* depends on genetic background but changes only in magnitude, not in sign. **c** | Sign epistasis can cause a change in rank fitness, depending on the genetic background. For example, the *a* to *A* transition is deleterious in the *b* genetic background, but it is advantageous in the *B* background. **d** | Without epistasis, there is no relationship between the selective advantage of a mutation and the fitness of the genetic background in which it occurs. **e** | With negative epistasis, new beneficial mutations confer a smaller fitness increase when they occur in a genetic background with relatively higher fitness. **f** | With sign epistasis, only very specific mutations in specific genetic backgrounds increase fitness; the same mutations would decrease fitness if they occurred in other genetic backgrounds. Panels **a**–**c** are modified, with permission, from Ref. 79 © (2005) Macmillan Publishers Ltd. All rights reserved. Panels **e** and **f** are reproduced, with permission, from Ref. 82 © (2011) American Association for the Advancement of Science.

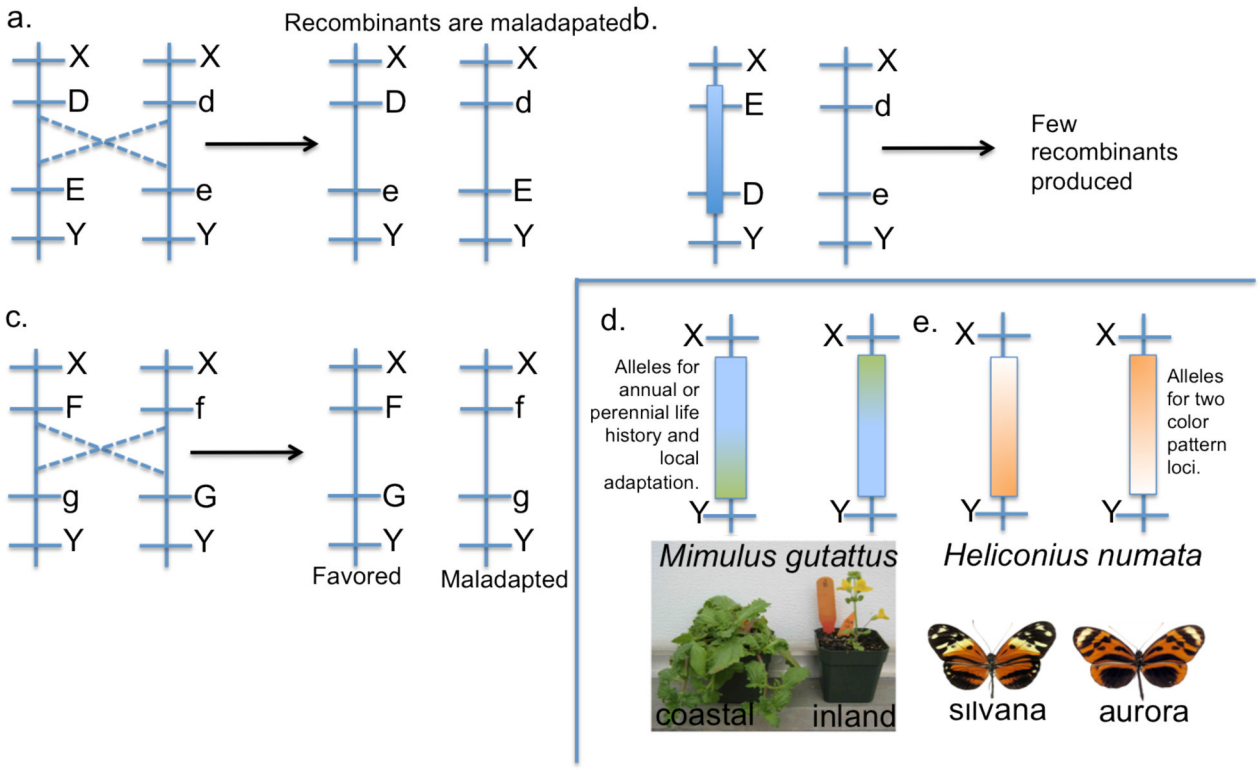

Figure 5. The dual nature of recombination

a | Consider two ecologically important genes, *D* and *E*, segregating for alleles that are adapted to different environments. Alleles *D* and *E* are best suited to environment 1, and alleles *e* and *d* are best in environment 2. Finally, genes *X* and *Y* are neutral loci. Mating between *DE* and *de* can produce maladaptive haplotypes *De* and *dE*. **b** | An inversion on the *DE* haplotype will repress recombination between these loci and increase fitness of these alleles in the ir favored environment. **c** | If alleles *f* and *g* are maladapted, then recombination between them will produce the high fitness *FG* haplotype. In this case, recombination aids in the emergence of adaptive haplotype *FG*. **d** | The inland, annual ecotype of *Mimulus guttatus* occurs in seasonally dry habitats and flowers early in the spring, whereas the sympatric coastal, perennial form is found in wetter areas and is dormant in the early spring and flowers later. Hybridization between these ecotypes would produce offspring that are less fit in either habitat. Traits that confer local adaptation to these distinct environments are located on an inversion (shown as a long rectangle) that preserves these phenotypic combinations[81]. **e** | *Heliconius* butterflies are a classic example of **Mullerian mimicry**. Many species of the genus *Heliconius* (for example, *Heliconius numata silvana* and *Heliconius numata aurora*) mimic the wing patterns of *Melinae* spp. to avoid predators. Each of these wing patterns requires a distinct combination of alleles that influence color and shape, and recombinants between these distinct types are maladapted. The different *Heliconius* mimics are closely related and occur sympatrically, yet hybrids are rarely found in nature. It has been shown that two phenotypically distinct mimics have an inversion that harbors at least two color-pattern loci[97]. Photographic images in panels **d** and **e** were provided by David Lowry (University of Texas at Austin, USA) and Mathieu Joron (Muséum National d'Histoire Naturelle, Paris, France), respectively.