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

ABSTRACT

Understanding how genetic and phenotypic variation is maintained is a major problem 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 in a manner that depends primarily on the epimutation and back-epimutation rates, and for certain fitness parameters, alleles with high fitness can facilitate the maintenance of recessive deleterious alleles that would otherwise be purged from the population. This demonstrates 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, and epigenetic modifications are likely candidates. This has 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 maintain a core tenet that epigenetics is something distinct from the DNA sequence that 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 (Jones and Baylin 2002; Feinberg 2004). Further, epigenetic modifications also play a major role in phenotypic plasticity in response to particular environmental conditions (Kelly et al. 2012), and mitosis can maintain the plastic response to the environment (Duncan et al. 2014). In general, however, the epigenetic state of an organism's cells, even if it is heritable mitotically, 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). Notably, because *D. melanogaster* and *C. elegans* lack DNA cytosine methylation, mechanisms typically involve histone modifications and small RNA.

The persistence of epigenetic modifications across generations that influences phenotype raises the possibility they 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 the environment 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* for the purpose of analyzing 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 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 equilibrium in epimutation/back-epimutation balance. This means that within a genotype, multiple stable epialleles can coexist. Across genotypes, an advantageous epiallele can, for certain fitness parameters, facilitate the maintenance of a deleterious genetic allele, or vice versa, and this can also facilitate the maintenance of multiple genetic alleles simultaneously. 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, with the assumption of no back-mutation. The epimutation rate, ε, is the rate of epimutation of A_1 or a_2 to A_3 or a_4, respectively. In this case, consistent with the limited inheritance potential of epimutations, the back-epimutation rate (β) models the rate at which A_3 or a_4 revert back to A_1 or 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.

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 consider a mutation rate of 10^{-7}, epimutation rates ranging from 10^{-5} to 0.5, and back-epimutation rates of 0.33, 0.1, and 0.01 for our preliminary analyses and examine potential circumstances that allow epialleles to remain in long-term stable equilibrium. The back-epimutation rates used are meant to represent whole allele rates rather nucleotide-specific rates. 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 situation. 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 ignore the possibility 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. Subsequently, for completeness, we analyze a larger parameter space ranging from 10^{-5} to 0.5 for both back-epimutation and epimutation rates, but we note that very low back-epimutation rates and very high epimutation rates are less likely to occur in nature.

Epimutations correspond to changes in epigenetic modifications, potentially impacting gene expression to affect fitness. Similarly, mutations can also impact gene expression to affect fitness. As a baseline, we determine the dynamics when selection coefficients are the same for all genotypes. We then interrogate four main scenarios for which epimutations and mutations impact gene expression and fitness in different ways. To analyze our model, we consider scenarios for which 1) epimutations and mutations both act to silence gene expression (Scenario 1), 2) both act to promote gene expression (Scenario 2), or 3) act in opposite directions, with epimutations compensating for a mutation (Scenarios 3 and 4, Figure 3). We consider how expression might map onto fitness when there is either a minimum or maximum threshold of expression for optimal gene function (Figure 3) and focus primarily on situations in which epialleles offer a selective advantage, but also consider some examples of when they are deleterious.

ANALYSIS

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. Frequency equilibria were determined in *Mathematica* using the NSolve function by setting $p_x(t+1) = p_x(t)$ in Eqs (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 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:

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

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.

When calculating equilibria for Scenarios 1 and 2, we set all fitness values equal to 1, set $\frac{dp_2}{dt} = 0$, and set $p_1 = 0$ and $p_3 = 0$, because these were not maintained for these scenarios and solved p_2 in terms of ε and β. For Scenario 3, we performed an analogous calculation to determine that the relationship between A_1 and A_3 which yielded a nearly identical approximation. Specifically, we set all fitness values equal to 1, set $\frac{dp_3}{dt} = 0$, and set $p_2 = 0$ and $p_1 = 1 - p_3$. We then solved for p_3 in terms of ε and β. Mutation is present in this scenario but only minimally affects the equilibrium frequency. When mutation is ignored, the equilibrium frequency of p_3 solved here is exactly that of p_4 in equation 8.

For all 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 w_{11}, w_{22}, w_{33}, w_{44}, w_{12}, w_{13}, w_{14}, w_{23}, w_{24}, w_{34}, according to the specific scenario being analyzed as indicated in Figure 3. When fitness values were not one of these, it is indicated on the figure. 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 1-3. To analyze the full parameter space of epimutation and back-epimutation rates (*i.e.,* the shaded red graphs in Figures 4-6), for each scenario, 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 2B, and Supplementary 3B.

RESULTS

Concordant effects of mutation and epimutation facilitate the maintenance of epigenetic variation

In our neutral scenario, we find that epialleles within a genotype, a_2 and a_4, are maintained in a manner that depends on the epimutation and back-epimutation rate (Figure 3, Scenario 0, Supplementary Figure 1). This highlights that it is possible to maintain epigenetic variation in the absence of fitness differences, given sufficiently high epimutation rate and low back-epimutation rate. Moving away from neutrality, when mutations and epimutations generate fitness differences via silencing of gene expression (Figure 3, Scenario 1), we find a single stable long-term equilibrium across the range of epimutation rates in each case, with both a_2 and a_4 being maintained. At low epimutation rates, a_2 is primarily maintained, but as epimutation rate increases and back-epimutation rate decreases, we see increased maintenance of a_4 (Figure 4A). Like the neutral scenario, only epigenetic variation is maintained. Analyzing the full parameter space reveals that these patterns largely hold; that is, only a_2 and a_4 are stably maintained, and the frequency of a_4 increases with higher epimutation rates and lower back-epimutation rates. Throughout most of the parameter space, there is a single stable equilibrium, but there is a small region for which there are two stable equilibria (Figure 4B). These patterns are also largely hold when fitness is varied; stable equilibria only maintain a_2 and a_4, and there is a region for which two stable equilibria are present, though again, most sets of parameters yield only a single, stable equilibrium (Figure 4C). Similarly, variability is maintained under different silencing affects (Scenario 1b; Figure 4D) and when the epiallele is deleterious (Scenario 1c, Figure 3, and Supplementary Figure 2).

We then analyze the case in which mutations and epimutations act to increase gene expression when activation affects fitness (Figure 3, Scenario 2) by considering a model in which being above the threshold confers a certain fitness level (Figure 3, Scenarios 2a and 2b). This generates a dominance-like effect in many cases (e.g., in Scenario 1 for which silencing was beneficial, a_4a_4 alone had the highest fitness, while in Scenario 2, for which activation is beneficial, all four diploids that include a_4 have the same high fitness).

Despite differences in gene function, the dynamics for activation are very similar to those seen for silencing. We again only find long-term stable equilibria for a_2 and a_4, in a manner that depends on epimutation and back-epimutation (Scenarios 2a and 2b; Figure 4E-F). Taken together, we find that when mutations and epimutations confer concordant changes in gene expression (either silencing or activation) that leads to epigenetic-dependent differences in fitness, multiple epialleles but not genetic alleles can be maintained. For the deleterious activation scenario (Figure 3, Scenario 2c, Supplementary Figure 2), and we gain find that epigenetic but not genetic variation is maintained.

Because we observed relatively simple dynamics that largely depended on the epimutation and back-epimutation rate, we approximate the relationship between the long-term equilibrium frequency of the reversible epiallele (in this case, a_4) and the epimutation and back-epimutation rate when the selection coefficient is small. Since we know that epialleles within one genotype are not maintained (in this case, A_1 and A_3), we can set p_1 and p_3 equal to 0, and solve $\frac{dp_2}{dt} = 0$ for p_2 when there are no fitness differences (all fitness parameters are equal to 1). This results in the following relationship for the long-term equilibrium of p_4:

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

Equation 8: Approximation for the long-term equilibrium frequency of an epiallele that arises at frequency ε and reverts back its baseline state at frequency β each generation. This is the directly analogous usual equilibrium derived for models with forward and back genetic mutation (Wright 1931).

Fitness effects only minimally affect the long-term frequencies of a_2 and a_4, as seen in Figure 4. Therefore, in estimating the propensity for maintenance of epigenetic variation, the main considerations are the frequency at which new epialleles are generated (ε) and lost (β).

We also analyzed the symmetric cases for Scenarios 1 and 2, for which fitness for genotypes including A_1 and A_3 alleles were swapped with fitness associated with genotypes including a_2 and a_4 alleles, respectively. Due to the asymmetry in the genetic mutation rate, we see that at low epimutation rates, it is possible for maintenance of A_1 and a_2 to occur when A_3A_3 has a fitness advantage (the symmetric case for Scenarios 1a and 1b) or when a_4a_4 has a fitness disadvantage (Scenario 1c). However, for most epimutation

rates, maintenance of either a_2 and a_4, or A_1 and A_3 occurs. In either case, the maintenance of epigenetic variation occurs, and this is primarily driven by the dynamics described in Equation 8 rather than fitness.

Discordant effects of mutation and epimutation create the opportunity for maintenance of genetic variation

Given the results for concordant changes in gene expression, we next analyzed instances for which mutation and epimutation confer discordant changes in gene expression and fitness. In Scenario 3a-c (Figure 3), a mutation reduces gene expression and fitness, but epimutation compensates for the deleterious mutation and thereby increases fitness. When a_2a_2 has low fitness with all other combinations of alleles have the same high fitness (Scenario 3a), we see dynamics similar to the first two scenarios in that maintenance only occurs within a single genotype, with only a single stable equilibrium is present, resulting in primarily the maintenance of A_1 and A_3 without the maintenance of a_2 and a_4 (Figure 5A). The maintained frequency of epiallele A_3 increases as the epimutation rate increases and back-epimutation rate decreases, similar to a_2 in Scenario 1. In the symmetrical scenario (with fitness values for genotypes that include A_1 and A_3 alleles swapped with those that include a_2 and a_4), we see the same outcome, but with maintenance of only a_2 and a_4, as expected. In either case, only epigenetic variation within a genotype is maintained for the focal back-epimutation rates.

While the predominant outcome is the maintenance of multiple epialleles within a single genotype, it is possible to maintain both genetic alleles as well. Specifically, when the back-epimutation rate is low and the epimutation rate is high, a_4 is maintained at high frequency (Figure 5A-B). Thus, it is possible to maintain genetic variation at the boundary of parameter space for which both A_3 and a_4 are present, but only under assumptions for epimutation and back-epimutation rates that are unlikely to occur in nature.

Moving to the next threshold model (Scenario 3b), in which a_2a_2 has the lowest fitness, any genotype including A_3 or a_4 alleles have the highest fitness, and other genotypes have intermediate fitness yields results that are qualitatively similar to Scenario 3a, with regions in which A_3 and a_4 are maintained. Here, the parameter space that allows multiple equilibria is larger than for Scenario 3a, as it possible to attain when the epimutation rate is low (Figure 5C-D). Nevertheless, we again do not see any long-term maintenance of a_2 for any parameter set.

Next, when a_2a_2 has the lowest fitness, any genotype including the a_4 allele has the highest fitness, and all other genotypes have intermediate fitness, we find a very unique outcome: the maintenance of the deleterious genetic allele alongside the favored allele at appreciable frequency. Specifically, while a_2 is not maintained when the back-epimutation rate is low (*i.e.,* at rates more consistent with mutation than observed for transient epimutations), a_2 can generally be maintained for all epimutation rates in a manner dependent on the back-epimutation rate (Figure 5E-F). Specifically, a_2 can be maintained when the back-epimutation rate is high (greater than 0.1), so long as the epimutation rate is moderately high as well (Figure 5F). Previously we found that only epigenetic variation could be maintained, while here we find the maintenance of both epigenetic and genetic variation. This particularly occurs at the boundary of parameter space that allows a_2 to be maintained, such that, as long as a_2 is maintained, it is possible to have long-term maintenance of all four epialleles at a single locus simultaneously. For Scenarios 3a-c, the symmetric fitness comparisons resulted in the similar dynamics – maintenance of epigenetic variation for Scenarios 3a-b and maintenance of genetic variation, including the deleterious genetic allele, for Scenario 3c.

We next considered the possibility that mutation causes deleterious activation and epimutation causes advantageous silencing (Scenario 4a-c). Again, we consider different thresholds for silencing that are beneficial for fitness. When any genotype containing the a_2 allele has low fitness and all other genotypes have high fitness (Scenario 4a), we find that the maintenance of A_1 and A_3 is the predominant patten, with two exceptions (Figure 6B). First, there is a small region for which two stable equilibria are present, with the other stable equilibrium facilitating maintenance of a_2 and a_4. Second, as in previous scenarios, we also find that maintenance of a_4 can occur when epimutation rates are high and back-epimutation rates are low. However, this small region is the only region for which genetic variation can be maintained and is not a scenario likely to occur in nature. For the case when any genotype containing the a_2 allele has the lowest fitness, remaining genotypes containing A_1 have intermediate fitness, and any remaining genotypes containing A_3 or a_4 have the highest fitness (Scenario 4b), there is a small region for which two or three equilibria are present, though one stable equilibrium occurs throughout most of the parameter space (Figure 6C-D). When more than one equilibrium is present, it is possible for a_2 to be maintained at appreciable frequencies along with a_4. Otherwise, it is not maintained. As in Scenario 4a, maintenance of A_1 and A_3 is the dominant pattern, although maintenance of a_2 and a_4 is also possible. In either case, genetic variation is not maintained. Maintenance of

genetic variation is possible under very low back-epimutation rates, but this makes epimutations behave essentially like genetic mutations, which is unlikely to be the case in nature.

The final threshold considered for which epialleles have a fitness advantage is Scenario 4c, in which any genotype that includes a_2 has the lowest fitness, a_4a_4 has the highest fitness, and all other genotypes have intermediate fitness. Note that this has parallels with Scenario 3c. In Scenario 3c, a_2a_2 had a recessive disadvantage (reduced expression when it is beneficial to be expressed) while any genotype including a_4 had a dominant advantage (high expression from one allele compensated for the other). Here, a_2 has a dominant disadvantage (high expression from one allele when it is beneficial to be silenced) while a_4a_4 has a recessive advantage (both alleles are silenced). Like Scenario 3c, it is possible for the deleterious allele a_2 to be maintained in long-term equilibrium, when it would not be maintained without the incorporation of epigenetic information into the model. We find that there are often two stable equilibria and as many as three under certain conditions (Figure 6E-F). Like Scenario 3c, there is a boundary for which a_2 is maintained, but it is a sharper boundary such that back-epimutation rates must be low or epimutation rates high in order for maintenance to occur. Because of this sharp boundary, A_1 and a_2 are not maintained simultaneously as in 3c.

Because we observed that the maintenance of a deleterious allele in 3c and 4c is dependent on the fitness advantage of epialleles, we wanted to assess the dynamics under the conditions of the epialleles having fitness disadvantage but the corresponding genetic allele being advantageous (Figure 3, Scenarios 3d and 4d). These results further highlight the dynamics of the maintenance of genetic variation and deleterious genetic variation. While Scenario 3c promoted the maintenance of a deleterious genetic allele (a_2) *and* genetic variation (also maintaining A_1), Scenario 4c promoted the maintenance of the deleterious allele (a_2) but not while A_1 was also maintained. When the epialleles are deleterious, this dynamic is reversed (Figure 7A, Supplemental Figure 3). In Scenario 3d, the deleterious epiallele a_4 can be maintained, but not while A_3 is also present. However, for Scenario 4d, the deleterious epiallele facilitates the maintenance of multiple genetic alleles in equilibrium simultaneously (Figure 7A). What do Scenarios 3c and 4d have in common to promote the maintenance of genetic variation? In either case, an allele or epiallele acts dominantly to increase fitness, while its corresponding epiallele within the same genotype has the lowest fitness and acts in a recessive manner. In Scenario 3c, it is the epiallele a_4 that is dominant, while in 4d, it is the allele a_2, and in either case, the other epiallele is recessive (and the symmetric situations for A_3 and A_1 also show these same dynamics;

see Supplementary Figure 3). 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.

Since there is a situation in which deleterious genetic variation can be maintained in long-term equilibrium alongside the favored allele, we assessed how the fitness of the deleterious allele affects these dynamics. We held the epimutation rate constant (at 10^{-5} and 0.01) but allowed the fitness of a_2a_2 (w_{22}) to vary for Scenario 3c. 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 7B-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

The results of this model lead to two major conclusions. First, the maintenance of *epigenetic* variation occurs in all scenarios examined, suggesting that it may be relatively common if epigenetic variation arises frequently. Second, the maintenance of *genetic* variation in the presence of epigenetic effects occurs more rarely. Specifically, it only occurs when the alternative epiallele of a recessive deleterious allele has a dominant fitness advantage, or similarly, when a dominant allele has an alternative epiallele that is recessively deleterious. In other words, epialleles within a genotype must be at the extremes of the fitness distribution for the collection of alleles at a locus, and the advantaged allele must be dominant. When epialleles within a genotype are at the extremes of the fitness distribution, the deleterious allele/epiallele can be maintained, but not alongside the advantaged allele. We discuss these two major results below. In addition, we consider the potential role of two factors not considered here, particularly drift and the role of 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.,* it has a high back-epimutation rate). This is true across all scenarios examined, including when mutation and epimutation of an allele have similar effects on gene activity and fitness and when they act in opposite directions. It is also robust to small differences in fitness. 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, they must also be generated at relatively high frequency (~0.001 to 0.01). 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 for which epimutation rates are higher. It has been frequently 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. 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 in these scenarios 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, given that the environment is constant and the population is sufficiently large for the present model to be applicable.

The maintenance of deleterious genetic variation is possible but likely rare

While epialleles were maintained for certain parameters in every scenario analyzed, deleterious genetic variation was typically not maintained. This means that either A_1 and A_3 were maintained, or a_2 and a_4 were maintained, but not typically combinations of $A_{1/3}$ and $a_{2/4}$. Sometimes, we find that A_3 and a_4 are maintained

together (see the boundary between A_3 and a_4 in Figures 5 and 6), but this is in a part of the parameter space for which the back-epimutation rate is very low, making the epimutation behave more like a regular mutation. In these cases, it is relatively unsurprising that an advantageous epiallele that is not reversible 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. Even if a genetic mutation occurs, the mutation may not realize its effect if the silenced epigenetic state is consistently inherited. However, in the context of highly reversible epigenetic variants, deleterious variation can only be maintained when the deleterious allele has an alternative epiallele with higher fitness than other alleles in the system (see Scenario 3c in Figures 5E-F and Scenario 4c in Figure 6E-F). In both cases, the extremes of the fitness distribution include the alternative epialleles of a – with a_2a_2 genotypes 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 (either silencing or activating a gene with a mutation that causes the opposite effect), 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). 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 of things 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

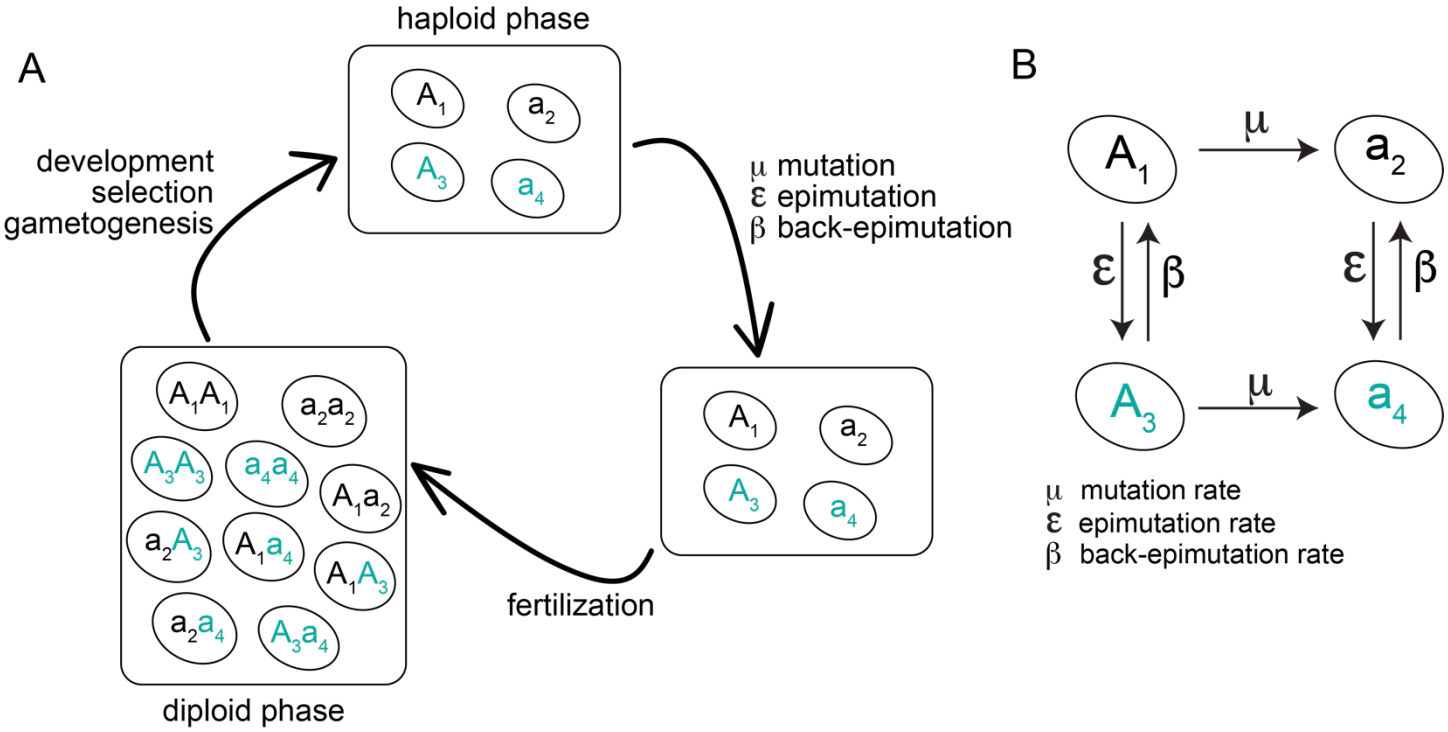


Figure 1: Life cycle diagram and transition probabilities for population epigenetic model.

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

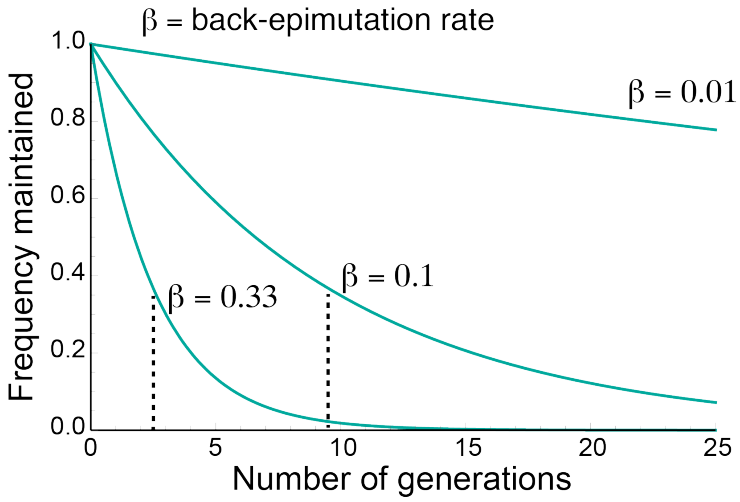


Figure 2: The back-epimutation rate models the maintenance of an epiallele without selection. In the absence of other factors, including new mutation, epimutation, or selection, the back-epimutation rate models the expected number of generations an alternative epiallele (A_3 or a_4) would be expected to be maintained. The equations plotted here are p = (1 – β)^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 scenarios 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: Schematics for how mutation and epimutation map to gene expression and fitness

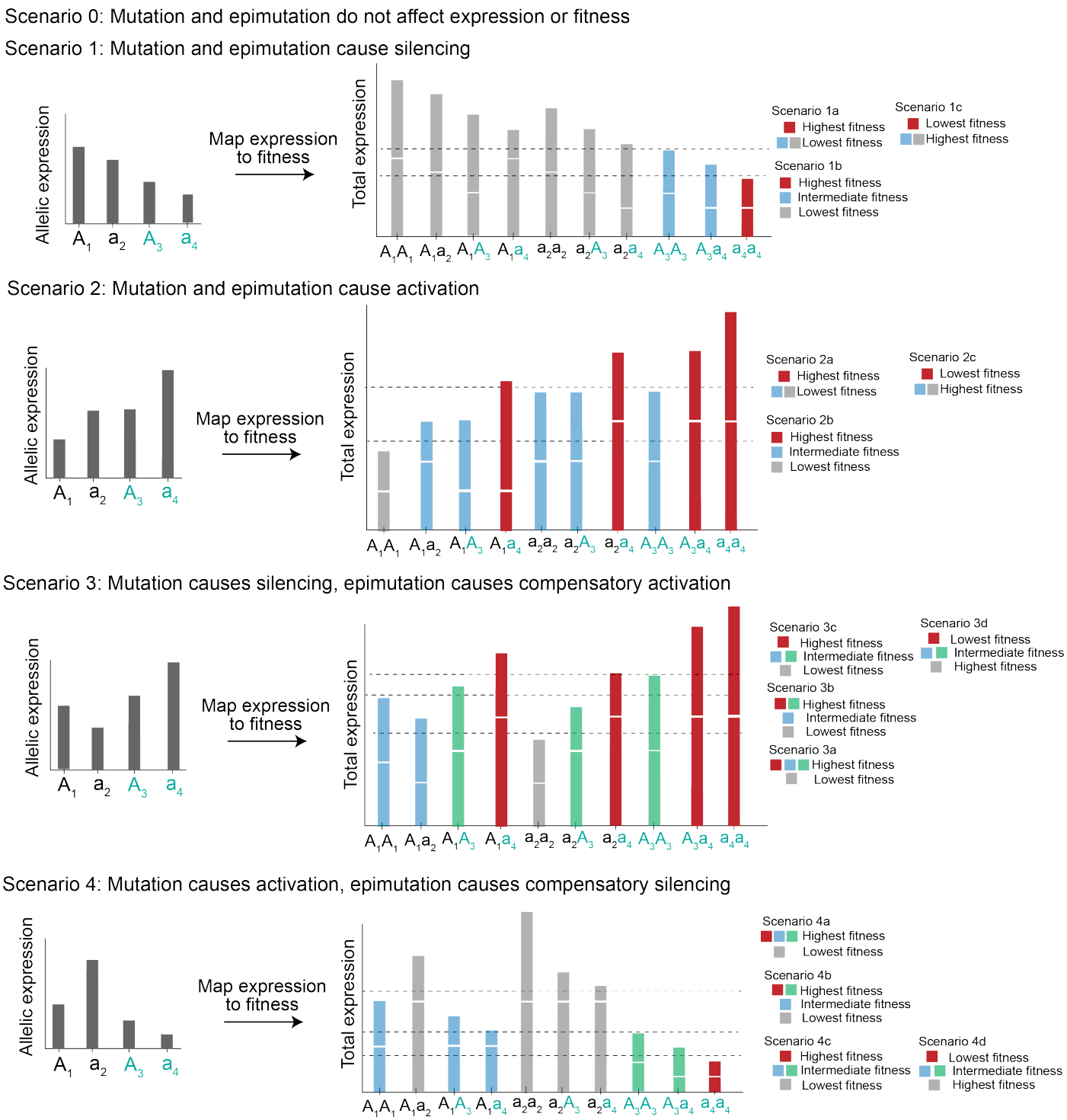


Figure 3: Schematics for how mutation and epimutation map to gene expression and fitness. Mutation and epimutation can both impact fitness in the same or opposing directions, and these expression changes can map onto fitness in different ways. These scenarios will be analyzed in the theoretical model. The symmetric scenarios (in which fitness levels of genotypes including A_1 and A_3 were swapped with those containing a_2 and a_4, respectively) were also analyzed to ensure that overall patterns persist despite the asymmetry in the

mutation rate. In Scenario 1, both mutation and epimutation cause gene silencing, and silencing increases fitness. We consider two different levels of silencing related to fitness, and both alleles in a diploid must be silenced to affect fitness. In Scenario 2, mutation and epimutation both cause activation of gene expression, and activation relates to fitness. We consider two different levels of activation related to fitness, but only one allele in a diploid must be above the threshold, consistent with dominance. In Scenario 3, mutation causes silencing, but epimutation cause gene activation, and these have opposite effects on fitness. We consider three levels of activation related to fitness, and only one allele must be above the threshold, consistent with dominance. Finally, we consider Scenario 4, in which mutation causes activation and epimutation causes silencing, and these have opposite effects on fitness. We consider three levels of silencing related to fitness, but both alleles must be below the threshold.

Figure 4: Concordant changes in expression and fitness by mutations and epimutations (scenarios 1 and 2) facilitate the maintenance of epigenetic but not genetic variation.

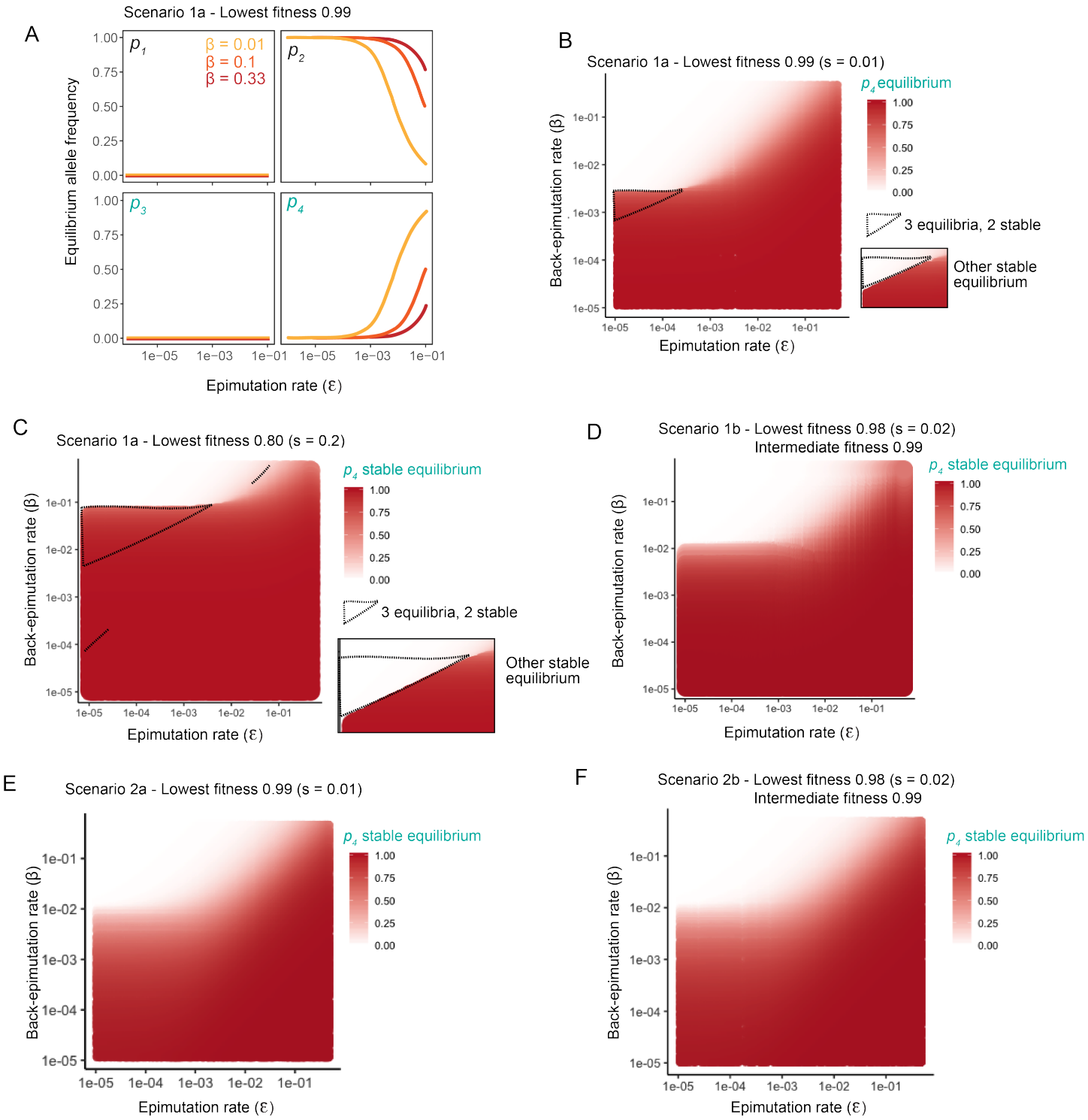


Figure 4: Concordant changes in expression and fitness by mutations and epimutations (scenarios 1 and 2) facilitate the maintenance of epigenetic but not genetic variation.

A. Stable equilibrium allele frequency for Scenario 1a when the lowest fitness is 0.99 and for back-epimutation rates of 0.01, 0.1, and 0.33 for p_1 through p_4.

B-F. Stable equilibria for p_4 for Scenarios 1 and 2. In all cases, p_1 and p_3 are not maintained, and $p_2 = 1 - p_4$.

B. Stable equilibria for Scenario 1a when the lowest fitness is 0.99 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 the lowest fitness is 0.80.

D. Stable equilibrium for Scenario 1b when the lowest fitness is 0.98 and intermediate fitness is 0.99 for a range of epimutation and back-epimutation rates.

E. Stable equilibrium for Scenario 2a when the lowest fitness is 0.99 for a range of epimutation and back-epimutation rates.

F. Stable equilibrium for Scenario 2b when the lowest fitness is 0.98 and intermediate fitness is 0.99 for a range of epimutation and back-epimutation rates.

Figure 5: Compensatory epimutations promoting gene activation (Scenario 3) facilitate the maintenance of epigenetic and genetic variation.

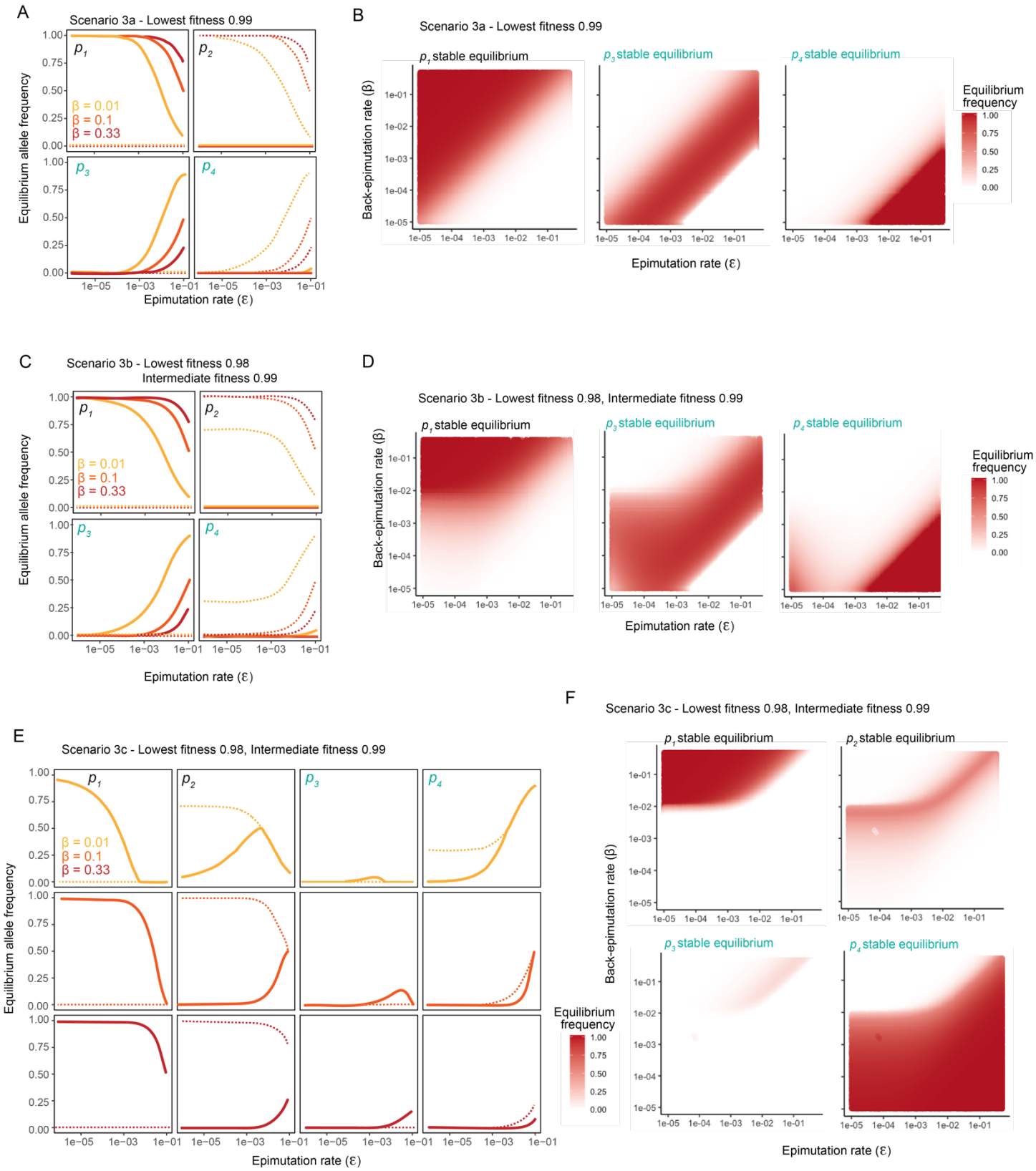


Figure 5: Compensatory epimutations promoting gene activation (Scenario 3) facilitate the maintenance of epigenetic and genetic variation.

A. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) for Scenario 3a when the lowest fitness is 0.99 and for back-epimutation rates of 0.01, 0.1, and 0.33.

B. Stable equilibrium for Scenario 3a when the lowest fitness is 0.99 for a range of epimutation and back-epimutation rates. p_2 is not shown because it is not maintained in stable equilibrium throughout the parameter space.

C. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) for Scenario 3b when the lowest fitness is 0.98 and intermediate fitness is 0.99 for back-epimutation rates of 0.01, 0.1, and 0.33.

D. Stable equilibrium for Scenario 3b when the lowest fitness is 0.98 and intermediate fitness is 0.99 for a range of epimutation and back-epimutation rates. p_2 is not shown because it is not maintained in stable equilibrium throughout the parameter space.

E. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) for Scenario 3c when the lowest fitness is 0.98 and intermediate fitness is 0.99 for back-epimutation rates of 0.01, 0.1, and 0.33.

F. Stable equilibrium for Scenario 3c when the lowest fitness is 0.98 and intermediate fitness is 0.99 for a range of epimutation and back-epimutation rates.

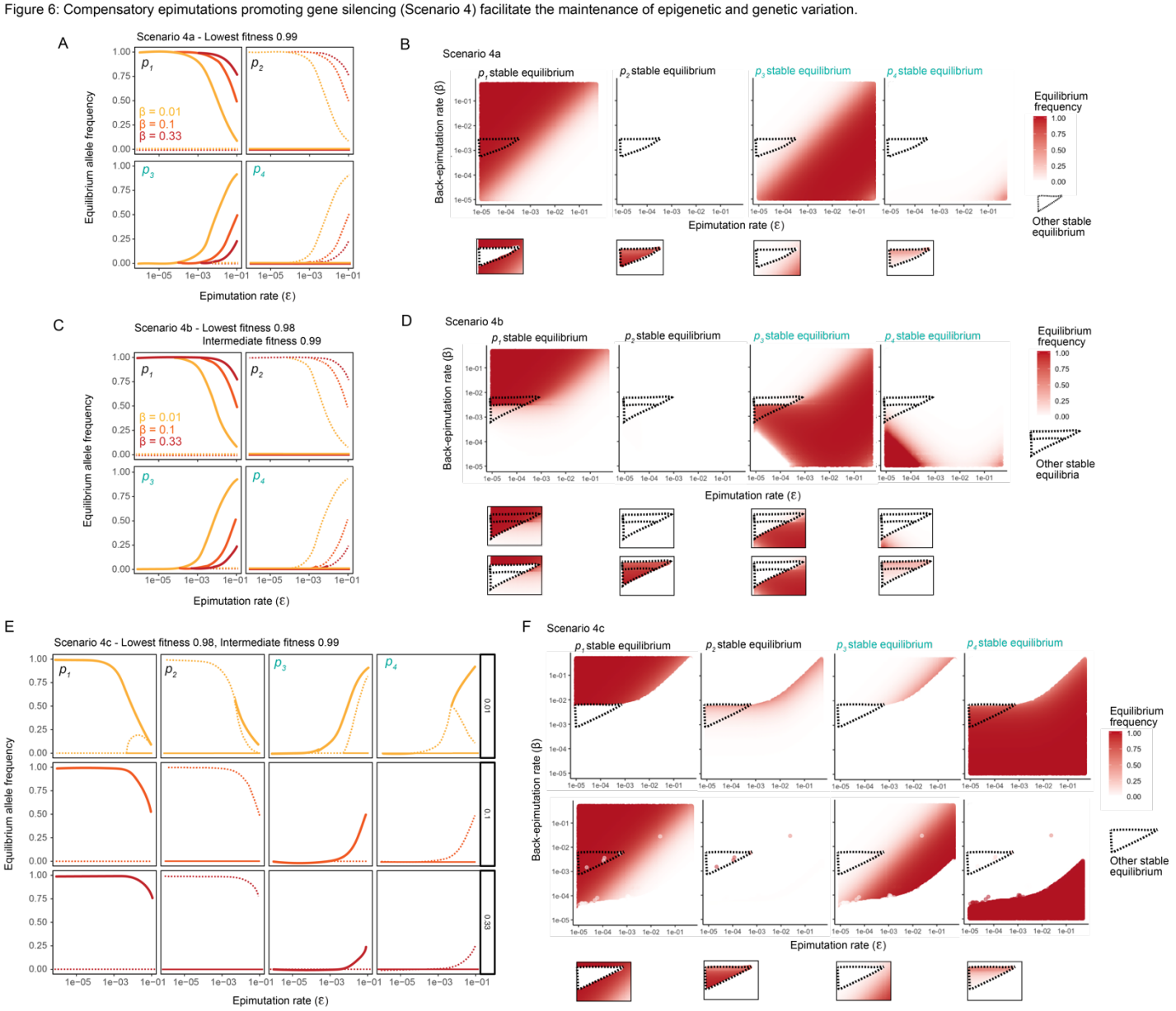


Figure 6: Compensatory epimutations promoting gene silencing (Scenario 4) facilitate the maintenance of epigenetic and genetic variation.

A. Stable equilibrium allele frequencies (solid lines) and unstable equilibrium allele frequencies (dashed lines) for Scenario 4a when the lowest fitness is 0.99 and for back-epimutation rates of 0.01, 0.1, and 0.33.

B. Stable equilibria for Scenario 4a when the lowest fitness is 0.99 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) for Scenario 4b when the lowest fitness is 0.98, intermediate fitness is 0.99, and back-epimutation rates are 0.01, 0.1, and 0.33.

D. Stable equilibria for Scenario 4b when the lowest fitness is 0.98 and intermediate fitness is 0.99 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) for Scenario 4c when the lowest fitness is 0.98, intermediate fitness is 0.99, and back-epimutation rates are 0.01, 0.1, and 0.33.

F. Stable equilibria for Scenario 4c when the lowest fitness is 0.98 and intermediate fitness is 0.99 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 7: Discordant effects of mutation and epimutation facilitate maintenance of genetic variation, depending on fitness dynamics

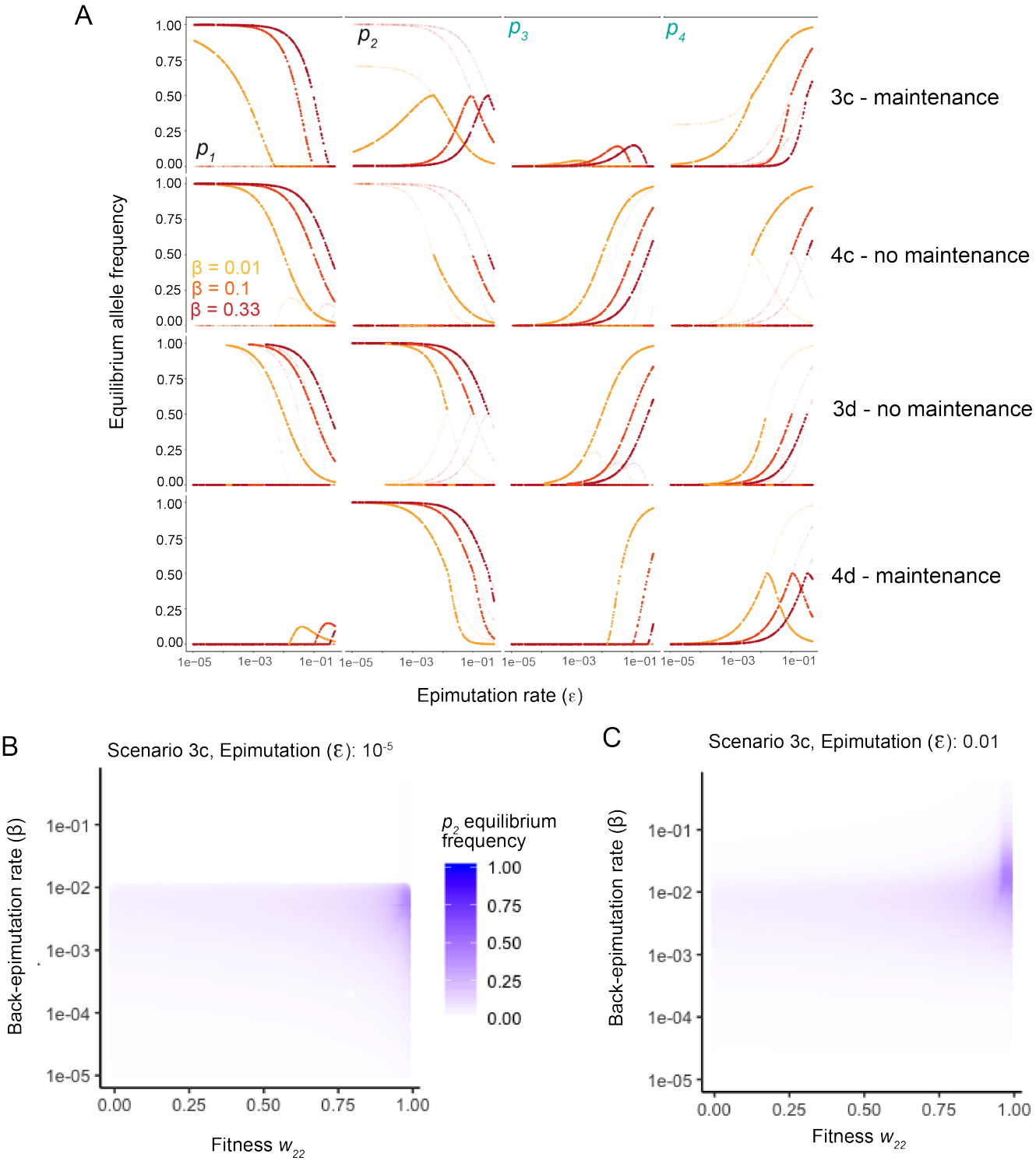


Figure 7: Discordant effects of mutation and epimutation facilitate maintenance of genetic variation, depending on fitness dynamics

A. Equilibrium frequencies for Scenarios 3c-d and 4c-d. Opaque points represent stable equilibria, while faded points are unstable equilibria. Scenarios 3c and 4d each have a single stable equilibrium that facilitates

maintenance of two genetic alleles simultaneously. Scenarios 3d and 4c each have up to 2 stable equilibria, and one of these allows for the maintenance of a deleterious allele, but without the alternate genetic allele simultaneously.

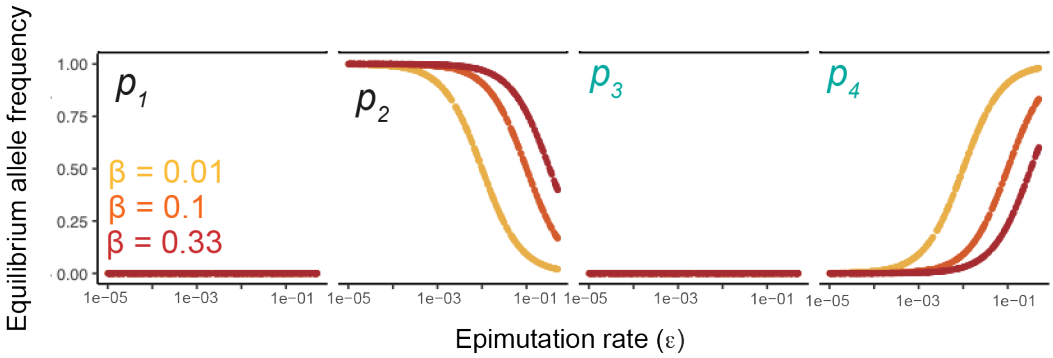
B-C. w_{22} (the lowest fitness) was allowed to take any value between 0 and 0.98 and back-epimutation rate was allowed to take any value between 10^{-5} and 0.5. The long-term equilibrium of p_2 (the frequency of a_2) is indicated by the shading of blue.

B. Scenario 3c when the epimutation rate is held constant at 10^{-5}.

C. Scenario 3c when the epimutation rate is held constant at 0.01.

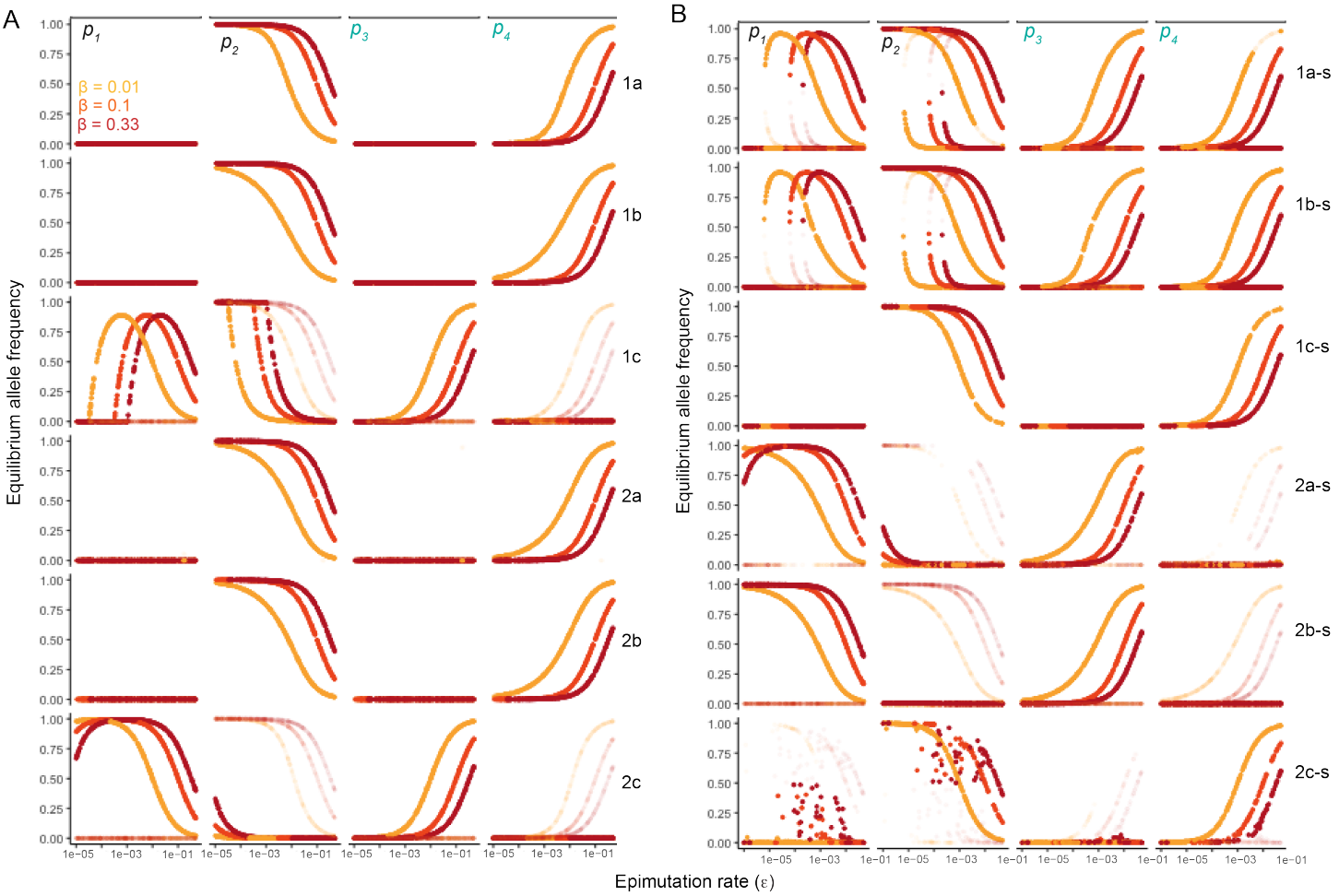
SUPPLEMENTARY FIGURES

Supplementary Figure 1: Epigenetic variation is maintained under neutral conditions



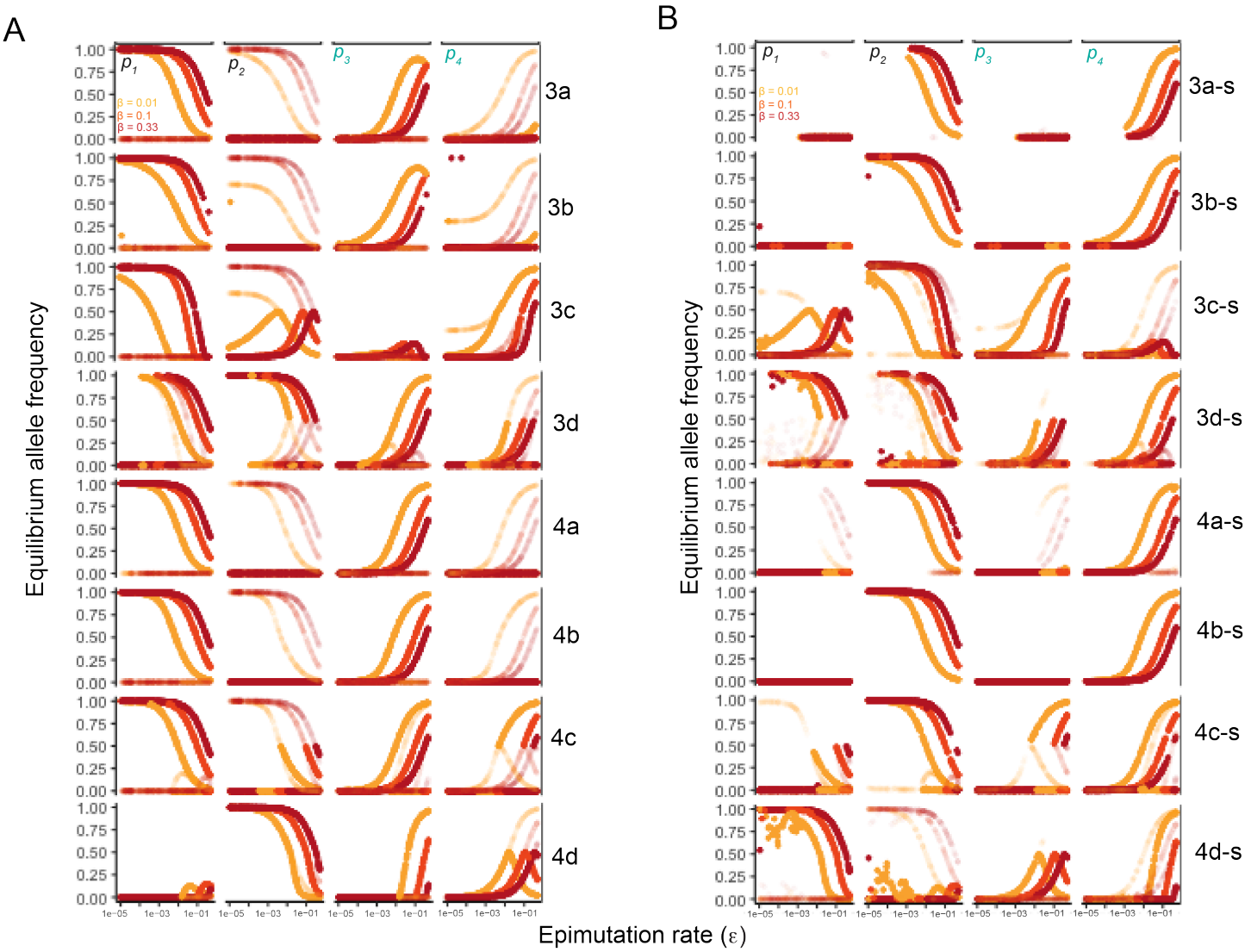
Supplementary Figure 1: Epigenetic variation is maintained under neutral conditions. When all fitness levels are set to 1, a_2 and a_4 alleles are found in equilibrium at frequencies that depend on the epimutation and back-epimutation rate, according to Equation 8.

Supplementary Figure 2: The dominant pattern for scenarios 1 and 2, including symmetric and deleterious sub-scenarios, is the maintenance of a pair of epialleles



Supplementary Figure 2: The dominant pattern for scenarios 1 and 2, including symmetric and deleterious sub-scenarios, is the maintenance of a pair of epialleles. 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 from Scenarios 1a, 1b, 1c, 2a, 2b, and 2c. B. From top to bottom facet, raw output from the symmetric scenarios to those in A. Symmetric is indicated by '-s' after the scenario identifier. Scenarios 1a-s, 1b-s, 1c-s, 2a-s, 2b-s, and 2c-s are shown.

Supplementary Figure 3: Discordant effects of mutation and epimutation (Scenarios 3 and 4) can yield the maintenance of epigenetic or genetic variation, depending on fitness dynamics



Supplementary Figure 3: Discordant effects of mutation and epimutation (Scenarios 3 and 4) can yield the maintenance of epigenetic or genetic variation, depending on fitness dynamics. 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 from Scenarios 3a, 3b, 3c, 3d, 4a, 4b, 4c, and 4d. B. From top to bottom facet, raw output from the symmetric scenarios to those in A. Symmetric is indicated by '-s' after the scenario identifier. Scenarios 3a-s, 3b-s, 3c-s, 3d-s, 4a-s, 4b-s, 4c-s, and 4d-s are shown.

SUPPLEMENTARY FILES

Mathematica scripts for the full parameter space and focal back-epimutation rates are available on Github (https://github.com/amykwebster/2022_EpigeneticModels). In addition, csv files and R scripts used to generate plots are available.

ACKNOWLEDGEMENTS

We thank members of the Phillips lab for critical feedback. Funding was provided by NIH grant number R35GM131838.

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**.

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.

Duncan EJ, Gluckman PD, Dearden PK. 2014. Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *J Exp Zool B Mol Dev Evol* **322**: 208-220.

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.

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.

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.