

# Heterozygote advantage as a natural consequence of adaptation in diploids

Diamantis Sellis<sup>a</sup>, Benjamin J. Callahan<sup>b</sup>, Dmitri A. Petrov<sup>a,1</sup>, and Philipp W. Messer<sup>a,1</sup>

Departments of <sup>a</sup>Biology and <sup>b</sup>Applied Physics, Stanford University, Stanford, CA 94305

Edited\* by Boris I. Shraiman, University of California, Santa Barbara, CA, and approved November 2, 2011 (received for review September 7, 2011)

Molecular adaptation is typically assumed to proceed by sequential fixation of beneficial mutations. In diploids, this picture presupposes that for most adaptive mutations, the homozygotes have a higher fitness than the heterozygotes. Here, we show that contrary to this expectation, a substantial proportion of adaptive mutations should display heterozygote advantage. This feature of adaptation in diploids emerges naturally from the primary importance of the fitness of heterozygotes for the invasion of new adaptive mutations. We formalize this result in the framework of Fisher's influential geometric model of adaptation. We find that in diploids, adaptation should often proceed through a succession of short-lived balanced states that maintain substantially higher levels of phenotypic and fitness variation in the population compared with classic adaptive walks. In fast-changing environments, this variation produces a diversity advantage that allows diploids to remain better adapted compared with haploids despite the disadvantage associated with the presence of unfit homozygotes. The short-lived balanced states arising during adaptive walks should be mostly invisible to current scans for long-term balancing selection. Instead, they should leave signatures of incomplete selective sweeps, which do appear to be common in many species. Our results also raise the possibility that balancing selection, as a natural consequence of frequent adaptation, might play a more prominent role among the forces maintaining genetic variation than is commonly recognized.

Adaptation by natural selection is the key process responsible for the fit between organisms and their environments. The invasion of new adaptive mutations is an essential component of this process that fundamentally differs between haploid and diploid populations. In diploids, while a new mutation is still rare, natural selection acts primarily on the mutant heterozygote (1). As a result, only those adaptive mutations that confer a fitness advantage as heterozygotes have an appreciable chance of invading the population ("Haldane's sieve"). However, if the heterozygote of an invading adaptive mutation is fitter than the mutant homozygote, the mutation will not be driven to fixation but, instead, maintained at an intermediate, balanced frequency.

We argue that heterozygote advantage should be very common during adaptation in diploids if selection is stabilizing and at least some mutations are large enough to overshoot the optimum. Consider, for example, adaptation through changes in gene expression (Fig. 1A), which is an important, if not the dominant, mechanism of adaptation (2). Here, mutations of small effect (Fig. 1B) will be adaptive when they modify expression in the adaptive direction whether the organism is haploid or diploid. However, mutations of large effect (Fig. 1C) can be nonadaptive in haploids, because they overshoot the optimum, yet be adaptive in diploids, because their phenotypic effect is moderated when heterozygous. These mutations will have heterozygote advantage and lead to balancing selection in diploids.

The components of the simple model in Fig. 1, stabilizing selection and availability of mutations of various sizes, are well established by empirical data in the case of gene expression. Pervasive stabilizing selection is indicated by the lack of large gene expression differences between and within species despite the abundance of mutations that change gene expression (3–7). In

addition, expression-altering mutations are known to come in a variety of sizes ranging from subtle changes to dramatic ones of tens of fold or even hundreds of fold (4, 5, 8).

The model in Fig. 1, albeit instructive, may lack generality because it incorporates only a single trait. However, adaptation might often involve mutations that have complex pleiotropic effects in an effectively multidimensional phenotypic space. The classic model that incorporates this key feature of adaptation is Fisher's geometric model (9–13). In this single-locus model, phenotypes are vectors in an abstract geometric space (Fig. 2A), with the orthogonal axes representing different, independent traits, such as color and height. Mutations move phenotypes in a random direction, with the size of the step corresponding to the effect size of the mutation. The fitness landscape is peaked around a single optimal phenotype, capturing our intuition of stabilizing selection. Below, we extend Fisher's geometric model to diploidy and confirm the intuition from Fig. 1 that adaptation should generically lead to heterozygote advantage.

## Results

Consider an allele  $a$  with homozygous phenotype  $r_{aa}$  in Fisher's model. Mutations are modeled by adding a mutation vector  $m$  to the phenotype of the mutated allele (Fig. 2A). The direction of the mutation vector is chosen uniformly, and its size ( $m$ ) is drawn from a specified distribution  $P(m)$ . In diploids, the phenotype is determined by its two constituent alleles. The geometric nature of Fisher's model offers a straightforward mapping between the phenotype of the mutant heterozygote ( $r_{ab}$ ) and the homozygote phenotypes ( $r_{aa}$  and  $r_{bb} = r_{aa} + m$ ) in terms of the weighted average:  $r_{ab} = r_{aa} + hm$ , where the weight  $h$  specifies the phenotypic dominance of the new mutation. For convenience, we assign the phenotype of a diploid homozygote to be equal to that of a haploid carrying the same allele (i.e.,  $r_{ii} = r_i$ ).

Fitness  $w(r)$  decreases with the distance between the organismal phenotype and the phenotypic optimum. In haploids, adaptive mutations,  $w(r_b) > w(r_a)$ , are those that bring the mutant allele closer to the optimum (i.e., the mutant falls inside the sphere  $\alpha_{\text{hap}}$  of radius  $r_a$  centered at the optimum) (Fig. 2A). In diploids, it is the fitness of the mutant heterozygote  $r_{ab}$  that primarily determines the probability of successful invasion; therefore, we define the area in which mutations are adaptive in diploids ( $\alpha_{\text{dip}}$ ) by the condition  $w(r_{ab}) > w(r_{aa})$ .

Let us first focus on the instructive case of strict phenotypic codominance,  $h = 1/2$ . Here, mutations in diploids have half the initial phenotypic effect that they would have in a haploid.

Author contributions: D.A.P. and P.W.M. designed research; D.S., B.J.C., and P.W.M. performed research; D.S., B.J.C., and P.W.M. contributed new reagents/analytic tools; D.S., B.J.C., D.A.P., and P.W.M. analyzed data; and D.S., B.J.C., D.A.P., and P.W.M. wrote the paper.

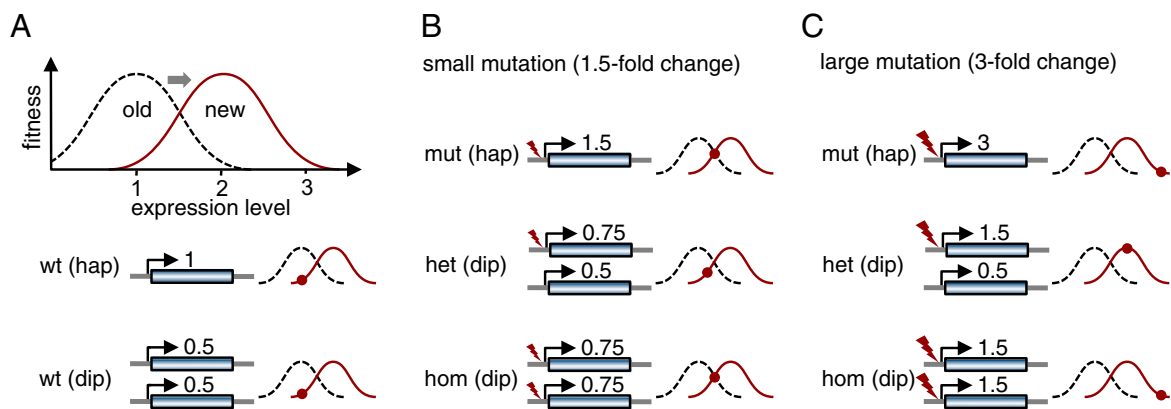
The authors declare no conflict of interest.

\*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. E-mail: dpetrov@stanford.edu or messer@stanford.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1114573108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1114573108/-DCSupplemental).



**Fig. 1.** Adaptation to a change in the optimal level of gene expression. (A) In both haploids (hap) and diploids (dip), the wild type (wt) is perfectly adapted to the original fitness function (dashed black curve). After an external change, the optimal expression level becomes twice the original level (solid red curve). Note that we are assuming phenotypic codominance; thus, the two individual gene copies in a diploid each contribute expression level 0.5, such that overall expression is 1. (B) Fitness effects of a small mutation that increases expression level by 1.5-fold. The mutant heterozygote (het) is less fit than both the haploid mutant (mut) and the mutant homozygote (hom). (C) Effects of a large mutation that increases expression by threefold. In this case, the mutant heterozygote effectively has only twofold increased expression, and thus lands right at the new fitness optimum. In contrast, both the haploid mutant and the mutant homozygote “overshoot” the optimum.

Mutations of small effect thus have reduced initial selective effects in diploids compared with in haploids, but the area in which mutations are adaptive in diploids ( $\alpha_{\text{dip}}$ ) is significantly enlarged (Fig. 2A). If the supply of mutations is restricted to those of small size compared with the distance to the optimum, adaptive mutations will invade a diploid population at a slower rate than they would a haploid population because of the reduced initial selection (SI Text and Figs. S1 and S2). In contrast, once mutations of large enough size are available, diploids start reaping the benefits of the larger space of adaptive mutants available to them. In this case, adaptive mutations will invade at a higher rate in diploids than in haploids, even after controlling for the difference in mutation rate that results from a diploid population having twice as many chromosomes as a haploid population of equal size (14).

The larger range of adaptive mutations available to diploids comes with a catch, however; many of these adaptive mutations display heterozygote advantage, and thus will not simply go to fixation. The fraction  $\delta_u$  of adaptive mutations with heterozygote advantage [ $w(r_{aa}) < w(r_{ab}) > w(r_{bb})$ ] in Fisher’s model is the fraction of mutations that fall into the sphere  $\alpha_{\text{dip}}$  [ $w(r_{ab}) > w(r_{aa})$ ] but not the sphere  $\gamma$  [ $w(r_{bb}) > w(r_{ab}) > w(r_{aa})$ ] (Fig. 2A). In the small-mutation limit,  $\delta_u$  goes to zero because the non-overlapping space between  $\alpha_{\text{dip}}$  and  $\gamma$  is not reachable by mutations that fall in the immediate neighborhood of  $r_u$ . In the limit in which mutations of all sizes are equally likely, however, most adaptive mutations show heterozygote advantage ( $\delta_u \geq 1/2$ ), except for the special cases of perfect dominance or perfect recessiveness (SI Text).

The crucial criterion determining the probability of heterozygote advantage during adaptation in diploids is the availability of “large” mutations. In Fisher’s model, for heterozygote advantage to be common, the average size of mutations ( $\langle m \rangle$ ) has to be at least of order  $r_{aa}/\sqrt{d}$ , the distance to the optimum divided by the square root of the number of phenotypic dimensions (SI Text). However, even when mutations are initially small compared with this distance, during an adaptive walk, a population gradually approaches a fitness optimum via successive adaptations (11). Thus, at some point on the walk, this condition will be met; thereafter, heterozygote advantage will be likely.

We performed simulations of adaptive walks in Fisher’s model to test these theoretical predictions and further investigate the consequences of frequent heterozygote advantage in such walks. Specifically, we simulated a single locus in a 2D phenotypic space

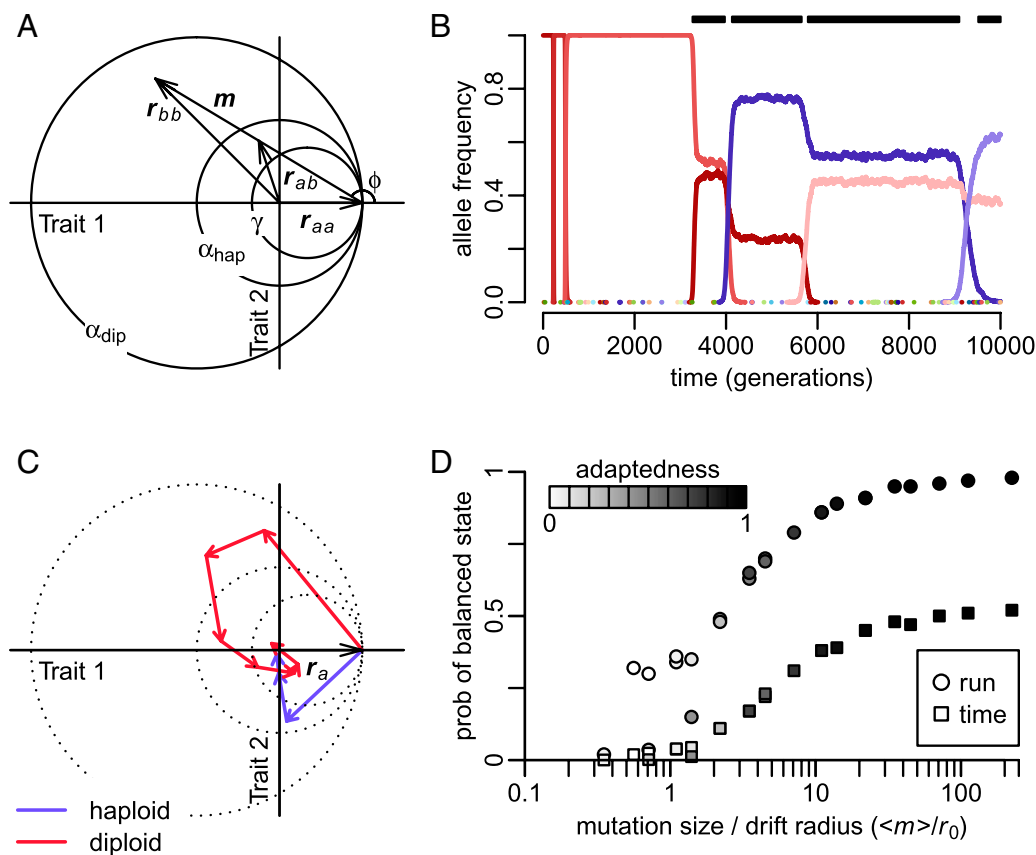
under phenotypic codominance in a Wright–Fisher framework using an exponential distribution of mutation sizes and a symmetrical Gaussian fitness landscape (Materials and Methods).

We find that adaptive walks in diploids typically involve the succession of many intermediate balanced polymorphisms that tend to be ephemeral; they are quickly displaced by new adaptive alleles, themselves often displaying heterozygote advantage (Fig. 2B and C). These dynamics contrast sharply to those in haploid populations, where adaptive walks proceed by successive sweeps of adaptive mutations and populations are generally monomorphic between sweeps (Fig. 2C).

We confirm that balanced states are likely during adaptive walks (Fig. S3), provided that mutation sizes meet our theoretical conditions and selection is strong enough to maintain balanced states despite the stochastic fluctuations arising from genetic drift (SI Text). Fig. 2D shows the probability of observing balanced states during adaptive walks under various parameter settings (Table S1). Note that the probability of observing balanced states correlates very strongly with the overall probability of observing successful adaptation. For example, in all our scenarios where the population eventually traversed at least 90% of the initial fitness distance to the optimum, balanced states were observed in  $\geq 90\%$  of the runs and were present  $\geq 40\%$  of the time during walks.

Simulated haploid populations approach the optimum faster, on average, than simulated diploid populations despite diploid populations generally containing the most fit individual (Fig. S4). This is because diploid populations suffer from a high genetic load, which can be identified as the segregation load attributable to pervasive heterozygote advantage. The balanced polymorphisms that result from the invasion of adaptive mutations with heterozygote advantage also cause high levels of standing variance in both phenotype and fitness to persist during the adaptive walk (Fig. S4).

The maintenance of genetic variation via heterozygote advantage during adaptation is a striking difference between the adaptive walks in haploids and diploids. Although the load resulting from the maintained variation is costly during adaptation to a constant environment, could this effect be advantageous in a changing environment? Environmental changes that displace the phenotypic optimum can convert dominance variance of fitness into additive variance, which would then fuel adaptation. To investigate this possibility, we incorporated a randomly



**Fig. 2.** Fisher's geometric model of adaptation in two dimensions. (A) Two orthogonal axes represent independent character traits. Fitness is determined by a symmetrical Gaussian function centered at the origin. Consider a population initially monomorphic for the wild-type allele  $r_{aa} = (2,0)$ . A mutation  $m$  gives rise to a mutant phenotype vector  $r_{bb} = r_{aa} + m$ . The phenotype of the mutant heterozygote assuming phenotypic codominance ( $h = 1/2$ ) is  $r_{ab} = r_{aa} + m/2$ . The different circles specify the areas in which mutations are adaptive in haploids ( $\alpha_{hap}$ ), adaptive in diploids ( $\alpha_{dip}$ ), and replacing in diploids ( $\gamma$ ). (B) Frequency trajectories of all alleles present during a representative adaptive walk in a diploid population with  $N = 5 \cdot 10^4$ ,  $r_{aa} = (2,0)$ , and  $\langle m \rangle = \sigma_w = 1$ . Different colors represent different alleles. The black bars over the graph indicate the periods during which a balanced polymorphism was present. (C) Representative adaptive walks in a haploid population and a diploid population. Vectors depict the successive mutations that led to the prevalent allele at the end of the walk. The haploid walk consists of a single lineage of successive mutations, each conferring a selective advantage over the previous one. In the diploid walk, the first mutation overshoots the fitness optimum, generating a sequence of intermediate balanced states. Note that the areas  $\alpha_{hap}$ ,  $\alpha_{dip}$ , and  $\gamma$  (dotted circles) from A apply only to the first mutation in the walk when the population is still monomorphic for  $r_a$ . (D) Probability of observing balanced polymorphism during adaptive walks toward a fixed fitness optimum as a function of mutation sizes scaled by effective drift radius  $r_0$  (SI Text) for the various settings of  $N$ ,  $\sigma_w$ , and  $\langle m \rangle$  specified in Table S1. Circles show the probability of at least one balanced state arising over the course of a walk, and squares show the fraction of time during which balanced states were present. Coloration indicates the average "adaptedness" achieved during a walk, defined by the improvement in mean population fitness over the walk ( $\langle w_{end} \rangle - \langle w_{start} \rangle$ ) relative to the maximally possible improvement ( $1 - \langle w_{start} \rangle$ ).

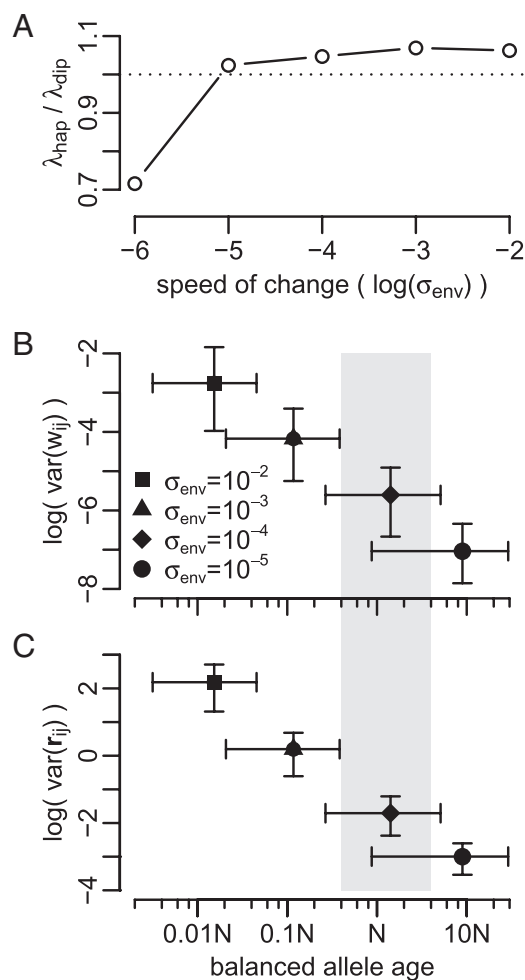
moving optimum into our simulations (*Materials and Methods*). The optimum moves every generation in a random direction with step sizes sampled from a Gaussian distribution, the variance of which ( $\sigma_{env}^2$ ) determines the speed of optimum movement.

Pervasive balanced polymorphism remains a feature of adaptation in diploid populations in the moving optimum scenario (typical trajectories are shown in Fig. S5). In simulations with a fast-moving optimum ( $\sigma_{env}/\sigma_w = 10^{-2}$ ), polymorphisms at frequencies  $0.05 < x < 0.95$  are present around 30% of the time when averaged over replicate runs. More than 80% of these polymorphisms are balanced. Polymorphisms are observed less frequently ( $\sim 5\%$ ) when environmental change becomes very slow ( $\sigma_{env}/\sigma_w = 10^{-5}$ ), because populations are well adapted most of the time. The fraction of polymorphisms that are balanced, however, remains substantial ( $\sim 58\%$ ) in this scenario. Table S2 shows the percentages of time during which polymorphisms were observed in our simulations and the fractions of those polymorphisms that were balanced for a wide range of values of  $\sigma_{env}/\sigma_w$ . We also demonstrate that many of the balanced polymorphisms eventually fix such that a substantial

proportion of substitutions (60–70%) pass through a balanced state (Fig. S6 and Table S2). These substitutions tend to be generated by larger phenotypic effect mutations than the substitutions that go to fixation without an intermediate balanced state (Fig. S6).

To evaluate how effectively populations followed the moving fitness optimum, we measured the mean population fitness averaged over a walk ( $\langle \bar{w} \rangle$ ). The difference,  $\lambda = 1 - \langle \bar{w} \rangle$ , is then the average lag in fitness between the population and the optimum. Fig. 3A shows the ratio  $\lambda_{hap}/\lambda_{dip}$  between haploid and diploid populations for different values of  $\sigma_{env}/\sigma_w$ . In slow-changing environments ( $\sigma_{env}/\sigma_w < 10^{-4}$ ), haploids follow the moving optimum more closely than diploids, replicating the constant environment result. However, in fast-changing environments ( $\sigma_{env}/\sigma_w > 10^{-4}$ ), it is diploids that prevail.

The diversity advantage of diploid populations derives from the greater variation maintained during adaptation. Specifically, this greater variation should lead to an increased range of mutations starting from multiple balanced alleles as well as the ability of diploids to adapt by fast adjustment of the frequencies



**Fig. 3.** Statistics of adaptive walks under a moving fitness optimum. (A) Ratio of the average lag in fitness ( $\lambda$ ) between the population and the optimum in haploids (hap) and diploids (dip) as a function of the speed of environmental change. In fast-changing environments, diploid populations follow the moving optimum more closely than haploids ( $\lambda_{hap}/\lambda_{dip} > 1$ ). (B) Fitness variance attributable to balanced polymorphisms (frequency  $0.05 < x < 0.95$ ) and the age of the balanced polymorphism for different values of  $\sigma_{env}$ . Both quantities are estimated from the balanced polymorphisms that were present at the end of simulation runs. The age of a balanced polymorphism is defined as the time since the most recent common ancestor of its constituent alleles. Data points are medians over  $10^3$  runs, and error bars specify the 10% and 90% quantiles. The gray-shaded area ( $0.04N < \text{age} < 4N$ ) indicates the expected age range of common neutral polymorphisms at frequencies between  $0.05 < x < 0.95$ . (C) Same as in B but phenotypic variance is shown instead of fitness variance.

of balanced alleles without waiting for de novo mutations. Note that this diversity advantage of diploids is different in kind from the advantages based on the differences in number of chromosomes and the efficacy of selection in haploids and diploids discussed previously (14–16).

The characteristics of the genetic variation in diploid populations depend strongly on the speed of environmental change. Fig. 3 B and C show the joint distributions of allele age and the fitness or phenotypic variance of balanced polymorphisms at intermediate population frequencies as a function of  $\sigma_{env}/\sigma_w$ . In slow-changing environments, balanced alleles generally persist over long periods of time, and thus tend to be much older than neutral alleles of comparable population frequencies. These old balanced alleles are similar phenotypically, and their segregation generates only limited variation in fitness (Fig. 3B) and

phenotype (Fig. 3C). In contrast, in fast-changing environments, balanced alleles are much younger than comparably frequent neutral alleles. These young balanced alleles are often very distinct phenotypically, and thus produce high levels of standing variation in fitness (Fig. 3B) and phenotype (Fig. 3C).

## Discussion

Our finding that heterozygote advantage emerges naturally in Fisher's geometric model if mutations are sufficiently large is very robust to the details of the model, such as the number of dimensions; the choice of phenotypic dominance rules, including under-, over-, and incomplete dominance; mutation rate and size distribution; population size; and flatness of the fitness function (SI Text, Figs. S1–S3, S7, and Table S1).

The features that underlie pervasive heterozygote advantage in Fisher's model are also likely to apply very generally in nature: (i) mutations in diploids should initially segregate as heterozygotes, (ii) selection should be stabilizing for some traits, and (iii) some invading mutations should be large enough to overshoot the local optimum. Indeed, rare mutations are generally heterozygous, except for cases of exceptionally strong inbreeding. It is also well established that stabilizing selection acts on many phenotypic traits (17, 18). Even if "more is always better" holds for certain traits, as long as adaptive mutations generally influence at least one trait under stabilizing selection, we still expect heterozygote advantage to be frequent. Finally, mutations of large phenotypic effect have been observed in many organisms, and it is becoming increasingly apparent that such large mutations do contribute to adaptation. For example, adaptive mutations of large effect have occurred in the domestication of maize (19) and dogs (20), the evolution of sticklebacks to fresh water habitats (21, 22), pesticide resistance in insects (8), coat color in mammals (23, 24), and several studies of experimental evolution (25, 26). Furthermore, analyses of mutation accumulation lines have demonstrated the availability of mutations causing multiple-fold changes in gene expression (4, 5). Because the expression of most genes is known to be constrained by stabilizing selection over much narrower ranges (3–7), we must conclude that gene expression mutations that overshoot the optimum are likely abundant.

The classic model of adaptation holds that adaptation is driven by adaptive mutations that sweep quickly to fixation. Although our model also predicts such fast fixation events, it additionally predicts that many adaptive mutations will initially only sweep to intermediate frequencies, where they are then maintained for a period of time by balancing selection, before continuing on to either fixation or loss. Both models thus predict an elevated rate of fixation at functional sites compared with the neutral expectation and a local reduction of genetic diversity around adaptive sites, as have been observed in a range of organisms (27–30). However, in contrast to the classic model, we also expect the presence of many incomplete selective sweeps. Indeed, the genomic signatures of such incomplete sweeps appear to be plentiful in a number of organisms (31–33). Incomplete sweeps also appear to be common in experimental evolution in *Drosophila*, where virtually no classic sweeps have been detected after 600 generations of evolution despite abundant evidence of phenotypic adaptation over this period (34).

The abundance of incomplete sweeps in natural and experimental populations is consistent with but, unfortunately, not uniquely predictive of adaptation-driven balancing selection. Other scenarios, such as frequency-dependent selection, adaptation to specific subhabitats, and polygenic adaptation, also predict incomplete sweeps. The only way to test the hypothesis of heterozygote advantage explicitly is to measure fitness of the homozygotes and heterozygotes for the putatively balanced alleles directly. Such measurements are difficult but not impossible and can now be carried out systematically in laboratory systems of artificial selection, such as yeast or *Drosophila* (26, 34, 35).

However, in many other systems, direct fitness measurements are not feasible and balanced polymorphisms have to be identified by alternative approaches. The standard scans for balanced polymorphisms are inappropriate for our purposes because they typically search for very ancient balanced alleles (36–38), whereas the balanced alleles predicted by our model are often short-lived. To identify young balanced alleles specifically, one can search for polymorphisms that maintain their frequencies in the face of sharp bottlenecks (39). A particularly powerful system in this context is provided by human genetics because of the multiple recent bottlenecks associated with human migrations. Other opportune systems are island species recovering from natural disasters or man-made relocations of species to new but environmentally similar locations, such as the relocation of *Euphydryas gillettii* from Wyoming to Colorado (40).

Adaptive evolution is generally thought to be antithetical to the maintenance of genetic variation. In contrast, in our model, pervasive adaptation systematically generates genetic variation by promoting balanced polymorphisms. These balanced polymorphisms are expected to segregate at high frequencies yet can affect both phenotype and fitness substantially. This is very different from the common view that frequent polymorphisms should be neutral or subject to only weak selective forces.

We argue that adaptation-driven balanced polymorphisms can be an important source of consequential genetic variation. In particular, we believe that the balanced polymorphisms predicted by our model can be associated with human disease. Some of the common disease variants could be mutations that are maintained at high population frequencies because of strong heterozygote advantage, although they are very harmful as homozygotes.

Balancing selection and, specifically, the prevalence of heterozygous advantage was once considered the dominant force maintaining variation in natural populations (41–43). This view fell out of favor with the rise of the neutral theory in the 1960s (10). The neutral theory postulates that only a small proportion of substitutions are adaptive and that these substitutions are fixed very quickly; this, in turn, implies that practically all polymorphisms should be either neutral or slightly deleterious. However, recent genomic evidence has suggested that the rate of adaptation is substantial in some organisms, with, for example, ~50% of all amino acid substitutions in *Drosophila* driven by positive selection (28). Here, we argue that such a high rate of adaptation in diploids should also lead to a high rate at which balanced polymorphisms are driven into the population. We thus advocate that the balance theory of genetic variation should be given new life and reassessed using all the modern genomic tools at our disposal.

## Materials and Methods

**Monte Carlo Simulations of a Single Adaptive Mutation in Fisher's Model.** We investigate the evolution of a single locus in Fisher's geometric model (9). Alleles are represented by vectors ( $r$ ) in an abstract,  $d$ -dimensional Euclidean phenotype space. Mutant alleles are obtained by adding a mutation vector ( $m$ ) to the parental allele:  $r_b = r_a + m$ . The directions of mutation vectors are distributed uniformly; mutation sizes ( $m$ ) are distributed according to a probability distribution  $P(m)$  with an average mutation size  $\langle m \rangle$ . We consider two such distributions in detail, the uniform distribution  $P(m) \propto \text{uniform}(0, 2\langle m \rangle)$  and the exponential distribution  $P(m) \propto \exp(-m/\langle m \rangle)$ . Haploid organismal phenotypes are equal to the allelic phenotype. In diploids, the phenotype is a weighted average of its two constituent alleles. In the case of the heterozygous mutant, the phenotype can be expressed as  $r_{ab} = r_a + hm$ , with  $h$  being the phenotypic dominance of the mutant allele.

The fitness of a phenotype is determined by its distance from the fitness optimum  $r^*$ . Following precedent (10, 11), we use a Gaussian fitness function:  $w(r) = \exp[-(r-r^*)^2/(2\sigma_w^2)]$ . For convenience, we set the origin of the phenotype space to be at  $r^*$  and choose the scale of the space such that  $\sigma_w^2 = 1$ . We begin by considering mutations arising in a population that is monomorphic for the wild-type allele  $r_a = (2, 0, \dots, 0)$ . The invasion probability of new mutations is then approximately  $\pi_{\text{hap}}(m) = 2[w(r_b)/w(r_a) - 1]$  in haploids (10). In diploids, we assume that mutants initially exist only as heterozygotes, therefore:  $\pi_{\text{dip}}(m) = 2[w(r_{ab})/w(r_a) - 1]$ .

Results in Figs. S1 and S2 were obtained from Monte Carlo simulations of the above model based on  $10^7$  randomly drawn mutations per data point. The ratios  $u_{\text{dip}}/u_{\text{hap}}$  of the rates at which adaptive mutations occur in diploids vs. haploids were estimated by counting the overall numbers of mutations where  $r_b \in \alpha_{\text{hap}}$  in haploids or, respectively,  $r_{ab} \in \alpha_{\text{dip}}$  in diploids. For the ratios  $v_{\text{dip}}/v_{\text{hap}}$  of the rates at which adaptive mutations invade the population, each adaptive mutation was additionally weighted by its respective invasion probability,  $\pi_{\text{hap}}$  or  $\pi_{\text{dip}}$ . Among successfully invading mutations, the expectation values,  $\langle w(r_{ab}) - w(r_a) \rangle$  in diploids and  $\langle w(r_b) - w(r_a) \rangle$  in haploids, were measured to estimate the ratios  $\langle \Delta w_{\text{dip}} \rangle / \langle \Delta w_{\text{hap}} \rangle$ . Estimates of  $\delta_u$  were obtained by counting the fraction of adaptive mutations with heterozygote advantage in the diploid scenario. For  $\delta_v$ , each adaptive mutation was thereby additionally weighted by its respective invasion probability.

**Simulation of Adaptive Walks Toward a Fixed Fitness Optimum.** To investigate adaptive walks toward a fixed fitness optimum, we simulated the full stochastic population dynamics in the above scenario under an infinite alleles assumption. We focused on the instructive case of a 2D Fisher's model with complete phenotypic codominance ( $h = 1/2$ ). The phenotype of a heterozygous diploid is then always the coordinate-wise average of its two alleles:  $r_{ab} = (r_a + r_b)/2$ .

Mutations are modeled by a Poisson process with rate  $\mu = 2.5 \cdot 10^{-7}$  per individual and generation. Mutation directions are drawn uniformly, and mutation sizes are sampled from an exponential distribution with mean  $\langle m \rangle = 1$ . Population sizes are  $N_{\text{hap}} = 10^5$  for haploids and  $N_{\text{dip}} = 5 \cdot 10^4$  for diploids, ensuring that new mutations arise at equal overall rates in the two populations ( $\Theta = 2cN\mu = 0.05$ , where  $c$  is ploidy).

The state of the population at any given time point is specified by the set of alleles  $\{r_i\}$  present in the population and their associated population frequencies  $\{x_i\}$ . Allele frequency dynamics are modeled in a Wright-Fisher framework with selection (44). For haploids, we use the standard Wright-Fisher sampling procedure in which allele frequencies  $x_i(t+1)$  in the next generation are drawn from a multinomial distribution  $P(N, \{p_i^{\text{hap}}(t)\})$  with selection-adjusted probabilities:  $p_i^{\text{hap}}(t) \propto w(r_i)x_i(t)$ . In the case of diploids, we first convert allele frequencies into genotype frequencies (assuming Hardy-Weinberg equilibrium) to calculate the selection-adjusted probabilities:  $p_i^{\text{dip}}(t) \propto \sum_j w(r_{ij})x_j(t)x_i(t)$ . Allele frequencies  $x_i(t+1)$  are then drawn from  $P(2N, \{p_i^{\text{dip}}(t)\})$ . In both cases,  $p_i$  and  $x_i$  are normalized such that  $\sum_i p_i = \sum_i x_i = 1$ .

As specified above, simulations start from a population that is monomorphic for the wild type  $r_a = (2, 0)$  with the optimal phenotype located at the origin, yielding an initial population average fitness of  $w(r_a) \sim 0.13$ . Populations are then evolved for  $10^4$  generations, which typically suffices to approach the fitness optimum closely ( $\langle w \rangle > 0.96$  at end of a run; Fig. S4A).

**Simulations Under a Moving Fitness Optimum.** For the analysis of the moving optimum scenario, we adjust our simulation as follows. At the start of the simulation, the population is initialized to be monomorphic for the optimal phenotype:  $r^*(t=0) = r_a = (0, 0)$ . In each subsequent generation, the optimum  $r^*(t)$  moves one step in a random direction and the size of the mutation is sampled from the positive half of a Gaussian distribution with variance  $\sigma_{\text{env}}^2$ . In a single simulation run, the population is evolved for  $10^7$  generations ( $\sim 100N$ ). We exclude the first  $10^5$  generations of each run from our analysis as a "burn-in" period so as to remove the influence of the initial state of the population.

**Ascertainment of Balanced Polymorphisms During Adaptive Walks.** Balanced polymorphisms can consist of several alleles (45, 46). We determine the presence of a balanced polymorphism at a given time point in our simulation runs using Kimura's analytic conditions (47). Assume that  $n$  alleles  $\{r_1, \dots, r_n\}$  are present in the population with diploid fitness values given by  $w(r_{ij})$ . Let  $T$  be the matrix defined by  $T_{ij} = w(r_{ij}) - w(r_{i1}) - w(r_{jn}) + w(r_{nn})$  ( $i, j = 1, \dots, n$ ). Let  $\Delta_i$  be the determinant obtained when substituting all elements in the  $i$ th column of the fitness-matrix  $w(r_{ij})$  ( $i, j = 1, \dots, n$ ) with 1. The necessary and sufficient conditions for the existence of a stable equilibrium with all individual population frequencies  $x_i$  of the alleles being nonzero are then that  $T$  is negative definite and that  $(-1)^{n-1} \Delta_i > 0$  for all  $i = 1, \dots, n$ . Geometrically, these first two conditions specify a peak in  $n$ -dimensional fitness space, and only one such peak is allowed for all alleles to coexist (47).

For heterozygote advantage to be consequential (i.e., to be capable of effectively stabilizing a balanced polymorphism against the stochastic fluctuations arising from random genetic drift), the fitness advantages of a heterozygote over its two homozygote have to be at least of order  $1/N$  (10). Because we are only interested in such consequential cases of heterozygote advantage, we thus require, as a third condition, that for at least one pair of alleles in a balanced polymorphism, it holds that  $w(r_{ij}) > \max[w(r_{ii}), w(r_{jj})] + 1/N$ .

In our simulations, we evaluate these three conditions for the fitness matrix of all alleles with frequencies  $0.05 < x_i < 0.95$ . Negative definiteness of  $T$  is tested by numerically calculating eigenvalues using symmetrical bidualization with the QR reduction method (48) and checking for the negativity of all eigenvalues. Signs of determinants  $\Delta_i$  are estimated using numerical LU decompositions (48).

All source code is openly available online at: <http://sourceforge.net/projects/fgm>. Simulations were run on the Bio-X<sup>2</sup> cluster at Stanford University.

- Haldane JBS (1927) A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Math Proc Camb Philos Soc* 23:838–844.
- Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. *Cell* 134(1):25–36.
- Lemos B, Meiklejohn CD, Caceres M, Hartl DL (2005) Rates of divergence in gene expression profiles of primates, mice, and flies: Stabilizing selection and variability among functional categories. *Evolution* 59:126–137.
- Denver DR, et al. (2005) The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat Genet* 37:544–548.
- Rifkin SA, Houle D, Kim J, White KP (2005) A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* 438:220–223.
- Gilad Y, Oshlack A, Rifkin SA (2006) Natural selection on gene expression. *Trends Genet* 22:456–461.
- Bedford T, Hartl DL (2009) Optimization of gene expression by natural selection. *Proc Natl Acad Sci USA* 106:1133–1138.
- Daborn PJ, et al. (2002) A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* 297:2253–2256.
- Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).
- Kimura M (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ Press, Cambridge, UK).
- Orr HA (1998) The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* 52:935–949.
- Hartl D, Taubes C (1998) Towards a theory of evolutionary adaptation. *Genetica* 102–103:525–533.
- Orr HA (2005) The genetic theory of adaptation: A brief history. *Nat Rev Genet* 6(2):119–127.
- Otto SP, Gerstein AC (2008) The evolution of haploidy and diploidy. *Curr Biol* 18:R1121–R1124.
- Goldstein DB (1992) Heterozygote advantage and the evolution of a dominant diploid phase. *Genetics* 132:1195–1198.
- Orr HA, Otto SP (1994) Does diploidy increase the rate of adaptation? *Genetics* 136:1475–1480.
- Haldane JBS (1959) Natural selection. *Darwin's Biological Work*, ed Bell PR (Cambridge Univ Press, Cambridge, UK), pp 101–149.
- Wright S (1977) *Evolution and the Genetics of Populations* (Univ of Chicago Press, Chicago).
- Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* 386:485–488.
- Sutter NB, et al. (2007) A single IGF1 allele is a major determinant of small size in dogs. *Science* 316:112–115.
- Colosimo PF, et al. (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol* 2:E109.
- Chan YF, et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327:302–305.
- Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci USA* 100:5268–5273.
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313:101–104.
- Dunham MJ, et al. (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99:16144–16149.
- Kao KC, Sherlock G (2008) Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nat Genet* 40:1499–1504.
- Eyre-Walker A (2006) The genomic rate of adaptive evolution. *Trends Ecol Evol* 21:569–575.
- Sella G, Petrov DA, Przeworski M, Andolfatto P (2009) Pervasive natural selection in the *Drosophila* genome? *PLoS Genet* 5:e1000495.
- Cai JJ, Macpherson JM, Sella G, Petrov DA (2009) Pervasive hitchhiking at coding and regulatory sites in humans. *PLoS Genet* 5:e1000336.
- Halligan DL, Oliver F, Eyre-Walker A, Harr B, Keightley PD (2010) Evidence for pervasive adaptive protein evolution in wild mice. *PLoS Genet* 6:e1000825.
- Clark RM, et al. (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317:338–342.
- Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA (2008) High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol* 6:e251.
- Coop G, et al. (2009) The role of geography in human adaptation. *PLoS Genet* 5:e1000500.
- Burke MK, et al. (2010) Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467:587–590.
- Zhang M, Azad P, Woodruff RC (2011) Adaptation of *Drosophila melanogaster* to increased NaCl concentration due to dominant beneficial mutations. *Genetica* 139:177–186.
- Asthana S, Schmidt S, Sunyaev S (2005) A limited role for balancing selection. *Trends Genet* 21:30–32.
- Bubb KL, et al. (2006) Scan of human genome reveals no new loci under ancient balancing selection. *Genetics* 173:2165–2177.
- Andres AM, et al. (2009) Targets of balancing selection in the human genome. *Mol Biol Evol* 26:2755–2764.
- Makinen HS, Cano JM, Merila J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations. *Mol Ecol* 17:3565–3582.
- Holdren CE, Ehrlich PR (1981) Long range dispersal in checkerspot butterflies: Transplant experiments with *Euphydryas gillettii*. *Oecologia* 50:125–129.
- Lerner IM (1954) *Genetic Homeostasis* (Wiley, New York), p 134.
- Dobzhansky T (1952) Nature and the origin of heterosis. *Heterosis: A Record of Researches Directed Toward Explaining and Utilizing the Vigor of Hybrids* (Iowa State College Press, Ames, IA), pp 218–223.
- Crow JF (1987) Muller, Dobzhansky, and overdominance. *J Hist Biol* 20:351–380.
- Ewens WJ (2004) *Mathematical Population Genetics: I. Theoretical Introduction* (Springer, New York), p 417.
- Hastings A, Hom CL (1989) Pleiotropic stabilizing selection limits the number of polymorphic loci to at most the number of characters. *Genetics* 122:459–463.
- Burger R, Gimelfarb A (2002) Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. *Genet Res* 80:31–46.
- Kimura M (1956) Rules for testing stability of a selective polymorphism. *Proc Natl Acad Sci USA* 42:336–340.
- Galassi M (2009) *GNU scientific library: Reference manual* (Network Theory, Bristol, UK).