Heritable epigenetic variation facilitates long-term maintenance of epigenetic and genetic variation

Amy K. Webster[1*] and Patrick C. Phillips[1]

1. Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403

*Corresponding author: awebst10@uoregon.edu

ABSTRACT

The maintenance of genetic and phenotypic variation has long been one of the fundamental questions in population and quantitative genetics. A variety of factors have been implicated to explain the maintenance of genetic variation in some contexts (*e.g.* balancing selection), but the potential role of epigenetic regulation to influence population dynamics has been understudied. It is well recognized that epigenetic regulation, including histone methylation, small RNA expression, and DNA methylation, helps to define differences between cell types and facilitate phenotypic plasticity. In recent years, empirical studies have shown the potential for epigenetic regulation to also be heritable for at least a few generations without selection, raising the possibility that differences in epigenetic regulation can act alongside genetic variation to shape evolutionary trajectories. Like genetic mutation, heritable differences in epigenetic regulation can arise spontaneously; these are termed 'epimutations'. Epimutations differ from genetic mutations in two key ways – they occur at a higher rate and the loci at which they occur often revert back to their original state within a few generations. Here, we present an extension of the standard population-genetic model with selection to incorporate epigenetic variation arising via epimutation. Our model assumes a diploid, sexually reproducing population with random mating. In addition to spontaneous genetic mutation, we included parameters for spontaneous epimutation and back-epimutation, allowing for four potential epialleles at a single locus (two genetic alleles, each with two epigenetic states), each of which affect fitness. We then analyzed the conditions under which stable epialleles were maintained. Our results show that highly reversible epialleles can be maintained in long-term equilibrium under neutral conditions in a manner that depends on the epimutation and back-epimutation rates, which we term epimutation-back-epimutation equilibrium. On the other hand, epialleles that compensate for deleterious mutations cause deviations from the expectations of mutation-selection balance by a simple factor that depends on the epimutation and back-epimutation rate. We also numerically analyze several sets of fitness parameters for which large deviations from mutation-selection balance occur. Together, these results demonstrate that transient epigenetic regulation may be an important factor in the maintenance of both epigenetic and genetic variation in populations.

INTRODUCTION

Epigenetic regulation is integral to the genotype-phenotype map. Indeed, the concept of the 'epigenotype' was originally defined to include anything 'between genotype and phenotype' (Waddington C 1942). However, modern usage is more limited to include either 1) specific heritable (mitotically or meiotically) modifications to DNA or RNA and/or 2) heritable changes in phenotype that are not dependent on an underlying nucleotide sequence. Studies employing the first usage commonly profile a population of cells or organisms for a specific epigenetic modification (Mehrmohamadi et al. 2021), under the assumption that epigenetic marks represent mitotically-heritable states and may impact gene expression. In contrast, the latter, more classical, definition does not require a specific type of modification but does require that a phenotypic change is inherited independently of genetic variation, with epigenetic modifications being likely candidates. This mode of transmission been used to denote transgenerational epigenetic inheritance in which an environmentally-induced phenotype is inherited across multiple generations without the initial trigger, even when the epigenetic mechanism is not known (Jobson et al. 2015; Webster et al. 2018). Typically, these examples of transgenerational epigenetic inheritance require that a phenotypic change is inherited three or more generations to ensure that inheritance is not driven by a maternal effect (*e.g.,* yolk provisioning) (Burton and Greer 2022). Both usages of epigenetics maintain a core tenet that something distinct from the DNA sequence is inherited. Here I will use the definition of epigenetics in which specific modifications impacting gene expression, and subsequently phenotype, occur and are inherited.

Epigenetic modifications often change the way that DNA is packaged, altering its accessibility to factors that promote or inhibit transcription of RNA. Modifiers that are traditionally considered 'epigenetic' include DNA methylation (particularly cytosine methylation), RNA methylation, histone modifications (including methylation, acetylation, phosphorylation, ubiquitination, etc.), and nucleosome positioning, along with larger-scale changes to genome architecture (Handy et al. 2011; Mishra and Hawkins 2017). In addition, certain RNAs are themselves epigenetic, as they can affect their own transcription or transcription of other genes, and they can also impact post-transcriptional regulation of mRNA (*e.g.,* small RNAs in RNA interference mechanisms) (Wei et al. 2017; Seroussi et al. 2022). Collectively, these epigenetic modifications each contribute to alterations in gene expression at the transcriptional or post-transcriptional level. However, notably, not all processes altering transcription or post-transcriptional regulation are epigenetic modifications (*e.g.,* transcription factor binding).

Epigenetic modifications are inherited, but this can be either mitotically (across cell divisions in the soma) or meiotically (across generations via the germline). Mitotically-heritable epigenetic modifications are one mechanism by which unique cell types within the same individual are maintained, despite containing the same underlying DNA sequence (Almouzni and Cedar 2016). In addition, mitotically-heritable epigenetic states are common in cancer cells and play a role in cancer progression (Jones and Baylin 2002; Feinberg 2004; Baxter et al. 2014). Further, epigenetic modifications play a major role in phenotypic plasticity in response to particular environmental conditions (Kelly et al. 2012), and epigenetic mechanisms can maintain a plastic response in adulthood well after the initial trigger (Werner et al. 2023). In general, however, the epigenetic state of an organism's cells, even if it is heritable mitotically and maintained throughout the lifespan, is not faithfully inherited meiotically through the germline, though there are some differences depending on the species. At one extreme, plant species including *Arabidopsis* have been observed to inherit cytosine methylation faithfully across meiosis for many generations, and differential methylation can be used to map quantitative traits to an 'epigenetic quantitative trait locus', which differs only in epigenetic modifications but not in underlying DNA sequence (Johannes et al. 2009; Cortijo et al. 2014; Kooke et al. 2015; Furci et al. 2019a; Liegard et al. 2019). Similarly, yeast have been shown to exhibit meiotically stable epigenetic states over many generations (Grewal and Klar 1996). For mammals, however, the notion of inheriting epigenetic modifications across generations has been more contentious, due to relatively few *bona fide* examples and the difficulty of studying mammals compared to other lab systems (Daxinger and Whitelaw 2012; Horsthemke 2018). In contrast, there are multiple examples of epigenetic inheritance in animal model systems including *Drosophila melanogaster* and *Caenorhabditis elegans* (Ruden and Lu 2008; Kelly 2014; Camacho et al. 2018; Moore et al. 2019; Baugh and Day 2020; Casas and Vavouri 2020; Evans et al. 2023). Notably, because *D. melanogaster* and *C. elegans* lack DNA cytosine methylation, mechanisms typically involve histone modifications and small RNAs.

The persistence of epigenetic modifications across generations that influence phenotype raises the possibility that epigenetic modifications can act like genetic variation, at least to some extent, in shaping the evolution of a population over time. Arguably, it is most interesting to investigate the role of epigenetic modifications in evolutionary trajectories when they act to some extent independently of genetic variation, acting as 'drivers' of phenotypic change rather than 'effectors' solely downstream of genetic variation (Sarkies

2020). That is, how do epigenetic modifications impact evolution when they arise spontaneously *a la* genetic mutations, and, while physically linked to genetic variants, impact fitness independently? This is distinct from asking how epigenetic modifications that arise because of environmental perturbations via plasticity (either within-generation or multigenerational plasticity) impact evolution. This is because plasticity can function like any other trait for which the genotype-phenotype map is of interest. Even in cases of multigenerational plasticity, it has been shown theoretically that heritable epigenetic responses to the environment can evolve given that the ancestral environment is predictive of the descendant environment (Herman et al. 2014). Epigenetic variants thar arise due to the environment (acting as 'effectors') can also impact evolutionary trajectories (subsequently acting as 'drivers'), which has also been theoretically analyzed in several models (Geoghegan and Spencer 2012; Geoghegan and Spencer 2013).

Epimutation rates, used here to mean spontaneous epigenetic modifications (Le Goff et al. 2021), in contrast to plasticity-driven epigenetic marks, have been measured in a variety of species, and the consensus has consistently been that epimutations occur at higher rates than genetic mutations (Horsthemke 2006; Becker et al. 2011; Schmitz et al. 2011; Gravina et al. 2015; van der Graaf et al. 2015; Carja et al. 2017). Studies that have estimated epimutation rates tend to focus on CpG methylation, rather than other forms of epigenetic modifications. However, unlike germline genetic mutations that are faithfully inherited across generations, studies evaluating the inheritance patterns of epimutations without selection have been more limited, particularly in animals. Recently, experimental evolution was performed in *C. elegans* to analyze epimutation rates and the persistence of epimutations (Beltran et al. 2020). Results suggest that epimutations occur at higher rates than genetic mutations, can be maintained a few (2-10) generations without selection and exhibit a bias in the type of genes they target. Because it is known that genetic mutations exhibit a bias in terms of where they occur in the genome (which is often influenced by epigenetic factors) (Chen et al. 2012; Schuster-Bockler and Lehner 2012; Li et al. 2013; Li and Luscombe 2020; Monroe et al. 2022), a key functional difference between genetic mutations and epimutations is their duration of inheritance. It has been suggested that due to their limited duration, epimutations can only be maintained in populations long term if they offer a selective advantage (Beltran et al. 2020). Theoretical studies that have examined the inheritance of spontaneous epialleles have been relatively limited. Non-genetic inheritance has been broadly analyzed by

extending the Price equation (Day and Bonduriansky 2011), and these results highlight the potential for the interaction of genetic and epigenetic alleles to affect evolutionary outcomes.

Here, we extend classic population genetic models to incorporate epimutations that impact fitness. Our results show that epialleles that would persist only a few generations without selection or ongoing epimutation can be maintained in long-term in epimutation-back-epimutation equilibrium. We show how this equilibrium changes when one epiallele has a selective advantage. Across genotypes, a compensatory or advantageous epiallele facilitates the maintenance of a deleterious genetic allele at a higher frequency than expected under mutation-selection balance. Collectively, these results suggest that a relatively simple accounting of epialleles into population genetic models can help explain both genetic and epigenetic variation in a population.

MODEL

We consider two genetic alleles (A_1 and a_2) at a locus, each with an alternative epiallelic states (A_3 corresponding to A_1 and a_4 corresponding to a_2). Therefore, there are a total of four alleles at the locus, but only two underlying genotypes: A_1 and A_3 are the same genotype but can have distinct fitness values due to their epigenetic modifications, and a_2 and a_4 are also the same genotype but can have distinct fitness values. Each generation, the mutation rate, μ, is the rate at which A_1 or A_3 mutate to a_2 or a_4, respectively, and we assume there is no back-mutation. The epimutation rate, ε, is the rate of epimutation of A_1 to A_3 and a_2 to a_4. In this case, consistent with the limited inheritance potential of epimutations, the back-epimutation rate (β) models the rate at which A_3 and a_4 revert to A_1 and a_2. We consider diploid individuals in an infinite population with discrete generations. Mutation, epimutation, and back-epimutation occur in the gametes, while viability selection occurs in diploids. A schematic of the life cycle is shown in Figure 1A. Using $p_1(t), p_2(t), p_3(t)$, and $p_4(t)$ as the frequency of A_1, a_2, A_3, and a_4 at time t in generations and representing the genotypic fitness values of the various diploid allelic combinations as w_{11}, w_{22}, w_{33}, w_{44}, w_{12}, w_{13}, w_{14}, w_{23}, w_{24}, and w_{34}, the recursion equations for each generation are given by:

$$p_1(t+1) = \frac{p_1(t)(p_1(t)w_{11}+p_2(t)w_{12}+p_3(t)w_{13}+p_4(t)w_{14})(1-\mu-\varepsilon)+\beta p_3(t)(p_3(t)w_{33}+p_1(t)w_{13}+p_2(t)w_{23}+p_4(t)w_{34})}{\bar{w}(t)} \quad (1)$$

$$p_2(t+1) = \frac{p_2(t)(p_1(t)w_{12}+p_2(t)w_{22}+p_3(t)w_{23}+p_4(t)w_{24})(1-\varepsilon)+\mu p_1(t)(p_1(t)w_{11}+p_2(t)w_{12}+p_3(t)w_{13}+p_4(t)w_{14})+\beta p_4(t)(p_3(t)w_{34}+p_1(t)w_{14}+p_2(t)w_{24}+p_4(t)w_{44})}{\bar{w}(t)} \quad (2)$$

$$p_3(t+1) = \frac{p_3(t)(p_3(t)w_{33}+p_1(t)w_{13}+p_2(t)w_{23}+p_4(t)w_{34})(1-\mu-\beta)+\varepsilon p_1(t)(p_1(t)w_{11}+p_2(t)w_{12}+p_3(t)w_{13}+p_4(t)w_{14})}{\bar{w}(t)} \quad (3)$$

$$p_4(t+1) = 1 - (p_1(t+1) + p_2(t+1) + p_3(t+1)) \quad (4)$$

where $\bar{w}(t)$ is the average fitness at time *t*. Transition probabilities from one epiallele to another, accounting for mutation, epimutation, and back-epimutation, are shown in Figure 1B. Under neutral conditions all fitness values are equal to 1, but when selection is occurring, each w_{nk} (for *n*, *k* equal to 1, 2, 3, or 4) can be written as 1-s_l for selection coefficient s_l. For each fitness value, the corresponding selection coefficient can be written as w_{11} = 1-s_0, w_{22} = 1-s_1, w_{33} = 1-s_2, w_{44} = 1-s_3, w_{12} = 1-s_4, w_{13} = 1-s_5, w_{14} = 1-s_6, w_{23} = 1-s_7, w_{24} = 1-s_8, and w_{34} = 1-s_9.

While epialleles are biologically distinct from genetic alleles, they are analogous to a multiallelic genetic model in which certain alleles have a high probability of 'mutating' (in our case, epimutating) to other, specific alleles. Equilibria for systems with *k* alleles were considered early in population genetics research (Kimura 1956; Mandel 1959), then subsequently expanded to include selection and/or mutation (Tallis 1966; Karlin 1980; Hadeler 1981; Varga and Zubiri 1993). These approaches have been based upon the assumption that mutation rates were low or equal across alleles, as is appropriate for approximately uniform changes in DNA sequence. The major difference with our analysis relative to other multiallelic models is that we allow epimutation and back-epimutations rates to be much higher than DNA-sequence mutation rates, which generates asymmetries in the equilibria that would not appear in sequence-based mutation approaches alone. We do make the simplifying assumption that multiple types of mutation do not occur in the same generation (*e.g.,* it is not possible to have both mutation and back-epimutation occur at once).

We derive analytical solutions when possible, but a complete analysis requires numerical analysis in many cases. For numerical analysis, we use a mutation rate of 10^{-7} and epimutation rates ranging from 10^{-5} to 0.5. The rates used represent whole allele rates rather nucleotide-specific rates. We note that very low back-epimutation rates and very high epimutation rates are less likely to occur in nature. While epimutation rates are typically higher than mutation rates genome wide, the upper value we chose is likely significantly higher than a genome-wide rate. However, because there can be bias for which genomic loci experience epimutation, it is possible for the rate of epimutation to be higher than the genome-wide rate for a given allele (Beltran et al. 2020), and this high rate is meant to account for this circumstance. Finally, high back-epimutation rates are largely the parameter that distinguish a model of epigenetic inheritance from other models. While genetic

mutations are unlikely to be truly reversible (and we assume that that is very unlikely here), epimutations are typically reversible on short timescales due to the rapid decay of epigenetic marks or modifiers. Based on the limited studies on the persistence of epimutations in animals, we have chosen two high back-epimutation rates of 0.33 and 0.1, which correspond to the maintenance of epialleles without selection for ~3 or ~10 generations, respectively (Sarkies 2020) (Figure 2). Based on the more extended persistence of epimutations observed in plants and yeast (Grewal and Klar 1996; Jacobsen and Meyerowitz 1997; Kakutani 2002), we have also included a back-epimutation rate of 0.01 (~100 generations, Figure 2). Thus, we consider 0.33, 0.1, and 0.01 our 'focal' back-epimutation rates.

ANALYSIS METHODS

We used *Mathematica* version 12.2.0.0 to perform analysis of models and used R version 3.5.1 and Adobe Illustrator 2021 for data visualization. The Solve function was used to find solutions when a system could be solved analytically. For numerical solutions, equilibria were determined in *Mathematica* using the NSolve function by setting $p_x(t+1) = p_x(t)$ in equations 1-3 above for a given parameter set, with $p_4 = 1 - (p_1 + p_2 + p_3)$. For a given solution, working precision was set to 10 and solutions were restricted to be bounded by 0 and 1, inclusive.

Stability of equilibrium numerical solutions was determined by calculating the Jacobian matrix of the system via approximating the recursion equations above by the following continuous differential equations:

$$\frac{dp_1}{dt} = \frac{p_1(p_1w_{11}+p_2w_{12}+p_3w_{13}+(1-(p_1+p_2+p_3))w_{14})(1-\mu-t_2)+\beta p_3(p_3w_{33}+p_1w_{13}+p_2w_{23}+(1-(p_1+p_2+p_3))w_{34})}{\bar{w}} - p_1 \quad (5)$$

$$\frac{dp_2}{dt} = \frac{p_2(p_1(t)w_{12}+p_2w_{22}+p_3w_{23}+(1-(p_1+p_2+p_3))w_{24})(1-t_2)+\mu p_1(p_1w_{11}+p_2w_{12}+p_3w_{13}+(1-(p_1+p_2+p_3))w_{14})+\beta p_4(p_3w_{34}+p_1w_{14}+p_2w_{24}+(1-(p_1+p_2+p_3))w_{44})}{\bar{w}} - p_2 \quad (6)$$

$$\frac{dp_3}{dt} = \frac{p_3(p_3w_{33}+p_1w_{13}+p_2w_{23}+(1-(p_1+p_2+p_3))w_{34})(1-\mu-\beta)+t_2p_1(p_1w_{11}+p_2w_{12}+p_3w_{13}+(1-(p_1+p_2+p_3))w_{14})}{\bar{w}} - p_3 \quad (7)$$

Allowing $f_1 = \frac{dp_1}{dt}$, $f_2 = \frac{dp_2}{dt}$, and $f_3 = \frac{dp_3}{dt}$, the Jacobian matrix is given by:

$$J = \begin{bmatrix} \frac{\partial f_1}{p_1} & \frac{\partial f_1}{p_2} & \frac{\partial f_1}{p_3} \\ \frac{\partial f_2}{p_1} & \frac{\partial f_2}{p_2} & \frac{\partial f_2}{p_3} \\ \frac{\partial f_3}{p_1} & \frac{\partial f_3}{p_2} & \frac{\partial f_3}{p_3} \end{bmatrix}$$

If all eigenvalues of this matrix were negative for a given parameter set, the equilibrium was considered stable; if at least one was positive, the equilibrium was considered unstable.

For all numerical analyses, the mutation rate was held constant at 10^{-7}. We typically used fitness values of 1, 0.99, or 0.98 for values of each w_{nk} (for n,k equal to 1, 2, 3, or 4), but w_{nk} could lie anywhere between 0 and 1, inclusive, depending on the specific analysis and as indicated in each figure and in the text. For analyses that focused on the focal back-epimutation rates of 0.01, 0.1, and 0.33, epimutation rates were randomly drawn from a log-uniform distribution for rates between 10^{-5} and 0.5, then analyzed for each of the three back-epimutation rates. Results were exported into a .txt file, organized in Microsoft Excel as a .csv file, then analyzed in R. For these figures, solid and smooth lines were then drawn in Adobe Illustrator along points. Original graphs without lines drawn are available in Supplemental Figures 1-2. To analyze the full parameter space of epimutation and back-epimutation rates (*i.e.,* the shaded red graphs in Figures 5-7), 300 random values were selected from a log-uniform distribution for both epimutation rates and back-epimutation rates between 10^{-5} and 0.5, then each combination of epimutation and back-epimutation rate (90,000 in total) was used to determine equilibrium solutions and stability. In rare cases, Mathematica failed to find equilibria for certain combinations of epimutation and back-epimutation rates. In other rare cases, the same equilibrium would be identified as an equilibrium multiple times for a given set of parameters. When these duplicates were identified, they were removed from the resulting data prior to plotting. If the source of an apparent irregularity could not be decisively determined, it was included in the final plot, such as some components of Figure 6F, Supplementary Figure 1B, and Supplementary Figure 2B.

RESULTS

Simplified models recapitulate standard population genetic solutions in the epigenetic context

Under neutral conditions (w_{nk} = 1 for all n,k), we find that both epialleles within a genotype (in this case, a_2 and a_4,) are maintained in long-term equilibrium in a manner that depends only on the epimutation and back-epimutation rates (Figure 3). Notably, this is exactly equivalent to genetic models that incorporate both forward and back genetic mutation (Wright 1931). Because back-mutation of a specific allele is rare, however, back-mutation is typically excluded from population genetic models that make simplifying assumptions. This

results in the following relationship for the long-term equilibrium of p_4, which we will refer to epimutation-back-epimutation equilibrium (EBE):

$$p_4 = \frac{\varepsilon}{\beta + \varepsilon} \tag{8}$$

Although biologically distinct from genetic mutations, epimutations should behave like genetic mutations as the back-epimutation rate (β) approaches zero. That is, we should be able to derive 'epimutation-selection balance' that is equivalent to mutation-selection balance in standard population genetic models. We focus again on a_2 and a_4, which are epialleles within the same genotype, and exclude the other genotype (A_1 and A_3). We thus have only three fitness values: s_1, s_3, and s_8. We first solve equations for when a_2 is dominant. Solving the recursion equations for the case when $\beta = 0$, s_1 =0, and s_8 = 0, yields the equilibrium frequency of p_4:

$$p_4 = \sqrt{\frac{\varepsilon}{s_3}} \tag{9}$$

Similarly, incorporating a dominance coefficient, h, when $\beta = 0$, s_1 =0, and s_8 = 1- hs_3 yields the equilibrium:

$$p_4 = \frac{2\varepsilon}{hs_3(1+\varepsilon)+\sqrt{s_3(4\varepsilon-8h\varepsilon+h^2s_3(1+\varepsilon)^2)}} \approx \frac{\varepsilon}{hs_3} \tag{10}$$

Equations 9 and 10 are likely biologically relevant for some plants, in which epialleles can be maintained long-term much like genetic mutations. Note that we can similarly obtain back-epimutation-selection balance if $\varepsilon = 0$, s_3 =0, and s_8 = 0, yielding the following equilibrium frequency of p_4:

$$p_4 = 1 - \sqrt{\frac{\beta}{s_1}} \tag{11}$$

Because we typically expect epigenetic variants to be easily reversible, making β larger than the selection coefficient, in keeping with our intuition and in stark contrast to standard genetic mutation-selection balance, it is difficult to maintain an epigenetic allele unless it is also re-generated via epimutation.

Compensatory epialleles cause deviations from mutation-selection balance by a common factor

Our main goal is to understand how all forces—epimutation, back-epimutation, mutation, and selection—interact in a system with four epialleles across two genotypes. We cannot analytically solve the full system with all selection coefficients as parameters, but we can solve cases that are direct extensions of classical mutation-selection balance. First, consider a case in which a reference genetic allele (A_1) mutates to a deleterious recessive allele (a_2) while the associated epialleles for each genotype, A_3 and a_4, are neutral with

respect to A_1. In other words, the a_4 epiallele, while the same genotype as a_2, has higher fitness than its counterpart. Biologically, this could model a common situation in which two genotypes each have an active and silent state (*e.g.* the locus is methylated or unmethylated). One genotype is deleterious when homozygous in its active state but is neutral with respect to the other alleles if at least one allele is silenced. The fitness of the other genotype does not depend on whether it is active or silenced. In our model this means all selection coefficients are 0 except for s_1. Solving for the equilibrium for all four epialleles across both genotypes yields:

$$p_1 = 1 - \sqrt{\frac{\mu}{s_1}} - \frac{\varepsilon - \varepsilon\mu}{\varepsilon+\beta-\mu} \approx 1 - \sqrt{\frac{\mu}{s_1}} - \frac{\varepsilon}{\varepsilon+\beta} = \frac{\beta}{\varepsilon+\beta} - \sqrt{\frac{\mu}{s_1}} \tag{12}$$

$$p_2 = \sqrt{\frac{\mu}{s_1}} \tag{13}$$

$$p_3 = \frac{\varepsilon(\beta+\mu\varepsilon-\mu)}{\beta(\beta+\varepsilon-\mu)} - \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} \approx \frac{\varepsilon}{\varepsilon+\beta} - \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} \tag{14}$$

$$p_4 = \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} + \frac{\varepsilon\mu(1-\beta-\varepsilon)}{\beta(\beta+\varepsilon-\mu)} \approx \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} \tag{15}$$

Recall that from a DNA sequence point of view, the observed allelic frequency within the population is the sum the of the modified and unmodified allelic states:

$$p_1 + p_3 \approx \frac{\beta}{\varepsilon+\beta} + \frac{\varepsilon}{\varepsilon+\beta} - \sqrt{\frac{\mu}{s_1}} - \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} = 1 - \sqrt{\frac{\mu}{s_1}} - \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} = 1 - (1 + \frac{\varepsilon}{\beta})\sqrt{\frac{\mu}{s_1}} \tag{16}$$

$$p_2 + p_4 \approx \sqrt{\frac{\mu}{s_1}} + \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} = (1 + \frac{\varepsilon}{\beta})\sqrt{\frac{\mu}{s_1}} \tag{17}$$

Under standard mutation-selection balance, we expect the dominant allele to be present at a frequency of $1 - \sqrt{\frac{\mu}{s_1}}$, and the recessive allele to be present at frequency $\sqrt{\frac{\mu}{s_1}}$. (This is also the case in the present model when s_1, s_3, and s_8 all equal s_1). However, in this case there is an alternative epigenetic state of the recessive deleterious allele that compensates for this deleterious effect. As a result, the equilibrium at a genotypic level is increased from mutation-selection balance by a factor of $(1 + \frac{\varepsilon}{\beta})$. This quantity readily reveals the important factors affecting the maintenance of genetic variation. At one extreme, if $\varepsilon > \beta$ we expect a large increase over mutation-selection balance, which could occur if new epigenetic variants are frequently generated and/or if they persist multiple generations. On the other hand, if new epigenetic variants arise infrequently and/or are readily reversed each generation ($\varepsilon < \beta$), then this will only minimally affect mutation-selection balance. And as

expected, when no epigenetic variation is generated, the equilibria are exactly those derived under regular mutation-selection balance.

Alternatively, we can ask how these equilibria deviate from expectations from the EBE by summing modified and unmodified states across alleles:

$$p_1 + p_2 \approx \frac{\beta}{\varepsilon+\beta} - \sqrt{\frac{\mu}{s_1}} + \sqrt{\frac{\mu}{s_1}} = \frac{\beta}{\varepsilon+\beta} \tag{18}$$

$$p_3 + p_4 \approx \frac{\varepsilon}{\varepsilon+\beta} - \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} + \frac{\varepsilon}{\beta}\sqrt{\frac{\mu}{s_1}} = \frac{\varepsilon}{\varepsilon+\beta} \tag{19}$$

This shows there are no deviations from EBE by the introduction of a selection coefficient affecting only w_{22}.

The analytical solution presented above applies in a case in which a deleterious mutation (a_2) is completely recessive relative to the reference genetic allele (A_1) as well as to alternative epialleles of either genotype (A_3 or a_4). However, it may be the case that a_2 is not completely recessive, such as if silencing only one allele is not sufficient to mediate the negative consequences of mutation. Although we were unable to solve the system analytically when a general dominance coefficient (h) was incorporated, we can find a solution in the additive case when h=0.5. That is, s_4, s_7, and s_8 all equal hs_1 when h=0.5 we find that solutions again deviate from mutation-selection balance – in this case $\frac{2\mu}{s_1}$ – by approximately a factor of $(1+\frac{\varepsilon}{\beta})$. For the more general case of arbitrary dominance, we examined part of the parameter space numerically and examined whether our epigenetic frequency escalation term worked as a general approximation. We let s_1 = 0.01 and s_4, s_7, s_8 = hs_1 and randomly selected values of h from 0 to 1. In all cases examined, the estimate of $(1+\frac{\varepsilon}{\beta})\frac{\mu}{hs_1}$ was a reasonable approximation for the sum of p_2 and p_4 and a better estimate than mutation-selection balance alone (Figure 4). Thus, these equations give us broad insight into how epigenetic variants that compensate from a deleterious mutation—whether that mutation is dominant, recessive, or with varying levels of dominance—cause higher equilibrium allele frequencies than those expected by mutation-selection balance alone, by deriving a fairly simple multiplier that is dependent on the relative magnitudes of ε and β.

Epialleles under strong selection cause deviations from epimutation-back-epimutation equilibrium

We next asked how population dynamics are affected when a single epiallele yields a selective advantage relative to other alleles that are neutral with respect to each other, rather than compensating for a deleterious allele. While advantageous genetic alleles would be expected to be maintained at high frequency in

a population, the high reversibility of epialleles makes it difficult to know how selection affects their long-term persistence. In our model we consider when the a_4 epiallele has both a recessive and dominant fitness advantage relative to all other alleles. In the absence of any selective advantage, epialleles within a genotype will be in EBE, and by incorporating a selection coefficient of 0.01, our results show only mild deviations from EBE (Figure 5A-C). Throughout most of the parameter space, a single equilibrium exists, although in the recessive case there are sometimes two stable equilibria (Figure 5B). In the symmetric case when A_3 has a selective advantage, we find that due to the asymmetry in the genetic mutation rate, it is possible for simultaneous maintenance of A_1 and a_2 to occur at low epimutation rates (Supplementary Figure 1), but for higher rates, patterns are consistent with results seen when a_4 has a selective advantage. Notably, when s=0.2, a_4 is maintained at high frequency even when the back-epimutation rate is high, highlighting the opportunity for an advantageous epiallele to significantly alter the dynamics of EBE. To systematically assess the role of selection coefficient on the dynamics of EBE, we randomly selected selection coefficients between 0 and 1 and allowed a_4 to have a dominant or recessive selective advantage relative to all other alleles for specific epimutation and back-epimutation rates. In all cases, EBE is a good approximation for the equilibrium as long as the selection coefficient is small (less than 0.01), but it becomes an underestimate as the selection coefficient increases (greater than 0.1). EBE also is more likely to yield an accurate estimate if $\varepsilon < \beta$. Thus, population dynamics of small-effect variants that are inherited a limited number of generations – likely the dominant class of epigenetic variants in nature—are dominated by epimutation and back-epimutation rates and are well estimated by EBE. Nonetheless, large-effect epigenetic variants that are less readily reversible, if they arise, can significantly alter these dynamics.

Certain advantageous epialleles maintain deleterious genetic alleles at high frequency

Thus far, we have examined population dynamics when an epiallele compensates for a deleterious allele (solution in equations 12-15) and when an epiallele has a selective advantage over other alleles that are neutral with respect to each other (Figure 5). Here, we combine these two cases and ask: how are population dynamics affected if an epiallele has higher fitness than both a reference and deleterious allele? For instance, methylated alleles could have higher fitness than an unmethylated alleles, regardless of the underlying DNA sequence, while the fitness of an allele when it is unmethylated is dependent upon its genetic sequence. For

comparison's sake, we first show the case of an epiallele compensating for a recessive deleterious allele numerically when s_1 = 0.01 (Figure 6A-B). We then show how these dynamics change when, instead of merely compensating for a deleterious allele, A_3 or a_4 epialleles have a selective advantage compared to other alleles in a dominant manner (Figures 6C-D). This leaves A_1A_1, A_1a_2, and a_2a_2 with intermediate fitness and a_2a_2 alone with the lowest fitness. The primary difference in the resulting equilibria between these two cases is for A_3 and a_4 equilibria when ε and β are both low. Thus, for high β, equations 12-15 are a good approximation of equilibria. However, these results change substantially if only a_4 has a dominant selective advantage, but A_3 does not (Figure 6E-F). Here, s_1=0.02, s_3, s_6, s_8, s_9 = 0, and all other s=0.01. We find that the frequency of a_2, despite being recessive and deleterious, increases to high frequency, well beyond the expectation of mutation-selection balance. All four epialleles can be maintained at appreciable frequency simultaneously, which is unique among situations considered thus far. In this case, the driving factor is likely that a_2 and a_4 – epialleles within a genotype – correspond to the extremes of the fitness distribution. This maintenance of all four epialleles across both genotypes also occurs if we look at the symmetric case in which diploids containing A_1 and A_3 are at the extremes of the fitness distribution (Supplementary Figure 2), highlighting that this phenomenon is not merely a function of the unidirectional mutation rate.

When the deleterious allele is instead dominant, and the compensatory epiallele is recessive, we expect the genetic allele to be maintained at a frequency of $(1+\frac{\varepsilon}{\beta})\frac{\mu}{s_1}$. We show this result numerically in Figure 7A-B for s_1, s_4, s_7, s_8 =0.01 when all other s=0. When only diploids containing epialleles A_3 and a_4 have the highest fitness (s_2,s_3,s_9=0), and those containing A_1 are intermediate, we find differences primarily in the equilibria of A_3 and a_4 when ε and β are both low (Figure 7C-D). This is analogous to Figure 6 for the recessive case and our derived deviation from mutation-selection balance, $(1+\frac{\varepsilon}{\beta})$, gives a reasonable approximation of equilibria for high β. An exception again occurs when a single epiallele exhibits a recessive advantage relative to all other alleles (Figure 7E-F). Here, we find multiple equilibria even for high β, and one of these equilibria includes the maintenance of the deleterious a_2 at high frequency. Unlike the recessive case shown in Figure 6E-F, all four epialleles are not simultaneously present at high frequency. Instead, the mix of epialleles consists primarily of A_1 and A_3 or a_2 and a_4. Nonetheless, it is striking that a recessive advantageous epiallele can potentially cause such deviations from expectation for a deleterious dominant mutation. As before, this is likely due to the epialleles within a genotype corresponding to the extremes of the fitness distribution.

Because we twice observed that maintaining a deleterious allele (a_2) at high frequency is dependent on the fitness advantage of its corresponding epiallele (a_4), we wanted to further assess the parameters for which this is possible. When a_2 is recessive deleterious and a_4 is dominant advantageous, a_2 is maintained at high frequency alongside all other epialleles simultaneously (Figure 6E-F, Figure 8A). When a_2 is dominant deleterious and a_4 is recessive advantageous, we find that a_2 can be maintained at high frequency but only alongside a_4 (Figure 7E-F, Figure 8A). When the alternative epiallele is deleterious, this dynamic is reversed (Figure 8A). That is, when a_2 is recessive advantageous and a_4 is dominant deleterious, only a_2 and a_4 are maintained at appreciable frequency, but when a_2 is dominant advantageous and a_4 is recessive deleterious, all four epialleles are maintained (Figure 8A). In both cases for which all four epialleles are maintained, one epiallele acts dominantly to increase fitness, while its corresponding epiallele within the same genotype has the lowest fitness and acts in a recessive manner (and the symmetric situations for A_3 and A_1 also show these same dynamics; see Supplementary Figure 2). In contrast, when deleterious alleles are maintained, but not in a manner that promotes multiple genetic alleles simultaneously, this is due to a recessive fitness advantage and dominant deleterious fitness. Because deleterious genetic variation can be maintained in long-term equilibrium alongside the favored allele, we assessed numerically how the fitness of the deleterious allele affects these dynamics. We held the epimutation rate constant (at either 10^{-5} and 0.01) but allowed the fitness of a_2a_2 (w_{22}) to vary. While the magnitude of the equilibrium frequency for a_2 varies depending on the fitness, the parameters (particularly, the back-epimutation rate) for which a_2 is maintained do not (Figure 8B-C). As a result, in order to maintain deleterious genetic variation when the back-epimutation rate is high (0.1 or greater), the epimutation rate must also be high, and this is true regardless of the fitness of the deleterious allele.

DISCUSSION

Here we developed and analyzed an extension of a population genetic model that incorporates epigenetic alleles. We solved relevant simplifications of the full system analytically and special cases of the full system numerically. First, the neutral case (all $w_{nk} = 1$) shows that multiple epigenetic states within a single genotype can be maintained in a manner that depends only on epimutation and back-epimutation rates, called epimutation-back-epimutation equilibrium (EBE). The neutral case is directly analogous to the maintenance of genetic variation by mutation and back-mutation, although epimutation and back-epimutation occur at much

higher rates than genetic mutation in nature and thus can drive substantial epigenetic variation. Second, we show that when the back-epimutation rate is 0 (meaning epigenetic variants are not reversible and thus act similarly to genetic mutations) and selection acts against new epiallales, we derive the equation for epimutation-selection balance that is directly analogous to mutation-selection balance. Third, we show that when an epimutation compensates for a recessive or dominant deleterious genetic allele, we find deviations from mutation-selection balance by a common factor, $(1+\frac{\varepsilon}{\beta})$. Finally, we use numerical analysis to analyze specific cases of the full model when both genetic and epigenetic variation are present. Of particular interest, we find instances of dramatic deviation from mutation-selection balance when the two epialleles of the same genotype are at the extremes of the fitness distribution. Collectively, this demonstrates the importance of considering epigenetic variation to understand evolutionary dynamics of populations.

The incorporation of epigenetic alleles in population genetic models has important implications for the maintenance of epigenetic and genetic variation. First, the maintenance of *epigenetic* variation occurs in all situations examined and is approximated well by epimutation-back-epimutation equilibrium when selection coefficients are small. This suggests the maintenance of epigenetic variation may be relatively common if the epimutation rate is sufficiently high. Second, the maintenance of *genetic* variation occurs when a deleterious allele is compensated for by an alternative epiallele. This likely does not represent a large deviation in most cases, but it highlights the possibility for large perturbations that may occur for high ε and low β. We discuss these major results below. In addition, we consider the potential role of two factors not considered here -- drift and the environment.

The maintenance of epigenetic variation may be common

Our results suggest that when an epigenetic variant arises, it may be maintained in long-term stable equilibrium even if it is highly reversible (*i.e.,* even if it has a high back-epimutation rate). This is true regardless of whether mutation and epimutation of an allele result in similar or distinct fitness effects. However, it is notable that for highly reversible epimutations (those maintained 3 to 10 generations without selection) to be maintained long-term at appreciable frequency in the absence of strong selection, they must also be generated at relatively high frequency (~0.001 to 0.01 probability each generation). While epimutations occur at higher frequency than genetic mutations, this range is higher than expected under non-stressful, baseline conditions

and when considering a genome-wide rate. As a result, it will be important for future empirical work to determine 1) whether there are environmental conditions for which the epimutation rate may be higher, and 2) whether there are epimutation hotspots in the genome for which epimutation rates are higher. It has been observed that certain stresses can act to increase phenotypic variation, possibly due to increasing epigenetic variation, and transgenerational epigenetic effects have also been identified in response to stress. In addition, the observation that transgenerational effects are more robust if individuals are exposed to environmental stressors for multiple generations lends credence to the idea the epimutation rates may be higher under stress (Vogt and Hobert 2023). However, to our knowledge, experimental evolution in an ongoing stressful environment with monitoring of epimutations across generations has not been performed. In addition, the potential role of drift remains unclear, though the maintenance of epialleles within a given population is likely to largely depend on the population size, which can be explored in future theoretical studies.

Design of experiments to assess the maintenance of epigenetic variation

In measurements of epigenetic regulation across individuals in a population, it is often unclear whether differences across individuals are due to technical variation, 'random' noise, or meaningful biological differences. Our equation approximating the maintenance of an epiallele (Equation 8) may be a useful consideration when attempting to differentiate these possibilities. Epimutation rates can be empirically measured, as can their reversibility. Given these two parameters, it is straightforward to calculate how much variation is expected to be present at a given locus. On the other hand, these parameters allow for more precise experimental evolution paradigms to proceed. If epimutation rates are very low, then epigenetic variation is likely not to be maintained, assuming that the environment is constant and the population is sufficiently large for the present model to be applicable.

Compensatory epimutations cause predictable deviations from mutation-selection balance

In standard population genetics, the equilibrium of a deleterious recessive or dominant allele is well known. Here, we introduce an epiallele of a deleterious allele that compensates for this effect and find that the genotype containing both the deleterious and compensatory epialleles deviates from the expectation of mutation-selection balance by a factor of $(1 + \frac{\varepsilon}{\beta})$. Thus, in cases in which epimutation and back-epimutation

are present and afford some degree of compensation for a deleterious mutation, mutation-selection balance will be an underestimate of the deleterious allele in the population. It will likely be difficult to assess whether epimutation and back-epimutation are playing a role in natural populations, given that these parameters may depend on an environment that will be difficult to recapitulate in the lab. However, at least one clue that these may be playing a role is if significant variation in the phenotype is present for a deleterious allele. For example, if the average fitness of an alternative genotype is lower than the reference genotype, but there is significant variability across individuals in the alternative genotype within a single environment, this may indicate that cryptic epialleles exist within the genotype. Directly associating an epigenetic mark, such as cytosine or histone methylation, with a phenotype can provide support to this idea in this case.

The maintenance of deleterious genetic variation at high frequency is possible but likely rare

While epialleles were maintained for certain parameters in every situation analyzed, deleterious genetic variation was typically not maintained at high frequency. Sometimes, we find that A_3 and a_4 are maintained together at high frequency (see the boundary between A_3 and a_4 in Figures 6 and 7), but this is in a part of the parameter space for which the back-epimutation rate is very low, causing the epimutation to behave more like a genetic mutation. In these cases, it is relatively unsurprising that an advantageous and minimally-reversible epiallele is maintained. Such a scenario might be applicable when a gene silencing state via DNA methylation is required for fitness, and the silenced state is inherited; in this case, if a genetic mutation occurs, the mutation may not realize its effect. However, in the context of highly reversible epigenetic variants, deleterious variation can only be maintained at high frequency when the deleterious allele has an alternative epiallele with higher fitness than other alleles in the system (see Figures 6E-F, Figure 7E-F, Figure 8). In both cases, the extremes of the fitness distribution include the alternative epialleles of a – with a_2a_2 having the lowest fitness and a_4a_4 having the highest fitness. While it is not difficult to imagine that an epiallele could compensate for a deleterious genetic allele resulting in the deviation from mutation-selection balance discussed above, it is more difficult to imagine that an epimutation would make the deleterious allele *more* fit than the non-deleterious allele. However, again, the role of the environment and drift were not considered in the present study, and analyzing both of these possibilities could lend additional insight into the feasibility of such a scenario in nature. For example, if a deleterious allele is only deleterious in a given environment, is it possible that occasionally

encountering a more favorable environment can facilitate maintenance of the deleterious allele? In addition, is it possible that deleterious variation will be maintained if the population is sufficiently small? Such questions can be answered by extensions to the present model and will inform future empirical studies.

CONCLUSIONS

Understanding the forces that maintain variation within natural populations represents a very long term and central part of evolutionary genetics (Lewontin 1974; Kimura 1983; Marion and Noor 2022). One of the earliest observations in the field is that there is far more variation at the molecular level than would be expected via purifying selection. Our results demonstrate that epigenetic modifications can be added to the long list of other potential factors such as balancing selection that can elevate allele frequencies above the mutation-selection balance expectation. There has long been interest in the potential for non-genetic inheritance to have an influence on evolutionary outcomes (Kirkpatrick and Lande 1989; Mousseau and Fox 1998; Jablonka and Raz 2009; Richards et al. 2012). The last decade has led to rapid increase in knowledge of the molecular underpinnings of epigenetic variation and the inheritance of these effects for a number of generations (Silveira et al. 2013; Öst et al. 2014; Schott et al. 2014; Siklenka et al. 2015; Ciabrelli et al. 2017; Klosin et al. 2017; Yu et al. 2018; Furci et al. 2019b; Moore et al. 2019; Kaletsky et al. 2020; Beck et al. 2021). While the pace of discovery on the molecular side is accelerating, the population genetic implications of these mechanisms have yet to be fully explored. Here, we examine a case in which epimutations are “random” with respect to any induction by the environment. In this case, variation is generated at the “phenotypic” level via the stable maintenance epialleles, and in some cases at the genetic level via the maintenance of deleterious alleles that whose effects would otherwise appear cryptic in the absence of specific information about potential epigenetic modification. Indeed, these epigenetic effects exist in the liminal space between genotype and phenotype in the classical models of genetic inheritance, and, depending on their prevalence, hold the potential to further complicate the already complex landscape functional variation within natural populations.

FIGURES

Figure 1: Life cycle diagram for population epigenetic model

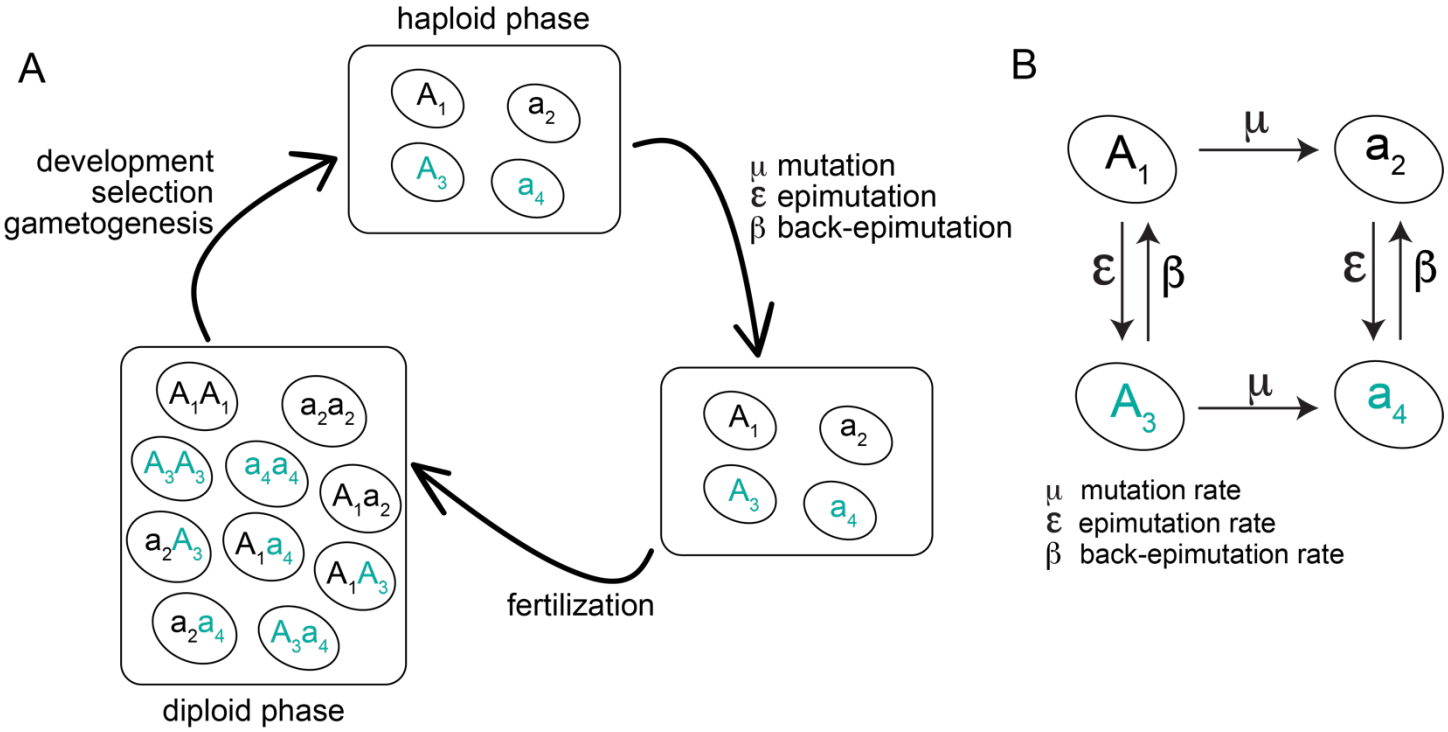


Figure 1: Life cycle diagram and transition probabilities for population epigenetic model.

A. Four epialleles are possible: A_1 and a_2 are genetic alleles, while A_3 and a_4 represent alternative epiallelic states. Mutation, epimutation, and back-epimutation occur in the haploid gametes, and diploid combinations of each pair of epialleles are possible upon fertilization. Selection occurs on diploids, which then produce haploid gametes, completing the life cycle.

B. Transition probabilities for mutation, epimutation, and back-epimutation for each type of epiallele to the other. Mutation occurs from A_1 to a_2 or from A_3 to a_4. Epimutation occurs from A_1 to A_3 or from a_2 to a_4. Finally, back-epimutation occurs from A_3 to A_1 or from a_4 to a_2. There is no parameter for back-mutation. Within a single generation, only one type of transition occurs, *i.e.,* mutation, back-epimutation, and epimutation cannot happen sequentially or simultaneously in the same generation.

Figure 2: The back-epimutation rate (β) models the maintenance of an epiallele without selection

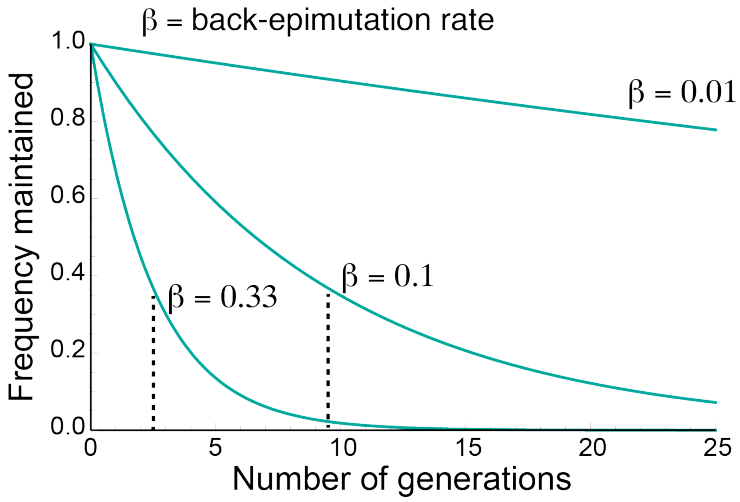


Figure 2: The back-epimutation rate models the maintenance of an epiallele without selection. In the absence of other factors, including new mutation, epimutation, or selection, the back-epimutation rate models the expected number of generations an alternative epiallele (A_3 or a_4) would be expected to be maintained. The equations plotted here are $p = (1 - \beta)^t$, where t represents the number of generations, β is the back-epimutation rate, and p is the frequency of the epiallele that is maintained, which starts at 1. Examples are shown for three values of β: 0.33, 0.1, and 0.01, which model three different expected numbers of generations an epiallele would be maintained without selection: 2.5, 9.5, and 99.5, respectively. The expected value is indicated by the dotted line for β = 0.33 and 0.1.

Figure 3: Epigenetic variation is maintained in epimutation-back-epimutation equilibrium under neutral conditions

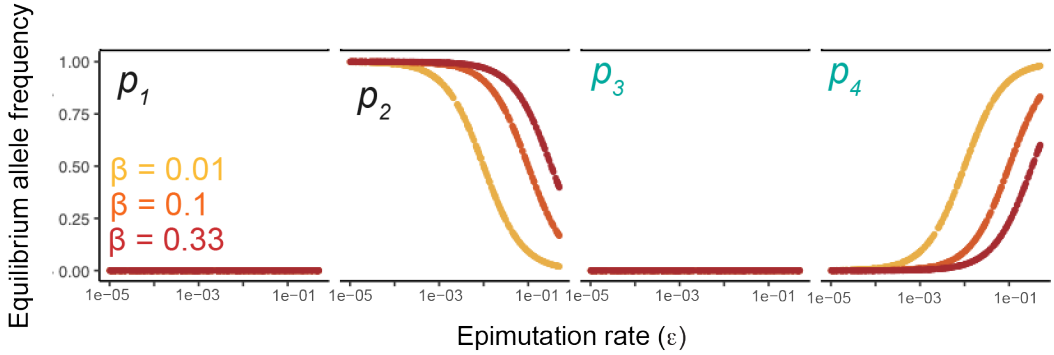


Figure 3: Epigenetic variation is maintained in epimutation-back-epimutation equilibrium under neutral conditions. When all fitness levels are set to 1, a_2 and a_4 alleles are found in equilibrium at frequencies that depend on the epimutation and back-epimutation rate, according to Equation 8.

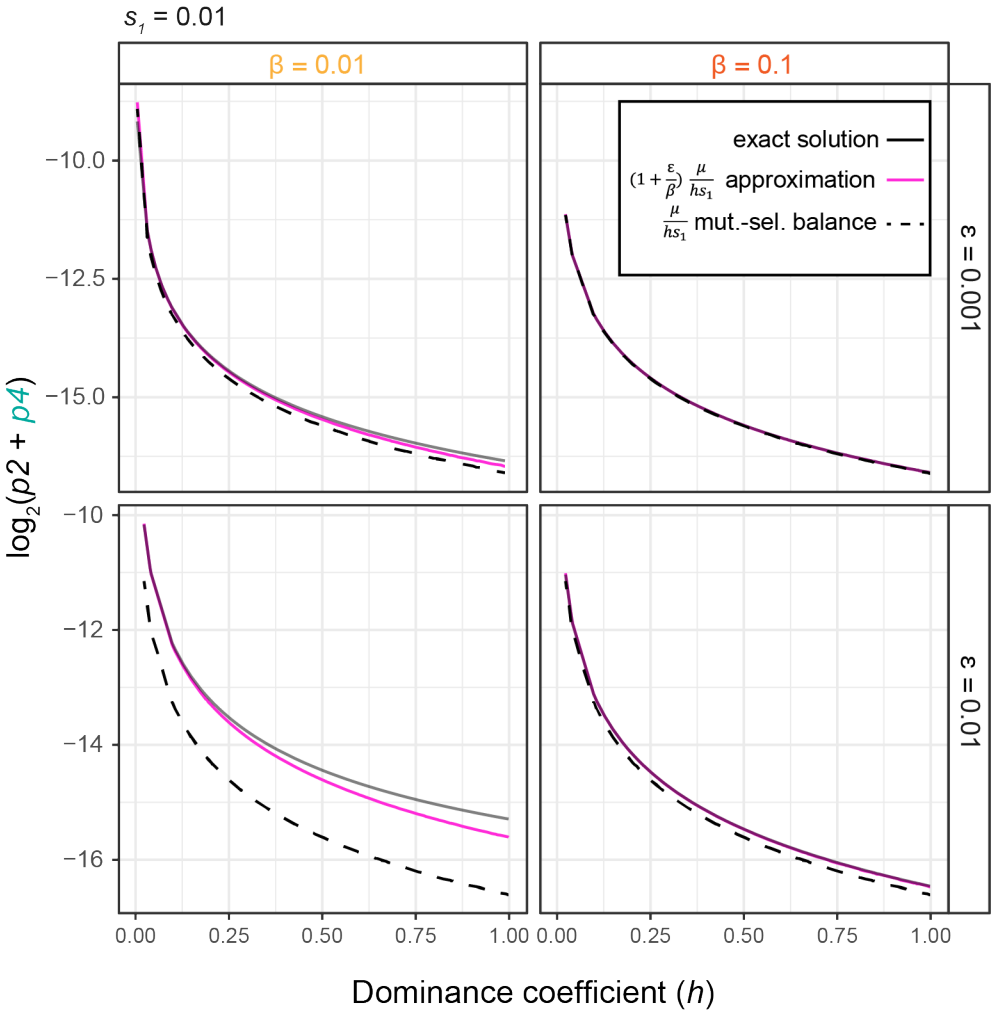


Figure 4: Numeric analysis suggests approximation with dominance coefficient accurately estimates epiallele equilibrium frequencies and deviation from mutation-selection balance

Numerical solutions were obtained by solving the system of equations when s_1=0.01, s_4,s_7,s_8= hs_1, β =0.01 or 0.1 and ε=0.001 or 0.01, for varying values of h. Exact numeric solutions (black) are shown alongside estimates based on the equation $(1+\frac{\varepsilon}{\beta})\frac{\mu}{hs_1}$ (pink) and estimates based on mutation-selection balance (dashed black).

Figure 5: Strong selection drives deviations from epimutation-back-epimutation equilibrium

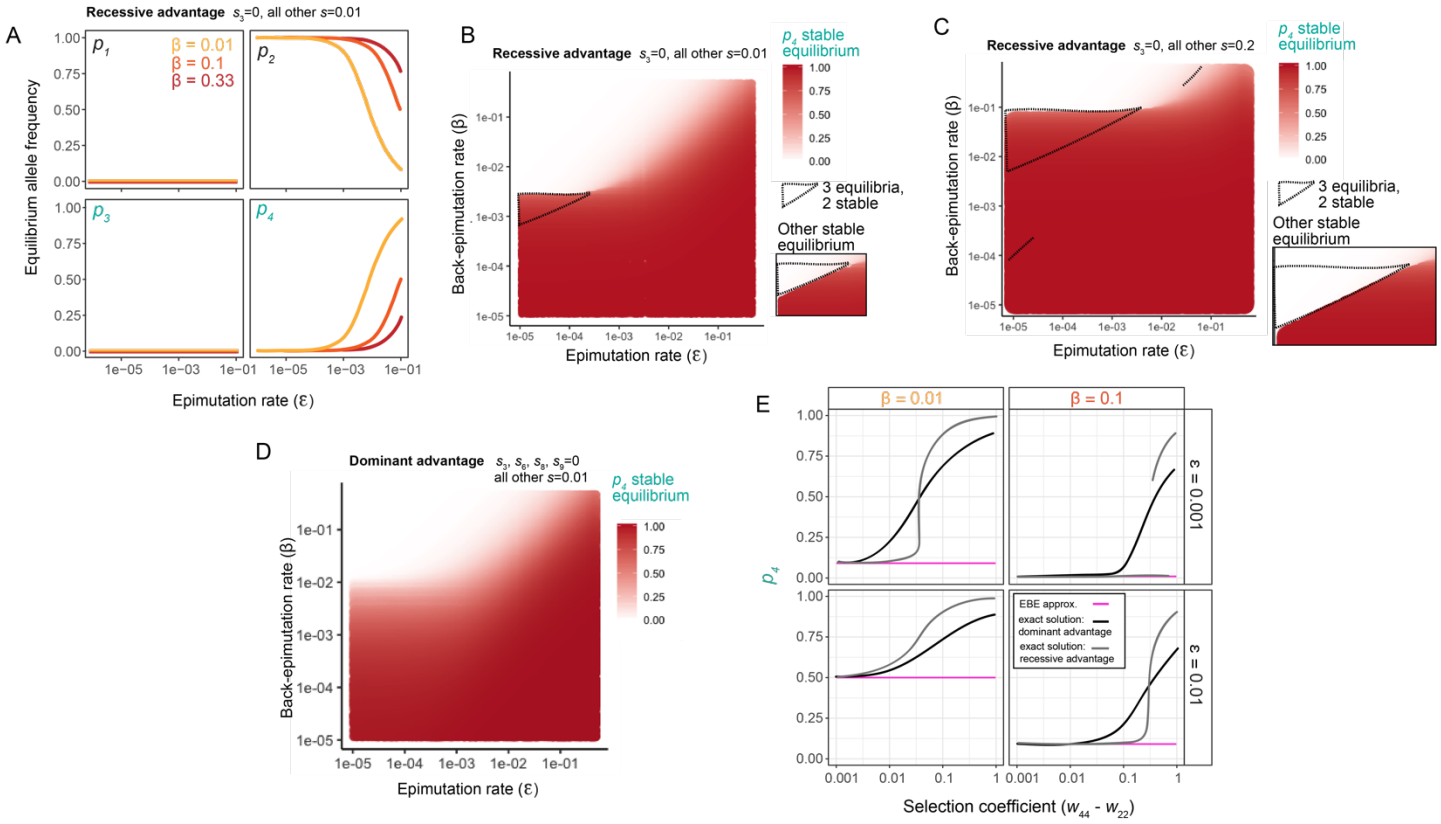


Figure 5: Strong selection drives deviations from epimutation-back-epimutation equilibrium

A. Stable equilibrium allele frequency for p_1 through p_4 when s_3=0 and other s=0.01 for back-epimutation rates of 0.01, 0.1, and 0.33.

B-D. Stable equilibria for p_4. In all cases, p_1 and p_3 are not maintained, and $p_2 = 1 - p_4$.

B. Stable equilibrium of p_4 when s_3=0 and other s=0.01 for a range of epimutation and back-epimutation rates. 3 equilibria are present within the dotted triangle (2 stable), and the other stable equilibrium is in the inset box.

C. Same as (B) but other s=0.2.

D. Stable equilibrium of p_4 when s_3, s_6, s_8, s_9=0 and other s=0.01 for a range of epimutation and back-epimutation rates.

E. Stable equilibrium of p_4 when either s_3=0 (recessive case) or s_3, s_6, s_8, s_9=0 (dominant case) and all other s vary between 0.001 and 1 for β =0.01 or 0.1 and ε=0.001 or 0.01.

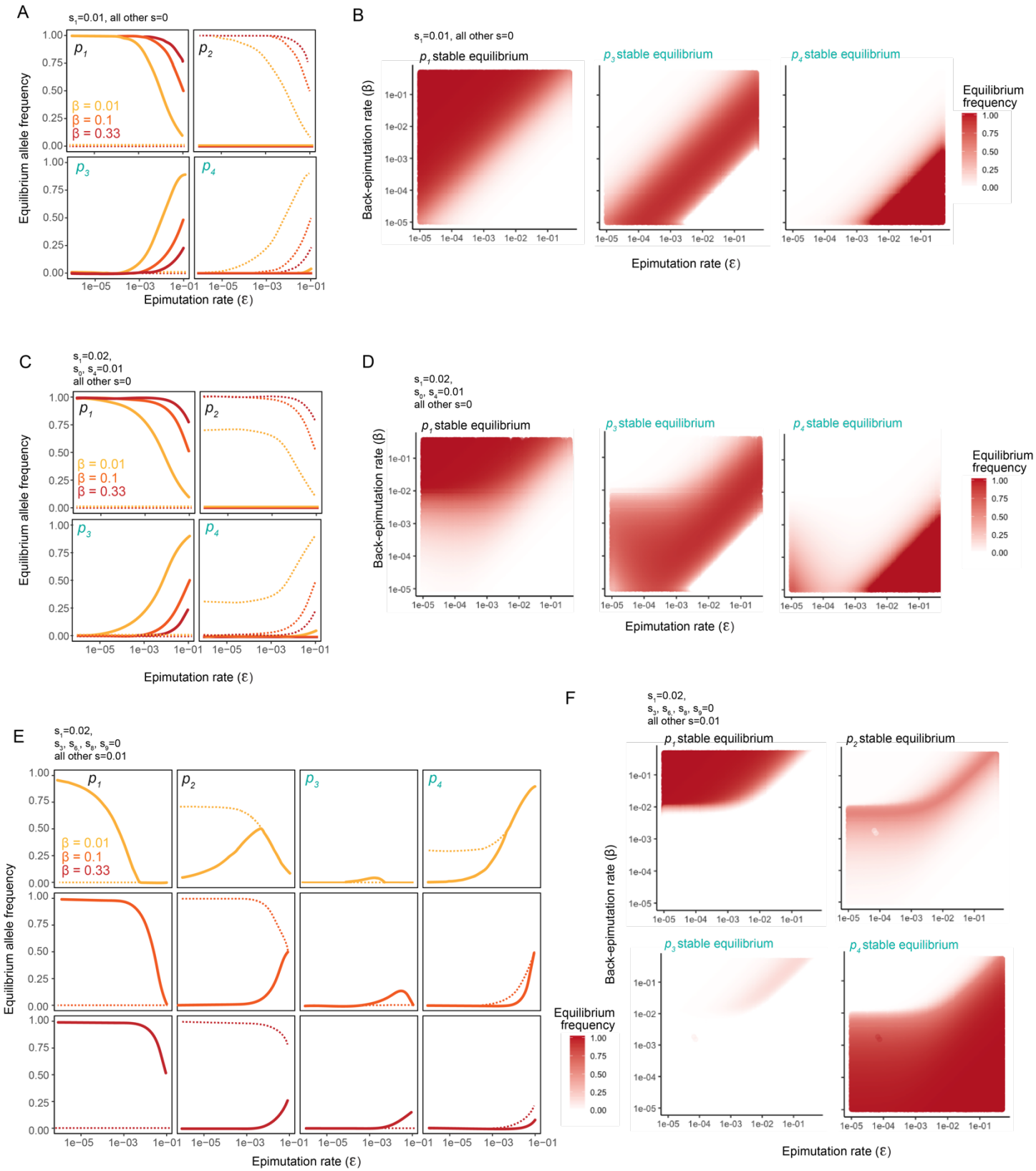


Figure 6: Epialleles compensating for deleterious recessive alleles drive deleterious alleles to high frequency

A. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) when s_1=0.01 and all other s=0, for back-epimutation rates of 0.01, 0.1, and 0.33.

B. Stable equilibrium when s_1=0.01 and all other s=0 for a range of epimutation and back-epimutation rates. p_2 is not shown because it is very close to 0 and not visible at the scale shown.

C. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) when s_1=0.02, s_0,s_4=0.01, and all other s=0 for back-epimutation rates of 0.01, 0.1, and 0.33.

D. Stable equilibrium when s_1=0.02, s_0,s_4=0.01, and all other s=0 for a range of epimutation and back-epimutation rates. p_2 is not shown because it is close to 0.

E. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) when s_1=0.02, s_3,s_6,s_8,s_9=0, and all other s=0.01 for back-epimutation rates of 0.01, 0.1, and 0.33.

F. Stable equilibrium when s_1=0.02, s_3,s_6,s_8,s_9=0, and all other s=0.01 for a range of epimutation and back-epimutation rates.

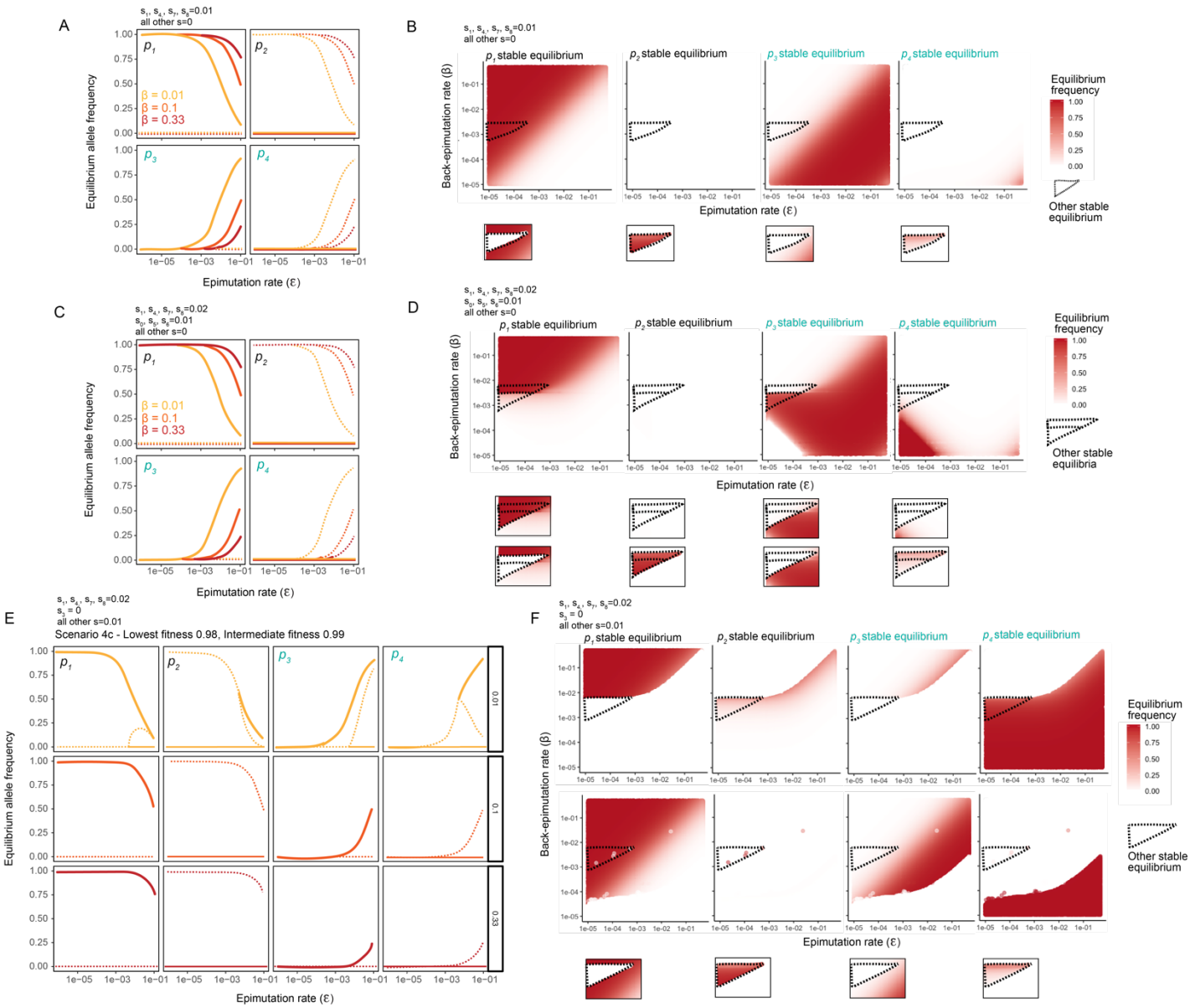


Figure 7: Epialleles compensating for deleterious dominant alleles drive deleterious alleles to high frequency

A. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) when s_1, s_4, s_7, s_8 = 0.01 and all other s = 0 for back-epimutation rates of 0.01, 0.1, and 0.33.

B. Stable equilibria when s_1, s_4, s_7, s_8 = 0.01 and all other s = 0 for a range of epimutation and back-epimutation rates. Within the dotted triangle, a second stable equilibrium is present.

C. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) when s_1, s_4, s_7, s_8 = 0.02, s_0, s_5, s_6 = 0.01, all other s = 0, and back-epimutation rates are 0.01, 0.1, and 0.33.

D. Stable equilibria when s_1, s_4, s_7, s_8 = 0.02, s_0, s_5, s_6 = 0.01, and all other s = 0 for a range of epimutation and back-epimutation rates. Within the smallest dotted triangle, 3 stable equilibria are present. Within the larger dotted triangle but excluding the smaller triangle, 2 stable equilibria are present.

E. Stable equilibria allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) when s_1, s_4, s_7, $s_8 = 0.02$, $s_3 = 0$, all other $s = 0.01$, and back-epimutation rates are 0.01, 0.1, and 0.33.

F. Stable equilibria when s_1, s_4, s_7, $s_8 = 0.02$, $s_3 = 0$, and all other $s = 0.01$ for a range of epimutation and back-epimutation rates. The full parameter space is shown for two stable equilibria (each row is a different equilibrium) because both are present throughout much of the space. A third stable equilibrium is found within the dotted triangle.

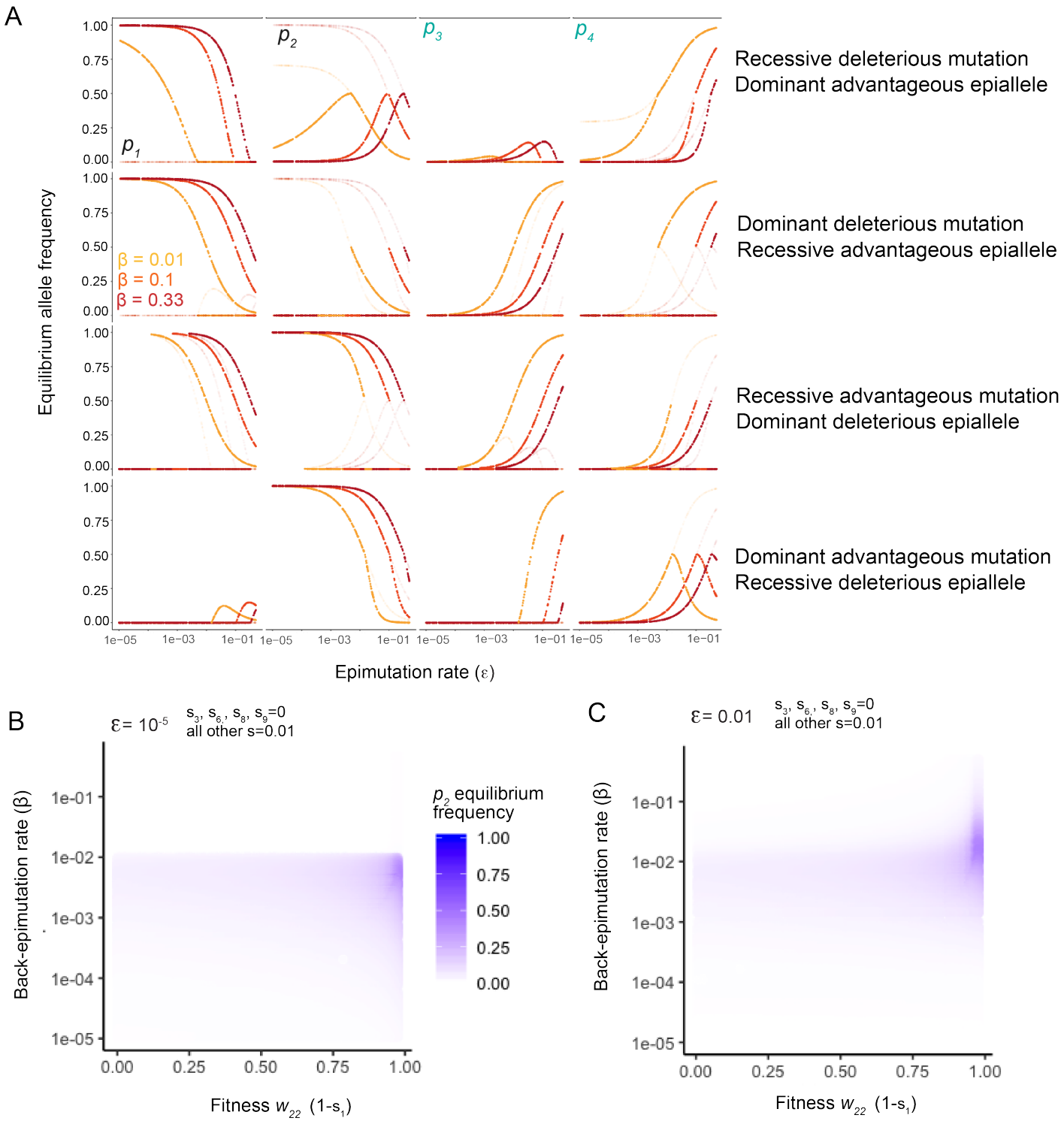


Figure 8: Deviations from mutation-selection balance depend on dominance status of mutations and compensatory epialleles

A. Equilibrium frequencies for p_1 through p_4. Opaque points represent stable equilibria, while faded points are unstable equilibria.

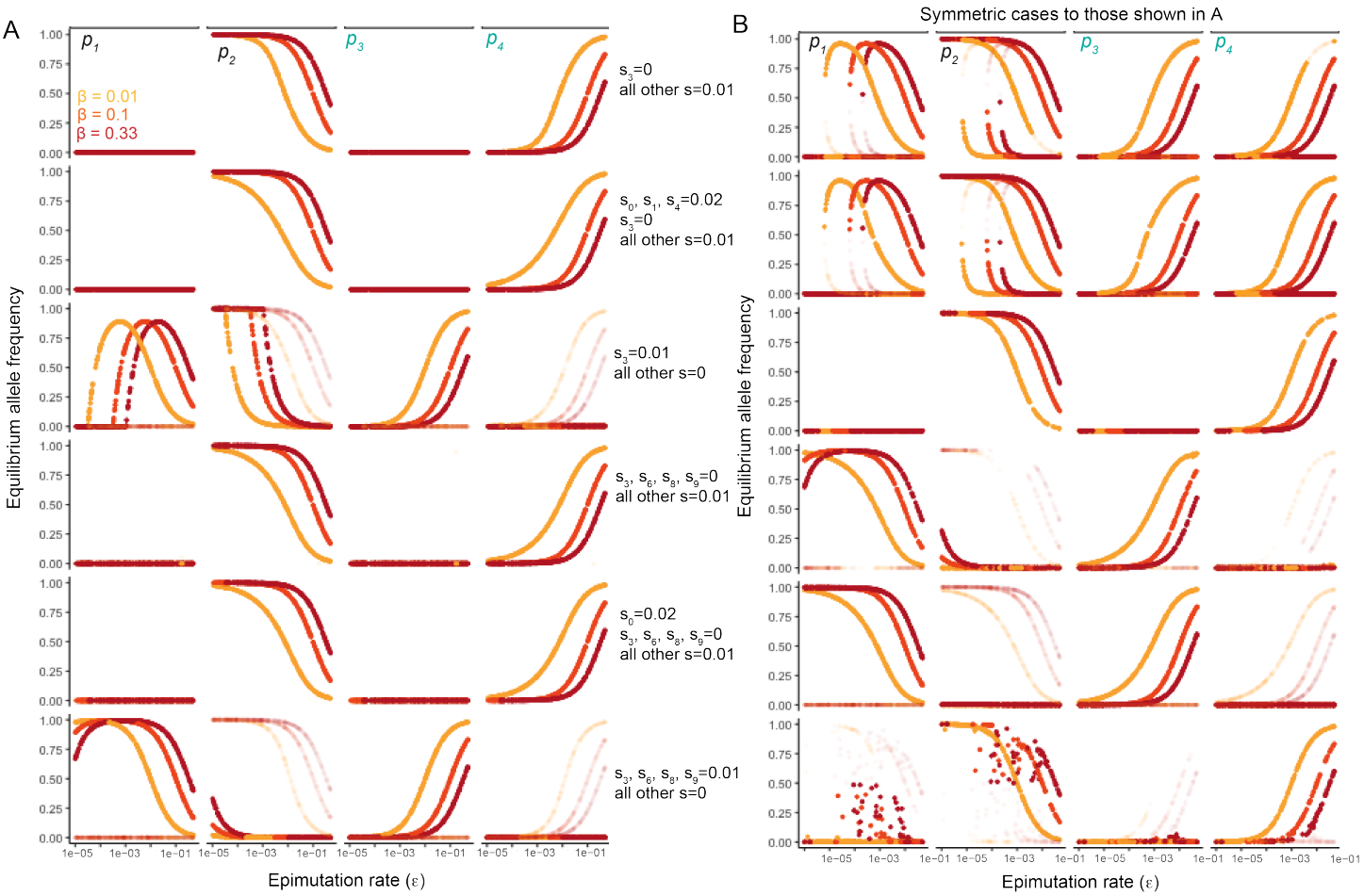
B-C. w_{22} (the lowest fitness) was allowed to take any value between 0 and 0.98 and back-epimutation rate was allowed to take any value between 10^{-5} and 0.5. The long-term equilibrium of p_2 (the frequency of a_2) is indicated by the shading of blue.

B. The epimutation rate is held constant at 10^{-5}.

C. The epimutation rate is held constant at 0.01.

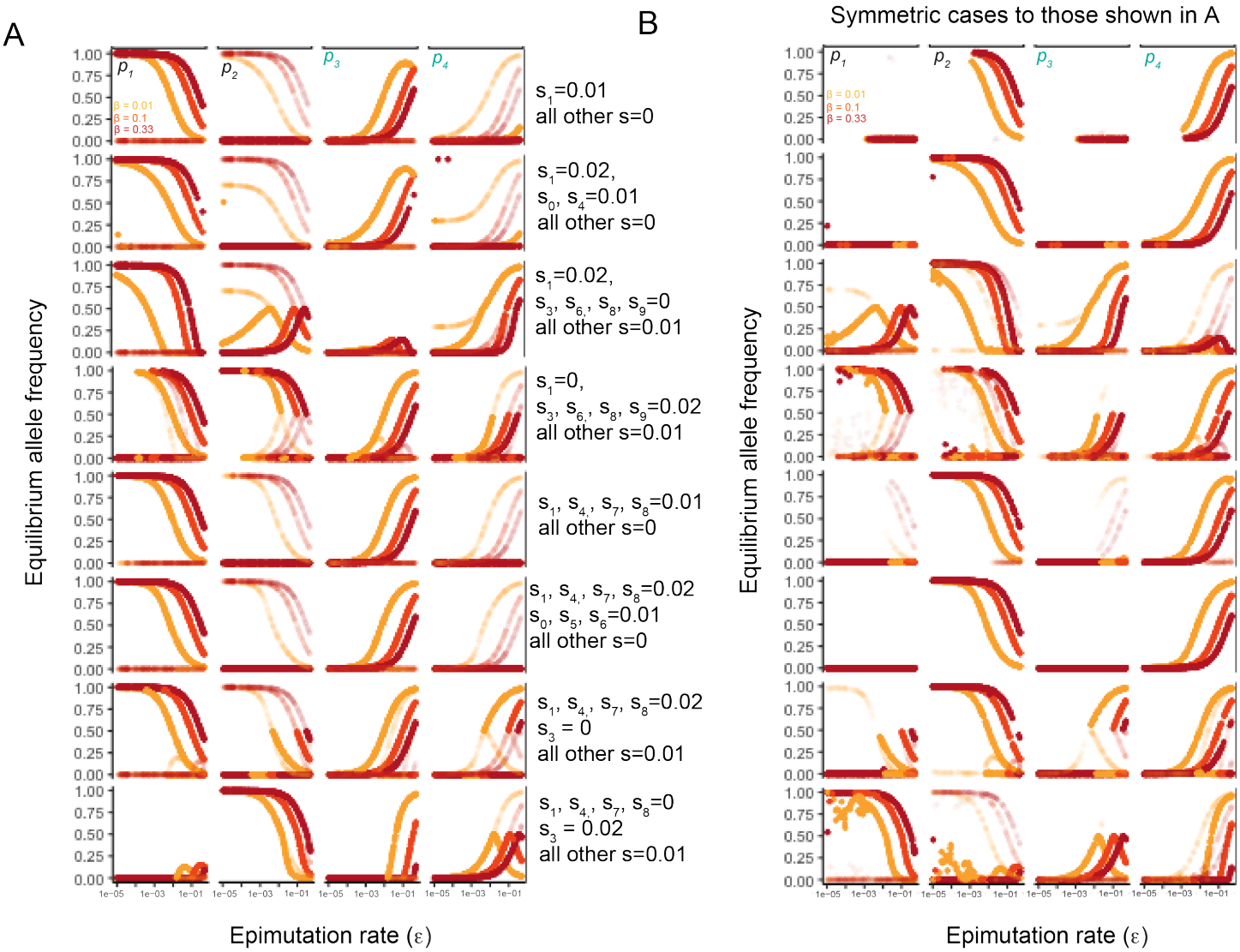
SUPPLEMENTARY FIGURES

Supplementary Figure 1: Exact solutions and symmetrical cases corresponding to Figure 4



Supplementary Figure 1: Exact solutions and symmetric cases corresponding to Figure 5. A-B. Equilibrium allele frequencies for p_1 through p_4. Stable equilibria are completely opaque points, while unstable equilibria are faded. A. From top to bottom facet, raw output for the cases shown in Figure 4 and referenced in the corresponding text with corresponding selection coefficients. B. From top to bottom facet, raw output for symmetric values to those shown A. Symmetric means that selection coefficients corresponding to A_1 and A_3 are swapped with a_2 and a_4, respectively.

Supplementary Figure 2: Exact solutions and symmetrical cases corresponding to Figures 5 and 6



Supplementary Figure 2: Exact solutions and symmetric cases corresponding to Figures 5 and 6. A-B. Equilibrium allele frequencies for p_1 through p_4. Stable equilibria are points with complete opacity, while unstable equilibria are faded. A. From top to bottom facet, raw output for cases shown in Figures 5 and 6 and referenced in the text, with corresponding selection coefficients. B. From top to bottom facet, raw output for symmetric values to those in A. Symmetric means that selection coefficients corresponding to A_1 and A_3 are swapped with a_2 and a_4, respectively.

DATA AVAILABILITY STATEMENT

Mathematica scripts for the full parameter space and focal back-epimutation rates, in addition to csv files and R scripts used to generate plots, are available on Github (https://github.com/amykwebster/2022_EpigeneticModels). Upon publication, the final version on Github will be used to generate a unique DOI on Zenodo.

ACKNOWLEDGEMENTS

We thank members of the Phillips lab and anonymous reviewers for critical feedback. Funding for this work was provided by NIH grant number R35GM131838 to PCP and F32GM146402 to AKW.

REFERENCES

Almouzni G, Cedar H. 2016. Maintenance of Epigenetic Information. *Cold Spring Harb Perspect Biol* **8**.

Baugh LR, Day T. 2020. Nongenetic inheritance and multigenerational plasticity in the nematode C. elegans. *Elife* **9**.

Baxter E, Windloch K, Gannon F, Lee JS. 2014. Epigenetic regulation in cancer progression. *Cell Biosci* **4**: 45.

Beck D, Ben Maamar M, Skinner MK. 2021. Integration of sperm ncRNA-directed DNA methylation and DNA methylation-directed histone retention in epigenetic transgenerational inheritance. *Epigenetics & chromatin* **14**: 1-14.

Becker C, Hagmann J, Muller J, Koenig D, Stegle O, Borgwardt K, Weigel D. 2011. Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* **480**: 245-249.

Beltran T, Shahrezaei V, Katju V, Sarkies P. 2020. Epimutations driven by small RNAs arise frequently but most have limited duration in Caenorhabditis elegans. *Nat Ecol Evol* **4**: 1539-1548.

Burton NO, Greer EL. 2022. Multigenerational epigenetic inheritance: Transmitting information across generations. *Semin Cell Dev Biol* **127**: 121-132.

Camacho J, Truong L, Kurt Z, Chen YW, Morselli M, Gutierrez G, Pellegrini M, Yang X, Allard P. 2018. The Memory of Environmental Chemical Exposure in C. elegans Is Dependent on the Jumonji Demethylases jmjd-2 and jmjd-3/utx-1. *Cell Rep* **23**: 2392-2404.

Carja O, MacIsaac JL, Mah SM, Henn BM, Kobor MS, Feldman MW, Fraser HB. 2017. Worldwide patterns of human epigenetic variation. *Nat Ecol Evol* **1**: 1577-1583.

Casas E, Vavouri T. 2020. Mechanisms of epigenetic inheritance of variable traits through the germline. *Reproduction* **159**: R251-R263.

Chen X, Chen Z, Chen H, Su Z, Yang J, Lin F, Shi S, He X. 2012. Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. *Science* **335**: 1235-1238.

Ciabrelli F, Comoglio F, Fellous S, Bonev B, Ninova M, Szabo Q, Xuéreb A, Klopp C, Aravin A, Paro R. 2017. Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. *Nature genetics* **49**: 876-886.

Cortijo S, Wardenaar R, Colome-Tatche M, Gilly A, Etcheverry M, Labadie K, Caillieux E, Hospital F, Aury JM, Wincker P et al. 2014. Mapping the epigenetic basis of complex traits. *Science* **343**: 1145-1148.

Daxinger L, Whitelaw E. 2012. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* **13**: 153-162.

Day T, Bonduriansky R. 2011. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. *Am Nat* **178**: E18-36.

Evans AD, Fairbanks RA, Schmidt P, Levine MT. 2023. An epigenetic basis of adaptive plasticity in Drosophila melanogaster. *bioRxiv* doi:10.1101/2022.10.11.511590: 2022.2010.2011.511590.

Feinberg AP. 2004. The epigenetics of cancer etiology. *Semin Cancer Biol* **14**: 427-432.

Furci L, Jain R, Stassen J, Berkowitz O, Whelan J, Roquis D, Baillet V, Colot V, Johannes F, Ton J. 2019a. Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in Arabidopsis. *Elife* **8**.

Furci L, Jain R, Stassen J, Berkowitz O, Whelan J, Roquis D, Baillet V, Colot V, Johannes F, Ton J. 2019b. Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in Arabidopsis. *Elife* **8**: e40655.

Geoghegan JL, Spencer HG. 2012. Population-epigenetic models of selection. *Theor Popul Biol* **81**: 232-242.

Geoghegan JL, Spencer HG. 2013. Exploring epiallele stability in a population-epigenetic model. *Theor Popul Biol* **83**: 136-144.

Gravina S, Ganapathi S, Vijg J. 2015. Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. *Nucleic Acids Res* **43**: e93.

Grewal SI, Klar AJ. 1996. Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis. *Cell* **86**: 95-101.

Hadeler KP. 1981. Stable Polymorphisms in a Selection Model with Mutation. *SIAM Journal on Applied Mathematics* **41**: 1-7.

Handy DE, Castro R, Loscalzo J. 2011. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. *Circulation* **123**: 2145-2156.

Herman JJ, Spencer HG, Donohue K, Sultan SE. 2014. How stable 'should' epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution* **68**: 632-643.

Horsthemke B. 2006. Epimutations in human disease. *Curr Top Microbiol Immunol* **310**: 45-59.
Horsthemke B. 2018. A critical view on transgenerational epigenetic inheritance in humans. *Nat Commun* **9**: 2973.
Jablonka E, Raz G. 2009. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly review of biology* **84**: 131-176.
Jacobsen SE, Meyerowitz EM. 1997. Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. *Science* **277**: 1100-1103.
Jobson MA, Jordan JM, Sandrof MA, Hibshman JD, Lennox AL, Baugh LR. 2015. Transgenerational Effects of Early Life Starvation on Growth, Reproduction, and Stress Resistance in Caenorhabditis elegans. *Genetics* **201**: 201-212.
Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, Bulski A, Albuisson J, Heredia F, Audigier P et al. 2009. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet* **5**: e1000530.
Jones PA, Baylin SB. 2002. The fundamental role of epigenetic events in cancer. *Nat Rev Genet* **3**: 415-428.
Kakutani T. 2002. Epi-alleles in plants: inheritance of epigenetic information over generations. *Plant Cell Physiol* **43**: 1106-1111.
Kaletsky R, Moore RS, Vrla GD, Parsons LR, Gitai Z, Murphy CT. 2020. C. elegans interprets bacterial non-coding RNAs to learn pathogenic avoidance. *Nature* **586**: 445-451.
Karlin S. 1980. The number of stable equilibria for the classical one-locus multiallele selection model. *J Math Biol* **9**: 189-192.
Kelly SA, Panhuis TM, Stoehr AM. 2012. Phenotypic plasticity: molecular mechanisms and adaptive significance. *Compr Physiol* **2**: 1417-1439.
Kelly WG. 2014. Transgenerational epigenetics in the germline cycle of Caenorhabditis elegans. *Epigenetics Chromatin* **7**: 6.
Kimura M. 1956. Rules for Testing Stability of a Selective Polymorphism. *Proc Natl Acad Sci U S A* **42**: 336-340.
Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press.
Kirkpatrick M, Lande R. 1989. The Evolution of Maternal Characters. *Evolution* **43**: 485-503.
Klosin A, Casas E, Hidalgo-Carcedo C, Vavouri T, Lehner B. 2017. Transgenerational transmission of environmental information in C. elegans. *Science* **356**: 320-323.
Kooke R, Johannes F, Wardenaar R, Becker F, Etcheverry M, Colot V, Vreugdenhil D, Keurentjes JJ. 2015. Epigenetic basis of morphological variation and phenotypic plasticity in Arabidopsis thaliana. *Plant Cell* **27**: 337-348.
Le Goff A, Allard P, Landecker H. 2021. Heritable changeability: Epimutation and the legacy of negative definition in epigenetic concepts. *Stud Hist Philos Sci* **86**: 35-46.
Lewontin RC. 1974. *The genetic basis of evolutionary change*. Columbia University Press New York.
Li C, Luscombe NM. 2020. Nucleosome positioning stability is a modulator of germline mutation rate variation across the human genome. *Nat Commun* **11**: 1363.
Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li GM. 2013. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. *Cell* **153**: 590-600.
Liegard B, Baillet V, Etcheverry M, Joseph E, Lariagon C, Lemoine J, Evrard A, Colot V, Gravot A, Manzanares-Dauleux MJ et al. 2019. Quantitative resistance to clubroot infection mediated by transgenerational epigenetic variation in Arabidopsis. *New Phytol* **222**: 468-479.
Mandel SPH. 1959. The stability of a multiple allelic system. *Heredity* **13**: 289-302.
Marion SB, Noor MAF. 2022. Interrogating the Roles of Mutation-Selection Balance, Heterozygote Advantage, and Linked Selection in Maintaining Recessive Lethal Variation in Natural Populations. *Annu Rev Anim Biosci* doi:10.1146/annurev-animal-050422-092520.
Mehrmohamadi M, Sepehri MH, Nazer N, Norouzi MR. 2021. A Comparative Overview of Epigenomic Profiling Methods. *Front Cell Dev Biol* **9**: 714687.
Mishra A, Hawkins RD. 2017. Three-dimensional genome architecture and emerging technologies: looping in disease. *Genome Med* **9**: 87.
Monroe JG, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, Klein M, Hildebrandt J, Neumann M, Kliebenstein D et al. 2022. Mutation bias reflects natural selection in Arabidopsis thaliana. *Nature* **602**: 101-105.
Moore RS, Kaletsky R, Murphy CT. 2019. Piwi/PRG-1 Argonaute and TGF-beta Mediate Transgenerational Learned Pathogenic Avoidance. *Cell* **177**: 1827-1841 e1812.

Mousseau TA, Fox CW. 1998. *Maternal effects as adaptations*. Oxford University Press.
Öst A, Lempradl A, Casas E, Weigert M, Tiko T, Deniz M, Pantano L, Boenisch U, Itskov PM, Stoeckius M. 2014. Paternal diet defines offspring chromatin state and intergenerational obesity. *Cell* **159**: 1352-1364.
Richards CL, Verhoeven KJ, Bossdorf O. 2012. Evolutionary significance of epigenetic variation. In *Plant Genome Diversity Volume 1*, pp. 257-274. Springer.
Ruden DM, Lu X. 2008. Hsp90 affecting chromatin remodeling might explain transgenerational epigenetic inheritance in Drosophila. *Curr Genomics* **9**: 500-508.
Sarkies P. 2020. Molecular mechanisms of epigenetic inheritance: Possible evolutionary implications. *Semin Cell Dev Biol* **97**: 106-115.
Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR. 2011. Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**: 369-373.
Schott D, Yanai I, Hunter CP. 2014. Natural RNA interference directs a heritable response to the environment. *Scientific reports* **4**: 1-10.
Schuster-Bockler B, Lehner B. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**: 504-507.
Seroussi U, Li C, Sundby AE, Lee TL, Claycomb JM, Saltzman AL. 2022. Mechanisms of epigenetic regulation by C. elegans nuclear RNA interference pathways. *Semin Cell Dev Biol* **127**: 142-154.
Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, Cohen T, Xia J, Suderman M, Hallett M. 2015. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350**: aab2006.
Silveira AB, Trontin C, Cortijo S, Barau J, Del Bem LEV, Loudet O, Colot V, Vincentz M. 2013. Extensive natural epigenetic variation at a de novo originated gene. *PLoS Genetics* **9**: e1003437.
Tallis GM. 1966. Equilibria under Selection for k Alleles. *Biometrics* **22**: 121-127.
van der Graaf A, Wardenaar R, Neumann DA, Taudt A, Shaw RG, Jansen RC, Schmitz RJ, Colome-Tatche M, Johannes F. 2015. Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proc Natl Acad Sci U S A* **112**: 6676-6681.
Varga Z, Zubiri JA. 1993. Equilibrium and stability of a Fisher type selection-mutation model. *Abstracta Botanica* **17**: 241-244.
Vogt MC, Hobert O. 2023. Starvation-induced changes in somatic insulin/IGF-1R signaling drive metabolic programming across generations. *Sci Adv* **9**: eade1817.
Waddington C H. 1942. The epigenotype. *Endeavour* **1**: 18-20.
Webster AK, Jordan JM, Hibshman JD, Chitrakar R, Baugh LR. 2018. Transgenerational Effects of Extended Dauer Diapause on Starvation Survival and Gene Expression Plasticity in Caenorhabditis elegans. *Genetics* **210**: 263-274.
Wei JW, Huang K, Yang C, Kang CS. 2017. Non-coding RNAs as regulators in epigenetics (Review). *Oncol Rep* **37**: 3-9.
Werner MS, Loschko T, King T, Reich S, Theska T, Franz-Wachtel M, Macek B, Sommer RJ. 2023. Histone 4 lysine 5/12 acetylation enables developmental plasticity of Pristionchus mouth form. *Nat Commun* **14**: 2095.
Wright S. 1931. Evolution in Mendelian Populations. *Genetics* **16**: 97-159.
Yu R, Wang X, Moazed D. 2018. Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. *Nature* **558**: 615-619.