Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift

Meike J. Wittmann[a,b,c,1], Alan O. Bergland[a,d], Marcus W. Feldman[a], Paul S. Schmidt[e], and Dmitri A. Petrov[a,1]

[a]Department of Biology, Stanford University, Stanford, CA 94305; [b]Fakultät für Mathematik, Universität Wien, 1090 Wien, Austria; [c]Fakultät für Biologie, Universität Bielefeld, 33615 Bielefeld, Germany; [d]Department of Biology, University of Virginia, Charlottesville, VA 22904; and [e]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104-6313

Edited by M. T. Clegg, University of California, Irvine, CA, and approved October 3, 2017 (received for review March 8, 2017)

Most natural populations are affected by seasonal changes in temperature, rainfall, or resource availability. Seasonally fluctuating selection could potentially make a large contribution to maintaining genetic polymorphism in populations. However, previous theory suggests that the conditions for multilocus polymorphism are restrictive. Here, we explore a more general class of models with multilocus seasonally fluctuating selection in diploids. In these models, the multilocus genotype is mapped to fitness in two steps. The first mapping is additive across loci and accounts for the relative contributions of heterozygous and homozygous loci—that is, dominance. The second step uses a nonlinear fitness function to account for the strength of selection and epistasis. Using mathematical analysis and individual-based simulations, we show that stable polymorphism at many loci is possible if currently favored alleles are sufficiently dominant. This general mechanism, which we call "segregation lift," requires seasonal changes in dominance, a phenomenon that may arise naturally in situations with antagonistic pleiotropy and seasonal changes in the relative importance of traits for fitness. Segregation lift works best under diminishing-returns epistasis, is not affected by problems of genetic load, and is robust to differences in parameters across loci and seasons. Under segregation lift, loci can exhibit conspicuous seasonal allele-frequency fluctuations, but often fluctuations may be small and hard to detect. An important direction for future work is to formally test for segregation lift in empirical data and to quantify its contribution to maintaining genetic variation in natural populations.

temporal heterogeneity | cyclical selection | genetic diversity | marginal overdominance | balancing selection

Ever since biologists were first able to detect population genetic variation at the molecular level, they have been puzzled by its abundance in natural populations (1). Dispute over the underlying reasons gave rise to two scientific schools (2, 3). Proponents of the "(neo)classical" school claim that the bulk of genetic variation is due to neutral or weakly deleterious mutations present at an equilibrium between mutation, genetic drift, and selection. The neoclassical view admits that selection may maintain alleles at intermediate frequency at some loci, but argues that such loci are exceedingly rare on a genomic scale (2). By contrast, the "balance" school posits that a substantial fraction of variation is maintained by some form of balancing selection [with some controversy over the meaning of "substantial" (3)]—for example, heterozygote advantage (overdominance), negative frequency-dependent selection, and spatial or temporal variability in selection pressures (4).

Fifty years later, the debate has not been conclusively settled (5, 6), although the majority view is that (nearly) neutral mutations cause most genetic variation, with overdominance playing a relatively minor part, perhaps acting at only tens of loci per species (7–9). A mechanism considered more common and powerful is spatial environmental heterogeneity. Temporal heterogeneity, by contrast, is believed to be of limited importance (10), despite widespread temporal fluctuations in the strength and direction of selection, both on phenotypes (11) and genotypes (12). In fact, most organisms with multiple generations per year experience a particular type of temporal heterogeneity: seasonality, for example, in temperature, rainfall, resource availability, or in the abundance of predators, competitors, or parasites. Even tropical populations usually experience some seasonality. For example, flowering and fruiting in tropical forests is often synchronized within and between tree species, leading to seasonal changes in food availability for animals (13). Often, there are life-history trade-offs across seasons (14, 15). For example, seasons with abundant resource supply might select for investment in reproduction, whereas stressful seasons may select for investment in survival. Since such life-history traits are usually polygenic, many organisms should experience seasonally fluctuating selection at a large number of loci.

With discrete generations, the fates of genotypes under temporally fluctuating selection depend on their geometric mean fitnesses over time (16). In haploids, two alleles generally cannot coexist because one will have a higher geometric mean fitness and eventually go to fixation (ref. 16, but see refs. 17 and 18). In diploids, polymorphism at a single locus is stable if heterozygotes have the highest geometric mean fitness ("marginal overdominance"), although in any particular generation, one of the homozygotes might be fittest (16, 19, 20).

Significance

A key question in evolutionary biology is: What maintains the abundant genetic variation observed in natural populations? Many organisms experience some seasonality in their habitats, and, if they have multiple generations per year, seasonally fluctuating selection is a potentially powerful mechanism to maintain polymorphism. However, previous research has argued that this occurs rarely. Inspired by recent empirical findings, we reevaluate the potential of seasonally fluctuating selection to simultaneously maintain polymorphism at many loci in the genome. We obtain a more general condition for the maintenance of multilocus polymorphism by seasonally fluctuating selection. This condition may plausibly be satisfied for many species and does not suffer from problems of previous models.

Author contributions: M.J.W., A.O.B., M.W.F., P.S.S., and D.A.P. designed research; M.J.W. performed analyses and simulations; A.O.B., M.W.F., and D.A.P. gave input on all aspects of the analyses; P.S.S. gave input on the paper; and M.J.W, A.O.B., M.W.F., and D.A.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Source code underlying the analyses in this manuscript has been deposited in the figshare repository (available at https://doi.org/10.6084/m9.figshare.5142262).

[1]To whom correspondence may be addressed. Email: meike.wittmann@uni-bielefeld.de or dpetrov@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1702994114/-/DCSupplemental.

Extending these results to the multilocus case is nontrivial, and, so far, only two cases are well-understood: (*i*) multiplicative selection across loci and (*ii*) temporally fluctuating selection on a fully additive trait. Under multiplicative selection in an infinite population with free recombination, the allele-frequency dynamics at a focal locus are independent of those at other loci. Thus, polymorphism is stable if heterozygotes have the highest geometric mean fitness, as in the single-locus case. However, deviations from multiplicative selection appear to be the rule. In particular, beneficial mutations often exhibit diminishing-returns epistasis (21–23). Additionally, there is the potential problem of genetic load. Genetic load is commonly defined as the difference between the population's average fitness and the fitness of the fittest possible genotype. Lewontin and Hubby (1) noticed that this value can become unsustainably high if there is strong heterozygote advantage at many loci. This was a conundrum for the neoclassical school, which was worried that with high genetic load, single individuals would have to produce an astronomically large number of offspring. Others have dismissed this concern, arguing, for example, that selection does not generally act on all loci independently or that only relative fitness differences within the population are relevant, not fitness relative to some optimum genotype that might not even exist (24–27). However, debate continues over whether genetic load should be an important consideration (28, 29).

The second previously studied scenario is seasonally fluctuating selection on a trait to which loci contribute additively (30, 31). These models generally assume additivity also within loci, such that the contribution of heterozygotes is exactly intermediate between the contributions of the two homozygotes. Temporally fluctuating selection can then cause intermediate trait values to be best in the long run (32), i.e., select against variance in fitness. Effectively, this is stabilizing selection on the temporal average. As such, it can generally maintain polymorphism at only one locus (33, 34), or two loci if their effect sizes are sufficiently different (35) or if they are closely linked (30). The reason is that with multiple loci and additivity within and between loci, there are multiple genotypes with intermediate phenotypes. For two loci, for example, there is the double heterozygote ("heterozygous intermediate") and the genotype homozygous at both loci, but for alleles with opposite effects ("homozygous intermediate"). These genotypes may all have the same high fitness. However, matings between heterozygous intermediates produce a range of different genotypes, some of which are less fit than their parents. By contrast, matings between homozygous intermediates only produce new homozygous intermediates. Homozygote intermediates can therefore go to fixation and eliminate all polymorphism.

In summary, multiplicative seasonal selection is a powerful mechanism to maintain multilocus polymorphism, but the assumed independence across loci and the associated load call into doubt its plausibility. On the other hand, selection on additive traits can maintain polymorphism at only a few loci. So far, there has been little need for further exploration because there were no clear empirical examples to challenge the view that temporal heterogeneity rarely maintains variation. This is now changing, however, as advances in sequencing technology allow detailed studies of genetic variation across time and space. For instance, by sampling the same temperate population of *Drosophila melanogaster* at several time points, Bergland et al. (36) detected seasonal allele-frequency fluctuations at hundreds of sites in the genome. Many of the SNPs are also shared with African populations of *D. melanogaster* and some even with the sister species *Drosophila simulans*, indicating that some of them may be ancient balanced polymorphisms. More generally, recent population genomic data appear to suggest that balancing selection contributes more to maintaining genetic variation than previously assumed (37) and that mutation-selection-drift balance alone is not sufficient to reconcile evidence from population genomics and quantitative genetics (38). Thus, we need to reconsider the potential of temporally fluctuating selection to maintain multilocus polymorphism.

As explained above, the conditions for multilocus polymorphism under seasonally fluctuating selection have been examined mostly in two narrow cases. Here, we examine a more general class of seasonal selection models with various forms of dominance and epistasis. Using deterministic mathematical analysis and stochastic simulations, we show that multilocus polymorphism is possible if the currently favored allele at any time is sufficiently dominant, with dominance measured by using a scale on which contributions across loci are additive. This mechanism, which we call "segregation lift," can maintain polymorphism at a large number of loci across the genome, is robust to many model perturbations, and does not require single individuals to have too many offspring. Depending on the parameter values, allele-frequency fluctuations can be large and readily detectable, or subtle and hard to discern.

Basic Model

We consider a diploid, randomly mating population in a seasonally fluctuating environment. While asymmetry in various model parameters will be explored later, we start with a fully symmetric model having a yearly cycle with g generations of winter followed by g generations of summer. The genome consists of L unlinked loci with two alleles each: one summer-favored and one winter-favored allele. For a given multilocus genotype, let n_s and n_w be the number of loci homozygous for the summer and winter allele, respectively, and n_{het} the number of heterozygous loci, with $n_s + n_w + n_{het} = L$.

In the basic model, loci are interchangeable in their effects (see *Stochastic Simulations* for a more general model), and the fitness of a multilocus genotype can be computed as a function of n_s, n_w and n_{het}. In the simplest case, fitness depends only on $n_s + 0.5 \cdot n_{het}$ in summer and $n_w + 0.5 \cdot n_{het}$ in winter, i.e., half the number of currently favored allele copies. To allow for dominance effects, we generalize this simple scenario and assume that fitness in summer depends on the summer score

$$z_s := n_s + d_s \cdot n_{het} \quad [1]$$

and fitness in winter depends on the winter score

$$z_w := n_w + d_w \cdot n_{het}. \quad [2]$$

The parameters d_s and d_w quantify the dominance of the currently favored allele in summer and winter, not with respect to fitness, but with respect to the seasonal scores z_s and z_w. Because we are interested in whether temporally fluctuating selection can maintain polymorphism in the absence of other stabilizing mechanisms, we only consider values of d_s and d_w between 0 and 1, and do not allow values >1, which would correspond to standard heterozygote advantage.

The relationship between the seasonal score z ($z = z_s$ in summer and $z = z_w$ in winter), and fitness, w, is given by a monotonically increasing fitness function $w(z)$. This function specifies the strength of selection and accounts for epistasis. With discrete generations, the allele-frequency dynamics at a focal locus are driven by the relative fitnesses of the three possible genotypes at that locus—for example, the ratio of the fitness of homozygotes and heterozygotes. We say that there is no epistasis if these ratios and thus the strength of selection are independent of the number of other loci and their contributions to z. This is the case when fitness is multiplicative across loci:

$$w = \prod_{i=1}^{L} w_i \Leftrightarrow \ln(w) = \sum_{i=1}^{L} \ln(w_i), \quad [3]$$

where w_i is the fitness value at locus i. In our model, this is achieved by setting $w(z) = \exp(z)$ because then Eq. **3** is fulfilled with $w_i = \exp(1)$ if locus i is homozygous for the currently favored allele, $w_i = \exp(d_s)$ or $w_i = \exp(d_w)$ if it is heterozygous, and $w_i = \exp(0) = 1$ if it is homozygous for the currently disfavored allele. With $v(z) := \ln(w(z))$, the multiplicative model is characterized by $v''(z) = 0$. We therefore use the second derivative of the logarithm of fitness $v''(z)$ as a measure of epistasis (see ref. 39 for a similar definition of epistasis). Under positive or synergistic epistasis ($v''(z) > 0$), the logarithm of fitness increases faster than linearly with z, and thus selection at a focal locus increases in strength with increasing contribution of the other loci to z. By contrast, under negative or diminishing-returns epistasis ($v''(z) < 0$), selection at a focal locus becomes weaker with increasing contribution of other loci to z. We focus on two classes of fitness functions (Fig. 1). The first is of the form

$$w(z) = (1 + z)^y \quad \textbf{[4]}$$

with a positive parameter y. Although for $y > 1$ in Eq. **4**, fitness increases faster than linearly with increasing z (Fig. 1*A*), transformation to the logarithmic scale $v(z) = y \cdot \ln(1 + z)$ reveals that epistasis is negative for all y (Fig. 1*B*). Epistasis (v'') becomes more negative with increasing y, but v' and thus the selective advantage of having an additional favored allele still increases with y for all z. Thus, larger y values correspond to overall stronger per-locus selection.

The second class of fitness functions is of the form

$$w(z) = \exp(z^q) \quad \textbf{[5]}$$

with $q \geq 1$. This function reduces to the multiplicative model with $q = 1$ (cyan lines in Fig. 1) and has positive epistasis with $q > 1$ (e.g., magenta lines in Fig. 1).

In summary, fitness is computed in two steps. The first maps the multilocus genotype onto a seasonal score z to which loci contribute additively, essentially a generalized counter of the number of favored alleles (Eqs. **1** and **2**), and the second maps z to fitness (Fig. 1). This two-step process disentangles dominance (step 1) from selection strength and epistasis (step 2).

In our model, genotypes with many summer alleles have a high summer score but a low winter score and vice versa, a form of antagonistic pleiotropy. Previous theoretical studies suggest that antagonistic pleiotropy is most likely to maintain polymorphism if for each trait affected by a locus the respective beneficial allele is dominant (40, 41). Such "reversal of dominance" also facilitates the maintenance of polymorphism in single-locus models for temporally fluctuating selection (42). Hypothesizing that reversal of dominance would also help to maintain polymorphism under multilocus temporally fluctuating selection, we assume that d_s and d_w in Eq. **1** and **2** take the same value,

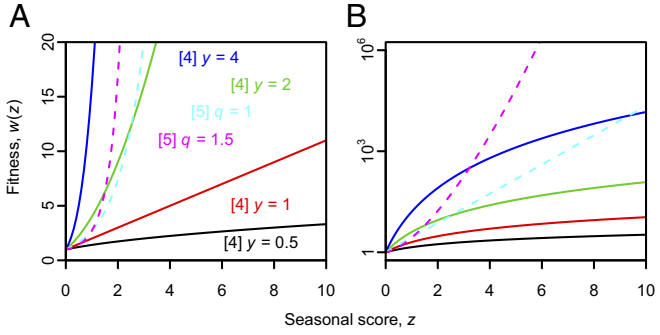

Fig. 1. Examples for fitness functions generated by Eq. **4** or Eq. **5** with various parameters. In *A*, fitness is shown on a linear scale, and in *B*, on a logarithmic scale.

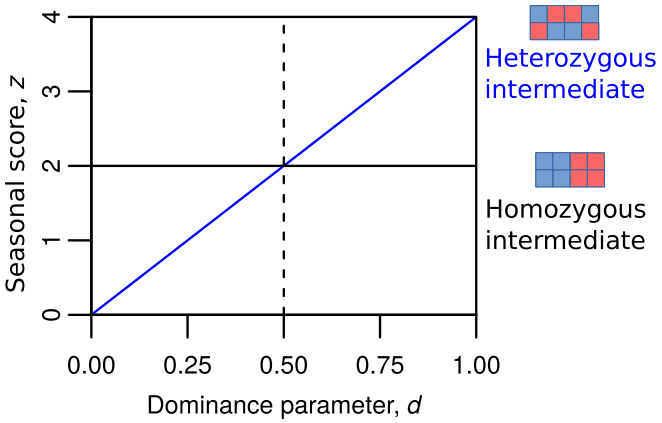

Fig. 2. Values of the seasonal score, *z*, as a function of the dominance parameter, *d*, for two example four-locus genotypes: a heterozygous intermediate (blue line) and a homozygous intermediate (black solid line).

d, which means that dominance switches between seasons (see *Stochastic Simulations* for a more general model). For $d < 0.5$, we have "deleterious reversal of dominance" (41), and the currently favored allele is always recessive, whereas for $d > 0.5$, we have "beneficial reversal of dominance" (41), and the currently favored allele is always dominant. If $d = 0.5$, the seasonal score z is additive, not just between loci, but also within loci. Importantly, the value of d also determines the relative fitness of heterozygous intermediates, multilocus genotypes with the same number of summer and winter alleles and at least some heterozygous loci, compared with homozygous intermediates, which also have the same number of summer and winter alleles, but are fully homozygous (Fig. 2). For $d < 0.5$, heterozygous intermediates have a lower seasonal score, z, and therefore a lower fitness in both seasons than homozygous intermediates. For $d = 0.5$, heterozygous and homozygous intermediates have the same score and fitness. Finally, for $d > 0.5$, heterozygous intermediates have a higher score and fitness. Interestingly, because d measures dominance not at the scale of fitness but at the scale of the seasonal score, z, beneficial reversal of dominance for fitness is neither sufficient nor necessary for $d > 0.5$ (*SI Appendix*, Fig. S1).

Multiple mechanisms could underlie a seasonal reversal of dominance. For example, metabolic control theory suggests that deleterious mutations affecting multistep metabolic pathways are generally recessive (43). If selection is fluctuating such that each allele is favored during one season and deleterious in the other season, we might thus expect a beneficial reversal of dominance. Alternatively, changes in dominance could be mediated by seasonal changes in gene expression. But even without changes in the genotype–phenotype map, seasonal changes in dominance are possible. In the example scenario in Fig. 3, the additive seasonal score, z, is a composite phenotype, a weighted average of two (also additive) traits—for example, starvation tolerance and fecundity—and there is antagonistic pleiotropy. Although the allelic effects on the two traits remain constant throughout the year, $d > 0.5$ because the relative importance of the traits changes between seasons. This scenario requires changes (though not necessarily a reversal; *SI Appendix*, Fig. S2*A*) in dominance with respect to the pleiotropic effects of a locus on the two traits. For example, the winter allele (blue) in Fig. 3 produces higher starvation tolerance and is dominant for this trait, whereas it leads to smaller fecundity and is recessive there. One specific way in which such changes in dominance across pleiotropic effects can arise is via branched enzyme pathways with saturation or feedbacks (44).

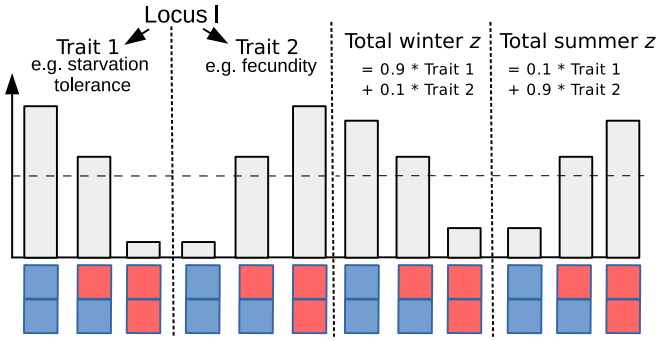

Fig. 3. Potential mechanistic underpinning for beneficial reversal of dominance. There is antagonistic pleiotropy for two traits, and the seasonal scores for winter and summer are computed as weighted averages of traits 1 and 2, with the relative importance of the two traits switching between seasons. The dashed line indicates the average of the two homozygote traits. If the heterozygotes are closer to the fitter homozygote with respect to both traits 1 and 2, there is a beneficial reversal of dominance at the level of the seasonal score, z. See *SI Appendix*, Fig. S2 for alternative scenarios.

Deterministic Analysis

In this section, we assume that population size is so large that genetic drift does not play a role. We also assume that mutations are rare enough that the allele-frequency dynamics will equilibrate before a new mutation arises at one of the L loci. This simple deterministic framework allows us to develop an intuitive understanding of the conditions for stable polymorphisms for various genotype-to-fitness maps. The intuitions developed here will then be checked and extended with stochastic simulations in the next section.

We will first confirm that the conditions under which seasonally fluctuating selection can maintain polymorphism are restrictive when contributions to the seasonal score z are additive within loci ($d = d_s = d_w = 0.5$ in Eqs. **1** and **2**). Then, $z_s + z_w = L$ for all possible genotypes, and the mean z over time is $z^* = L/2$. The long-term success of a genotype depends on its geometric mean fitness, or, equivalently, on the arithmetic mean of the logarithm of fitness, $v(z)$. Jensen's inequality or simple geometric considerations (Fig. 4) tells us that the arithmetic mean of $v(z_s)$ and $v(z_w)$ for a given genotype will be smaller than or equal to $v(z^*)$ if $v'' < 0$ everywhere (Fig. 4*A*), equal to $v(z^*)$ if $v'' = 0$ (Fig. 4*B*), and larger than or equal to $v(z^*)$ if $v'' > 0$ (Fig. 4*C*).

The interannual allele-frequency dynamics (e.g., from summer to summer or from winter to winter) with multiplicative fitness ($v'' = 0$) and $d = 0.5$ are thus neutral. No balancing selection emerges. With positive epistasis ($v'' > 0$), extreme types with either only summer or only winter alleles have the highest geometric mean fitness. Therefore, the population ends up in a state where all loci are fixed for the summer allele or all for the winter allele. With negative epistasis ($v'' < 0$), the genotypes with the highest geometric mean fitness are those with the same number of summer and winter alleles and thus $z_s = z_w$. There are always some genotypes heterozygous at one or more loci that fulfill this condition (heterozygous intermediates; Fig. 2). For an even number of loci, $z_s = z_w$ is also true for genotypes homozygous for the summer allele at half of the loci and homozygous for the winter allele at the other half (homozygous intermediates). When one of the homozygous intermediates fixes in the population, it cannot be invaded by any mutant starting at small frequency (under the assumptions of the deterministic model; see *SI Appendix*, section S1 for a detailed proof), and all polymorphism is eliminated. For an odd number of loci, homozygous intermediates do not exist, and some polymorphism may be maintained, at least at one locus. This case, which we will examine in more detail below, appears to be the only way in which seasonally fluctuating selection can maintain polymorphism under additivity ($d = 0.5$).

Next, we explore whether deviations from additivity ($d \neq 0.5$) can facilitate multilocus polymorphism. Under multiplicative selection, i.e., without epistasis, the conditions for polymorphism at one locus are not affected by the dynamics at other loci. Thus, given the fitness values for individual loci ($\exp(d)$ for heterozygotes, $\exp(1)$ and $\exp(0)$ for currently favored and disfavored homozygotes, respectively, see text below Eq. **3**), we can conclude that polymorphism is possible if

$$\exp(d)^2 > \exp(1) \cdot \exp(0) \Leftrightarrow d > 0.5. \quad [6]$$

That is, there must be a reversal of dominance with respect to z, such that at any time the currently favored allele is dominant.

Now, we explore whether such a beneficial reversal of dominance can also maintain polymorphism in the presence of epistasis. In each case, a necessary condition for polymorphism is that a population fixed for the fittest possible fully homozygous genotype can be invaded by mutants. As we have seen above, with synergistic epistasis ($v'' > 0$), there are two fully homozygous genotypes with maximum fitness, the one with the summer allele at all loci and the one with the winter allele at all loci. In both cases, the resident type has score L in one season and score 0 in the other season, whereas mutants differing in one position have scores $L - 1 + d$ and d. For mutants to invade, we thus need

$$v(d) + v(L - 1 + d) > v(L) + v(0). \quad [7]$$

Using our example class of fitness functions with positive epistasis, Eq. **5** with $q > 1$, we thus obtain the condition

$$d^q + (L - 1 + d)^q > L^q. \quad [8]$$

The critical value of d, d_{crit}, at which extreme types become invasible, satisfies

$$\left(\frac{d_{crit}}{L}\right)^q + \left(1 + \frac{d_{crit} - 1}{L}\right)^q = 1. \quad [9]$$

For $q = 2$, d_{crit} takes values 0.707, 0.954, and 0.995, with 1, 10, and 100 loci, respectively. For $q > 1$ in general, d_{crit} approaches 1 as the number of loci increases. To see this, note first that the condition in Eq. **9** is always fulfilled for $d = 1$ and thus $d_{crit} \leq 1$. Thus, as the number of loci, L, goes to infinity, $(d_{crit} - 1)/L$ becomes small, and we can approximate the second term on the left-hand side of Eq. **9** by a Taylor expansion around 1 to obtain

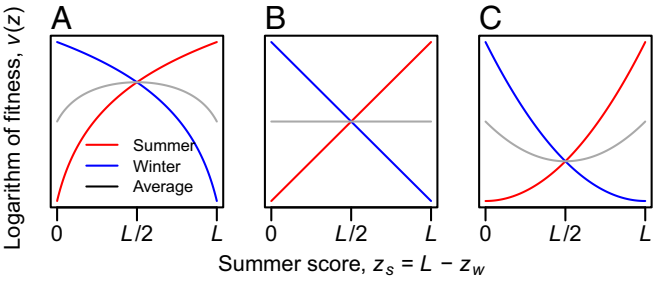

Fig. 4. The logarithm of summer fitness (red) and winter fitness (blue) and the average logarithm of fitness (gray) as a function of a genotype's summer score, z_s. Assuming $d = 0.5$, the winter score is $z_w = L - z_s$, leading to the mirror symmetry around $L/2$. If the fitness function is concave on the logarithmic scale, intermediate types have the highest average log-fitness (*A*); if the fitness function is log-linear, then all types have the same average log-fitness (*B*); and if the fitness function is convex on a logarithmic scale, extreme types have the highest average log-fitness and thus the highest geometric mean fitness (*C*).

$$\left(\frac{d_{crit}}{L}\right)^{q} + 1 + q \cdot \frac{d_{crit} - 1}{L} + \mathcal{O}\left(\frac{1}{L^2}\right) = 1. \quad [10]$$

Multiplying both sides by L and letting L go to infinity, we can conclude that d_{crit} is approximately 1 if the number of loci is large. Thus, for fitness functions of the type in Eq. **5** with positive epistasis, seasonally fluctuating selection can, in principle, maintain polymorphism at many loci, but the respective favored allele would have to be almost completely dominant, requiring large seasonal changes in dominance.

With diminishing-returns epistasis ($v'' < 0$), the fittest possible fully homozygous genotype carries the summer allele at half of the loci and the winter allele at the other half of the loci (assuming an even number of loci). Thus, a necessary condition for polymorphism is that this homozygous intermediate type can be invaded by mutants. The resident type has score $L/2$ in both seasons whereas mutants differing in one position have scores $L/2+d$ and $L/2-1+d$. Thus, the resulting necessary condition for polymorphism is

$$w(L/2+d) \cdot w(L/2-1+d) > w(L/2)^2. \quad [11]$$

Again, this condition is always fulfilled for $d = 1$. For fitness functions of the form Eq. **4** with any exponent y, the critical dominance coefficient d_{crit} at which homozygous intermediates become invasible satisfies

$$\left(1 + \frac{L}{2} + d_{crit}\right)\left(1 + \frac{L}{2} - 1 + d_{crit}\right) = \left(1 + \frac{L}{2}\right)^2. \quad [12]$$

This quadratic equation has a negative solution, which is not relevant for our model, and a positive solution

$$d_{crit} = \frac{1}{2}\left(-1 - L + (2+L)\sqrt{1 + \frac{1}{(2+L)^2}}\right). \quad [13]$$

From Eq. **13**, d_{crit} decreases as L increases and approaches 0.5 as L goes to infinity. The intuition here is that the second derivative of the logarithm of fitness $v''(z) = -y(1+z)^{-2}$ decreases with increasing z. Therefore, for large L, epistasis around the intermediate type with $z = L/2$ is weak, and the conditions for polymorphism approach those without epistasis. In other words, with increasing L, selection against temporal variation around the intermediate type becomes weaker, and a smaller change in dominance is sufficient to overcome it.

Our results so far suggest that for a broad class of fitness functions, seasonally fluctuating selection can maintain polymorphism if in both seasons the respective favored allele is sufficiently dominant. We call this mechanism "segregation lift" because it is based on a positive aspect of two alleles segregating at the same locus, as opposed to the negative aspect of segregation load. However, the preceding analysis does not tell us whether polymorphism will be maintained at all loci, or just one or a few of them. Also, it is still unclear how efficient segregation lift is at maintaining multilocus polymorphism in finite populations with genetic drift and recurrent mutations and whether genetic load is a problem. To address these questions, we now turn to stochastic simulations.

Stochastic Simulations

We use Wright–Fisher type individual-based forward simulations (see *SI Appendix*, section S2 for details). That is, for every individual in a generation independently, two individuals are sampled as parents in proportion to their fitnesses. We focus on diminishing-returns fitness functions of type Eq. **4** both because diminishing-returns epistasis appears to be more common and plausible (e.g., refs. 21–23) and because the above theoretical arguments suggest that it is more conducive to multilocus polymorphism than synergistic epistasis. Specifically, the critical dominance parameter, d_{crit}, for diminishing-returns epistasis in Eq. **13** is generally smaller than the one for synergistic epistasis in Eq. **9**. In addition, we run some simulations for the multiplicative model.

Table 1. Overview of model parameters and the ranges explored

Parameter	Explanation	Range explored
g	Number of generations per season; a year has $2g$ generations	1–20
N	Population size	100–10,000
L	Number of loci	1–500
μ	Mutation probability per allele copy per generation	$10^{-6} - 10^{-4}$
d	Dominance parameter	0–1
y	Exponent of the fitness function Eq. **4**	0.5–4

Additional parameters in the stochastic simulations are the symmetric mutation probability μ per allele copy per generation and the population size N. We generally keep population size constant, but also run supplementary simulations with seasonal changes in population size. Table 1 gives an overview of the model parameters and the ranges explored. In most simulated scenarios, selection and dominance effects are strong relative to mutation [$w'(z)$ and $d - 0.5$ are much larger than the mutation rate μ]. Although natural populations are often larger and mutation rates smaller than the values used here, many population genetic processes depend only on the product $N\mu$ (e.g., ref. 45). Thus, large populations with small mutation rates may be well approximated by computationally more manageable smaller populations with larger mutation rates.

In addition to the basic model, we design a "capped" model to assess the relevance of genetic load. In this model, each individual can be drawn at most 10 times as a parent of individuals in the next generation, i.e., contribute at most 10 gametes. Once an individual has reached that number, its fitness is set to 0 so that it cannot be drawn again. To better understand the role of offspring-number capping, we also run supplementary simulations with a cap of three, the smallest possible value that still allows for differences in offspring number between individuals in the population.

From the simulation output, we estimate an "effective strength of balancing selection" (*Materials and Methods* and *SI Appendix*, section S2), which tells us whether and how fast a rare allele increases in frequency over a full yearly cycle. As expected from the above theoretical arguments, additive contributions within loci ($d = 0.5$) are not conducive to multilocus polymorphism (Fig. 5). For even numbers of loci, i.e., situations where both homozygous and heterozygous intermediates exist (Fig. 2), the effective strength of balancing selection estimated from the simulations is negative, indicating that rare alleles tend to become even rarer. For small odd numbers of loci, the effective strength of balancing selection is positive, but only one or two loci at a time fluctuate at intermediate frequency (*SI Appendix*, Fig. S4). As the number of loci increases, the effective strength of balancing selection eventually becomes negative even for odd numbers, decreases overall in absolute value, and finally approaches zero (effective neutrality) from below (Fig. 5). This behavior is independent of the exponent, y, of the fitness function Eq. **4**. Also, as expected, effective balancing selection (*Materials and Methods*) emerges if the dominance parameter, d, is larger than a certain critical value, which decreases with the number of loci and is only weakly influenced by mutation rate (Fig. 6).

From now on, we will focus on scenarios with large numbers of loci. For the case of 100 loci, Fig. 7 shows example allele-frequency trajectories for three different dominance parameters, d. For small d, each locus is almost fixed either for the summer or winter allele. For large d, all loci fluctuate at intermediate

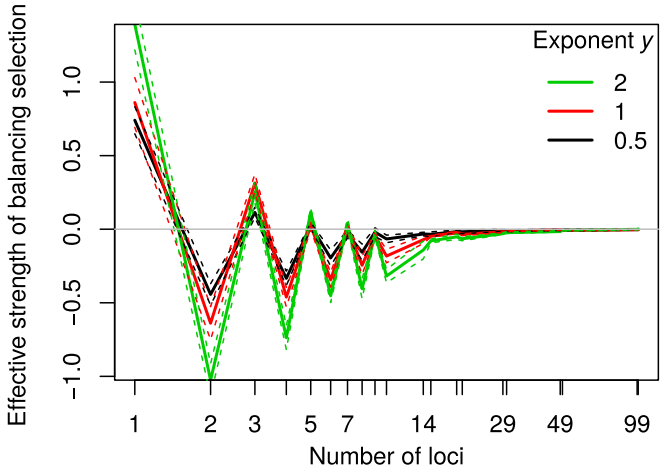

Fig. 5. Effective strength of balancing selection (b_e in Eq. **14** in *Materials and Methods*) in the additive case ($d = 0.5$) as a function of the number of loci. Solid lines indicate means and dashed lines indicate means ± two standard errors. Simulations are always run for successive odd and even numbers. $N = 1{,}000$, $g = 15$, $\mu = 0.0001$.

frequency. The critical dominance parameter, d_{crit}, with 100 loci is close to 0.5, independently of the exponent of the fitness function, y (Fig. 8*A*). For $d < 0.5$, i.e., if the currently favored allele is recessive, the effective strength of balancing selection is negative and polymorphism is unstable (Fig. 8*A*). As d increases beyond 0.5, i.e., as the currently favored allele becomes more dominant, effective balancing selection becomes stronger (Fig. 8*A*). Both the stabilizing and destabilizing effects increase with increasing exponent y (Fig. 8*A*).

A tendency for rare alleles to increase in frequency does not guarantee that the average lifetime of a polymorphism is larger than under neutrality (42, 46). This is particularly interesting for fluctuating selection regimes with positive autocorrelation where alleles regularly go through periods of low frequency (42). We therefore compute the so-called retardation factor (46), the average lifetime of a polymorphism in the selection scenario relative to the average lifetime under neutrality (see *SI Appendix*, section S2 for detailed methodology). The results for 100 loci are consistent with those for the effective strength of balancing selection: For $d > 0.5$, polymorphism under segregation lift is lost more slowly than under neutrality (Fig. 8*B*).

To quantify seasonal fluctuations, we compute an effective selection coefficient (*Materials and Methods* and *SI Appendix*, section S2). We also compute the predictability of fluctuations as the proportion of seasons over which the allele frequency changes in the expected direction, e.g., where the summer-favored allele increases over a summer season. Both the effective selection coefficient and the predictability of fluctuations have a maximum at intermediate values of d and increase with increasing exponent y of the fitness function Eq. **4** (Fig. 8 *C* and *D*). For even higher values of d, fluctuations are not as strong, presumably because heterozygotes are fitter, and therefore more copies of the currently disfavored allele enter the next generation. Also, effective strength of balancing selection, effective selection coefficient, and predictability of fluctuations increase with the number of generations per season (*SI Appendix*, Fig. S6).

With an offspring-number cap of 10, the results for the capped model generally match the results for the uncapped model in all respects, especially for $d > 0.5$ (Fig. 8). It appears that the capping mechanism only rarely takes effect because of relatively narrow distributions of the seasonal score, z, within a generation (*SI Appendix*, Fig. S7), and consequent relatively low variance in fitness (*SI Appendix*, Fig. S8). For example, for $d = 0.7$ and $y = 0.5$, the fittest individual in the population was on average 1.2 times fitter than the least fit individual and 4.9 times fitter for $y = 4$ (see *SI Appendix*, section S4 for a supporting heuristic analysis). When the offspring-number cap is set to three, substantial quantitative differences between uncapped and capped simulations are seen (*SI Appendix*, Fig. S9). However, 0.5 remains the critical dominance parameter. This result also holds for the multiplicative model (*SI Appendix*, Fig. S10), but with otherwise larger differences between capped and uncapped model versions, even with an offspring-number cap of 10. Also, the retardation factor for the multiplicative model is often below one, even when the effective strength of balancing selection is positive. The reason appears to be that there is larger variance in fitness under the multiplicative model (*SI Appendix*, Fig. S11) and that fluctuations are sometimes so large that alleles go to fixation (*SI Appendix*, Fig. S12). Both of these effects are weakened by offspring number capping, so that the capped multiplicative model behaves more similarly to the diminishing-returns model (*SI Appendix*, Fig. S10).

Additional simulations for the diminishing-returns model suggest that the finding of stable multilocus polymorphism for $d > d_{crit} \approx 0.5$ still holds under various forms of asymmetry, e.g., when one season has more generations than the other (*SI Appendix*, Fig. S13); when the exponent of the fitness function, y, differs between summer and winter (*SI Appendix*, Fig. S14); and when there are seasonal changes in population size (*SI Appendix*, Fig. S15). To explore the effects of asymmetry in the dominance parameters, we fixed the winter dominance parameter, d_w, at 0.4 and varied the summer dominance parameter, d_s. Stable polymorphism then arises for $d_s > 0.6$ (*SI Appendix*, Fig. S16), suggesting that stable polymorphism requires an arithmetic mean dominance parameter >0.5. Compared with the diminishing-returns model, the multiplicative model seems less robust to asymmetry (*SI Appendix*, Fig. S17).

With an increasing number of loci under the diminishing-returns model and with $d > 0.5$, the strength of balancing selection, the retardation factor, and the magnitude and predictability of fluctuations all decrease (Fig. 9). Population size hardly influences effective strength of balancing selection and effective

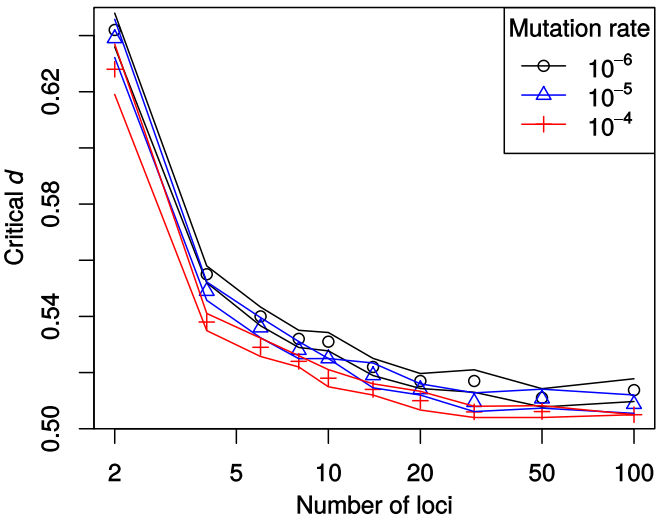

Fig. 6. Critical value of the dominance parameter, d_{crit}, such that the effective strength of balancing selection (b_e in Eq. **14** in *Materials and Methods*), is positive (stable polymorphism) if $d > d_{crit}$, and negative (unstable polymorphism) if $d < d_{crit}$. Symbols represent means across replicates, and lines represent averages ± two standard errors. Since the pattern for odd numbers of loci is more complex (Fig. 5), only even values for the number of loci are included here. $N = 1{,}000$, $y = 2$, $g = 15$.

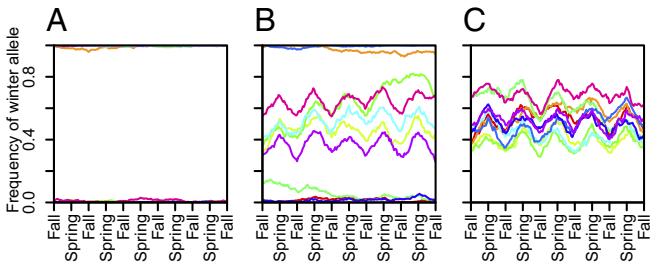

Fig. 7. Three examples of allele-frequency trajectories for $N = 1{,}000$, $L = 100$, $g = 15$, $y = 4$, $\mu = 10^{-4}$, and $d = 0.15$ (*A*), $d = 0.5$ (*B*), and $d = 0.65$ (*C*). Only 10 randomly selected loci (shown in different colors) out of 100 loci are shown for 5 years (150 generations) in the middle of the simulation run (years 301–305).

selection coefficient, measures which are based on average allele-frequency changes (Fig. 9 *A* and *C*), but large populations maintain polymorphism for longer (Fig. 9*B*) and have more predictable allele-frequency fluctuations (Fig. 9*D*). In small populations, polymorphism can even be lost slightly faster than under neutrality (Fig. 9*B*).

Finally, we consider a generalized model where parameters vary across loci and may be asymmetric between seasons. Independently, for each locus l, we draw four parameters: Summer effect size $\Delta_{s,l}$ and winter effect size $\Delta_{w,l}$ are drawn from a log-normal distribution. For this, we draw a pair of parameters from a bivariate normal distribution with mean 0, SD 1, and correlation coefficient 0.9, and then apply the exponential function to each of them. Summer and winter dominance parameters, $d_{s,l}$ and $d_{w,l}$, are drawn independently from a uniform distribution on [0,1]. Seasonal scores are then computed as $z = \sum_{l=1}^{L} c_l$, where the contribution c_l of locus l in summer is 0 for winter–winter homozygotes, $d_{s,l}\Delta_{s,l}$ for heterozygotes, and $\Delta_{s,l}$ for summer–summer homozygotes. Winter contributions are computed analogously. Because all effect sizes Δ are positive, all loci exhibit a trade-off between summer and winter effects. We use $y = 4$ here because it led to the most stable polymorphism in the basic model.

The results indicate that polymorphisms with different parameters can be maintained in the same population, with their allele frequencies fluctuating on various trajectories (Fig. 10*A*). With a sufficiently high total number of loci, hundreds of stable polymorphisms (positive expected frequency change of a rare allele; see *SI Appendix*, section S2 for details) can be maintained in populations of biologically plausible size (Fig. 10*B*). The number of loci classified as stable depends only weakly on population size. However, only a small proportion of the polymorphisms classified as stable also exhibit detectable allele-frequency fluctuations, defined as changes in the expected direction by at least 5% in at least half of the seasons (Fig. 10 *C* and *D*). The number of detectable polymorphisms is highest at an intermediate total number of loci and increases with population size (Fig. 10*D*). Detectable polymorphisms tend to have larger summer and winter effect sizes than polymorphisms that are only stable (Fig. 10*E*). Compared with unstable polymorphisms, stable polymorphisms are more balanced in their summer and winter effect sizes (two-sample t-test on $|\ln(\Delta_{s,l}/\Delta_{w,l})|$, $p < 2.2 \cdot 10^{-16}$; Fig. 10*E*; see also *SI Appendix*, Fig. S18). Many stable polymorphisms have asymmetric dominance parameters, but for almost all of them, detectable or not, the average dominance parameter across seasons is >0.5 (Fig. 10*F*).

Discussion

We study a simple model for seasonally fluctuating selection that maps the multilocus genotype to fitness in two steps. First, we count the number of loci homozygous for the currently favored allele and add the number of heterozygous loci weighted by a dominance parameter. The resulting seasonal score is then mapped to fitness via a monotonically increasing function which accounts for strength of selection and epistasis. The previously studied cases of multiplicative selection and selection on a fully additive phenotype are special cases of our model. We identify a general mechanism, segregation lift, by which seasonally fluctuating selection can maintain polymorphism at tens or hundreds of unlinked loci. Segregation lift requires that the average dominance parameter of the currently favored allele—the summer allele in summer and the winter allele in winter—is sufficiently large. Individuals with many heterozygous loci then have higher scores in both seasons than individuals with the same number of summer and winter alleles, but more homozygous loci. Unlike in previously studied additive models, fully homozygous types thus cannot fix in the population, and multilocus polymorphism is maintained. In some cases, segregation lift may also be interpreted as a type of phenotypic plasticity, where more heterozygous genotypes can better adjust to both summer and winter environments.

The critical value of the dominance parameter required to maintain polymorphism, d_{crit}, depends mostly on the type of epistasis and on the number of loci. Without epistasis, i.e., for multiplicative selection, d_{crit} is 0.5. With synergistic epistasis, it is close to one when there are multiple loci. With diminishing-returns epistasis, the type of epistasis that appears most common, d_{crit}, is substantially larger than 0.5 with few loci, but quickly approaches 0.5 as the number of loci increases.

Robustness and Plausibility of Segregation Lift as a Mechanism to Maintain Variation. Segregation lift requires that dominance changes over time such that the currently favored allele is on average at least slightly dominant with respect to the seasonal

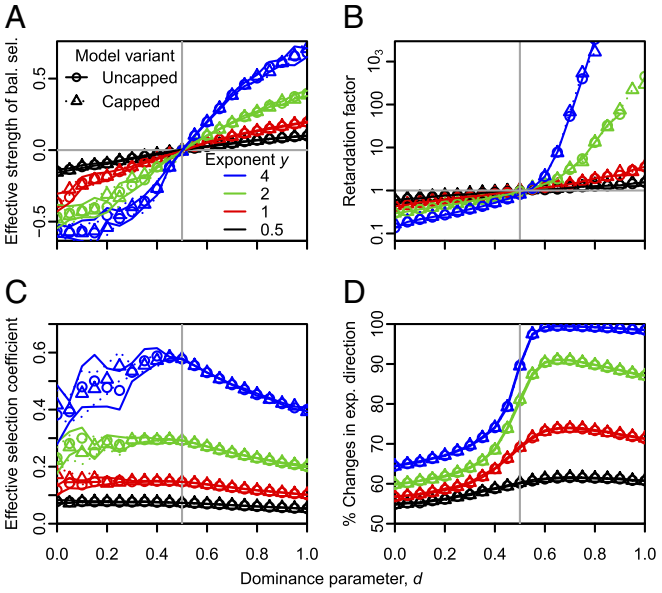

Fig. 8. Influence of the dominance parameter d on effective strength of balancing selection (*A*; b_e, Eq. **14**, *Materials and Methods*), retardation factor (*B*), magnitude of fluctuations (*C*; s_e, Eq. **15**, *Materials and Methods*), and predictability of fluctuations (*D*). Symbols indicate averages across replicates for the uncapped vs. capped model variant (often overlapping), and solid vs. dashed lines in *A*, *C*, and *D* indicate the respective means ± two standard errors. Lines in *B* simply connect maximum-likelihood estimates obtained jointly from all replicates. $N = 1{,}000$, $L = 100$, $g = 15$, $\mu = 10^{-4}$. See *SI Appendix*, Fig. S5 for more detailed information on the distribution and frequency dependence of seasonal allele frequency changes. The vertical gray lines are at $d = 0.5$. Bal. sel., balancing selection; exp., expected.

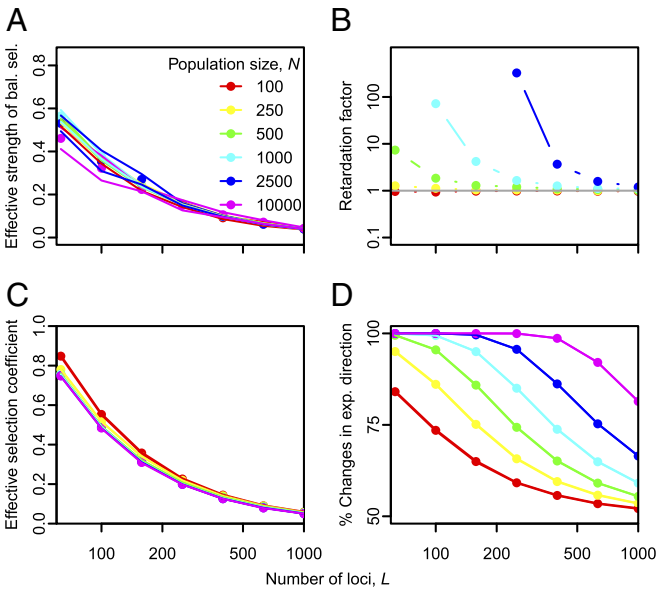

Fig. 9. Influence of population size and the number of seasonally selected loci on effective strength of balancing selection (*A*; b_e in Eq. **14**, *Materials and Methods*), retardation factor (*B*), magnitude of fluctuations (*C*; s_e in Eq. **15**, *Materials and Methods*), and predictability of fluctuations (*D*). Symbols indicate averages across replicates and lines in *A*, *C*, and *D* indicate means ± two standard errors (in *C* and *D*, standard errors are too small to be visible). Lines in *B* simply connect maximum-likelihood estimates obtained jointly from all replicates. Note that in *B*, some points are missing because the rate of loss of polymorphism was too small to be quantified. $d = 0.7$, $y = 4$, $g = 15$, $\mu = 10^{-4}$. Bal. sel., balancing selection; exp., expected.

score. As discussed above, there are several potential mechanisms that can plausibly produce such changes in dominance. Moreover, the required changes are small. Unfortunately, there have been only few relevant empirical studies so far. For instance, in the copepod *Eurytemora affinis*, there appears to be beneficial reversal of dominance for fitness across salinity conditions (47). In experimental *Drosophila* populations, changes in dominance for gene expression across environments appear to be common (48). More empirical and theoretical work is required to find out how common changes in dominance are, in particular on the relevant scale of the seasonal score. However, even if the required changes in dominance are rare on a per-site basis and the vast majority of polymorphisms are lost under fluctuating selection, there may still be many sites in the genome with appropriate reversal of dominance, and, as we show, those are then the ones that we should see as seasonally fluctuating polymorphisms.

In the focal diminishing-returns scenario, the conditions for stable polymorphism via segregation lift are surprisingly robust to changes in the mutation rate (Fig. 6) and to asymmetries in number of generations, strength of selection, or population size between summer and winter (*SI Appendix*, Figs. S13–S15), apparently more so than under the multiplicative model (*SI Appendix*, Fig. S17). When the dominance parameter differs between summer and winter, polymorphism is generally stable at those loci whose average dominance parameter across seasons is >0.5 (Fig. 10*F* and *SI Appendix*, Fig. S16). Segregation lift is also robust to variation in effect sizes and dominance parameters across loci (Fig. 10). In reality, the strength of seasonality likely varies in space and time, which could make the maintenance of polymorphism by segregation lift even more robust (*SI Appendix*, Fig. S2*B* and ref. 40). Future work needs to explore whether segregation lift is robust also to linkage between selected loci. Since our diminishing-returns fitness function has a particular relationship between epistasis and strength of selection, future work should also consider more general fitness functions allowing for various combinations of epistasis and selection strength.

Whenever there is balancing selection at a large number of loci, genetic load is a potential concern. In the case of segregation lift with diminishing-returns epistasis, however, genetic load does not appear to play an important role. The results for our capped model closely match the results for the original, uncapped model. Apparently, independent segregation at a large number of unlinked loci leads to relatively small variance in seasonal scores within the population and, together with the diminishing-returns fitness function, to relatively small variance in fitness. With a much smaller offspring number cap or with multiplicative selection, differences between the capped and uncapped model are more substantial, but even then, balancing

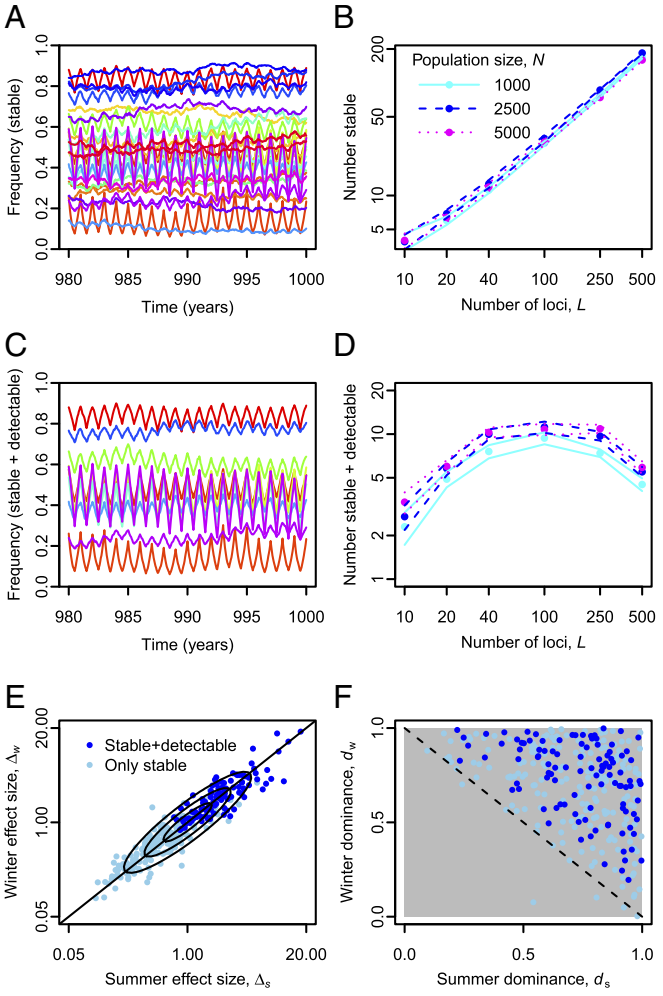

Fig. 10. Stability of polymorphism and detectability of allele-frequency fluctuations when parameters vary across loci and seasons. (*A*) Snapshot of allele-frequency trajectories for stable polymorphisms in one simulation run. (*B*) Average number of stable polymorphisms as a function of the total number of loci for different population sizes. (*C* and *D*) As in *A* and *B*, but only for polymorphisms that are also detectable. (*E*) Winter effect size, $\Delta_{l,w}$, vs. summer effect size, $\Delta_{l,s}$, for stable and detectable and only stable polymorphisms. The plot shows pooled results over ten simulation runs with independently drawn parameters. Oval isoclines indicate the shape of the original sampling distribution, with 75% of the sampling probability mass inside the outermost isocline. (*F*) Corresponding dominance parameters (see also *SI Appendix*, Fig. S19). The dominance parameters were originally drawn from a uniform distribution on the unit square. Parameters: $y = 4$, $g = 10$, $\mu = 10^{-4}$ and in *A*, *C*, *E*, and *F* $N = 10{,}000$, $L = 100$.

selection emerges for $d > 0.5$. Thus, unrealistically large offspring numbers are not required for stable multilocus polymorphism via segregation lift.

Magnitude and Detectability of Allele-Frequency Fluctuations. In addition to stable polymorphism, segregation lift can also produce strong and predictable seasonal fluctuations in allele frequencies. The magnitude of these fluctuations, however, decreases with the number of loci under selection. Thus, the number of detectable polymorphisms may be maximized at an intermediate number of loci (Fig. 10). In *SI Appendix*, section S3, we use a heuristic mathematical argument to explore the relationship between number of loci and magnitude of fluctuations in a population of infinite size. As the number of loci goes to infinity, the effective strength of selection at each locus is predicted to go to zero, i.e., effective neutrality. This is because more loci lead to higher overall seasonal scores, z, which under diminishing-returns epistasis leads to weaker average selection pressures at each locus. Thus, even if segregation lift contributes substantially to maintaining polymorphism at a large number of loci, it is not necessarily easy to detect individual selected loci based on their allele-frequency fluctuations. Future research will need to explore new ways of detecting such subtle seasonal allele-frequency fluctuations at many loci, perhaps based on their collective behavior rather than on patterns at individual sites.

Empirical Evidence, Alternative Hypotheses, and Potential Tests for Segregation Lift. As mentioned above, a recent pooled-sequencing study by Bergland et al. (36) detected strong seasonal allele-frequency fluctuations at hundreds of sites in a temperate population of *D. melanogaster*. At many sites, allele frequencies fluctuated by $\sim$10% over a single season of $\sim$10 generations, and many of the polymorphisms appear to be long-term stable. Based on our results, segregation lift could potentially explain these observations, but caution is warranted for several reasons. First, with hundreds of seasonal SNPs and only few chromosomes, there will necessarily be substantial linkage between some of the sites. Second, the distribution and dynamics of dominance effects at the seasonally selected loci are still unknown. Finally, it is not completely clear whether the observed magnitude of allele-frequency fluctuations can be explained by our segregation lift model, where fluctuations are often more subtle (Fig. 10 *C* and *D*, but see Fig. 7*C*). Based on our current knowledge, we therefore cannot claim that the empirical observations by Bergland et al. (36) are explained by segregation lift. Future work will need to empirically test this model and possible alternatives.

One alternative is that genetic variation is not really stably maintained, but simply induced by recurrent mutation, with selection responsible only for the seasonal fluctuations (31, 49), or by recurrent immigration from other subpopulations where either winter- or summer-favored alleles dominate. However, in the case of the *Drosophila* observations, this was considered unlikely (36). Alternative mechanisms that could lead to both fluctuations and long-term stability are (*i*) differential responses to fluctuating resource concentrations and population densities (17, 50–52) and (*ii*) a so-called "temporal storage effect" where genetic variation can be buffered by a long-lasting life-history stage on which selection does not act, or by some other protected state (51, 53, 54). However, these mechanisms are more commonly studied in ecology as mechanisms for species coexistence, and it is unclear whether they can maintain polymorphism at multiple loci in diploids.

In future empirical tests for segregation lift, a main challenge will be that the pivotal dominance parameter, d, is not relative to fitness but relative to the seasonal score, z, which mediates between multilocus genotype and fitness, and is itself not directly measurable. Since the shape of the fitness function, w, is also generally unknown, it is not possible to infer d from fitness measurements of different single-locus genotypes in a common genetic background. In an ideal situation, with fitness measurements for many different multilocus genotypes at different times, we could use statistical methods such as machine learning to jointly estimate parameters of the fitness function, effect sizes, and dominance parameters and thereby assess whether or not there is segregation lift. Such statistical approaches could also take into account the existence of several multiplicative fitness components, each with a set of contributing loci that might exhibit segregation lift and epistatic interactions. In practice, however, measuring fitness is challenging in itself. One productive direction could be to stock a large number of outdoor mesocosms, each with a different multilocus genotype, and track fitness over multiple seasons. However, apart from the logistic challenges, we do not know a priori which loci to focus on. Coming up with a meaningful and feasible way to empirically get at the scale of the seasonal score z and estimate the relevant dominance parameters is thus an important research direction arising from this study. An alternative approach is to make predictions for the genetic footprint of selection in linked neutral regions, e.g., look at diversity levels, site-frequency spectra, and patterns of linkage disequilibrium, and use these empirically more accessible patterns to distinguish between multiple possible models.

Conclusions

We identify segregation lift as a general mechanism by which seasonally fluctuating selection can maintain polymorphism at hundreds of unlinked loci in populations of biologically reasonable size. Segregation lift circumvents the problems associated with maintenance of polymorphism under stabilizing selection and does not require highly heterozygous individuals to have unrealistically many offspring. Given the ubiquity of environmental fluctuations, segregation lift could make a substantial contribution to genetic variation in natural populations of many taxa. An important question for future work is how we can use modern molecular biology and sequencing technologies to test for segregation lift and thus make progress on solving the puzzle of genetic variation.

Materials and Methods

For the basic model and the capped model, we assess stability of polymorphism by estimating an effective strength of balancing selection, b_e, from the year-to-year allele-frequency dynamics. For this, we fit a standard balancing selection model (55) of the form

$$\Delta_y x = b_e x(1-x)(1-2x) \quad [14]$$

to average changes in allele frequency over one yearly cycle, $\Delta_y x$ (see *SI Appendix*, section S2 for details). Positive values of b_e indicate that rare alleles tend to become more common in the long run, whereas negative values indicate that rare alleles tend to become even more rare. Second, we quantify the magnitude of fluctuations over individual seasons. For this, we fit a standard directional selection model (55)

$$\Delta_s x = s_e x(1-x), \quad [15]$$

with an effective selection coefficient s_e, to average allele-frequency changes over one season, $\Delta_s x$ (*SI Appendix*, section S2).

To obtain a measure for statistical uncertainty in our results, we run 10 replicates for every parameter combination and calculate effective strength of balancing selection, b_e, and effective selection coefficient, s_e, independently for each replicate. We do the same for the predictability of fluctuations, i.e., the proportion of seasons in which a locus changes its allele frequency in the expected direction. In all three cases, we report the mean over replicates $\pm$ two standard errors of the mean. To obtain retardation factors, we run 100 replicates until polymorphism is lost at one of the loci or a maximum time of 500 years is reached. From the times of loss for the replicates, we obtain maximum-likelihood estimators for the rate of loss of polymorphism (see *SI Appendix*, section S2 for details).

C++ simulation code and supporting R scripts are available at https://doi.org/10.6084/m9.figshare.5142262.

ACKNOWLEDGMENTS. For helpful discussion and/or comments on the manuscript, we thank Michael Desai, Joachim Hermisson, Oren Kolodny, Mike McLaren, Richard Nichols, Pleuni Pennings, Jitka Polechová, and members of D.A.P.'s laboratory, as well as two anonymous reviewers. Simulations were performed on Stanford's FarmShare Cluster and on the Vienna Scientific Cluster. M.J.W. was supported by fellowships from the Stanford Center for Computational Evolutionary and Human Genomics and from the Austrian Science Fund (M 1839-B29).

1. Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54:595–609.
2. Lewontin RC (1974) *The Genetic Basis of Evolutionary Change* (Columbia Univ Press, New York).
3. Beatty J (1987) Weighing the risks: Stalemate in the classical/balance controversy. *J Hist Biol* 20:289–319.
4. Hedrick PW (2007) Balancing selection. *Curr Biol* 17:R230–R231.
5. Barton N, Keightley P (2002) Understanding quantitative genetic variation. *Nat Rev Genet* 3:11–21.
6. Turelli M, Barton NH (2004) Polygenic variation maintained by balancing selection: Pleiotropy, sex-dependent allelic effects and G x E interactions. *Genetics* 166:1053–1079.
7. Asthana S, Schmidt S, Sunyaev S (2005) A limited role for balancing selection. *Trends Genet* 21:30–32.
8. Hedrick PW (2012) What is the evidence for heterozygote advantage selection? *Trends Ecol Evol* 27:698–704.
9. Croze M, Živković D, Stephan W, Hutter S (2016) Balancing selection on immunity genes: Review of the current literature and new analysis in *Drosophila melanogaster*. *Zoology* 119:322–329.
10. Hedrick PW, Ginevan ME, Ewing EP (1976) Genetic polymorphism in heterogeneous environments. *Annu Rev Ecol Syst* 7:1–32.
11. Siepielski AM, DiBattista JD, Carlson SM (2009) It's about time: The temporal dynamics of phenotypic selection in the wild. *Ecol Lett* 12:1261–1276.
12. Thurman TJ, Barrett RD (2016) The genetic consequences of selection in natural populations. *Mol Ecol* 25:1429–1448.
13. van Schaik CP, Terborgh JW, Wright SJ (1993) The phenology of tropical forests–Adaptive significance and consequences for primary consumers. *Annu Rev Ecol Syst* 24:353–377.
14. Schmidt PS, Conde DR (2006) Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in *Drosophila melanogaster*. *Evolution* 60:1602–1611.
15. Behrman EL, Watson SS, O'Brien KR, Heschel MS, Schmidt PS (2015) Seasonal variation in life history traits in two *Drosophila* species. *J Evol Biol* 28:1691–1704.
16. Haldane JBS, Jayakar SD (1963) Polymorphism due to selection of varying direction. *J Genet* 58:237–242.
17. Dean AM, Lehman C, Yi X (2017) Fluctuating selection in the Moran. *Genetics* 205:1271–1283.
18. Novak R, Barton NH (2017) When does frequency-independent selection maintain genetic variation? *Genetics* 207:653–668.
19. Dempster ER (1955) Maintenance of genetic heterogeneity. *Cold Spring Harb Symp Quant Biol* 20:25–32.
20. Gillespie JH (1973) Polymorphism in random environments. *Theor Popul Biol* 4:193–195.
21. Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332:1190–1192.
22. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332:1193–1196.
23. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM (2014) Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344:1519–1522.
24. Feller W (1967) On fitness and the cost of natural selection. *Genet Res* 9:1–15.
25. Milkman RD (1967) Heterosis as a major cause of heterozygosity in nature. *Genetics* 55:493–495.
26. Sved JA, Reed TE, Bodmer WF (1967) The number of balanced polymorphisms that can be maintained in a natural population. *Genetics* 55:469–481.
27. Turner JRG, Williamson MH (1968) Population size, natural selection and the genetic load. *Nature* 218:700.
28. Agrawal AF, Whitlock MC (2012) Mutation load: The fitness of individuals in populations where deleterious alleles are abundant. *Annu Rev Ecol Evol Syst* 43:115–135.
29. Charlesworth B (2013) Why we are not dead one hundred times over. *Evolution* 67:3354–3361.
30. Korol AB, Kirzhner VM, Ronin YI, Nevo E (1996) Cyclical environmental changes as a factor maintaining genetic polymorphism. 2. Diploid selection for an additive trait. *Evolution* 50:1432–1441.
31. Bürger R, Gimelfarb A (2002) Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. *Genet Res* 80:31–46.
32. Lande R (2008) Adaptive topography of fluctuating selection in a Mendelian population. *J Evol Biol* 21:1096–1105.
33. Wright S (1935) Evolution in populations in approximate equilibrium. *J Genet* 30:257–266.
34. Bürger R, Gimelfarb A (1999) Genetic variation maintained in multilocus models of additive quantitative traits under stabilizing selection. *Genetics* 152:807–820.
35. Nagylaki T (1989) The maintenance of genetic variability in 2-locus models of stabilizing selection. *Genetics* 122:235–248.
36. Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA (2014) Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet* 10:e1004775.
37. Gloss AD, Whiteman NK (2016) Balancing selection: Walking a tightrope. *Curr Biol* 26:R73–R76.
38. Charlesworth B (2015) Causes of natural variation in fitness: Evidence from studies of *Drosophila* populations. *Proc Natl Acad Sci USA* 112:1662–1669.
39. Desai MM, Weissman D, Feldman MW (2007) Evolution can favor antagonistic epistasis. *Genetics* 177:1001–1010.
40. Rose MR (1982) Antagonistic pleiotropy, dominance, and genetic variation. *Heredity* 48:63–78.
41. Curtsinger JW, Service PM, Prout T (1994) Antagonistic pleiotropy, reversal of dominance, and genetic polymorphism. *Am Nat* 144:210–228.
42. Hedrick PW (1976) Genetic variation in a heterogeneous environment. II. Temporal heterogeneity and directional selection. *Genetics* 84:145–157.
43. Kacser H, Burns JA (1981) The molecular basis of dominance. *Genetics* 97:639–666.
44. Keightley PD, Kacser H (1987) Dominance, pleiotropy and metabolic structure. *Genetics* 117:319–329.
45. Gillespie JH (1998) *Population Genetics—A Concise Guide* (The John Hopkins Univ Press, Baltimore).
46. Robertson A (1962) Selection for heterozygotes in small populations. *Genetics* 47:1291–1300.
47. Posavi M, Gelembiuk GW, Larget B, Lee CE (2014) Testing for beneficial reversal of dominance during salinity shifts in the invasive copepod *Eurytemora affinis*, and implications for the maintenance of genetic variation. *Evolution* 68:3166–3183.
48. Chen J, Nolte V, Schlötterer C (2015) Temperature stress mediates decanalization and dominance of gene expression in *Drosophila melanogaster*. *PLoS Genet* 11:e1004883.
49. Kondrashov A, Yampolsky L (1996) High genetic variability under the balance between symmetric mutation and fluctuating stabilizing selection. *Genet Res* 68:157–164.
50. Armstrong RA, McGehee R (1976) Coexistence of species competing for shared resources. *Theor Popul Biol* 9:317–328.
51. Chesson P (2000) Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* 31:343–366.
52. Yi X, Dean AM (2013) Bounded population sizes, fluctuating selection and the tempo and mode of coexistence. *Proc Natl Acad Sci USA* 110:16945–16950.
53. Svardal H, Rueffler C, Hermisson J (2015) A general condition for adaptive genetic polymorphism in temporally and spatially heterogeneous environments. *Theor Popul Biol* 99:76–97.
54. Gulisija D, Kim Y, Plotkin JB (2016) Phenotypic plasticity promotes balanced polymorphism in periodic environments by a genomic storage effect. *Genetics* 202:1437–1448.
55. Durrett R (2008) *Probability Models for DNA Sequence Evolution* (Springer, New York).

Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift Supplementary Information

Meike J. Wittmann[a,b,c,1], **Alan O. Bergland**[a,d], **Marcus W. Feldman**[a], **Paul S. Schmidt**[e], **and Dmitri A. Petrov**[a,1]

[a]Department of Biology, Stanford University, Stanford, CA 94305; [b]Fakultät für Mathematik, Universität Wien, 1090 Wien, Austria; [c]Fakultät für Biologie, Universität Bielefeld, 33615 Bielefeld, Germany; [d]Department of Biology, University of Virginia, Charlottesville, VA 22904; [e]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104-6313

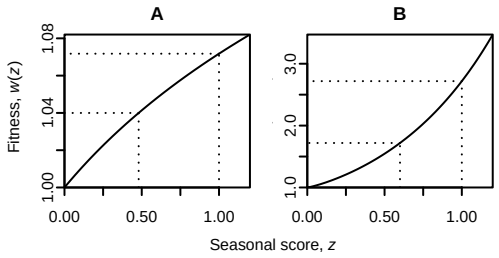

Fig. S1. Examples illustrating why a heterozygote fitness closer to the fitter homozygote is neither a sufficient nor a necessary condition for $d > 0.5$. If the fitness function is concave, the fitness of a heterozygote can be closer to the fitter homozygote even for $d < 0.5$ (A). On the other hand, if the fitness function is convex, the fitness of heterozygotes can be closer to the less fit homozygote even for $d > 0.5$ (B). It is assumed that the genetic background is the same for all genotypes and, for simplicity, that it makes contribution 0 to z. The dotted lines illustrate the mapping between $z = d$ and $z = 1$ and the respective fitnesses. A) Eq. (4) with $y = 0.1$. B) Eq. (5) with $q = 1.2$.

Appendix S1. Local stability analysis for the additive scenario ($d = 0.5$) with diminishing-returns epistasis

In this section, we consider a scenario where alleles contribute to the seasonal score z additively within and between loci ($d_s = d_w = 0.5$ in Eq. (1) and Eq. (2)). We show that in the deterministic model without recurrent mutation, a population fixed for a balanced haplotype with $L/2$ summer-favored alleles and $L/2$ winter-favored alleles (L even) cannot be invaded by any other haplotype. We give an intuitive rather than a mathematically formal outline of the stability analysis.

Assume that a certain balanced haplotype is at frequency close to 1 in the population. We will call it the resident haplotype. Let $\epsilon_{i,j}$ be the combined frequency of all haplotypes that have a winter allele at i positions where the resident haplotype has a summer allele and a summer allele at j positions where the resident haplotype has a winter allele. Thus an i,j-haplotype has $L/2 + i - j$ winter alleles, $L/2 - i + j$ summer alleles, and differs in $i + j$ positions from the resident haplotype.

Rare i,j-haplotypes will almost certainly occur in a genotype with a copy of the resident balanced haplotype. This has two consequences. First, recombination with the resident balanced haplotype will produce other haplotypes that differ in at most $i + j$ positions from the resident haplotype. Haplotypes with more than $i + j$ differences cannot be produced because if resident and invading haplotype have the same allele at a locus, then all offspring haplotypes will carry this allele as well. Given free recombination, the probability that one of the two recombinant haplotypes produced by an i,j-resident genotype (consisting of one i,j-haplotype and one resident haplotype) is of type i,j is $1/2^{i+j}$. Second, the relevant fitness values for the invasion of the i,j-haplotype are $w((L+i-j)/2)$ in winter and $w((L-i+j)/2)$ in summer. The population mean fitness will be approximately $w(L/2)$ in both seasons, namely the fitness of the resident-resident genotype.

To linear order, the frequency of the i,j-haplotype after one yearly cycle with g generations of summer and g generations of winter will be

$$\epsilon'_{i,j} = \epsilon_{i,j} \cdot \left(\frac{w((L+i-j)/2)}{w(L/2)}\right)^g \cdot \left(\frac{w((L-i+j)/2)}{w(L/2)}\right)^g \cdot \left(2 \cdot \frac{1}{2^{i+j}}\right)^{2g}, \quad [S1]$$

where the factor of two in the last term comes from the fact that each genotype contributes on average two haplotypes in the next generation. From the assumption that intermediate types are favored in the long run (negative epistasis), we can conclude that

$$w((L+i-j)/2) \cdot w((L-i+j)/2) \leq w(L/2)^2, \quad [S2]$$

with equality for $i = j$. Hence $\epsilon'_{i,j} < \epsilon_{i,j}$ for $i \neq j$, showing that unbalanced haplotypes cannot invade the population.

For a balanced haplotype ($i = j$),

$$\epsilon'_{i,j} = \epsilon_{i,j} \cdot 1 \cdot \left(2 \cdot \frac{1}{2^{2i}}\right)^{2g} < \epsilon_{i,j} \text{ for } i \geq 1. \quad [S3]$$

Intuitively, other balanced haplotypes cannot invade, because they differ from the resident balanced haplotype at more than one position and are therefore broken down by recombination. Hence we have shown that a population fixed for a certain balanced haplotype cannot be invaded by any other balanced or unbalanced haplotype.

Appendix S2. Detailed methods

Individual-based simulations. We assume discrete, non-overlapping generations with population size N and L loci. In each generation, the following events take place in this order: 1) The fitnesses of all individuals in the parent generation are calculated using the genotype-to-fitness map described in the main text (Eq. (1), Eq. (2), Eq. (4)). 2) Each individual in the offspring generation draws two parents, independently, with replacement (i.e. selfing is possible), and proportionally to parent fitness. 3) Each parent passes one set of alleles to each of its offspring. We assume unlinked loci such that at each of the L loci each of the two allele copies is passed on with

[1]To whom correspondence should be addressed. E-mail: meike.wittmann@uni-bielefeld.de or dpetrov@stanford.edu

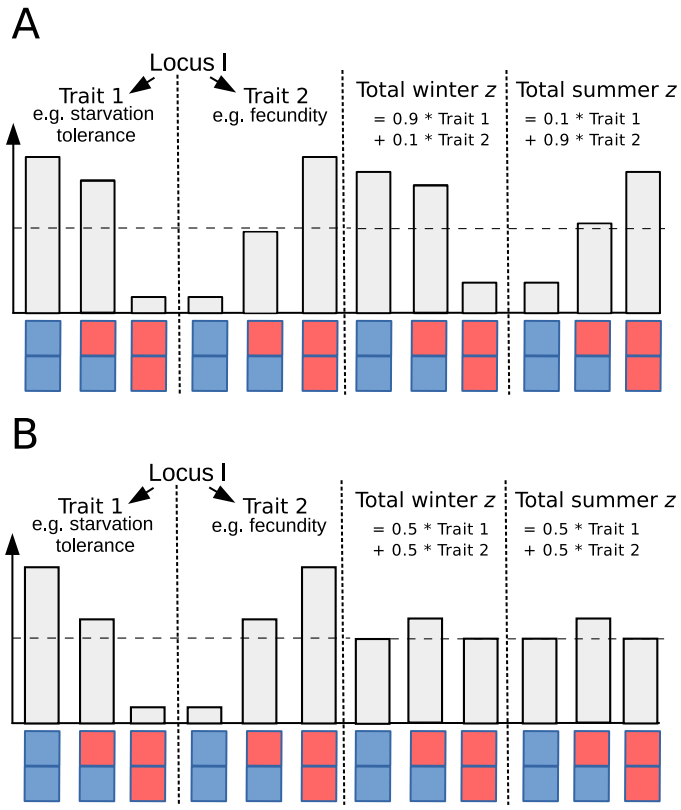

Fig. S2. Alternatives to the scenario in Fig. 3. A) If heterozygotes are relatively close to the fitter homozygote with respect to one of the traits and the beneficial allele is only slightly recessive for the other trait, we can still obtain a beneficial reversal of dominance with respect to the seasonal score, z. However, in this case, $d_s \neq d_w$. B) If the two traits are of similar importance in the two seasons, heterozygotes are fitter than either homozygote at all times.

equal probability, independently of the alleles passed on at other loci, and also independently of which alleles were passed on to other offspring of the same parent. 4) Independently and with probability μ, each of the allele copies passed to an offspring mutates to the respective other allele. 5) The parent generation is removed from the model and replaced by the individuals in the offspring generation.

To initialize a simulation run, we randomly assemble genotypes. Independently for each individual, locus, and allele copy, we draw the summer and the winter allele with equal probabilities. Each simulation runs for 500 years, corresponding to $500 \cdot 2 \cdot g$ generations. Each parameter combination is replicated 10 times. For each locus, we store the allele-frequency trajectory over time.

Stability of polymorphism and magnitude of fluctuations. To quantify and compare the dynamics of different simulation runs, we compute three summary statistics from the allele-frequency trajectories: an effective strength of balancing selection, b_e, as a measure of the stability of polymorphism, an effective selection coefficient, s_e, as a measure of the magnitude of fluctuations, and the proportion of allele frequency changes that go in the expected direction as a measure of predictability of fluctuations.

To compute the strength of balancing selection, we divide the allele-frequency interval between 0 and 1 into 25 equally-sized bins (from 0 to 0.04, from 0.04 to 0.08, ...). For each frequency bin, we compute the average allele-frequency change over the course of one year, from the middle of the cold season to the middle of the next cold season. We chose the middle of the cold season as a reference point because at this point the average frequency across all loci should be approximately 0.5. Choosing the middle of the warm season as the reference point should give the same results, but choosing either the beginning or end of the warm season would lead to asymmetric results. We average over all loci and times, for which the frequency at the reference point is in the respective frequency bin. Because we want to quantify dynamics at equilibrium, we only use data from the second half of each simulation run. We also exclude data points for which the allele frequency at the beginning of the year is exactly 0 or 1. For each bin, we then subtract an approximate average frequency change due to mutations $2 \cdot g \cdot (\mu \cdot (1-\bar{p}) - \mu \cdot \bar{p}) = 2 \cdot g \cdot \mu \cdot (1-2\bar{p})$, where $\bar{p}$ is the midpoint of the respective frequency bin (0.02, 0.06,...). Finally, we use the lm function in R [1] to fit a balancing-selection model of the form $b_e \cdot \bar{p}(1-\bar{p})(1-2\bar{p})$ to the mutation-adjusted average allele-frequency changes. The coefficient b_e is our effective strength of balancing selection. Example model fits are shown in Fig. S3 A.

To compute the effective selection coefficient, we use the same data subsets and frequency bins as for the effective strength of balancing selection. But now we compute the average frequency change of the currently favored allele for each season, i.e. of the summer allele from spring to fall or of the winter allele from fall to spring. We bin data points according to the mid-season frequency of the currently favored allele. After subtracting $g \cdot \mu \cdot (1-2\bar{p})$, the approximate expected frequency change due to mutations, we fit the model $s_e \cdot \bar{p}(1-\bar{p})$.

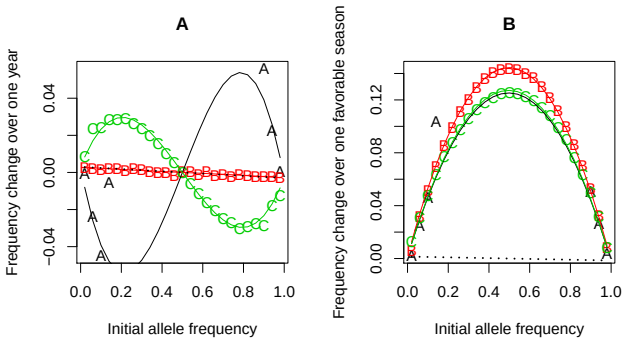

Fig. S3. Examples of model fits to estimate (A) the effective strength of balancing selection (Eq. (14)) and (B) the effective selection coefficient (Eq. (15)). The letters A, B, and C in the plots indicate the panel in Fig. 7 that depicts the corresponding time series.

The coefficient s_e is the effective selection coefficient. Example model fits are shown in Fig. S3 B.

Retardation factor. To obtain the retardation factor, we ran additional simulations without recurrent mutations ($\mu = 0$). We started at allele frequency 0.5 for all loci and simulated $n_{\text{rep}} = 100$ replicate populations. We stopped the simulation as soon as polymorphism was lost at any of the L loci, but at most after 500 years, corresponding to $t_{\max} = 500 \cdot 2 \cdot g$ generations. For each parameter combination, we ran a neutral control simulation, which was achieved by setting $y = 0$.

The relevant results for each parameter combination are the number of replicates, n_{lost}, in which the first polymorphism was lost before $t_{\max}$ and the times of loss for each of these replicates $t_1, t_2, ..., t_{n_{\text{lost}}}$.

For simplicity, we assume that polymorphism is lost with the same probability p in every generation such that the time to the first loss is geometrically distributed. We then adopt a Maximum-Likelihood approach to estimate p. The likelihood of p is

$$L(p) = (1-p)^{t_{\max}\cdot(n_{\text{rep}}-n_{\text{lost}})} p^{n_{\text{lost}}} \prod_{i=1}^{n_{\text{lost}}} (1-p)^{t_i - 1} \quad [S4]$$

$$= p^{n_{\text{lost}}} (1-p)^{a} \quad [S5]$$

with

$$a := t_{\max}(n_{\text{rep}} - n_{\text{lost}}) - n_{\text{lost}} + \sum_{i=1}^{n_{\text{lost}}} t_i. \quad [S6]$$

To look for extreme values, we take the first derivative with respect to p

$$L'(p) = p^{n_{\text{lost}}-1}(1-p)^{a-1}\left(n_{\text{lost}} - p(n_{\text{lost}} + a)\right) \quad [S7]$$

and obtain our estimator

$$\hat{p} = \frac{n_{\text{lost}}}{n_{\text{lost}} + a} = \frac{n_{\text{lost}}}{\sum_{i=1}^{n_{\text{lost}}} t_i + t_{\max}(n_{\text{rep}} - n_{\text{lost}})}. \quad [S8]$$

Eq. (S7) is positive for $p < \hat{p}$ and negative for $p > \hat{p}$. Therefore, $\hat{p}$ is a maximum of the likelihood function. To see that this estimator makes sense, consider an example where loss is so fast that all replicates lose a polymorphism before $t_{\max}$. In this case, $\hat{p}$ will be the inverse of the average time to loss.

For each parameter combination, we obtain the ML-estimator $\hat{p}_{\text{sel}}$ under seasonally fluctuating selection and the ML-estimator for the corresponding neutral scenario $\hat{p}_{\text{neutral}}$. Finally, we compute the retardation factor as $\hat{p}_{\text{neutral}}/\hat{p}_{\text{sel}}$. Values larger than one occur for selection scenarios that tend to maintain polymorphism longer than neutrality, whereas in scenarios with a retardation factor below one polymorphism is lost more quickly than under neutrality.

Stability in asymmetric scenarios. In our model with parameter variation across loci and seasons, and also in the basic model with different numbers of generations in summer and winter, the dynamics are generally asymmetric. That is, either the summer or the winter allele is more common on average. Hence the standard balancing selection model Eq. (14) does not fit anymore. We therefore use a different method to assess stability of polymorphism for individual loci. For each frequency bin between 0 and 0.5, we take all time points for which the frequency of the currently rare allele is in the frequency bin at the beginning of the cycle and compute the average change in its allele frequency over one cycle. As above, we subtract the expected input from new mutations. We do this separately for 10 replicate simulation runs (or for 10 disjoint subsets of the data if the actual replicates differ in the locus parameters) and for each frequency bin we compute the interval mean $\pm$ 2 standard errors. We then call a polymorphism stable if the rare allele has a positive expected frequency change in the most marginal frequency bin whose interval does not overlap zero.

Appendix S3. Magnitude and detectability of seasonal fluctuations with diminishing-returns epistasis

In this section, we use a heuristic mathematical argument to explore how the magnitude of allele-frequency fluctuations changes as the number of loci increases in a population of infinite size.

Let us assume that there are $L+1$ loci in total, with one focal locus whose dynamics we will now study while the remaining L loci form its "genetic background". Let us first consider the contribution, z_b, of the genetic background to the seasonal score, z, i.e. the total score minus the contribution from the focal locus. For simplicity, we will focus on the expected allele frequency change over one generation at one of the two time points at which the allele frequency equals the yearly average allele frequency. At this point, both the population mean and the variance of the score z_b among individuals in the population should be approximately proportional to the number of loci:

$$\bar{z_b} = \sum_{l=1}^{L} \bar{c_l} \approx k_1 L \quad [S9]$$

and

$$\mathbf{Var}(z_b) = \sum_{l=1}^{L} \mathbf{Var}(c_l) \approx k_2 L, \quad [S10]$$

where we have assumed free recombination. Here, c_l is the contribution of locus l to an individual's score, and k_1 and k_2 are positive constants that will depend on the distribution of locus parameters. For example, if all loci have symmetric parameters ($\Delta_{s,l} = \Delta_{w,l} =: \Delta_l$ and $d_{s,l} = d_{w,l} =: d_l$), all loci will have an allele frequency of approximately 0.5 in the middle of summer. Since mating is random, allele frequencies

will be in approximate Hardy-Weinberg equilibrium in every generation before selection and, with $\sim$ denoting averages across loci,

$$\bar{z_b} = \sum_{l=1}^{L} \left(\frac{1}{4} + \frac{1}{2}d_l\right)\Delta_l = \underbrace{\left(\frac{1}{4}\widetilde{\Delta} + \frac{1}{2}\widetilde{d\Delta}\right)}_{k_1} L. \quad \text{[S11]}$$

Similarly,

$$\mathbf{Var}(z_b) = \sum_{l=1}^{L} \mathbf{E}[c_l^2] - \mathbf{E}[c_l]^2 \quad \text{[S12]}$$

$$= \sum_{l=1}^{L} \left(\frac{1}{4} + \frac{1}{2}d_l^2\right)\Delta_l^2 - \left(\frac{1}{4} + \frac{1}{2}d_l\right)^2 \Delta_l^2 \quad \text{[S13]}$$

$$= \underbrace{\left(\frac{3}{16}\widetilde{\Delta^2} + \frac{1}{4}\left(\widetilde{\Delta^2 d^2} - \widetilde{\Delta^2 d}\right)\right)}_{k_2} L. \quad \text{[S14]}$$

Next, we quantify the fitnesses of the three genotypes (WW, WS, SS) at a focal locus averaged over the possible genetic backgrounds. For this, we expand the fitness function $w(z)$ as a second-order Taylor expansion around $\bar{z_b}$:

$$w(z) \approx w(\bar{z_b}) + w'(\bar{z_b})(z - \bar{z_b}) + \frac{1}{2}w''(\bar{z_b})(z - \bar{z_b})^2. \quad \text{[S15]}$$

Now, we need the values of the phenotype z for the three genotypes at the focal locus. Let us assume that the focal locus has parameters d and Δ. Because we consider the population in the middle of summer, the winter-winter homozygote (WW) at the focal locus does not contribute anything to the seasonal score and $z = z_b$. For the heterozygote (WS), $z = z_b + d \cdot \Delta$, and for the summer-summer homozygote, $z = z_b + \Delta$. With this, we obtain

$$\overline{w_{WW}} \approx w(\bar{z_b}) + \frac{1}{2}w''(\bar{z_b})\mathbf{Var}(z_b), \quad \text{[S16]}$$

$$\overline{w_{WS}} \approx w(\bar{z_b}) + w'(\bar{z_b})d \cdot \Delta + \frac{1}{2}w''(\bar{z_b})\mathbf{Var}(z_b), \quad \text{[S17]}$$

and

$$\overline{w_{SS}} \approx w(\bar{z_b}) + w'(\bar{z_b})\Delta + \frac{1}{2}w''(\bar{z_b})\mathbf{Var}(z_b). \quad \text{[S18]}$$

Finally, we compute the expected change in summer allele frequency, p, over one generation at the focal locus. The frequency of the summer allele in the next generation is

$$p' \approx \frac{p^2\overline{w_{SS}} + p(1-p)\overline{w_{WS}}}{p^2\overline{w_{SS}} + 2p(1-p)\overline{w_{WS}} + (1-p)^2\overline{w_{WW}}} \quad \text{[S19]}$$

and the proportional change in allele frequency is

$$\frac{p'-p}{p} \approx \frac{p(1-p)\overline{w_{SS}} + (1-p)(1-2p)\overline{w_{WS}} - (1-p)^2\overline{w_{WW}}}{p^2\overline{w_{SS}} + 2p(1-p)\overline{w_{WS}} + (1-p)^2\overline{w_{WW}}}. \quad \text{[S20]}$$

Substituting Eq. (S16)–Eq. (S18) and simplifying, we obtain

$$\frac{p'-p}{p} \approx \frac{(1-p)w'(\bar{z_b})\Delta(p + d(1-2p))}{w(\bar{z_b}) + \frac{1}{2}w''(\bar{z_b})\mathbf{Var}(z_b) + \Delta p w'(\bar{z_b})(p + 2(1-p)d)}. \quad \text{[S21]}$$

For the specific fitness function fitness function used in this paper

$$w(z) = (1+z)^y, \quad \text{[S22]}$$

with

$$w'(z) = y(1+z)^{y-1}, \quad \text{[S23]}$$

and

$$w''(z) = y(y-1)(1+z)^{y-2}, \quad \text{[S24]}$$

and using Eq. (S9) and Eq. (S10), we obtain

$$\frac{p'-p}{p} \approx \frac{\Delta(1-p)(p + d(1-2p))y(1+k_1L)^{y-1}}{\begin{array}{c}(1+k_1L)^y + \frac{1}{2}y(y-1)(1+k_1L)^{y-2}k_2L \\ +\Delta p y(1+k_1L)^{y-1}(p+2(1-p)d)\end{array}} \quad \text{[S25]}$$

$$= \frac{\Delta(1-p)(p + d(1-2p))y}{1 + k_1L + \frac{1}{2}y(y-1)\frac{k_2L}{1+k_1L} + \Delta p y(p + 2(1-p)d)} \quad \text{[S26]}$$

$$=: \phi(L). \quad \text{[S27]}$$

It follows that,

$$\lim_{L\to\infty} \phi(L) = 0. \quad \text{[S28]}$$

Moreover,

$$\frac{d}{dL}\phi(L) < 0 \quad \text{[S29]}$$

$$\Leftrightarrow \frac{d}{dL}\left(1 + k_1L + \frac{1}{2}y(y-1)\frac{k_2L}{1+k_1L}\right) > 0 \quad \text{[S30]}$$

$$\Leftrightarrow k_1 + \frac{1}{2}y(y-1)\frac{k_2}{(1+k_1L)^2} > 0. \quad \text{[S31]}$$

With $y > 1$, the second term is positive and the inequality is always fulfilled. With $y < 1$ it is fulfilled for sufficiently large L. Thus, with $y > 1$, the magnitude of allele-frequency change in an infinite population decreases monotonically as L increases and approaches zero for very large numbers of loci. With $y < 1$, allele-frequency change may first increase with increasing number of loci, but eventually decreases and also approaches zero. This finding indicates that there is a limit to the number of loci that can be detected to exhibit seasonal allele-frequency fluctuations. Intuitively, as the number of loci increases the average fitness of the genetic background becomes larger. Thus selection at a focal locus gets effectively weaker.

Appendix S4. Extreme-value statistics

To obtain a better intuitive feeling for the distribution of fitness within a population and for the effect of offspring-number capping, we will use a heuristic argument similar to the one presented by Ewens [2, p. 83–84]. Specifically, we will calculate the expected value of the seasonal score of the fittest and the least fit individual in the population and then compare the fitness of individuals with these seasonal scores. As an approximation, we will assume the seasonal scores within the population to be normally distributed with mean $\bar{z}$ and variance σ^2. Using Eq. (S11) and Eq. (S14), with L as the total number of loci,

$$\bar{z} = \left(\frac{1}{4} + \frac{d}{2}\right) \cdot L \quad \text{[S32]}$$

and

$$\sigma = \sqrt{\left(\frac{3}{16} + \frac{1}{4} \cdot (d^2 - d)\right) \cdot L} \quad \text{[S33]}$$

This assumes that all loci have the same effect size and dominance parameter and that we are studying the population in the middle of a season.

We now assume that the seasonal scores of N individuals (where N is the population size) are drawn independently from this distribution. Let Z_i be the seasonal score of individual i with $1 \leq i \leq N$ and let Y be the largest of the Z_i. Then the cumulative distribution function of Y is

$$F(y) = \Pr(Y \leq y) = \prod_{i=1}^{N} \Pr(Z_i \leq y) = \left(\frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{y - \bar{z}}{\sqrt{2}\sigma} \right) \right] \right)^N, \quad [S34]$$

which uses the standard formula for the cumulative distribution function of a normal distribution. Using Mathematica [3], the probability density function of Y is

$$f(y) = \frac{dF}{dy} = \frac{\sqrt{\frac{2}{\pi}} 0.5^N N e^{-\frac{(y-\bar{z})^2}{2\sigma^2}} \left(\operatorname{erf}\left(\frac{y-\bar{z}}{\sqrt{2}\sigma} \right) + 1 \right)^{N-1}}{\sigma}. \quad [S35]$$

and the expected seasonal score of the fittest individual in the population is

$$\mathbf{E}[Y] = \int_{-\infty}^{\infty} y \cdot f(y) dy. \quad [S36]$$

Since there does not seem to be an analytical solution, we evaluated this integral numerically using Mathematica [3]. For example, for $d = 0.7$ and $N = 1000$, $\mathbf{E}[Y] \approx 72$. That is, the average seasonal score of the fittest individual in the population is about 72. The average seasonal score in this situation is 60 and, because of symmetry, the average seasonal score of the least fit individual in the population must be approximately 48. Using the fitness function Eq. (4), an individual with seasonal score 72 is 1.2 times fitter than an individual with seasonal score 48 if $y = 0.5$ and 4.9 times fitter for $y = 4$. These values are in excellent agreement with those obtained from the individual-based simulations (see main text). The moderate fitness differences obtained from this analysis confirm that segregation lift does not require the fittest individuals in the populations to have excessively many offspring.

Appendix S5. Additional results

See Figs. S4–S19.

1. R Core Team (2014) *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, Vienna, Austria).
2. Ewens WJ (2004) *Mathematical population genetics.* (Springer, New York), 2nd edition.
3. Wolfram Research Inc. (2015) Mathematica, Version 10.3. Champaign, IL.

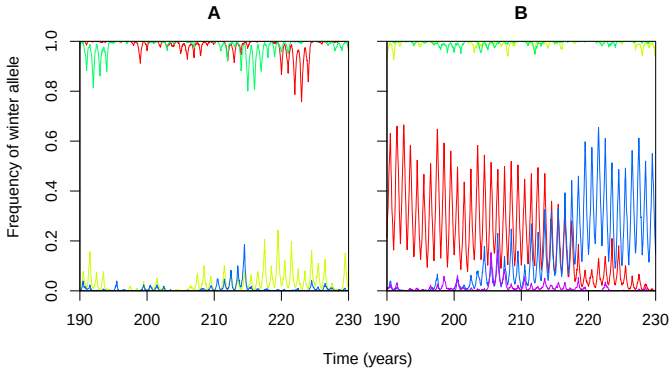

Fig. S4. Example time series for the additive model ($d = 0.5$) with (A) four loci, or (B) five loci. $N = 1000, g = 15, y = 1, \mu = 10^{-4}$.

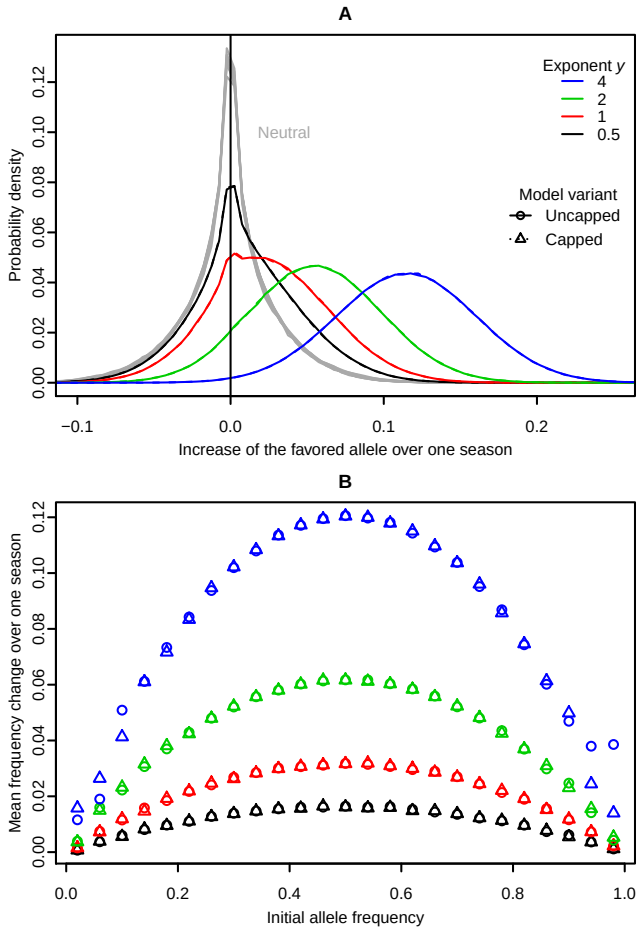

Fig. S5. Detailed information on allele frequency change over one season for the scenarios with $d = 0.7$ in Fig. 8. (A) The distribution of the change in frequency for all loci over the currently favored allele over one season compared to neutrality (gray curve), and (B) the average frequency change of the currently favored allele over one season for $d = 0.7$ as a function of the frequency at the beginning of the season. $N = 1000, L = 100, g = 15, \mu = 10^{-4}$. In (A) the results for the capped vs. uncapped model variants are virtually indistinguishable.

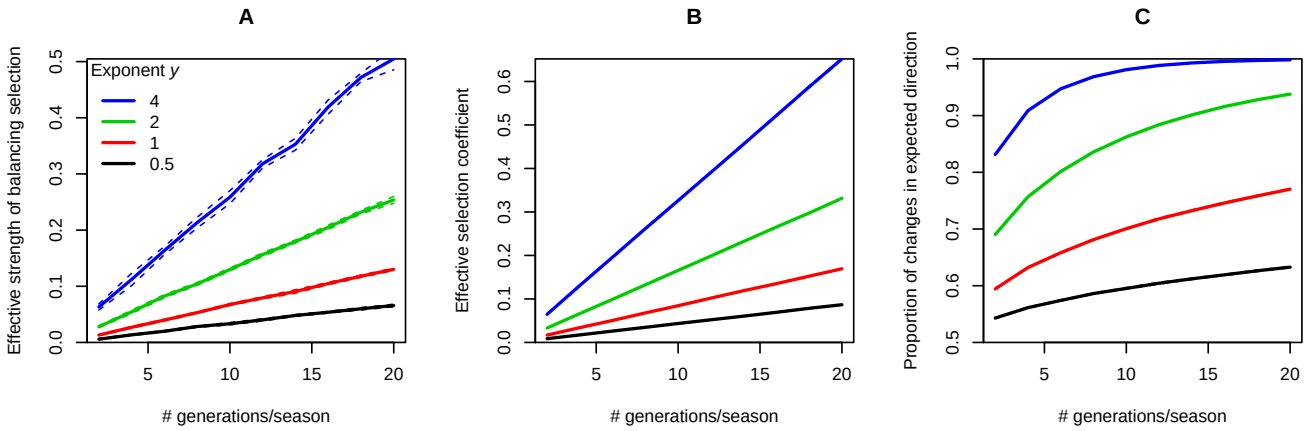

Fig. S6. Influence of the number of generations per season, g, on (A) stability of polymorphism, (B) magnitude and (C) predictability of fluctuations. Dashed lines indicate means $\pm$ two standard errors, but they coincide with the solid lines in B and C. $N = 1000, L = 100, \mu = 10^{-4}, d = 0.7$.

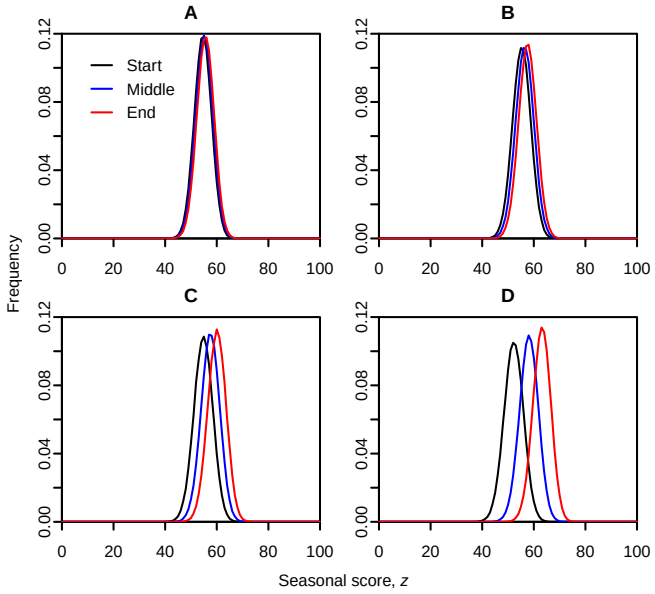

Fig. S7. Typical distributions of seasonal scores, z, within a generation at the start of the season (generation 1), in the middle of a season (generation 8), and at the end of the season (generation 15). The panels differ in the exponent of the fitness function, y: (A) $y = 0.5$, (B) $y = 1$, (C) $y = 2$, and (D) $y = 4$. $N = 1000, L = 100, g = 15, d = 0.7, \mu = 10^{-4}$. Offspring number is uncapped. Each distribution represents the average over ten replicates and 20 seasons per replicate.

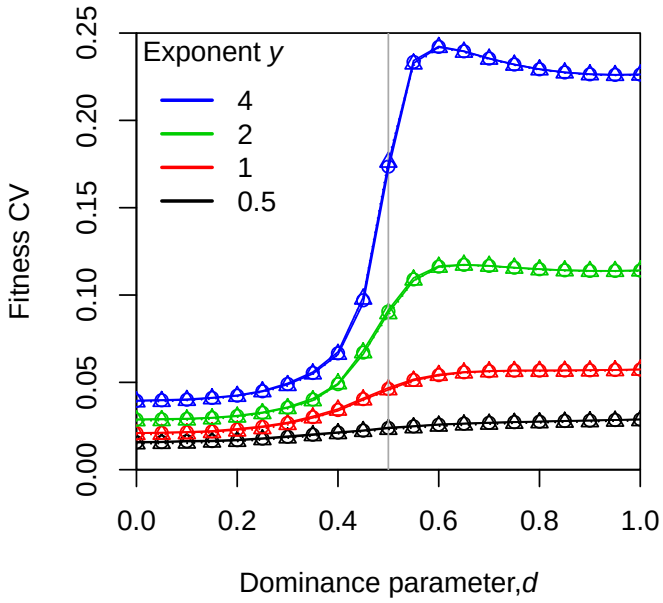

Fig. S8. Coefficient of variation in fitness among individuals in the population as a function of the dominance parameter d. $N = 1000, L = 100, g = 15, \mu = 10^{-4}$. Symbols indicate the mean across replicates for the uncapped vs. capped model variant (generally coinciding) and lines indicate means $\pm$ two standard errors for the uncapped case. Since the standard errors are so small, the lines are overlapping.

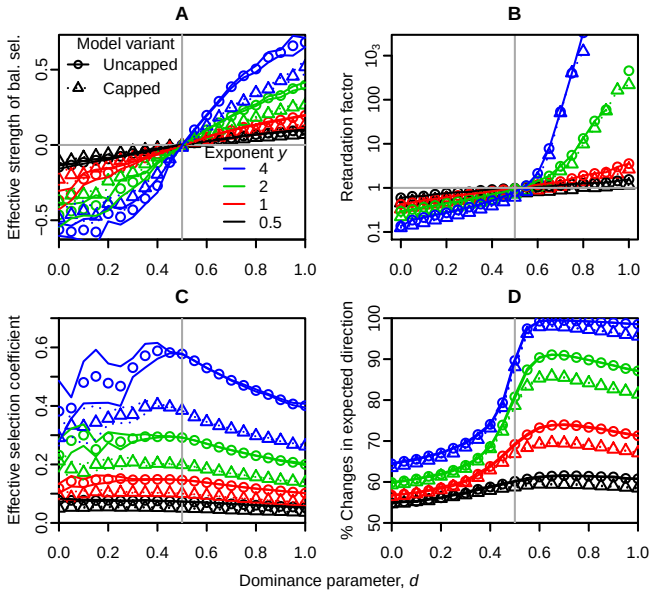

Fig. S9. Effect of capping offspring number at three, the smallest possible value. Influence of the dominance parameter d on (A) effective strength of balancing selection (b_e, Eq. (14), Methods), (B) retardation factor, (C) magnitude of fluctuations (s_e, Eq. (15), Methods), (D) predictability of fluctuations. Symbols indicate averages across replicates for the uncapped vs. capped model variant and solid vs. dashed lines in A, C, and D indicate the respective means $\pm$ two standard errors. Lines in B simply connect maximum-likelihood estimates obtained jointly from all replicates. $N = 1000, L = 100, g = 15, \mu = 10^{-4}$. The vertical grey lines are at $d = 0.5$.

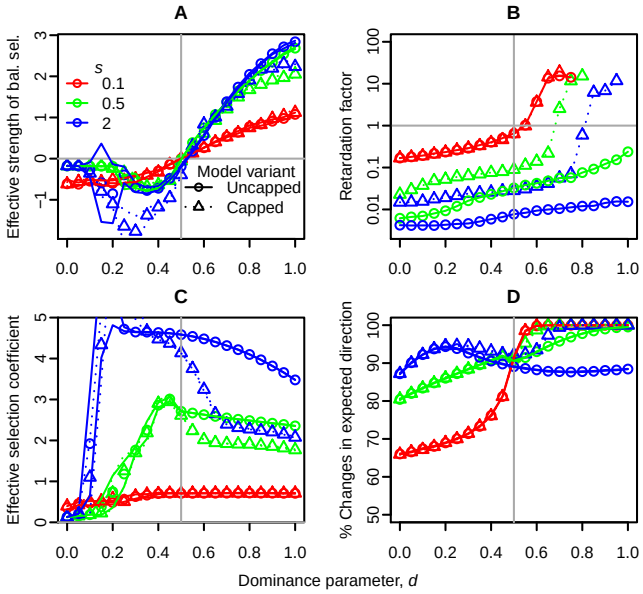

Fig. S10. Influence of the dominance parameter d under a multiplicative model ($w(z) = \exp(\ln(1+s)\cdot z) = (1+s)^z$) on (A) effective strength of balancing selection (b_e, Eq. (14), Methods), (B) retardation factor, (C) magnitude of fluctuations (s_e, Eq. (15), Methods), (D) predictability of fluctuations. Symbols indicate averages across replicates for the uncapped vs. capped model variant and solid vs. dotted lines in A, C, and D indicate the respective means $\pm$ two standard errors. Lines in B simply connect maximum-likelihood estimates obtained jointly from all replicates. Note that in B some points are missing because the rate of loss of polymorphism was too small to be quantified. $N = 1000, L = 100, g = 15, \mu = 10^{-4}$.

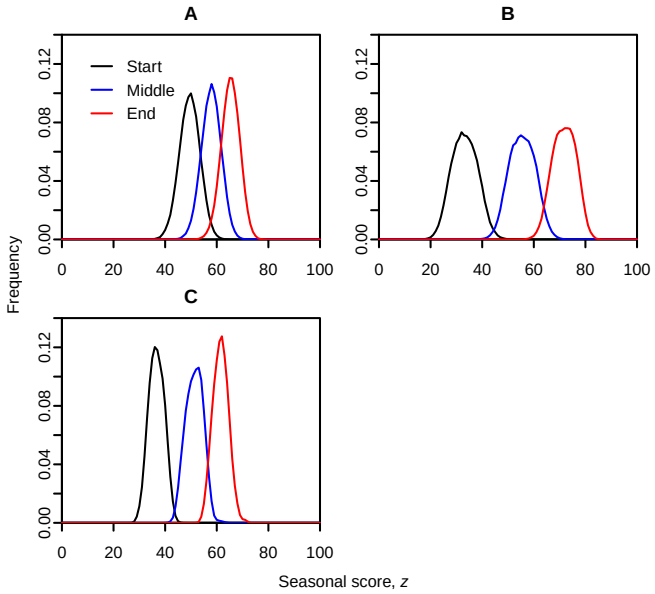

Fig. S11. Typical distributions of seasonal scores, z, under a multiplicative model ($w(z) = \exp(\ln(1+s)\cdot z) = (1+s)^z$). Shown are distributions within a generation at the start of the season (generation 1), in the middle of a season (generation 8), and at the end of the season (generation 15). The panels differ in the selection coefficient, s: (A) $s = 0.1$, (B) $s = 0.5$, (C) $s = 2$. Offspring number is uncapped. $N = 1000, L = 100, g = 15, d = 0.7, \mu = 10^{-4}$. Each distribution represents the average over ten replicates and 20 seasons per replicate. The fittest individual was on average 9.3 times fitter than the least fit individual for $s = 0.1$, $1.7 \cdot 10^3$ times fitter for $s = 0.5$, and $4.2 \cdot 10^3$ times fitter for $s = 2$.

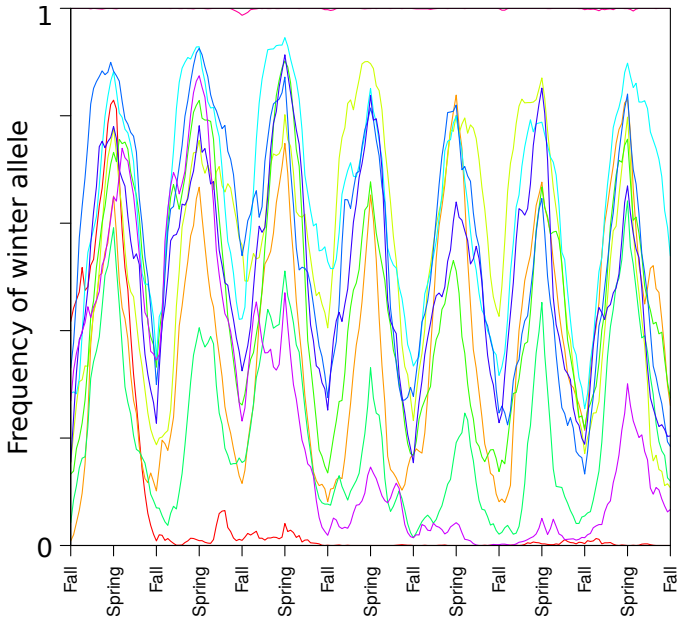

Fig. S12. Example time series for the uncapped multiplicative model ($w(z) = \exp(\ln(1+s) \cdot z) = (1+s)^z$) with $d = 0.74, s = 0.5$.

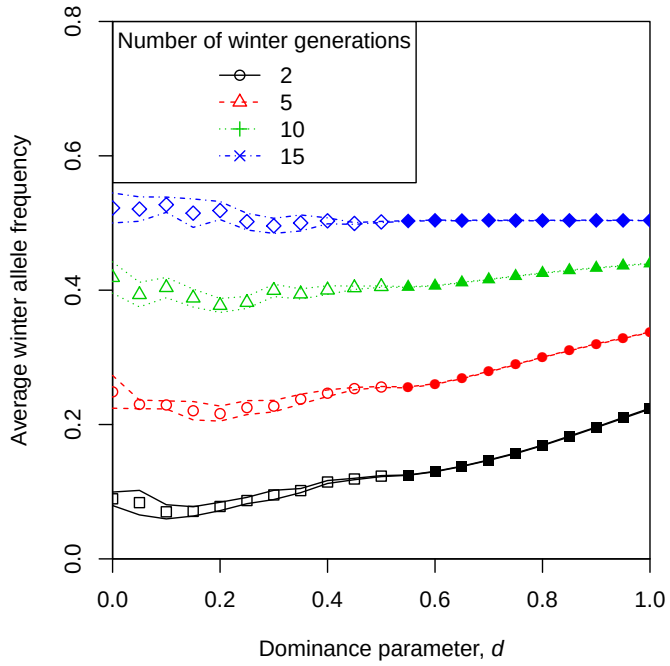

Fig. S13. Effects of asymmetry in the number of generations between summer and winter. There are 15 generations of summer. The y-axis shows the winter allele frequency in the middle of winter, averaged across time and loci. Closed symbols represent parameter combinations that were classified as having stable polymorphism (see section Stability in asymmetric scenarios), open symbols represent unstable polymorphism. Lines indicate means $\pm$ two standard errors. Other parameters: $N = 1000, L = 100, y = 4, \mu = 10^{-4}$.

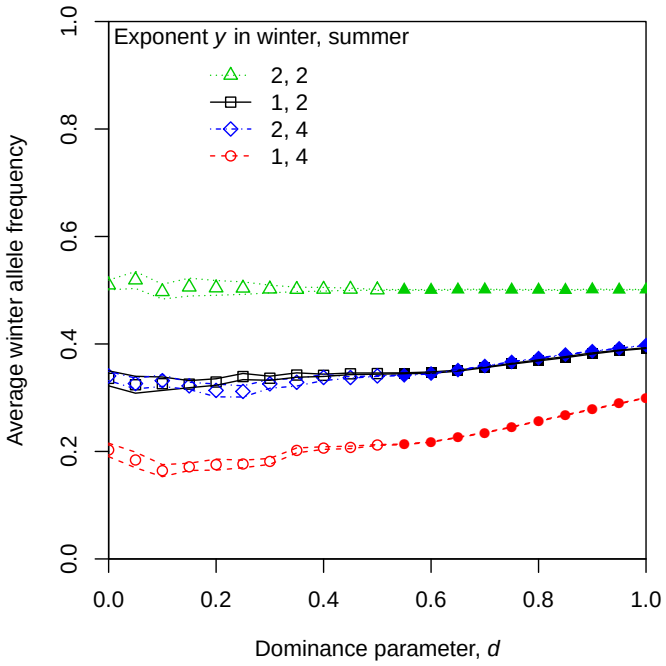

Fig. S14. Effects of asymmetry in the exponent of the fitness function, y, between summer and winter. The y-axis shows the winter allele frequency in the middle of winter, averaged across time and loci. Solid symbols represent parameter combinations that were classified as having stable polymorphism (see section Stability in asymmetric scenarios), open symbols represent unstable polymorphism. Dotted lines indicate means $\pm$ two standard errors. Other parameters: $N = 1000, L = 100, g = 15, \mu = 10^{-4}$.

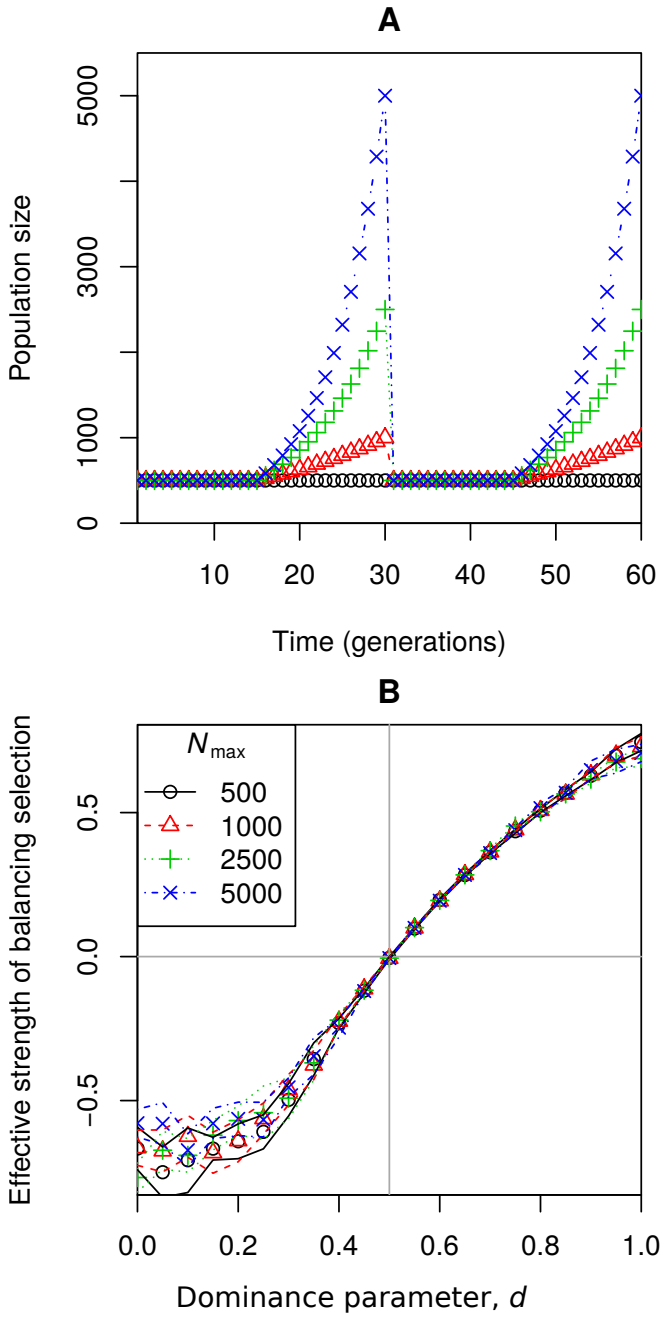

Fig. S15. Robustness of results to seasonal changes in population size. (A) Four population-size trajectories. The trajectories are deterministic and cyclic and only two years are shown. Population size is 500 during winter and grows exponentially to different final sizes, N_{max}, over summer (except for the case $N_{\text{max}} = 500$, where population size remains constant throughout the year). (B) Corresponding effective strength of balancing selection. $L = 100, g = 15, y = 4, \mu = 10^{-4}$.

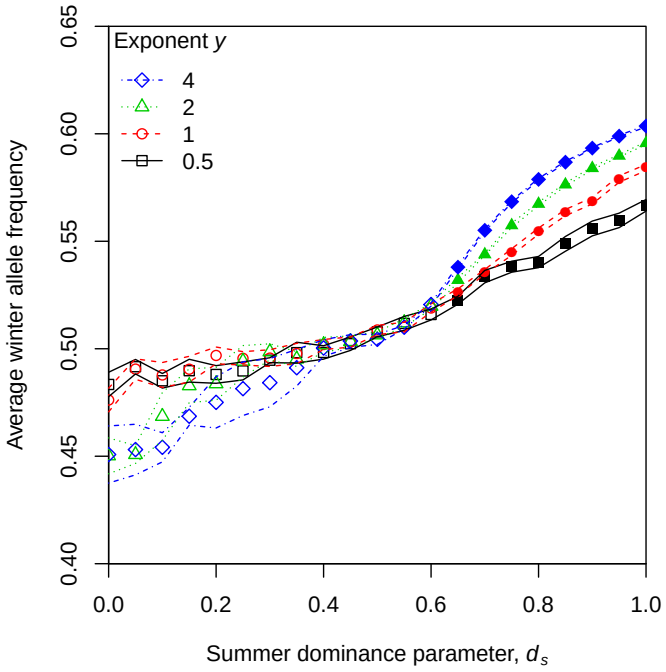

Fig. S16. Effects of asymmetry in the dominance parameter, d, between summer and winter. The winter dominance parameter, d_w, is set to 0.4 whereas the summer dominance parameter, d_s, is varied. The y-axis shows the winter allele frequency in the middle of winter, averaged across time and loci. Solid symbols represent parameter combinations that were classified as having stable polymorphism (see section Stability in asymmetric scenarios), open symbols represent unstable polymorphism. Dotted lines indicate means $\pm$ two standard errors. Other parameters: $N = 1000$, $L = 100$, $\mu = 10^{-4}$.

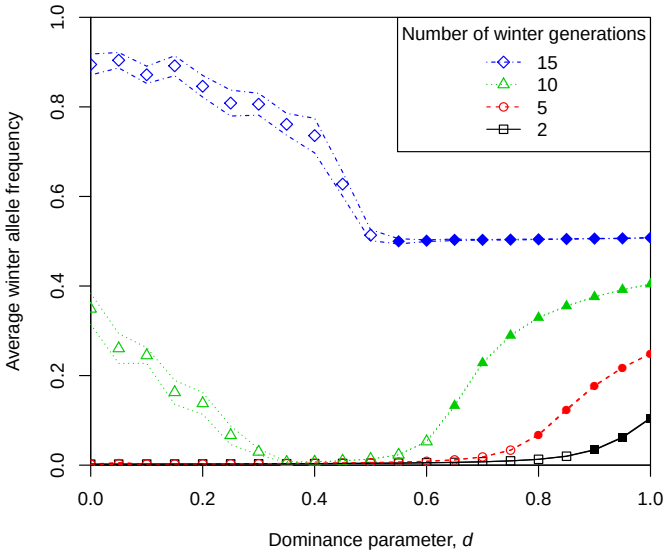

Fig. S17. Effects of asymmetry in the number of generations between summer and winter under a multiplicative model ($w(z) = \exp(\ln(1+s) \cdot z) = (1+s)^z$). There are 15 generations of summer. The y-axis shows the winter allele frequency in the middle of winter, averaged across time and loci. Closed symbols represent parameter combinations that were classified as having stable polymorphism (see section Stability in asymmetric scenarios), open symbols represent unstable polymorphism. Lines indicate means $\pm$ two standard errors. Other parameters: $N = 1000$, $L = 100$, $\mu = 10^{-4}$.

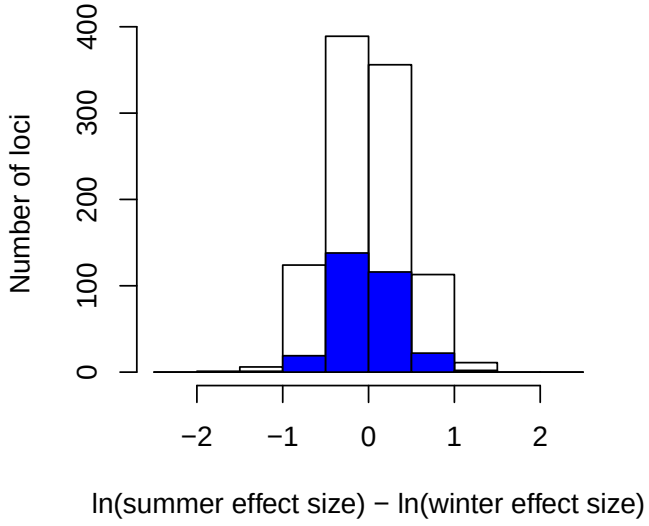

Fig. S18. Histograms of the difference in the logarithms of seasonal effect sizes for stable (blue bars) compared to all polymorphisms (white bars) in Fig. 10 E, i.e. for $N = 10,000, L = 100, g = 10, y = 4, \mu = 10^{-4}$.

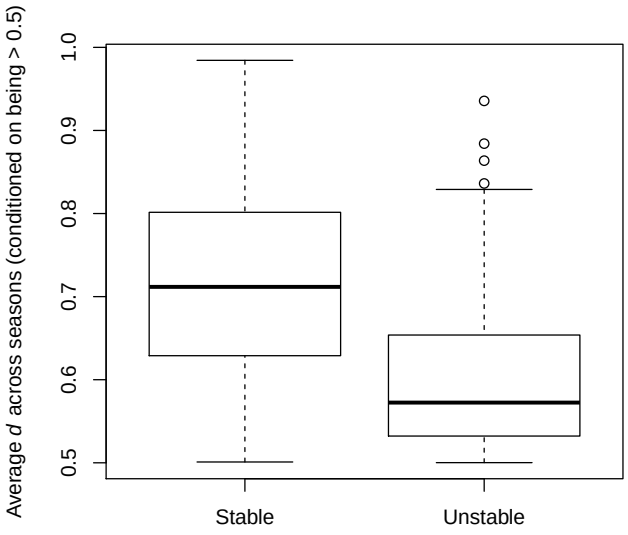

Fig. S19. Boxplots of $(d_s + d_w)/2$ for stable and unstable polymorphisms in Fig. 10 E, conditioned on $(d_s + d_w)/2 > 0.5$. $N = 10,000, L = 100, g = 10, y = 4, \mu = 10^{-4}$.