Original Article

A model of functionally buffered deleterious mutations can lead to signatures of positive selection

Abstract

Selective pressures on DNA sequences often result in signatures of departures from neutral evolution that can be captured by the McDonald-Kreitman (MK) test. However, the nature of such selective forces often remains unknown to the experimentalists. Amino acid fixations driven by natural selection in protein coding genes are often associated with genetic conflict or changing biological purposes, leading to proteins with new functionality. Here, we propose that amino acid changes can also be correlated with a return to the original function after a period of relaxed selective constraint. Dynamic environments and changing population sizes offer opportunities for fluctuations in selective constraints on genes. We suggest that an evolutionary process in which a period of relaxed selective constraint can allow for slightly deleterious mutations to fix in a population by drift, and increased selective constraint can result in positive selection for residues that bring the gene back to its optimal functional state. We designed a model to investigate this possibility using the functionally critical *bag of marbles* (*bam*) gene in *D. melanogaster*. *Bam*, along with other germline genes, may experience variation in selective constraint due to the presence or absence of the endosymbiont bacteria *Wolbachia*, which infects many arthropod species episodically. We use simulations to implement this model and find that signatures of positive selection for the amino acid changes resembling the original states after the loss of *Wolbachia* can be detected by the MK test. However, the proportion of adaptive amino acids and the power of the test are both significantly lower than seen in parallel simulations of a change-in-function model that favors proteins with diversified amino acids to escape *Wolbachia*'s manipulation of reproduction.

Introduction

Patterns of DNA sequence variation within and between species have been widely used to infer the evolutionary forces that have acted on genes and genomes. Over the past three decades, many tests of fit to a model of neutral evolution have been developed, with one of the most widely applied being that proposed by McDonald and Kreitman (McDonald and Kreitman 1991) and referred to as the McDonald-Kreitman (MK) test. The basis of this test is a comparison of the ratios of nonsynonymous and synonymous fixed differences between species to those segregating as polymorphisms within species using a 2x2 contingency test (e.g., Fisher's Exact Test, Chi-Square test, or G-test). Synonymous variation acts as a proxy for neutral

variation, while an excess of nonsynonymous fixed differences between species is typically interpreted as evidence that natural selection has accelerated the fixation of advantageous amino acid replacements. This pattern is often associated with natural selection fine tuning protein function, in response to a changing function, and/or an intra- or inter-genomic conflict (McLaughlin and Malik 2017). Although the MK test has been found to have low power in detecting positive selection, particularly when applied to single genes (Akashi 1999; Zhai, et al. 2009), there are many empirical reports in the literature of significant departures in the direction of positive selection (e.g., Eyre-Walker 2006). The obvious question from the experimentalist's perspective is what evolutionary mechanisms are driving such signatures of positive selection.

In some cases, a model illustrating the selective pressure for the observed positive selection is straightforward. For example, positive selection observed at immune genes are regularly attributed to host-pathogen conflict with support from functional data (Obbard et al. 2006, Obbard et al. 2009, Shultz and Sackton 2019). Likewise, it is well established that positive selection is common among genes involved in sexual reproduction (Swanson and Vacquier 2002, Clark et al. 2006, Turner and Hoekstra 2008). On the other hand, the pervasive positive selection seen in the piRNA pathway has been proposed to be due to an arms race with transposable elements, but there is still speculation on the way in which this occurs (Crosby et al. 2019).

We have studied the population genetics of two *Drosophila* germline stem cell genes, *bag-of-marbles* (*bam*) and *Sex-lethal* (*Sxl*), that show strong evidence of episodic positive selection (Bauer DuMont, et al. 2007; Flores, DuMont, et al. 2015; Bauer DuMont, et al. 2021; Bubnell et al. 2021). However, the typical selective pressures that we expect on reproductive genes, including sexual selection and conflict, do not obviously apply to *bam* and *Sxl* because these genes function so early in gametogenesis. Instead, the positive selection has been proposed to be due to a change in gene function and/or a genetic conflict with the endosymbiont bacteria *Wolbachia* that genetically interacts with both genes and resides in the germarium where they function (Bauer DuMont, et al. 2007; Flores, Bubnell, et al. 2015; Flores, DuMont, et al. 2015; Bubnell et al. 2021, Bauer DuMont et al 2021). The bursts of positive selection at *bam* in only certain lineages of *Drosophila* is consistent with the episodic nature of bacterial infections and the heterogeneous presence of *Wolbachia* reported across the genus (Richardson, et al. 2012; Turelli, et al. 2018; Meany, et al. 2019). Thus, an evolutionary conflict between the host

(*Drosophila melanogaster*) and *Wolbachia* over the control of reproduction may contribute to the observed positive selection at *bam* and *Sxl*.

Here, we propose an evolutionary dynamic based on the genetic interaction between germline genes and *Wolbachia* that offers an additional possible explanation for the observed positive selection at these genes. As *Wolbachia* partially rescues the fertility defects of four distinct single amino acid replacement hypomorphic mutations of *bam* and *Sxl* in *D. melanogaster* (Flores, Bubnell, et al. 2015; Star and Cline, 2002), we posit that in the presence of *Wolbachia*, slightly deleterious mutations may accumulate in these genes without significantly reducing their function. When *Wolbachia* is lost from the population, there is positive selection for new nonsynonymous mutations that return the *bam* and *Sxl* protein sequences to their initial, and assumed optimal, functional state. We term this dynamic the Buffering model, as the effects of deleterious mutations are "buffered" during periods of infection by *Wolbachia*.

To evaluate the potential of the Buffering model compared to established forces of selective pressures, we created a parallel model to simulate the classical hypothesis of genetic conflict between the host germline gene and *Wolbachia*. In this model, we assume that *Wolbachia* infection is detrimental to a germline reproductive process in which *bam* and *Sxl* is involved. For example, *Wolbachia* may manipulate *bam* and *Sxl* in a way counterproductive to the fitness of the fly. Under arms race dynamics, we would expect positive selection to favor diversifying amino acids in these genes that result in *Drosophila's* escape from the deleterious impact of *Wolbachia* on their fitness. We term this the Conflict model. Note that while we model this as an evolutionary conflict, selection associated with a strong directional shift in function would be similar in outcome in many ways.

We carry out evolutionary simulations to explore the population genetic consequences for sequence evolution both buffering and conflict interactions between the endosymbiont *Wolbachia* and the *Drosophila* host. We focus on modeling the *bam* gene alone because *Sxl* has a large and highly conserved RNA binding domain which would limit the sites available for adaptive evolution in our simulations (Bauer DuMont et al. 2021). With these models, we demonstrate that buffering type of interaction can result in signatures of positive selection detectable with the MK test. We also evaluate the robustness of the Buffering model to variation in different magnitudes of selection coefficients and duration of *Wolbachia* infections. We further show that it is disappointingly difficult to distinguish which model has led to the

observed positive selection at *bam* based on properties of the amino acid substitutions observed, though the statistically significant MK test results shown in the Conflict model are more consistent with the empirical observations at *bam*.

Materials and Methods

Simulation setup using SLiM 3.5

The evolution of *bam* was simulated under a Wright-Fisher model using nucleotide-based simulation in SLiM 3.5 (Haller, et al. 2019). We inferred the ancestral DNA sequence of the exons in *bam* (1338 nucleotides) for *Drosophila melanogaster* and *D. simulans* using maximum likelihood with codeML v4.8 (Yang 2007). Briefly, alignments of seven *Drosophila* coding sequences were made using PRANK v.170427 (Löytynoja 2014), including sequences of *D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta, D. eugracilis, and D. pseudoobscura*. The alignment was input in codeML and an ancestral sequence was estimated by maximum likelihood for the common ancestor of *D. melanogaster* and *D. simulans*, using the other species' sequences as outgroup references. The estimated common ancestral nucleotide sequence was then used in SLiM as the starting sequence for the entire population in each simulation.

The evolutionary parameters in the simulations were based on empirical estimates and then scaled to make the simulations run efficiently (summarized in Table 1). The reported evolutionary parameters from empirical observations were an effective population size (N_e) of 1e6 (Campos, et al. 2017), an overall mutation probability (μ) of 2.8e-9 per nucleotide per generation (Keightley, et al. 2014), an average recombination rate (ρ) of 1e-8 per nucleotide per generation (Comeron, et al. 2012) and a divergence time t of 25 million generations (2.5 million years assuming 10 generations per year) from the common ancestor to each species (Russo, et al. 1995). These parameters were then scaled down by 1,000 to run more efficient simulations by using a smaller population and shorter simulation time while keeping the key products, $N_e\mu$, N_er and N_e/t constant to approximate the same evolutionary process (Haller and Messer 2019). Hence, in each Wright-Fisher simulation, we have 1) a scaled population size N_e of 1,000 diploid individuals, 2) a mutation matrix representing a Kimura (1980) model with transition rate α and transversion rate β of 1.88e-6 and 0.47e-6 respectively (calculated from a scaled mutation rate μ=2.8e-6 and an observed 2:1 transition:transversion ratio in the sequences (Keightley, et al.

2009)), 3) a scaled recombination rate ρ of 1e-5 and 4) 25,000 scaled generations for the divergence time t. We simulated *bam* as a single contiguous exon, though in reality there are two short introns (61 and 64 bp). The effect of excluding these introns had a negligible impact on the rate of recombination.

We observed 85% of the codons in *bam* encode the same amino acids in both *D. melanogaster* and *D. simulans* reference sequences, likely due to functional constraints (example classifications are show in supplemental fig. S1). Thus, we used this metric as a baseline in our initial simulations and first randomly sampled 75% of the codons from the identical amino acids in the ancestral sequence to be constrained to the original amino acids. The rest of the identical amino acids (10% of the total amino acids) were assumed to be under completely neutral evolution, while the other 15% unidentical amino acids were subject to selection based on the setup of our models. A nonsynonymous mutation in the conserved codons was always assigned a selection coefficient s = -0.1, so that it would undergo strong purifying selection ($N_e s$ = -100 in our simulations). Note that the fitness of an individual in SLiM is calculated multiplicatively as $(1+s)$ when it carries a homozygous mutation of selection coefficient s and $(1+hs)$ when the mutation is heterozygous, where h is the dominance coefficient. In all our simulations, the dominance coefficient of any mutation was set to a constant of 0.5. A mutation in the neutral codons was always assigned a selection coefficient s = 0.

Each simulation run began with a neutral "burn-in" period of 20,000 simulation generations (=20×scaled N_e) to accumulate genetic variation consistent with an equilibrium state of mutation-drift balance before non-neutral dynamics started. During this neutral period, mutations occurring at the conserved sites were still assigned a selection coefficient of s = -0.1 to retain the functionally constrained amino acid positions. At the end of the neutral "burn-in" period all new variations (fixations and polymorphisms) were retained in the simulation. However, the "reference sequence" in SLiM (the sequence that stores fixations in the simulation population) was manually reset to the original estimated ancestral sequence inferred from PAML so the selection coefficients of subsequent new mutations would be based on the particular amino acid change encoded by the new mutation compared with the original inferred ancestral sequence.

For each selection phase of the simulations, the absolute value of selection coefficient $|s|$ for each positively or negatively selected mutation in the 15% of codon sites under selection was

fixed for the duration of each simulation. The beneficial mutations were assigned a selection coefficient of $s > 0$ while deleterious mutations had a selection coefficient $s < 0$. To determine the fitness effect of each mutation, we explored several correlated measures of amino acid substitutions (e.g., Grantham et al. 1974; Miyata et al. 1979; Henikoff and Henikoff 1992), and decided to use the amino acid matrix of Miyata et al. (1979), which captures the primary features of biochemical and physical differences between amino acid pairs but does not take empirical protein sequence conservation into account, since some of our changes were going to be positively selected, and not conserved. We henceforth refer to the pairwise measures from the Miyata matrix as "Miyata scores (MS)" and use them to determine whether a mutation is neutral (synonymous, no change in the encoded amino acid) or under positive or negative selection (nonsynonymous, MS between the original amino acid and the mutated amino acid $\neq 0$) in each selection scenario. Below, we will use a shorthand for Miyata score calculations as follows, with, for example, MS between the current amino acid and the mutated amino acid represented as $MS(AA_{cur}, AA_{mut})$.

Selection regimes

Wolbachia infections have been observed to be temporally dynamic in host populations, being lost at times and then regained (Richardson, et al. 2012; Turelli, et al. 2018; Meany, et al. 2019). Thus, each model has two selection phases: one with selection parameters to represent a period of *Wolbachia* infection and another to represent a period of *Wolbachia* absence in the population. There are four different selection schemes: 1a) *Wolbachia*-infection phase in the Buffering model, 1b) *Wolbachia*-absence phase in the Buffering model, 2a) *Wolbachia*-infection phase in the Conflict model, and 2b) *Wolbachia*-absence phase in the Conflict model. The phases of infection and absence of *Wolbachia* alternated in each model, which simulated the periodic occurrence of *Wolbachia* in natural populations. To keep the simulations simple, we assume that *Wolbachia* infection and loss is instantaneous throughout the entire population and that there are no other effects of *Wolbachia* on the host beyond which we are modeling.

The Buffering model is built on the observation that *Wolbachia* offers a functional buffer for deleterious mutations in *bam* and rescues what would be reduced fertility of its host in the absence of *Wolbachia*. During the *Wolbachia*-infection phase, *Wolbachia* functionally alleviates the deleterious effects of certain nonsynonymous mutations and makes them effectively neutral.

Under this scenario, mutations leading to divergent amino acids different from the ancestral state ($MS(AA_{anc}, AA_{mut}) > 0$) are all regarded as neutral (fig. 1, bottom) and thus can accumulate. If *Wolbachia* infection is lost, these previously buffered mutations are now deleterious, and new mutations leading to amino acids that converge back towards the initial amino acid ancestral states are favored as there is selective pressure for *bam* to regain its original optimal function. In this case, a mutation that converts the current amino acid to a mutated amino acid that is biochemically more similar to the ancestral amino acid ($MS(AA_{anc}, AA_{mut}) – MS(AA_{anc}, AA_{cur}) < 0$) is beneficial with a selection coefficient $s > 0$, while divergent mutations ($MS(AA_{anc}, AA_{cur}) – MS(AA_{anc}, AA_{cur}) > 0$) are deleterious with selection coefficients $s < 0$ (fig. 1, bottom). Empirical data on the nature of amino acid changes for which buffering by *Wolbachia* has been functionally documented are limited. There are three single amino acid substitution hypomorphs in *Sxl* (Starr and Cline 2002) and one in *bam* (Ohlstein et al 2000). The three in *Sxl* are Sxl^{f4} (a proline > serine, MS=0.56), Sxl^{f5} (a proline > leucine, MS=2.70) and Sxl^{F18} (a glycine > aspartic acid, MS=2.37) while the one hypomorph available for *bam* is bam^{BW} (a leucine > phenylamine replacement, MS=0.63).

The Conflict model was implemented based on a traditional arms race model. In the *Wolbachia*-infection phase, we assume that *Wolbachia*'s presence drives the positive selection of *Drosophila* by favoring the nucleotide changes that lead to biochemically more diversified amino acids to "escape" the present function which may be targeted by *Wolbachia*'s harmful impact. In this case, all the nonsynonymous mutations that give rise to different amino acids from the current states ($MS(AA_{cur}, AA_{mut}) > 0$) were positively selected for, with selection coefficients $s > 0$ (fig. 1, top). In the *Wolbachia*-absence phase of the Conflict model, there is no evolutionary conflict between *bam* and *Wolbachia* and thus no selective pressure on the host to adapt. Under these conditions, we assume that the current amino acid sequence functions adequately such that the DNA sequences in the population remain largely unchanged and the *bam* gene is under purifying selection to preserve the present amino acid sequences. In this case, nonsynonymous mutations leading to amino acid changes ($MS(AA_{cur}, AA_{mut}) > 0$) are considered deleterious with selection coefficients $s < 0$ (fig. 1, top).

The above models capture two types of potential driving forces behind the signatures of positive selection observed on the *bam* gene. These models are the simplest with regard to assigning selection coefficients and thus are regarded as the "Base" models. For instance, in the

Conflict model, any nonsynonymous mutation would be positively selected during the presence of *Wolbachia.* This allows substitutions to be dominated by amino acids of any type, including those of "small steps" as suggested by Bergman and Eyre-Walker (2019) to be prevalent in *Drosophila*. Additionally, we test "Complex" models that limit the mutations under positive selection under the preposition that mutations that lead to biochemically similar amino acids with homogeneous properties may not bear such strong selective advantages and could be regarded as almost neutral, while the mutations giving rise to extremely dissimilar amino acids may be deleterious. We set up "Complex" models based on empirical observations to determine whether a mutation would be neutral, beneficial, or deleterious. There are a few nonsynonymous substitutions predicted to have been fixed by positive selection in response to evolutionary conflict, with supporting functional evidence. Demogines et al. (2013) identified adaptively evolving sites in the transferrin receptor gene *TfR1* in wild rodents to include amino acids R, K, N, I, and T, which corresponds with pairwise Miyata scores ranging from 0.4 to 3.37. Likewise, Charron et al. (2008) proposed sites in the plant gene eIF4E to be in an arms race conflict with viral proteins, which includes those with amino acids L, P, and A that give pairwise Miyata scores ranging from 0.06 to 2.76. We found that the Miyata scores for all proposed positively selected residues in these studies ranged from 0.05 to 3.37, with the majority of scores falling between 1.5 and 2.5.

With the above proposition, we first parameterized the selection schemes in the complex Conflict model based on the empirical observations, and then adopted the same selection schemes in the complex Buffering model to represent one biologically plausible possibility for the *Wolbachia-bam* interaction. In the complex Conflict model, when *Wolbachia* is present, nonsynonymous mutations that give rise to biochemically similar amino acids ($0 < MS(AA_{cur}, AA_{mut}) \leq 1$) are regarded as neutral with selection coefficients $s = 0$ (fig. 2, top); mutations leading to mildly different amino acids ($1 < MS(AA_{cur}, AA_{mut}) \leq 3$) are considered beneficial with selection coefficients $s > 0$ (fig. 2, top); and mutations become deleterious with selection coefficients $s < 0$ when they generate extremely dissimilar amino acids ($MS(AA_{cur}, AA_{mut}) > 3$) (fig. 2, top), as they are likely to disrupt the biological function of *bam*. These cutoffs are consistent with the range of Miyata scores found at sites that are proposed to be adaptively evolving in response to an evolutionary conflict. In the *Wolbachia*-absence phase in the complex Conflict model, to preserve the current amino acid sequences, we still assume that mutations

leading to similar amino acid changes ($0 < MS(AA_{cur}, AA_{mut}) \leq 1$) are considered neutral, but any mutation that causes a dissimilar amino acid change ($MS(AA_{cur}, AA_{mut}) > 1$) is deleterious with a selection coefficient $s < 0$ (fig. 2, top).

When *Wolbachia* is present in the complex Buffering model, any mutation that gives rise to a mildly biochemically different amino acid from the ancestral state ($0 < MS(AA_{anc}, AA_{mut}) \leq 3$) is regarded as neutral with selection coefficients $s = 0$ (fig. 2, below) due to the protection by *Wolbachia*. However, mutations are considered deleterious with selection coefficients $s < 0$ when they generate extremely dissimilar amino acids ($MS(AA_{anc}, AA_{mut}) > 3$) (fig. 2, bottom), since they are likely to disrupt the biological function of *bam*. When *Wolbachia* is lost from the population, only mutations that converge back towards the biochemical characteristics of the initial ancestral state relative to the current amino acid are favored ($MS(AA_{anc}, AA_{mut}) - MS(AA_{anc}, AA_{mut}) < -1$) with a selection coefficient $s > 0$, while the more divergent mutations ($MS(AA_{anc}, AA_{mut}) - MS(AA_{anc}, AA_{cur}) > 1$) are deleterious with a selection coefficient $s < 0$ (fig. 2, bottom). Any mutation in between ($-1 \leq MS(AA_{anc}, AA_{mut}) - MS(AA_{anc}, AA_{cur}) \leq 1$) is considered neutral ($s = 0$) since it does not cause a radical functional change in the amino acid to increase or decrease the fitness of an individual. Below, we test the evolution of *bam* under both the "Base" and "Complex" models to investigate how implementation of different fitness parameterizations affect our simulation results.

Simulation parameters

We focused on investigating the impacts of two key parameters on the evolution of the *Drosophila* species in each of the proposed models: 1) the magnitude of the selection coefficient for both beneficial and deleterious mutations, and 2) the length of alternating *Wolbachia*-infection and *Wolbachia*-absence phases in each model in which the different selection phases occur. The absolute values of selection coefficients included $|s| = 0.1$, $|s| = 0.01$, and $|s| = 0.001$, resulting in $N_e|s| = 100$, $N_e|s| = 10$, and $N_e|s| = 1$ respectively, where $N_e|s| = 1$ can be considered effectively neutral. The lengths of different selection phases varied from equal periods of 12,500, 6,250, and 3,125 simulation generations (corresponding to one, two, and four *Wolbachia* infection-loss cycle(s) respectively in a total divergence time of 25,000 simulation generations). For each set of parameter combinations, we ran 50 independent simulations and performed downstream analyses including the MK test, inferences of α (the proportion of amino acid

fixations driven by positive selection; Smith and Eyre-Walker 2002), and average Miyata score differences between the ancestral and evolved sequences. All these downstream analyses were conducted every 3,125 simulation generations after the neutral burn-in period, by comparing the "reference sequence" in SLiM to the common ancestral sequence of *D. melanogaster* and *D. simulans*. For the MK test, 100 diploid individuals were randomly sampled from the population and nonsynonymous and synonymous fixations (relative to the inferred ancestral sequence) and polymorphisms present were tabulated.

Analyses of simulated sequences

The MK test was used to evaluate departures from an equilibrium neutral model consistent with positive selection and was implemented with a custom script modified from the iMKT package (Murga-Moreno, et al. 2019) to include mutations at 2-fold degenerate sites which the standard iMKT package implementation ignores. Polymorphisms and divergences found at 4-fold degenerate sites were considered synonymous and those found at 0-fold degenerate sites were considered nonsynonymous. If there was a polymorphism or divergence at a 2-fold site, the site was classified based on the synonymous or nonsynonymous nature of the resultant amino acid. Any sites in codons with a change at more than one position were rare in our simulations and ignored. The Fay, et al. (2001) correction for low frequency polymorphisms was applied, counting only polymorphisms > 5% frequency to avoid including deleterious variation segregating in the populations. Significance of the MK test was determined by Fisher's exact test. An estimate of the proportion of amino acid causing nonsynonymous substitutions driven to fixation by positive selection ($\alpha = 1 - \frac{D_s P_n}{D_n P_s}$) was calculated from the input values of the MK test following Smith and Eyre-Walker (2002). We also calculated the "true α" in the simulations by tracking the actual fraction of nonsynonymous substitutions in *bam* that were driven to fixation by positive selection in the simulation. Since the selection coefficient of a mutation could change as *Wolbachia* was gained and lost from the population, any mutation that once had a selection coefficient $s > 0$ and was eventually fixed in the population was regarded as being driven to fixation by positive selection. As with the estimated iMKT α, the true α is calculated for each simulation from the observed substitutions relative to the ancestral sequence. The average Miyata score calculated for each amino acid change between the simulated, evolved

sequence and the ancestral sequence was used as an assessment of physicochemical similarity between the two sequences.

Results

Buffering Models

We first simulated the Buffering Base model for a protein coding gene based on the *Drosophila bam* gene together with the cyclic pattern of *Wolbachia* infection and loss in the *Drosophila* population. When *Wolbachia* is present in the population, nonsynonymous mutations experienced neutral evolution since their deleterious effects are buffered by the bacteria. However, when *Wolbachia* is lost, mutations that brought the *bam* protein physicochemically closer to the ancestral state were positively selected, while any mutation that pushed the *bam* protein further away from the ancestry was selected against.

The Buffering Base model showed a positive true α that emerged in the first *Wobachia*-absence phase (fig 3A, row 1), consistent with our intuition of positive selection to revert *bam* back towards the functionally optimal status after it accumulated deleterious fixations buffered by *Wolbachia.* This pattern was most evident with the strongest selection: the true α's increased and stayed constant in the phases where positive selection was expected, with only a marginal decrease during the phases of neutral evolution. Weaker selection led to a smaller increase in the true alpha. Notably, longer *Wolbachia* infection periods resulted in larger true α's, presumably due to the greater time to accumulate buffered deleterious fixations in the presence of *Wolbachia*.

Regarding the estimated iMKT α, the iMKT α's were largely negative across the whole simulation in the Buffering Base model across all selection coefficients, However, changes in magnitude are evident in different selection phases, e.g., iMKT α decreased during neutral phases (*Wolbachia* present) but increased in phases with selection (*Wolbachia* absent) (fig. 3A, row 2). These observed negative iMKT estimates of α were due to the contributions from both D_n and P_n. In the initial *Wolbachia*-infection phase of the Buffering model, nonsynonymous polymorphisms were negatively selected in the constrained codons and neutrally buffered by *Wolbachia* in the codons under selection, with few such mutations in the latter category going to fixation, explaining the negative iMKT α's in the initial phases (fig 3A, row 4). Following this "buffering" period, a subset of nonsynonymous mutations was selected for. However, the

number of nonsynonymous mutations that could be positively selected in the Buffering Base model was limited, leading to a smaller D_n and thus a smaller (possibly < 0) iMKT α, even when positive selection was present as evidenced by the true α.

The boxplots of differences between the true and iMKT α's were used to evaluate the accuracy of iMKT α's. In general, iMKT α's systematically underestimate the true α's due to the presence of deleterious polymorphisms (Fay et al. 2001, Eyre-Walker and Keightley 2009, Messer and Petrov 2013), which is what we observe here (especially with $N_e|s|>1$ in the later *Wolbachia*-infection phases), with the boxplots distributed above 0. For the four MKT parameters, the final magnitude of D_n, D_s, P_n, and P_s observed at the end of the simulations were slightly impacted by the length of *Wolbachia* infection and absence periods and only P_n showed dramatic periodic fluctuations due to the cyclic infection and absence periods (fig. 3A&B, row 4).

Additionally, we looked at the distributions of p-values in iMKT (FWW correction, SNPs frequency > 5%) and the correlation between iMKT α's and their corresponding p-values at the end of the simulation. We find that even under the strongest selection in our simulations, the MK test could hardly detect any statistically significant signals of positive selection in the Buffering model, likely due in part to the modest length of the *bam* gene (fig 5A). Overall, smaller p-values were always associated with larger iMKT α's, and all the significant p-values (<0.05) were associated with iMKT α's close to 1.0 across all selection coefficients (data not shown).

All together, these results demonstrate that a modest number of amino acid fixations can occur due to selection for an optimal ancestral allele after a period of buffered mutations accumulate. However, the α estimated from iMKT did not reliably identify departures from neutrality in the direction of positive selection in the Buffering Base model.

To further explore the Buffering model and its potential to generate signals of positive selection, we also ran additional simulations with different parameterization for the selection schemes. Here, we narrowed down the Miyata score range for the positively selected nonsynonymous mutations by introducing neutral and deleterious ranges to allow for a more nuanced selection scheme, and termed the new model as the Buffering Complex model. The Buffering Complex models had lower true α's compared to Base models (fig 3A&B, row 1), with only a barely perceptible increase in true α when *Wolbachia* is lost from the population even in the strongest selection scenario of $N_e|s| = 100$. The patterns of iMKT α's and boxplots

for the difference between the true and iMKT α's were similar between Complex and Base models (fig 3A&B, row 2&3). D_n had a slight increase in the Buffering Complex model compared with the Buffering Base model, potentially due to the introduction of the neutral region leading to a small number of additional nonsynonymous fixations by genetic drift alone (fig 3A&B, row 4). The MK test still could not detect any statistically significant signals of positive selection in the Complex model (fig 5C), just like the case for the Base model.

These results suggest that restricting the degrees to which we positively select an amino acid based on their physicochemical change makes it harder to generate signatures of positive selection for *bam* in the Buffering models, especially with a limited infection period. The signatures of positive selection required both fixations of deleterious mutations during the infection and a significant amount of amino acids to be positively selected due to their ability to revert mutated proteins back to their optimal functions.

Conflict Models

To compare the results of our proposed Buffering model with that of a more established mode of interaction regarding genetic conflict between two species, we also designed and implemented a corresponding Conflict model in which positive selection acted on nonsynonymous mutations in the presence of *Wolbachia* and purifying selection was imposed on nonsynonymous mutations in the absence of *Wolbachia*.

We find that the patterns of true α's were clearly indicative of positive selection in the Conflict Base model phase with *Wolbachia* with the strongest selection (fig 4A, row 1). The elevated true α's persisted, though with a slow decline during the subsequent *Wolbachia*-free phase. Positive selection in the *Wolbachia* infection phases increased α or kept it as a constant, while purifying selection in the *Wolbachia* absence phases decreased α marginally.

The average of iMKT estimates of α were almost all positive for selection coefficients with $N_e|s|>1$ and showed clear periodic changes as *Wolbachia* comes in and out of the population across all three phase lengths (fig. 4A, row 2). Surprisingly, the magnitude of the iMKT α's increased in the phase without the imposed positive selection. This unexpected increase is explained by the change of nonsynonymous polymorphisms (P_n) in the population. In the *Wolbachia*-infection phase, both D_n and P_n accumulated due to the positive selection of nonsynonymous mutations, as expected; however, after the sudden change to the *Wolbachia*-

absence phase, D_n was largely unchanged while P_n experienced a sudden decrease as nonsynonymous mutations were all selected against (fig 4A, row 4). Given the equation for calculating the iMKT α ($\alpha = 1 - \frac{D_s P_n}{D_n P_s}$), the iMKT estimate of α therefore increased in the phase with the implemented purifying selection that followed the positive selection. For the effectively neutral case of $N_e|s| = 1$, the iMKT $\alpha's$ in the Conflict Base model fluctuated around 0.

For the Conflict Base model with effectively neutral evolution ($N_e|s|$=1), iMKT α's usually underestimated the true α's (fig 4A, row 3). However, under stronger selection ($N_e|s|$=10 or 100), iMKT α's underestimated the true α's only during the *Wolbachia*-infection phase; there was good accuracy in iMKT α's estimation when *Wolbachia* was lost, which reflected the delay in detecting selection based on changing P_n as previously explained. Besides, the pattern of the true α's and the iMKT α's was not dramatically influenced by the magnitude of the selection coefficient ($N_e|s|$=10 or 100) and the varying lengths of the infection/absence periods that we examined.

When we looked at the distributions of p-values in iMKT (FWW correction, SNPs frequency > 5%) and the correlation between iMKT α's and their corresponding p-values at the end of the simulation. In general, a statistically significant rejection of neutrality in the direction of positive selection was more likely to be detected with the iMKT in the Conflict Base model than in the Buffering Base model with the iMKT, since the values of the key MK test parameter D_n are generally much larger in both *Wolbachia*-infection and *Wolbachia*-absence phases in the Conflict model. This increased magnitude of D_n provided more statistical power in Fisher's exact test (fig. 5B). Overall, smaller p-values were always associated with larger iMKT α's, and all the significant p-values (<0.05) were associated with iMKT α's close to 1.0 across all selection coefficients (data not shown).

We also implemented the Conflict model with a more complex parameterization of selection schemes based on Miyata scores and termed it as the Conflict Complex model. The MK test results for these Conflict Complex models closely resembled those for the Base models. Since we narrowed down the Miyata score range for the positively selected nonsynonymous mutations in both complex models by introducing neutral and deleterious ranges, we observed lower true α's compared to their Base models (fig 4A&B, row 1). The patterns of iMKT α's and boxplots for the difference between the true and iMKT α's were similar between Complex and Base models (fig 4A&B, row 2&3). The total number of D_n did not reach the same magnitude at

the end of simulation for the Conflict Complex model as it did in Conflict Base model across different phase lengths and selection coefficients (fig 4A&B, row 4). However, the MK test still detects statistical signals of positive selection under the first two selection coefficients in the Complex model (fig 5D).

Distributions of Miyata scores under Buffering and Conflict models

We have observed that Buffering models could generate signatures of positive selection based on true α's, but the signals were harder to be detected compared to the cases in the Conflict models. We expect amino acid substitutions to be more diversified in the Conflict model than in the Buffering model in both the Base and Complex cases, as the former is based on the premise of a sequence evolving *away* from the ancestral sequence and the latter is based on the premise of a sequence evolving *toward* the ancestral sequence. To assess this, we calculated Miyata scores between each amino acid substitution and its ancestral amino acid to represent their physiochemical differences.

For both the Base and Complex cases, the distributions of Miyata scores per amino acid from the Conflict and Buffering models were most distinguishable from each other in the strongest selection scenario at $N_e|s| = 100$. Here, the interquartile ranges of Miyata score distributions of the two models were completely separated at the end of simulations (fig6A, 6B; row 1), but they were basically indistinguishable from each other throughout the simulations when the selection is the weakest at $N_e|s| = 1$ (fig6A, 6B; row 3).

For $N_e|s| = 10$, the distributions of Miyata scores overlap more in the Base case compared with the Complex case for both Buffering and Conflict models (fig6A, 6B; row 2) because the positively selected mutations had a higher concentration of Miyata scores between 1 and 3 in the Complex case, which made the differences between Miyata scores more prominent. The same results can be observed in the strongest selection simulations, where the distance between interquartile ranges was also larger in the Complex cases than in the Base cases of each model.

In summary, we are better able to distinguish between the Conflict and Buffering models using Miyata score distributions when there is strong selection. Our ability to discern these two models also depends on the specific parameterizations of the selection scheme. When the amino acids were positively selected based on similar degrees of change in physicochemical properties,

the distributions of Miyata scores overlap with each other to a significant extent; while the distributions would be more distinguishable if the Conflict models favored more diversified amino acids than the Buffering models. Different infection/absence phase lengths did not have a large impact on the average Miyata scores across the simulations.

Comparison with the empirical data

To evaluate which model in our analysis better captures *bam's* observed patterns of sequence evolution within and between natural populations of *Drosophila*, we first performed our MK test on a population sample (n=89) of *D. melanogaster* (Lack et al. 2015), using divergence to the predicted common ancestral sequence with *D. simulans* as the outgroup and a randomly sampled sequence as the reference sequence used in the iMKT estimate. Analysis of these data reject neutrality in the direction of positive selection using the MK test with a p = 0.00015 and α estimated to be 0.91 (FWW correction, SNPs > 5% only). We used the number of nonsynonymous substitutions per nonsynonymous site (dN) calculated from iMKT results as the summary statistic to tune selection parameters of the two simulation models with only one *Wolbachia* infection-loss cycle.

While examining polymorphism levels would seem important to distinguish between Buffering and Conflict models, these levels are very sensitive to the length of *Wolbachia* infection and absence as we have modeled it, for example, due to the strong purifying selection occurring in the Conflict model when *Wolbachia* is lost. Thus, determining the time point for sampling is problematic for the empirical data as we do not know for a species that is infected with *Wolbachia* how long it has been infected, nor do we know for uninfected species when the last time they were infected (or even if they were). The problematic effect of this timing choice on P_n and P_s can be seen in Figures 3 and 4. As D_n is less sensitive to the sampling time points and represents the number of amino acid changes in *bam*, we chose to only use this parameter to evaluate how well our models fit to empirical data.

Applying our custom iMKT script to our empirical sequence data, we found that iMKT D_n = 34 and dN = 0.033 for the empirical *D. melanogaster* population. We initially found that Conflict models always predicted a much higher D_n than the empirical observation, while Buffering models often exhibited a much lower D_n. Such results showed the initially assumed ratio of codons under selection (RS=15%) and ratio of codons under constraints (RC=75%)

could not reproduce similar results for the evolution of amino acids for either model. Thus, we chose to tune these two ratios of selected and constrained codons (RS and RC) under different strengths of positive selection ($N_e|s|$ = 100, 10, 1) and explore under which parameter settings could we fit the empirical dN in each of our proposed models. When we achieved a matching dN, we then compared the Miyata scores per amino acid change in the observed data and our simulation results and see whether a Conflict or a Buffering model is more similar to our observations.

For each selection coefficient *s*, we first ran simulations using RS and RC both sampled from a uniformly distributed grid of nine points ranging from 0 to 0.8, since the maximum proportion of conserved codons is 0.85, and assessed the resulting dN. Preliminary simulations revealed that for the Conflict models, dN was consistently more than two-fold overestimated for any proportion of selected sites greater than 0 (e.g., RS > 0, data not shown). Therefore, we refined the Conflict model grid search for RS to a uniform grid of 6 points from [0, 0.1], while keeping the full grid range for RC. For the Buffering models, we kept the full range of the RS grid as we did find parameters that fit the observed dN. For each pair of parameters for all models, we ran 50 simulations and calculated the mean of dN ($\bar{dN}$) across the runs. We then compared the difference between the empirical dN and $\bar{dN}$. The best pair of RS and RC was the one that led to the smallest difference between $\bar{dN}$ and the empirical dN under each selection coefficient *s*.

For the models with selection coefficient $N_e|s|$ = 1, all combinations of the two ratios reproduced similar results consistent with effective neutrality (fig 7). For moderate or strong selection, the best-fit parameters are shown in Table 2.

Analyses of Buffering and Conflict Models Best Fitting the Empirical Data

To evaluate how well the Buffering and Conflict models implemented with the best-fit pairs of RS and RC recapitulate the empirical data for *D. melanogaster*, we performed the same iMKT analysis and Miyata score analysis for the resulting simulations. Positive true α's were observed in the Buffering Base, Conflict Base, and Conflict Complex models across different phase lengths, indicating that positive selection was present under these scenarios. However, iMKT could only identify positive selection by α and statistically significant p-values in the two Conflict models with strong selection at $N_e|s|$ = 100. Moderate selection at $N_e|s|$ = 10 in the

Conflict models or any levels of selection in the two Buffering models was not detected by iMKT p-values or inferred α (fig. 8).

In addition, we calculated the empirical per-site Miyata scores between the current *D. melanogaster* sequences and their predicted common ancestral sequence shared with *D. simulans* (supplemental file S2), and compared it with the distributions of per-site Miyata scores simulated from the best-fitted RS and RC at different timepoints. The end of the simulations at 45,000 scaled generations represents the actual divergence time between the ancestral sequence and the extant *D. melanogaster* and *D. simulans* species. At this time point, the interquartile ranges of Miyata scores of the Buffering and Conflict models have separated from each other, with fully non-overlapping interquartile distributions in the Complex models. In all models, the per-site Miyata score of *D. melanogaster* are located closer to the center of the distributions from the Buffering models than to the center of distributions from the Conflict models (Figure 9).

In summary, the best-fit Conflict models with strong selection reproduced the most significant iMKT p-values and high estimates of α like those observed in the *D. melanogaster* sample, but the Miyata-score analysis indicated the Buffering models as a better fit for the evolution of the amino acids' biochemical properties. It is important to note that while the average iMKT α is extremely close to zero for all Buffering models, the lower whiskers on the box plots in fig 8A and 8C show that high iMKT inferred α's can, although infrequently, occur under the Buffering models as well.

Discussion

The increasing availability of DNA sequence datasets for diverse genes, genomes and organisms has led experimentalists to scan genes and genomes for footprints of natural selection. Using tests like the MK test, a striking number of cases of departures from neutrality have been revealed (Eyre-Walker 2006). In some cases, statistical evidence of strong positive selection can be easily associated with a proposed causal factor (e.g., genetic conflict associated with genes involved in antiviral immunity or in mating behavior; McLaughlin and Malik 2017). In other cases, the driving factor is less clear.

In this study, we proposed an evolutionary Buffering model to explore a novel hypothesis on drivers of positive selection based on periods of varying selective constraint. We implemented the model in simulations to examine its ability to generate signatures of positive selection. We

were motivated by the partial rescue by *Wolbachia* infection of fertility defects associated with mutations at *bam* and *Sxl* in *D. melanogaster*, and thus first discuss our findings with respect to our modeling of *bam*. We then discuss how are our findings may apply more broadly to other genes and organisms.

The Buffering model is based on the observations that *Wolbachia* protects the functions of *bam* and *Sxl* from the effects of deleterious mutations thereby allowing these mutations to accumulate during the *Wolbachia* infection phase by drift (equivalent to the relaxation of functional constraints for amino acid mutations). When *Wolbachia* is lost, the constraints are reimposed and amino acids similar to *bam*'s and *Sxl*'s ancestral state are selected for. We compared this model to the Conflict model, which is an implementation of traditional arms race dynamics to describe the interactions between competing symbiotic species and serves as a baseline model of positive selection for comparison.

We used simulations to study the evolutionary process involved in each model and found that both models can generate positive selection as measured by the true α. Therefore, the positive true α in the Buffering models reveals that *Wolbachia* need not function as a reproductive parasite in conflict with a gene like *bam* to drive positive selection in the host gene.

Importantly, in all simulations the iMKT α generally underestimated the true α. This underestimation had a minimal effect on our interpretation of evolution in the Conflict models, as the true α was very large in all simulations outside of those with the weakest selection. On the other hand, with a maximum true α of ~ 0.25 in the Buffering models' simulations, an underestimation led to a weak or absent signal of positive selection detectable by iMKT, which could further be confounded by statistical noise. Such findings highlight some limitations of the MK test that are consistent with the findings of others (Akashi 1999, Fay et al. 2001, Eyre-Walker and Keightley 2009, Zhai et al. 2009, Messer and Petrov 2013). Even with these limitations, the iMKT could still infer high α's, representing detection of positive selection, in some simulation runs under the Buffering models.

The Buffering models require the fixation of *Wolbachia*-buffered deleterious nonsynonymous mutations by drift for there to be resulting positive selection during a subsequent phase without *Wolbachia*. This effect is seen across the three different infection lengths that we simulated in the Buffering Base model. Longer *Wolbachia* infection phases increase the chance of detecting positive selection in a subsequent *Wolbachia* absence phase,

though never to the level resulting from Conflict models. The average length of *Wolbachia* infection time is unknown for *Drosophila*, but two independent studies that found the *w*Mel *Wolbachia* variant to have been in *D. melanogaster* for at least 79,000 and 80,000 *Drosophila* generations (Richardson et al. 2012; Choi and Aquadro, 2014). These time periods are shorter than what we have simulated, but there is evidence to suggest turnover of *Wolbachia* variants that could act as a longer standing infection period than currently documented (Riegler et al. 2005, Kriesner et al. 2013). Thus, *Wolbachia* infection of the length we have simulated, and with it a potential for subsequent positive selection, is not out of question.

To better evaluate the fit of the observed data from *D. melanogaster* to the predictions of the Buffering and Conflict models, we tuned the simulation selection parameters of both models to fit the observed nonsynonymous sequence divergence per nonsynonymous site (dN) between *D. melanogaster* and the inferred common ancestor with *D. simulans*. Explorative simulations are reflected in the empirically tuned simulations. Only the tuned Conflict model recapitulated the statistically significant positive iMKT α's that we observed for the *D. melanogaster* population. As in the general Buffering results discussed above, the tuned Buffering model can result in evidence of positive selection as indicated by a true α under certain conditions, but we can rarely detect it with the iMKT in *bam* with statistical significance. For the Miyata score analysis, we found that the Buffering models better fit our empirical data, as the Conflict models predicts greater amino acid diversity than we observe. We expect that a restriction of the positively selected amino acids in the Conflict model to only those physicochemically similar to the current states would bring the Miyata score analysis in line with our observed amino acid diversity without compromising the high iMKT α that is generated. Thus, combining these results, we suggest that the Buffering model is a possible, but unlikely, explanation behind the observed evolution in the *D. melanogaster bam* gene. This is particularly the case as a p-value less than or equal to 0.05 for the empirical MKT result is the typical criteria used by experimentalists to infer a departure from an equilibrium neutral model. Thus, with the current assumptions of our models, the Conflict model of an arms race between *Wolbachia* and *bam* is the better explanation for the signature of selection that we observe at *bam*.

We note that the best fit results for all models come with parameterizations that include a considerable proportion of neutral sites. This suggests that our model is missing important subtleties behind the evolution of *bam*. For instance, we have only used fixed selection

coefficients and strict Miyata score cutoffs throughout our simulations to model the selection coefficients for both beneficial and deleterious mutations, when they could be drawn from some distribution. It is also possible that a mixture of Buffering and Conflict models may be operating, with each driving evolution at a subset of sites.

With regard to resolving the evolutionary interactions between *bam* and *Wolbachia* in *Drosophila*, it will now be important to explore other experimental evidence with respect to potential conflict, change in function or buffering effects due to relaxed selective constraints. For example, we can further evaluate the contributions of buffering and conflict by testing for positive selection in *Wolbachia* genes, as positive selection in *Wolbachia* is expected under the Conflict, but not the buffering model. In conflict, *Wolbachia* would co-evolve with *bam* to continue its impact on *Drosophila* fertility. There is already some evidence of positive selection at a few genes across different *Wolbachia* strains of arthropods and nematodes (Baldo et al. 2002; Baldo et al. 2010) but a much more thorough analysis of closely related *Wolbachia* strains infecting *D. melanogaster* is needed. To test for a conflict-like interaction between germline stem cell genes and *Wolbachia*, signals of selection in *Wolbachia* need to be examined solely within the *Drosophila* genus, since *Wolbachia* has a very different relationship with its nematode hosts (Taylor, et al. 2005).

We also want to emphasize that a better fit to the Conflict compared to the Buffering model would not by itself imply a conflict drives the positive selection observed. A change in function that favors diversification of the protein-coding gene would also give similar results because, like the Conflict model, selection to refine a new function would likely favor positive selection for physicochemically different amino acids, which is selection for increasingly diverse Miyata scores. A recent analysis of CRISPR/Cas-9 generated nulls in five *Drosophila* species raises this possibility for *bam* (Bubnell, et al. 2021). Whether the observed changes in function are associated with conflict with *Wolbachia* remains an open question as the two are not mutually exclusive.

As we show the Buffering model can generate signals of positive selection in *bam*, we consider the applicability of the Buffering model beyond this case. Other cases of *Wolbachia* interacting with the *Drosophila* germline include the rescue of *Sxl* hypomorphs (Starr and Cline 2002) and one *mei-P26* hypomorph (unpublished). *Wolbachia* also increases the fecundity of *D. mauritiana* (Fast et al. 2011). These examples suggest *Wolbachia* could function in a “buffering”

manner to increase the fitness of its host across a variety of situations, especially given the high prevalence of *Wolbachia* across different species (Hilgenboecker et al. 2008; Zug et al. 2012; Weinert et al. 2015).

The Buffering model may also be pertinent to other facultative symbiotic relationships wherein one organism acts as a defensive symbiont by providing protection from a natural enemy for the other species (reviewed in Clay et al. 2014). In the presence of the defensive symbiont, the host could experience relaxed constraint on immune or other relevant genes, akin to the relaxed constraint we model on *D. melanogaster* reproductive genes in the presence of *Wolbachia*. The presence of a defensive symbiont may come at a cost to the other species and may be lost if the natural enemy is lost. This loss would lead to a reestablishment of selective constraint on the relevant genes resulting in selection favoring new mutations reversing or compensating for harmful mutations fixed during the period of relaxed constraint.

Species with smaller population sizes may present better opportunities for detection of positive selection from the Buffering model. Drift during the "buffering" phase is what allows the buffered deleterious nonsynonymous mutations to fix and the time to fixation of neutral mutations is approximately $4N_e$ generations (Kimura and Ohta 1969). Thus, a smaller N_e will allow for more chances of deleterious mutations fixing and subsequently being under selection to return to an ancestral state. Additionally, genes of a greater length are of interest as there is potentially more statistical power to detect positive selection resulting from the Buffering model.

We suggest that the Buffering model is a new entry to the suite of models that need to be considered in cases where molecular population genetic evidence is found for departures from selective neutrality consistent with positive selection. The Buffering model is a framework that could apply to populations that experience cycles of higher mutational loads, followed by positive selection. This is observed in seasonally small populations, where a drop in population size allows the fixation of some deleterious alleles that are subsequently purged from the population. Additionally, populations in changing environments may experience higher mutational loads at the onset of the change. This phenomenon would be similar to that of antagonistic pleiotropy, in which, for our case, one environment is more tolerant of various alleles, allowing some alleles to fix that would be considered deleterious in the subsequent environment. In the subsequent environment, positive selection favors new mutations that return the gene to its optimum sequence (Chen and Zhang 2020).

Supplemental Materials:

Supplemental materials are available at _____.

Data Availability

No new sequencing data were generated in this work; the employed data sets are listed throughout the text. Sequence alignments used are available as supplementary file S2, Supplementary material online. SLiM3 and python code for analyses used in this study are available online at github.com/runxi-shen/Modeling-Evolution-at-bam.

References

Akashi H. 1999. Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics 151:221-238.

Bauer DuMont VL, Flores HA, Wright MH, Aquadro CF. 2007. Recurrent positive selection at Bgcn, a key determinant of germ line differentiation, does not appear to be driven by simple coevolution with its partner protein bam. Mol Biol Evol 24:182-191.

Bauer DuMont VL, White SL, Zinshteyn D, Aquadro CF. 2021. Molecular population genetics of Sex-lethal (Sxl) in the Drosophila melanogaster species group: a locus that genetically interacts with Wolbachia pipientis in Drosophila melanogaster. G3 (Bethesda) 11.

Bubnell JE, Fernandez-Begne P, Ulbing CKS, Aquadro CF. 2021. Diverse wMel variants of Wolbachia pipientis differentially rescue fertility and cytological defects of the bag of marbles partial loss of function mutation in Drosophila melanogaster. G3 (Bethesda) 11.

Campos JL, Zhao L, Charlesworth B. 2017. Estimating the parameters of background selection and selective sweeps in Drosophila in the presence of gene conversion. Proc Natl Acad Sci U.S.A. 114:E4762-E4771.

Charron C, Nicolaï M, Gallois JL, Robaglia C, Moury B, Palloix A, Caranta C. 2008. Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. Plant J 54:56-68.

Chen P, Zhang J. 2020. Antagonistic pleiotropy conceals molecular adaptations in changing environments. Nat Ecol Evol 4:461-469.

Clark NL, Aagaard JE, Swanson WJ. 2006. Evolution of reproductive proteins from animals and plants. Reproduction 131:11-22.

Comeron JM, Ratnappan R, Bailin S. 2012. The Many Landscapes of Recombination in Drosophila melanogaster. PLOS Genetics 8:33-35.

Cosby RL, Chang N-C, Feschotte C. 2019. Host-transposon interactions: conflict, cooperation, and cooption. Genes Dev. 33:1098-116.

Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL. 2013. Dual Host-Virus Arms Races Shape an Essential Housekeeping Protein. PLoS Biol 11:e1001571.

Fay JC, Wyckoff GJ, Wu C-I. 2001. Positive and Negative Selection on the Human Genome. Genetics 158:1227.

Flores HA, Bubnell JE, Aquadro CF, Barbash DA. 2015. The Drosophila bag of marbles Gene Interacts Genetically with Wolbachia and Shows Female-Specific Effects of Divergence. PLOS Genetics 11:1-31.

Flores HA, DuMont VLB, Fatoo A, Hubbard D, Hijji M, Barbash DA, Aquadro CF. 2015. Adaptive evolution of genes involved in the regulation of germline stem cells in Drosophila melanogaster and D. simulans. G3 (Bethesda) 5:583-592.

Haller BC, Galloway J, Kelleher J, Messer PW, Ralph PL. 2018. Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Mol Ecol Resour 19:552-566.

Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. Genetics 196:313-320.

Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. Genome Res 19:1195-1201.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015. The drosophila genome nexus: A population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. Genetics 199:1229-1241.

Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. Methods in molecular biology (Clifton, N.J.) 1079:155-170.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654.

McLaughlin RN, Malik HS. 2017. Genetic conflicts: The usual suspects and beyond. J Expe Biol 220:6-17.

Meany MK, Conner WR, Richter SV, Bailey JA, Turelli M, Cooper BS. 2019. Loss of cytoplasmic incompatibility and minimal fecundity effects explain relatively low Wolbachia frequencies in Drosophila mauritiana. Evolution 73:1278-1295.

Miyata T, Miyazawa S, Yasunaga T. 1979. Two types of amino acid substitutions in protein evolution. J Mol Evol 12:219-236.

Murga-Moreno J, Coronado-Zamora M, Hervas S, Casillas S, Barbadilla A. 2019. iMKT: The integrative McDonald and Kreitman test. Nucleic Acids Res 47:W283-W288.

Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr Biol 16:580-85.

Obbard DJ, Welch JJ, Kim K-W, Jiggins FM. 2009. Quantifying adaptive evolution in the Drosophila immune system. PLoS Genet: e1000698.

Ohlstein B, Lavoie CA, Vef O, Gateff E, McKearin DM. 2000. The Drosophila cystoblast, differentiation factor benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics 155:1809-1819.

Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM. 2012. Population Genomics of the Wolbachia Endosymbiont in Drosophila melanogaster. PLoS Genet 8.

Russo CA, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol 12:391-404.

Shultz AJ, Sackton T. 2019. Immune genes are hotspots of shared positive selection across birds and mammals. Elife 8:e41815.

Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in Drosophila. Nature 415:1022-1024.

Starr DJ, Cline TW. 2002. A host-parasite interaction rescues Drosophila oogenesis defects. Nature 418:76-79.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat. Rev. Genet. 3:137-144.

Taylor MJ, Bandi C, Hoerauf A. 2005. Wolbachia bacterial endosymbionts of filarial nematodes. Adv Parasitol 60:245-284.

Turelli M, Cooper BS, Richardson KM, Ginsberg PS, Peckenpaugh B, Antelope CX, Kim KJ, May MR, Abrieux A, Wilson DA, et al. 2018. Rapid Global Spread of wRi-like Wolbachia across Multiple Drosophila. Curr Biol 28:963-971.e968.

Turner LM, Hoekstra HE. 2008. Causes and consequences of the evolution of reproductive proteins. Int J Dev Biol 52:769-780

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586-1591.

Zhai W, Nielsen R, Slatkin M. 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. Mol Biol Evol 26:273-283.

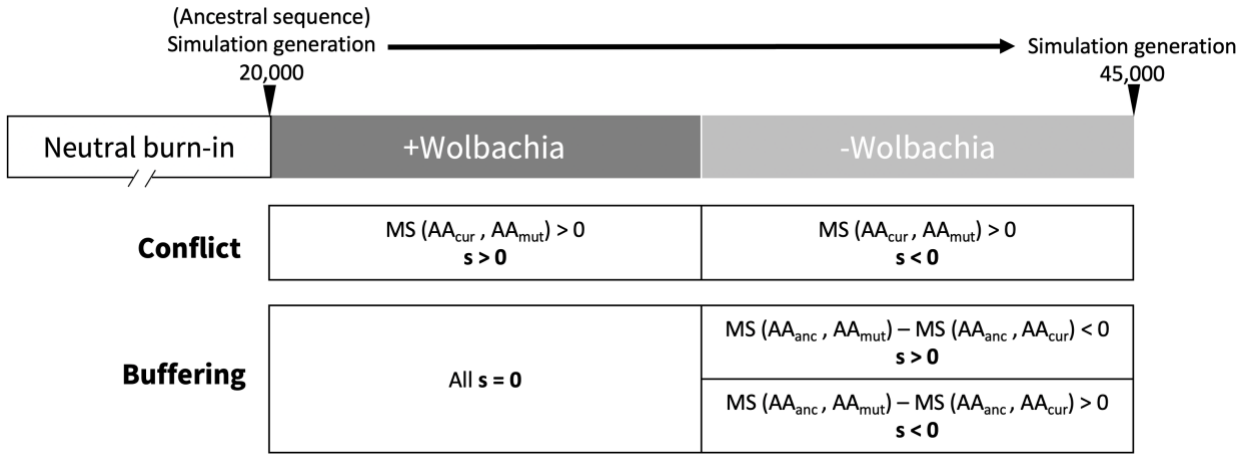


Fig. 1. Simulation setup for Conflict and Buffering Base models. Selection on new nonsynonymous mutations (mutated amino acid, AA_{mut}) is determined by their Miyata score (MS) to the appropriate reference amino acid (the current amino acid, AA_{cur} , or the ancestral amino acid, AA_{anc}).

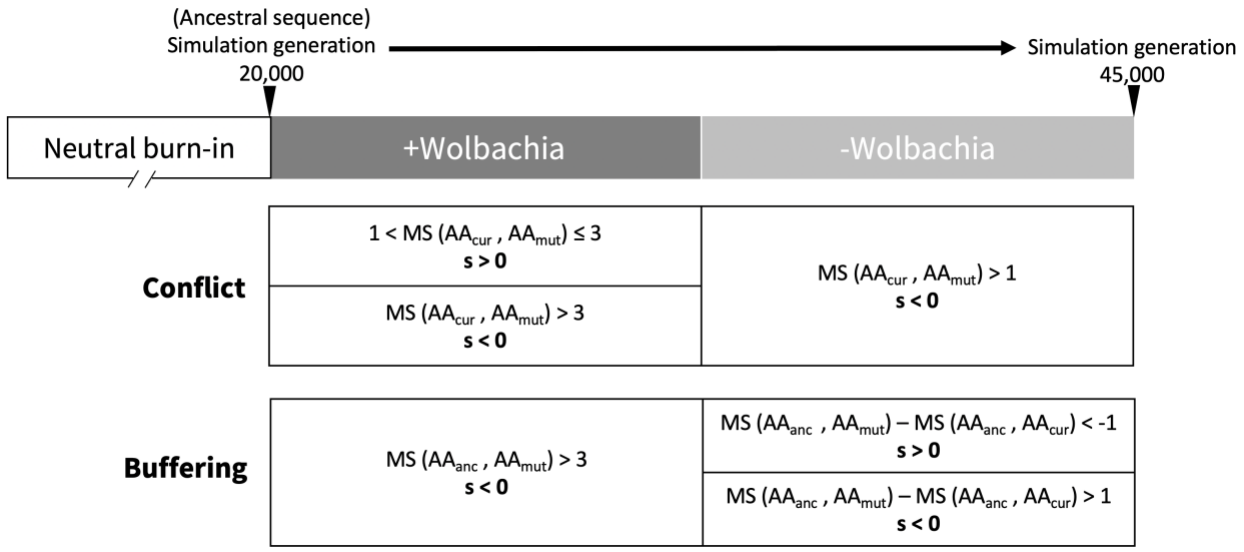
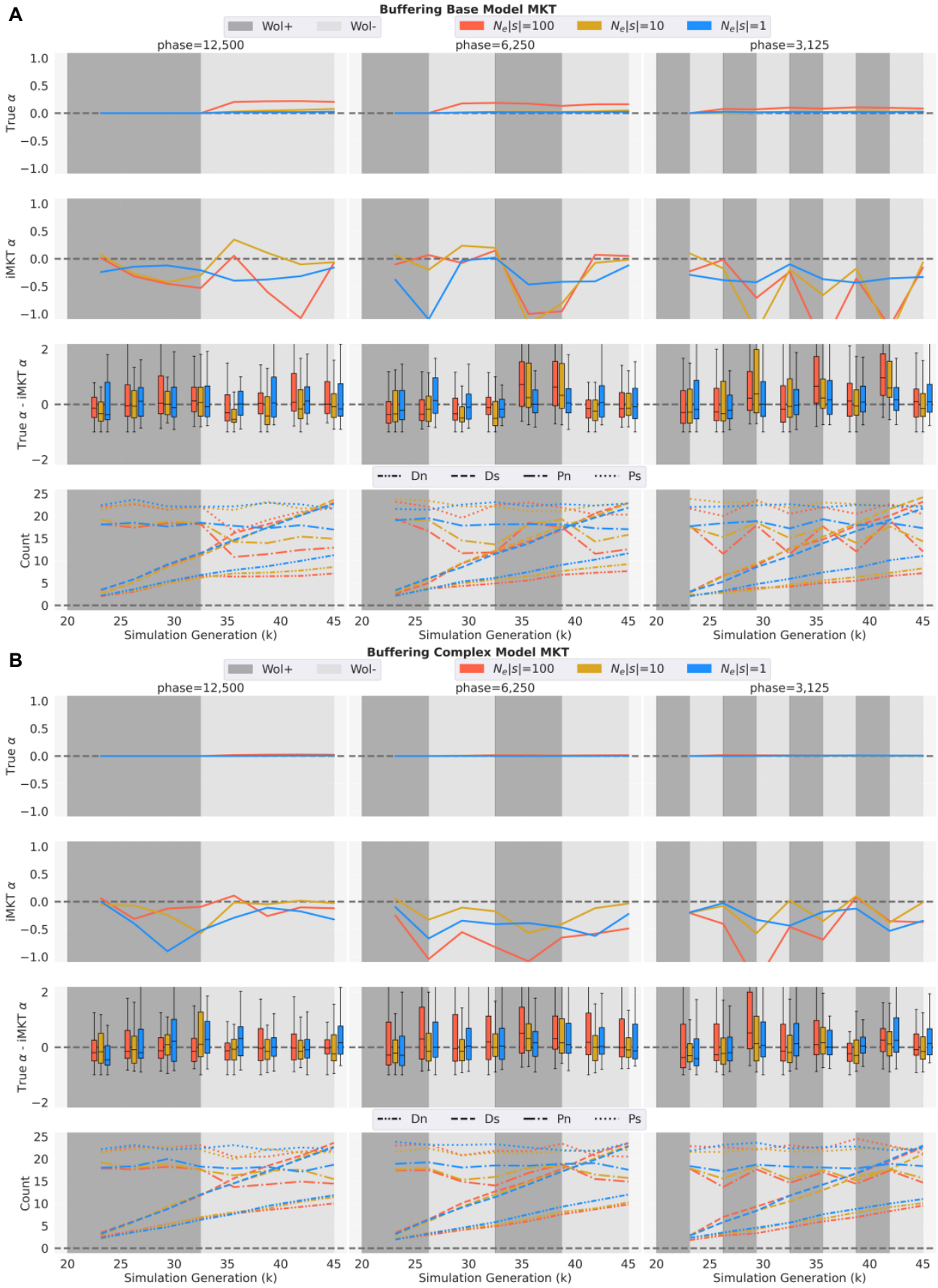


Fig. 2. Simulation setup for Conflict and Buffering Complex models. Selection on new nonsynonymous mutations (mutated amino acid, AA_{mut}) is determined by their Miyata score (MS) to the appropriate reference amino acid (the current amino acid, AA_{cur}, or the ancestral amino acid, AA_{anc}).


A
Buffering Base Model MKT
Wol+
Wol-
$N_e|s|=100$
$N_e|s|=10$
$N_e|s|=1$
phase=12,500
phase=6,250
phase=3,125
True α
iMKT α
True α - iMKT α
Dn
Ds
Pn
Ps
Count
Simulation Generation (k)
B
Buffering Complex Model MKT
Wol+
Wol-
$N_e|s|=100$
$N_e|s|=10$
$N_e|s|=1$
phase=12,500
phase=6,250
phase=3,125
True α
iMKT α
True α - iMKT α
Dn
Ds
Pn
Ps
Count
Simulation Generation (k)

807

Fig. 3. MK test results of simulations for Buffering Base and Complex models. (A) Buffering Base model; (B) Buffering Complex model. Each panel shows MKT analyses with different selection coefficients of $N_e|s|=100$, $N_e|s|=10$, and $N_e|s|=1$ graphed across alternating phases (phase length=12,500, 6,250, 3,125 and simulation generations) of *Wolbachia* infection (Wol+, dark grey) and *Wolbachia* absence (Wol-, light grey) post-burn-in period. In each panel, row 1: the average true α in the simulations; row 2: the average iMKT α in the simulations (FWW correction, SNPs frequency > 5% only); row 3: the distributions of differences between the true and iMKT α every 3,125 simulation generations; row 4: The average of each iMKT component (D_n, D_s, P_n, P_s).

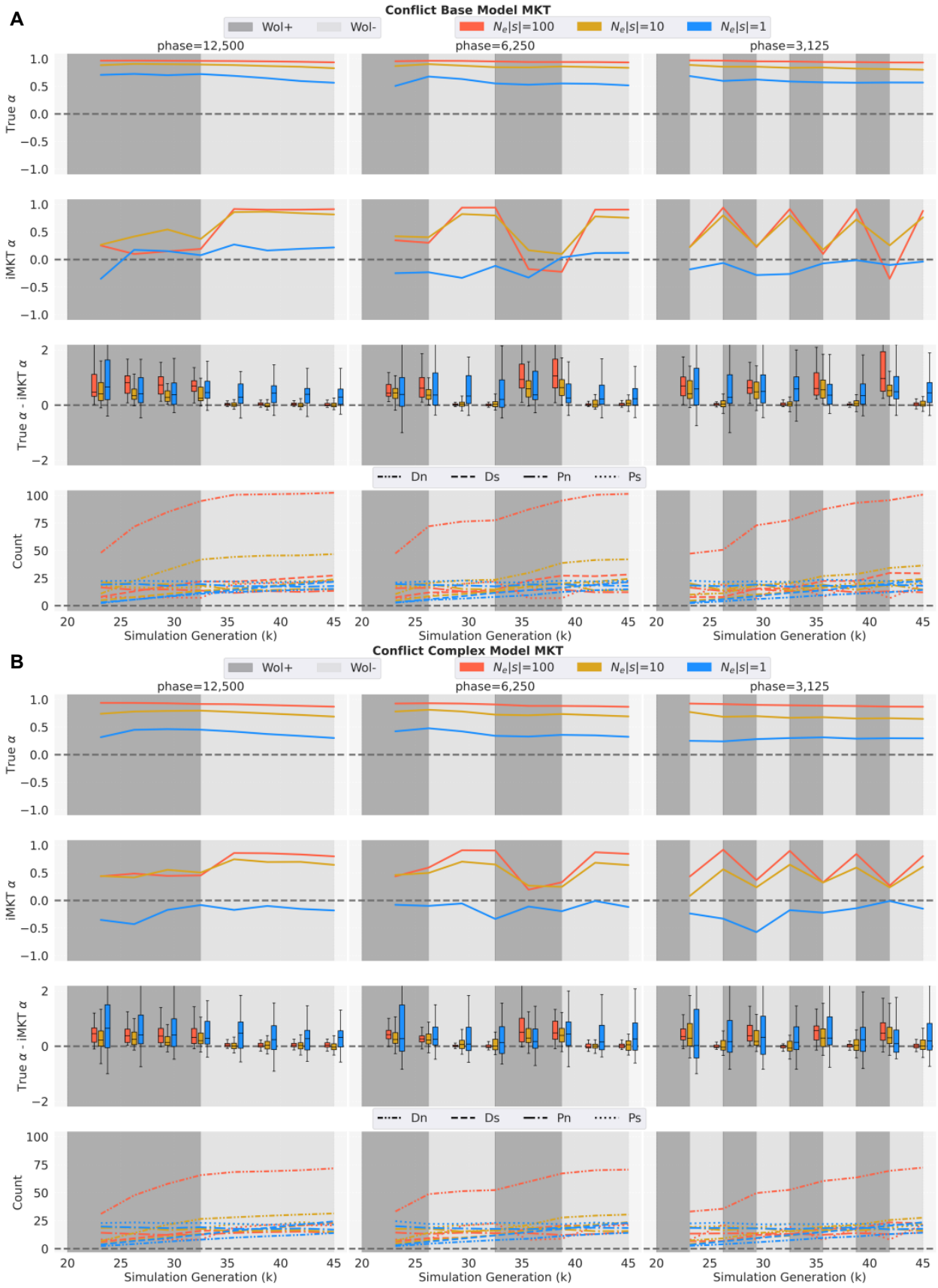

A
Conflict Base Model MKT
Wol+
Wol-
$N_e|s|=100$
$N_e|s|=10$
$N_e|s|=1$
phase=12,500
phase=6,250
phase=3,125
True α
iMKT α
True α - iMKT α
Dn
Ds
Pn
Ps
Count
Simulation Generation (k)
B
Conflict Complex Model MKT

Fig. 4. MK test results of simulations for Conflict Base and Complex models: (A) Conflict Base model, (B) Conflict Complex model. Each panel shows MKT analyses with different selection coefficients of $N_e|s|$=100, $N_e|s|$=10, and $N_e|s|$=1 graphed across alternating phases (phase length=12,500, 6,250, 3,125 and simulation generations) of *Wolbachia* infection (Wol+, dark grey) and *Wolbachia* absence (Wol-, light grey) post-burn-in period. In each panel, row 1: the average true α in the simulations; row 2: the average iMKT α in the simulations (FWW correction, SNPs frequency > 5% only); row 3: the distributions of differences between the true and iMKT α every 3,125 simulation generations; row 4: The average of each iMKT component (D_n, D_s, P_n, P_s).

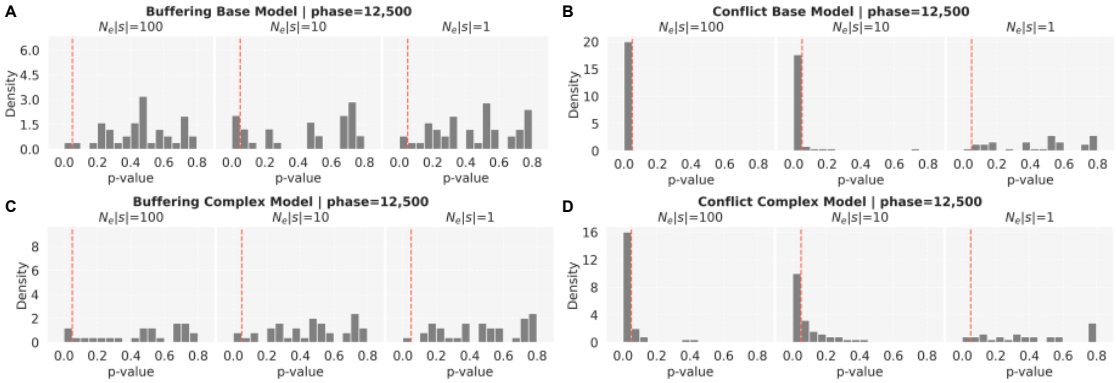
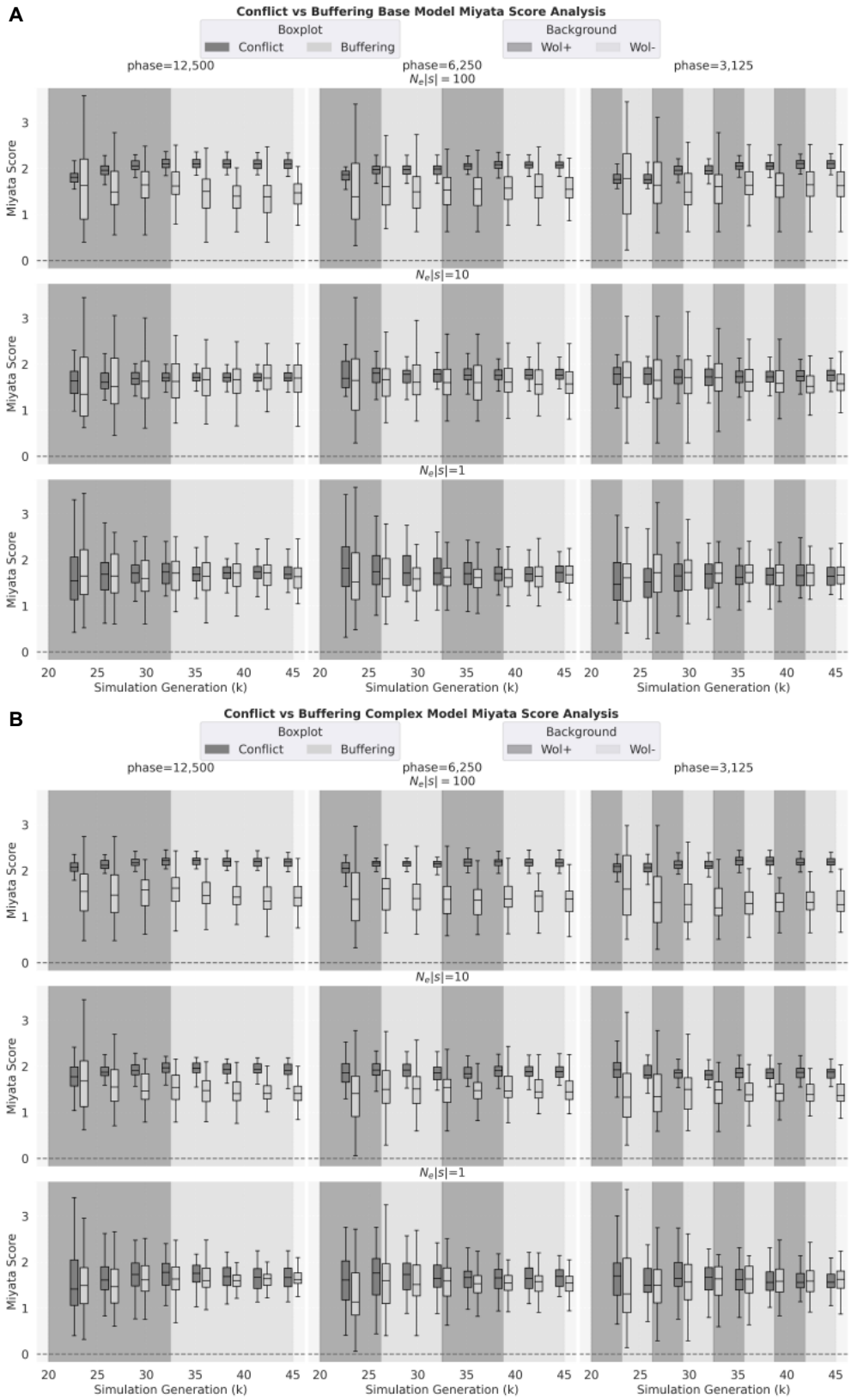
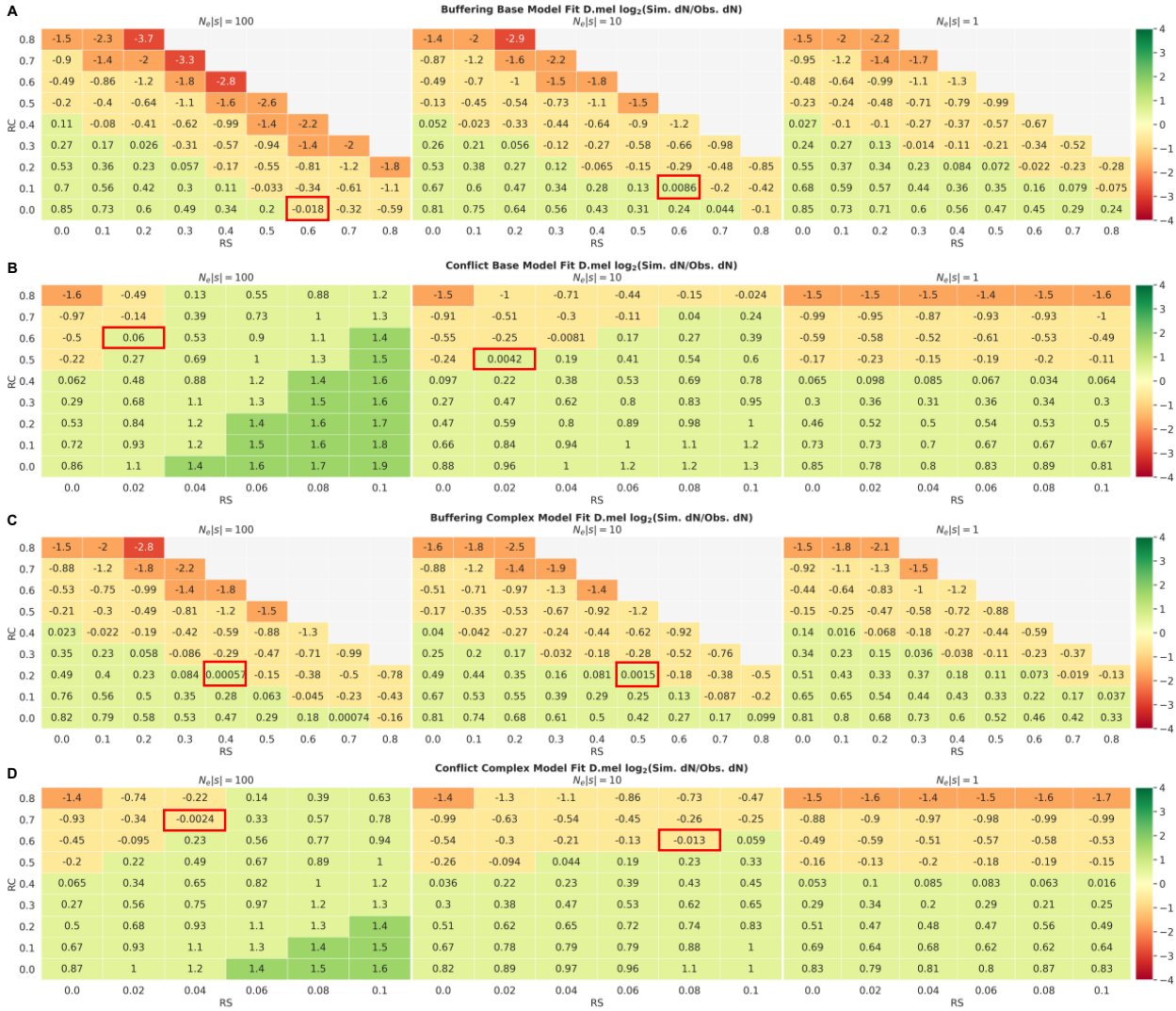


Fig. 5. Distributions of iMKT p-values. iMKT p-values (FWW correction, SNPs frequency > 5% only) for simulation runs with *Wolbachia* phase=12,500 simulation generations and $N_e|s|$=100, 10, 1 for the simulated models at 45,000 simulation generation. The vertical red line denotes p-value=0.05. Note that distributions are normalized to have an area of 1 under the histograms. (A) Buffering Base model; (B) Conflict Base model; (C) Buffering Complex model; (D) Conflict Complex model.


A
Conflict vs Buffering Base Model Miyata Score Analysis
Boxplot
Conflict
Buffering
Background
Wol+
Wol-
phase=12,500
phase=6,250
phase=3,125
$N_e|s| = 100$
$N_e|s|=10$
$N_e|s|=1$
Miyata Score
Simulation Generation (k)
B
Conflict vs Buffering Complex Model Miyata Score Analysis
Boxplot
Conflict
Buffering
Background
Wol+
Wol-
phase=12,500
phase=6,250
phase=3,125
$N_e|s| = 100$
$N_e|s|=10$
$N_e|s|=1$
Miyata Score
Simulation Generation (k)

835

Fig. 6. The distribution of Miyata scores per amino acid substitution. Miyata scores per amino acid substitution across multiple runs for substitutions between the consensus sequence at the end of a given simulation generation and the ancestral sequence. Data is shown for both the Conflict model (dark grey) and Buffering model (light grey) at every 3,125 simulated generations post burn-in for different phase lengths.



841

842 **Fig. 7. Heatmap of $log_2 \frac{Simulation\ dN}{Observed\ dN}$ for four models to fit the empirical data.** For each
843 model, average dN was calculated across 50 simulation runs per each pair of conserved-site ratio
844 and selection-site ratio to find the combination of parameters best fitting the empirical dN. Red
845 boxes highlight the pairs of parameters used to investigate which model is preferred to
846 recapitulate the observed data in iMKT and Miyata score analysis.

847

848

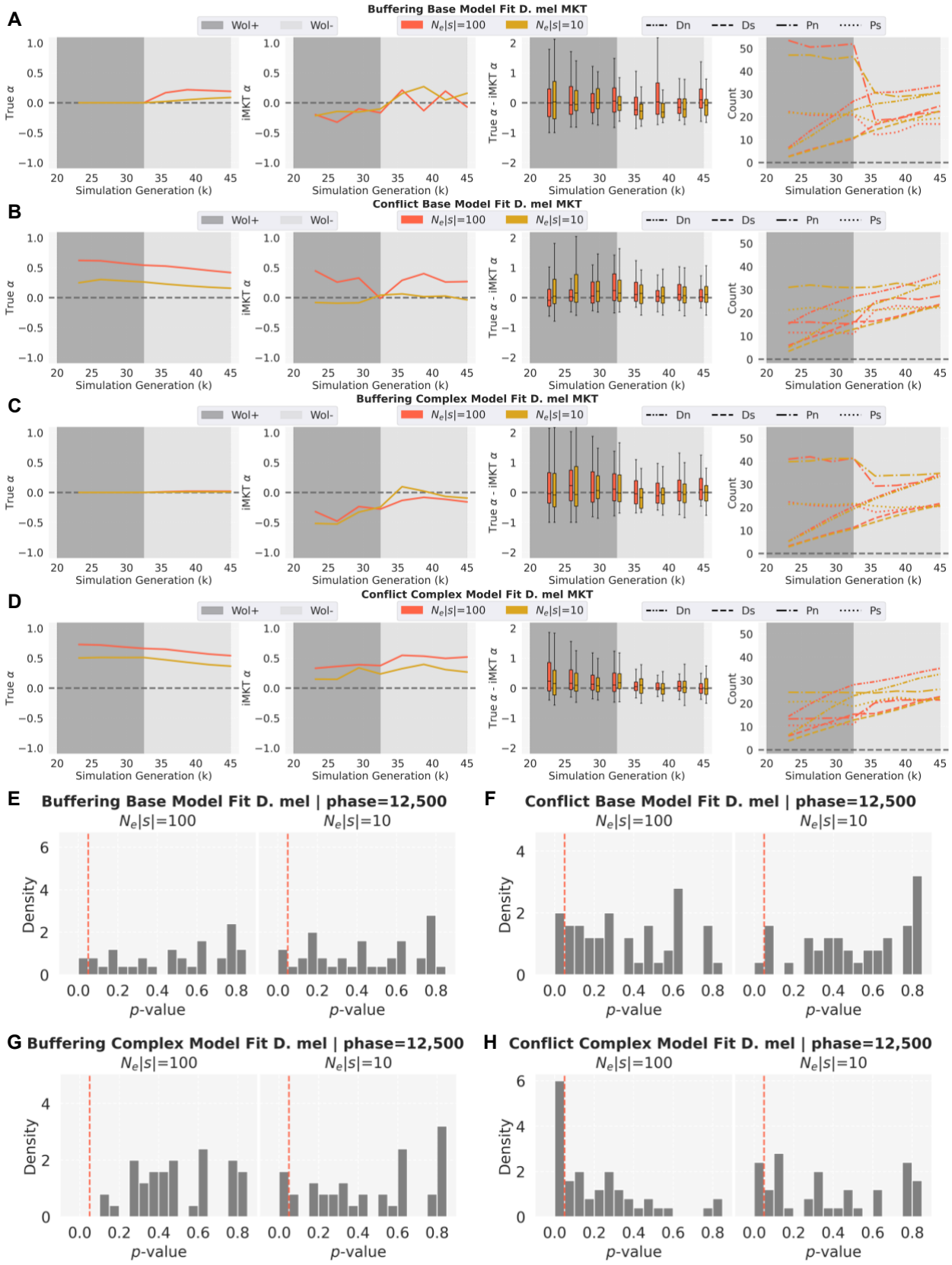

A
Buffering Base Model Fit D. mel MKT
B
Conflict Base Model Fit D. mel MKT
C
Buffering Complex Model Fit D. mel MKT
D
Conflict Complex Model Fit D. mel MKT
Wol+
Wol-
$N_e|s|=100$
$N_e|s|=10$
Dn
Ds
Pn
Ps
True α
iMKT α
True α - iMKT α
Count
Simulation Generation (k)
E
Buffering Base Model Fit D. mel | phase=12,500
F
Conflict Base Model Fit D. mel | phase=12,500
G
Buffering Complex Model Fit D. mel | phase=12,500
H
Conflict Complex Model Fit D. mel | phase=12,500
Density
p-value

Fig. 8. MK test results of simulations for best-fit models for *D. melanogaster*. (A-D) MK test α analysis of each model with different selection coefficients of $N_e|s|$=100, $N_e|s|$=10 graphed at phase length=12,500 simulation generations of *Wolbachia* infection (Wol+, dark grey) and *Wolbachia* absence (Wol-, light grey) post-burn-in period. In each panel, column 1: the average true α in the simulations; column 2: the average iMKT α in the simulations (FWW correction, SNPs frequency > 5% only); column 3: the distributions of differences between the true and iMKT α every 3,125 simulation generations; column 4: The average of each iMKT component (D_n, D_s, P_n, P_s). (E-H) Distributions of iMKT p-values. iMKT p-values (FWW correction, SNPs frequency > 5% only) for simulation runs with *Wolbachia* phase=12,500 simulation generations and $N_e|s|$=100 and 10 for the simulated models at 45,000 simulation generation. The vertical red line denotes p-value=0.05. Note that distributions are normalized to have an area of 1 under the histograms.

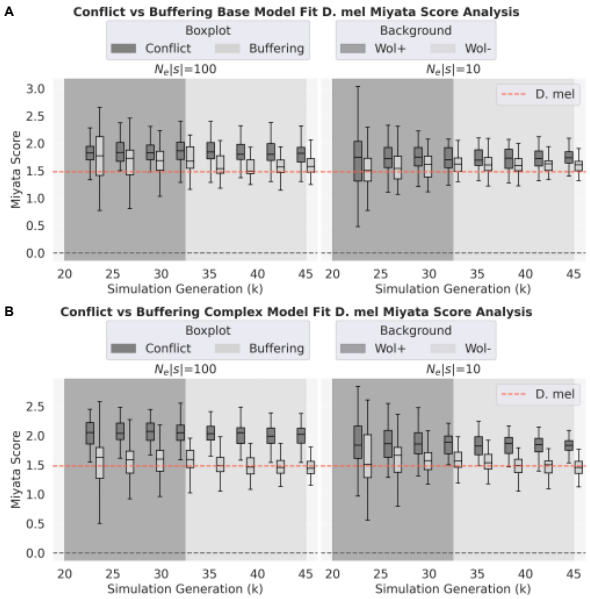


Fig. 9. Distributions of Miyata score for the *D. melanogaster* samples from the simulations with best-fit parameters. The boxplots are the distributions of Miyata scores for each model at every 3,125 generation. (A) Conflict and Buffering Base models; (B) Conflict and Buffering Complex model. The distributions are compared with the observed summary statistics of *D. melanogaster* empirical data (red horizontal line).

Table 1. Empirical (biological) estimates for evolutionary parameters and scaled estimates used for simulation.

Parameter	N_e	ρ	μ	t
Empirical Estimate	1e6	1e-8	2.8e-9	2.5e7
Scaled Estimate (simulation)	1e3	1e-5	2.8e-6	2.5e4

N_e: effective population size; ρ: recombination rate; μ: mutation rate; t: time in unit of generation

Table 2. Best fit RS and RC parameters for $N_e|s|$ =10 and 100 for Buffering and Conflict models.

	$N_e\|s\|$	RS	RC
Buffering	10	0.6	0.1
Base	100	0.6	0.0
Conflict	10	0.02	0.5
Base	100	0.02	0.6
Buffering	10	0.2	0.5
Complex	100	0.4	0.2
Conflict	10	0.08	0.6
Complex	100	0.04	0.7