Article

Methods

A model of functionally buffered deleterious mutations can lead to signatures of positive selection distinguishable from an evolutionary conflict model

Runxi Shen[†1], Miwa Wenzel[†2], Philipp W. Messer[1], Charles F. Aquadro[2]

[1] Department of Computational Biology, Cornell University, Ithaca, NY, USA

[2] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

[†] These authors contributed equally to this work.

*Corresponding author: Email: cfa1@cornell.edu

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 mostly remains unknown to the experimentalists. Here we use the *bag of marbles* (*bam*) gene in *Drosophila* to investigate different types of driving forces behind positive selection. We examine two evolutionary models for *bam.* The Conflict model originates from a conflict of fitness between *Drosophila* and *Wolbachia* that causes reciprocal adaptations in each, resulting in diversifying selection on the *bam* protein. In the alternative Buffering model, *Wolbachia* protects *bam* from deleterious mutations during an infection and thereby allows such mutations to accumulate and even fix in the population. If *Wolbachia* is subsequently lost from the species, mutations that revert the gene back towards its original biological function become advantageous. We use simulations to show that both models produce signals of positive selection, though the levels of positive selection under the Conflict model are more easily detected by the MK test. By fitting the two models to the empirical divergence of *D. melanogaster* from an inferred ancestral sequence, we found that the Conflict model reproduced strong signals of positive selection like those observed empirically, while the Buffering model better recapitulated the physicochemical signatures of the amino acid sequence evolution at *bam.* Our demonstration that the Buffering model can lead to positive selection suggests a novel mechanism that needs to be considered behind observed signals of positive selection on protein coding genes.

Introduction

Patterns of DNA sequence variation within and between species have increasingly been 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 (or at least variation that does not directly alter the protein sequence of genes), with an

excess of nonsynonymous fixed differences between species typically interpreted as evidence that some form of natural selection has accelerated the fixation of advantageous amino acid replacements. This pattern is often referred to as positive selection and is interpreted to mean that natural selection has been fine tuning the protein function of the gene or responding to a changing function, and/or that the gene is involved with intra- or inter-genomic conflict (McLaughlin and Malik 2017). Although the MK test has been found previously to have low power to detect 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.

We have studied the population genetics of two *Drosophila* genes that are critical for gametogenesis in *D. melanogaster* (*bag of marbles, bam*, and *Sex-lethal, Sxl*) and they both show evidence, using the MK test, of episodic bursts of positive selection (Bauer DuMont et al. 2007; Flores et al. 2015a; Bauer DuMont et al. 2021). These genes genetically interact with the endosymbiont bacteria *Wolbachia pipientis* in that *Wolbachia* rescues the reduced fertility of a partial loss-of-function *bam* mutant and the reduced egg production of multiple partial loss-of-function *Sxl* mutants (Starr and Cline 2002; Flores et al. 2015b; Bubnell et al. 2021). The bursts of positive selection in *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). These observations, along with the knowledge that functional divergence in *bam* is seen primarily in females, and that maternally-inherited *Wolbachia* physically resides in the germarium and manipulates reproduction of its host (Flores et al. 2015b) led Bauer DuMont et al. (2007), Flores et al. (2015a) and Flores et al. (2015b) to propose that evolutionary conflict between the host (*Drosophila melanogaster*) and *Wolbachia* over the control of reproduction may drive the observed positive selection at *bam*.

Here, we explore the dynamics of protein sequence evolution under two evolutionary models that we hypothesize could be responsible for the burst of positive selection we see at *bam.* The first is a classic arms race conflict model. In this model, we assume that *Wolbachia* infection is detrimental to a germline reproductive process in which *bam* is involved. For

example, *Wolbachia* may manipulate *bam* 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 *bam* 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.

Secondly, we propose and evaluate a new model of interaction that is based on the nature of the observed genetic interactions of *Wolbachia* with *bam* and *Sxl*. *Wolbachia* partially rescues the fertility defects of a single amino acid replacement hypomorphic mutant of *bam* (Flores et al. 2015b, Bubnell et al. 2021) as well as for several hypomorphic alleles of *Sxl* (Starr and Cline 2002) in *D. melanogaster.* We posit that in the presence of *Wolbachia*, slightly deleterious mutations may accumulate in the *bam* gene without significantly reducing *bam's* function. When *Wolbachia* is lost from the population, there is positive selection for new nonsynonymous mutations that return the *bam* protein sequence to its initial, and assumed optimal, functional state. We term this model the Buffering model, as the effects of deleterious mutations are "buffered" during periods of infection by *Wolbachia*.

We carry out evolutionary simulations to explore the population genetic consequences for sequence evolution at *bam* for both conflict and buffering interactions between the endosymbiont *Wolbachia* and the *Drosophila* host. With these models, we first reaffirm the resultant adaptive evolution of a theoretical arms race conflict and assess the likelihood that the alternative buffering type of interaction can also result in signatures of positive selection detectable with the MK test. We then evaluate the robustness of these models to variation in selection coefficients and duration of *Wolbachia* infection, and test our ability to distinguish between the two models using a measure of amino acid similarity (Miyata et al. 1979). Finally, using model parameters tuned to the observed sequences of *bam* in *D. melanogaster*, we evaluate evidence from the observed sequence divergence of *bam* in *D. melanogaster* compared to its common ancestor with *D. simulans* to see if we can discriminate as to which model (Conflict or Buffering) better captures the signatures of the bursts of positive selection observed at *bam*.

Materials and Methods

Simulation setup using SLiM 3.5

107 The evolution of *bam* was simulated under a Wright-Fisher model using nucleotide-based
108 simulation in SLiM 3.5 (Haller and Messer 2019; Haller et al. 2019). We inferred the ancestral
109 DNA sequence of the exons in *bam* (1338 nucleotides) for *Drosophila melanogaster* and
110 *Drosophila simulans* using maximum likelihood with codeML v4.8 (Yang 2007). Briefly,
111 alignments of seven *Drosophila* coding sequences were made using PRANK v.170427
112 (Löytynoja 2014), including sequences of *D. melanogaster, D. simulans, D. sechellia, D.*
113 *yakuba, D. erecta, D. eugracilis, and D. pseudoobscura*. The alignment was input in codeML
114 and an ancestral sequence was estimated by maximum likelihood for the common ancestor of *D.*
115 *melanogaster* and *D. simulans*, using the other species' sequences as outgroup references. The
116 estimated common ancestral nucleotide sequence was then used in SLiM as the starting sequence
117 for the entire population in each simulation.
118 The evolutionary parameters in the simulations were based on empirical estimates and
119 then scaled to make the simulations run efficiently (summarized in Table 1). The reported
120 evolutionary parameters from empirical observations were an effective population size (N_e) of
121 1e6 (Campos et al. 2017), an overall mutation probability (μ) of 2.8e-9 per nucleotide per
122 generation (Keightley et al. 2014), an average recombination rate (ρ) of 1e-8 per nucleotide per
123 generation (Comeron et al. 2012) and a divergence time t of 25 million generations (2.5 million
124 years assuming 10 generations per year) from the common ancestor to each species (Russo et al.
125 1995). These parameters were then scaled down by 1,000 to run more efficient simulations by
126 using a smaller population and shorter simulation time while keeping the key products, $N_e\mu$, N_er
127 and N_e/t constant to approximate the same evolutionary process (Haller and Messer 2019).
128 Hence, in each Wright-Fisher simulation, we have 1) a scaled population size N_e of 1,000 diploid
129 individuals, 2) a mutation matrix representing a Kimura (1980) model with transition rate α and
130 transversion rate β of 1.88e-6 and 0.47e-6 respectively (calculated from a scaled mutation rate
131 μ=2.8e-6 and an observed 2:1 transition:transversion ratio in the sequences (Keightley et al.
132 2009)), 3) a scaled recombination rate ρ of 1e-5 and 4) 25,000 scaled generations for the
133 divergence time t. We simulated *bam* as a single contiguous exon, though in reality there are two
134 short introns (61 and 64 bp). The effect of excluding these introns had a negligible impact on the
135 rate of recombination.
136 We observed 85% of the codons in *bam* encode the same amino acids in both *D.*
137 *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 among the 1338 nucleotide sites 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 under 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. Note that 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 that is used to track the substitutions in the population and to assign selection coefficients in the upcoming non-neutral phases of the simulation, was manually reset to the original estimated ancestral sequence inferred from PAML. We did this 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). We

169 used the amino acid matrix of Miyata et al. (1979), which captures the primary features of
170 biochemical and physical differences between amino acid pairs but does not take empirical
171 protein sequence conservation into account, since some of our changes were going to be
172 positively selected, and not conserved. We henceforth refer to the pairwise measures from the
173 Miyata matrix as "Miyata scores (MS)" and use them to determine whether a mutation is neutral
174 (synonymous, no change in the encoded amino acid) or under positive or negative selection
175 (nonsynonymous, MS between the original amino acid and the mutated amino acid $\neq 0$) in each
176 selection scenario. Below, we will use a shorthand for Miyata score calculations as follows, with,
177 for example, MS between the current amino acid and the mutated amino acid represented as
178 $MS(AA_{cur}, AA_{mut})$.
179
180 **Selection regimes**
181 *Wolbachia* infections have been observed to be temporally dynamic in host populations,
182 being lost at times and then regained, even from another species of *Drosophila* (Richardson et al.
183 2012; Turelli et al. 2018; Meany et al. 2019). Thus, each model has two selection phases: one
184 with selection parameters to represent a period of *Wolbachia* infection and another to represent a
185 period of *Wolbachia* absence in the population. There are four different selection schemes: 1a)
186 *Wolbachia*-infection phase in the Conflict model, 1b) *Wolbachia*-absence phase in the Conflict
187 model, 2a) *Wolbachia*-infection phase in the Buffering model, and 2b) *Wolbachia*-absence phase
188 in the Buffering model. The phases of infection and absence of *Wolbachia* alternated in each
189 model, which simulated the periodic occurrence of *Wolbachia* in natural populations. To keep
190 the simulations simple, we assume that *Wolbachia* infection and loss is instantaneous throughout
191 the entire population and that there are no other effects of *Wolbachia* on the host beyond which
192 we are modeling.
193 The Conflict model was implemented based on a traditional arms race model. In the
194 *Wolbachia*-infection phase, we assume that *Wolbachia*'s presence drives the positive selection of
195 *Drosophila* by favoring the nucleotide changes that lead to biochemically more diversified amino
196 acids to "escape" the present function which may be targeted by *Wolbachia*'s harmful impact. In
197 this case, all the nonsynonymous mutations that give rise to biochemically different amino acids
198 from the current states ($MS(AA_{cur}, AA_{mut}) > 0$) were positively selected for, with selection
199 coefficients $s > 0$ (fig. 1, top).

200 In the *Wolbachia*-absence phase of the Conflict model, there is no evolutionary conflict
201 between *bam* and *Wolbachia* and thus no selective pressure on the host to adapt. Under these
202 conditions, we assume that the current amino acid sequence functions adequately such that the
203 DNA sequences in the population remain largely unchanged and the *bam* gene is under purifying
204 selection to preserve the present amino acid sequences. In this case, nonsynonymous mutations
205 leading to amino acid changes ($MS(AA_{cur}, AA_{mut}) > 0$) are considered deleterious with selection
206 coefficients $s < 0$ (fig. 1, top).
207 The second evolutionary model, the Buffering model, is built on the observation that
208 *Wolbachia* offers a functional buffer for deleterious mutations in *bam* and rescues what would be
209 reduced fertility of its host in the absence of *Wolbachia*. During the *Wolbachia*-infection phase,
210 *Wolbachia* somehow functionally alleviates the deleterious effects of certain nonsynonymous
211 mutations and makes them effectively neutral. Under this scenario, mutations leading to
212 divergent amino acids different from the ancestral state ($MS(AA_{anc}, AA_{mut}) > 0$) are all regarded
213 as neutral (fig. 1, bottom) and thus can accumulate, albeit slowly due to drift alone. If *Wolbachia*
214 infection is lost, these previously buffered mutations are now deleterious, and new mutations
215 leading to amino acids that converge back towards the initial amino acid ancestral states are
216 favored as there is selective pressure for *bam* to regain its original optimal function. In this case,
217 a mutation that converts the current amino acid to a mutated amino acid that is biochemically
218 more similar to the ancestral amino acid ($MS(AA_{anc}, AA_{mut}) – MS(AA_{anc}, AA_{cur}) < 0$) is
219 beneficial with a selection coefficient $s > 0$, while divergent mutations ($MS(AA_{anc}, AA_{cur}) –$
220 $MS(AA_{anc}, AA_{cur}) > 0$) are deleterious with selection coefficients $s < 0$ (fig. 1, bottom).
221 The above models capture two types of potential driving forces behind the signatures of
222 positive selection observed on the *bam* gene. However, these models make many idealized
223 assumptions about amino acid evolution based on Miyata scores and thus are regarded as the
224 "Base" models. For instance, in the Conflict model, any nonsynonymous mutation would be
225 positively selected during the presence of *Wolbachia*. However, mutations that lead to
226 biochemically similar amino acids with homogeneous functions may not bear such strong
227 selective advantages compared to the original ones and could be regarded as almost neutral,
228 while the mutations giving rise to extremely dissimilar amino acids may completely lose their
229 original functionality and thus be deleterious. To incorporate these more realistic biological
230 assumptions, we set up additional "Complex" models based on empirical observations to

231 determine whether a mutation would be neutral, beneficial, or deleterious. For example,
232 Demogines et al. (2013) identified adaptively evolving sites in the transferrin receptor gene *TfR1*
233 in wild rodents to include amino acids R, K, N, I, and T, which corresponds with pairwise
234 Miyata scores ranging from 0.4 to 3.37. Likewise, Charron et al. (2008) propose sites in the plant
235 gene eIF4E to be in an arms race conflict with viral proteins, which includes those with amino
236 acids L, P, and A that give pairwise Miyata scores ranging from 0.06 to 2.76. We found that the
237 Miyata scores for all proposed positively selected residues in these studies ranged from 0.05 to
238 3.37, with the majority of scores falling between 1.5 and 2.5.

239 With the above proposition, in the complex Conflict model when *Wolbachia* is present,
240 nonsynonymous mutations that give rise to biochemically similar amino acids ($0 < MS(AA_{cur},$
241 $AA_{mut}) \leq 1$) are regarded as neutral with selection coefficients $s = 0$ (fig. 2, top); mutations
242 leading to mildly different amino acids ($1 < MS(AA_{cur}, AA_{mut}) \leq 3$) are considered beneficial
243 with selection coefficients $s > 0$ (fig. 2, top); and mutations become deleterious with selection
244 coefficients $s < 0$ when they generate extremely dissimilar amino acids ($MS(AA_{cur}, AA_{mut}) > 3$)
245 (fig. 2, top), as they are likely to disrupt the biological function of *bam*. These cutoffs are
246 consistent with the range of Miyata scores found at sites that are proposed to be adaptively
247 evolving in response to an evolutionary conflict. In the *Wolbachia*-absence phase in the complex
248 Conflict model, to preserve the current amino acid sequences, we still assume that mutations
249 leading to similar amino acid changes ($0 < MS(AA_{cur}, AA_{mut}) \leq 1$) are considered neutral, but
250 any mutation that causes a dissimilar amino acid change ($MS(AA_{cur}, AA_{mut}) > 1$) is deleterious
251 with a selection coefficient $s < 0$ (fig. 2, top).

252 Unlike the complex Conflict model, we have insufficient empirical observations to guide
253 our selection cutoffs for a complex Buffering model. For simplicity, we adopted the same
254 selection schemes used in the complex Conflict model to represent one biologically plausible
255 possibility for the interaction. When *Wolbachia* is infecting in the complex Buffering model, any
256 mutation that gives rise to a mildly biochemically different amino acid from the ancestral state (0
257 $< MS(AA_{anc}, AA_{mut}) \leq 3$) is regarded as neutral with selection coefficients $s = 0$ (fig. 2, below)
258 due to the protection by *Wolbachia*. However, mutations are considered deleterious with
259 selection coefficients $s < 0$ when they generate extremely dissimilar amino acids ($MS(AA_{anc},$
260 $AA_{mut}) > 3$) (fig. 2, bottom), since they are likely to disrupt the biological function of *bam*.
261 When *Wolbachia* is lost from the population, only mutations that converge back towards the

262 biochemical characteristics of the initial ancestral state relative to the current amino acid are
263 favored ($MS(AA_{anc}, AA_{mut}) - MS(AA_{anc}, AA_{mut}) < -1$) with a selection coefficient $s > 0$, while
264 the more divergent mutations ($MS(AA_{anc}, AA_{mut}) - MS(AA_{anc}, AA_{cur}) > 1$) are deleterious with a
265 selection coefficient $s < 0$ (fig. 2, bottom). Any mutation in between ($-1 \leq MS(AA_{anc}, AA_{mut}) -$
266 $MS(AA_{anc}, AA_{cur}) \leq 1$) is considered neutral ($s = 0$) since it does not cause a radical functional
267 change in the amino acid to increase or decrease the fitness of an individual. Below, we test the
268 evolution of *bam* under both the "base" and "complex" models to investigate how
269 implementation of different fitness parameterizations for the base and the complex Miyata score
270 ranges affect our simulation results.
271
272 **Simulation parameters**
273 We focused on investigating the impacts of two key parameters on the evolution of the
274 *Drosophila* species in each of the proposed models: 1) the magnitude of the selection coefficient
275 for both beneficial and deleterious mutations, and 2) the length of alternating *Wolbachia*-
276 infection and *Wolbachia*-absence phases in each model in which the different selection phases
277 occur. The absolute values of selection coefficients included $|s| = 0.1$, $|s| = 0.01$, and $|s| = 0.001$,
278 resulting in $N_e|s| = 100$, $N_e|s| = 10$, and $N_e|s| = 1$ respectively, where $N_e|s| = 1$ can be considered
279 effectively neutral. The lengths of different selection phases varied from equal periods of 12,500,
280 6,250, and 3,125 simulation generations (corresponding to one, two, and four *Wolbachia*
281 infection-loss cycle(s) respectively in a total divergence time of 25,000 simulation generations).
282 For each set of parameter combinations, we ran 50 independent simulations and performed
283 downstream analyses including the MK test, inferences of α (the proportion of amino acid
284 fixations driven by positive selection; Smith and Eyre-Walker 2002), and Miyata score
285 differences between the ancestral and evolved sequences. All these downstream analyses were
286 conducted every 3,125 simulation generations (the shortest phase length in our simulation setup)
287 after the neutral burn-in period, by comparing the "reference sequence" in SLiM to the common
288 ancestral sequence of *D. melanogaster* and *D. simulans*. For the MK test, 100 diploid individuals
289 were randomly sampled from the population and nonsynonymous and synonymous fixations
290 (relative to the inferred ancestral sequence) and polymorphisms present were tabulated.
291
292 **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

Conflict and Buffering Base Models

We first used the Conflict and Buffering Base models described above to simulate the cyclic pattern of *Wolbachia* infection and loss in the *Drosophila* population. In the Conflict Base model, nonsynonymous mutations underwent positive selection in the presence of *Wolbachia* infection but were negatively selected in the absence of *Wolbachia*. In the Buffering Base model,

323 nonsynonymous mutations experienced neutral evolution in the *Wolbachia*-infection phase but
324 were positively or negatively selected in the *Wolbachia*-absence phase.
325 The patterns of true α's were clearly indicative of positive selection in the Conflict Base
326 model phase with *Wolbachia* with the strongest selection ($N_e|s|$>100, fig. 3A, row 1). The
327 elevated true α's persisted though with a slow decline during the subsequent Wolbachia-free
328 phase. The Buffering Base model showed a positive, though much lower, α that emerges in the
329 first *Wobachia*-absence phase (fig 3B, row 1), consistent with our intuition of positive selection
330 to return to a more functional *bam* protein without the deleterious mutation buffering by
331 *Wolbachia.* This pattern was also most evident with strong selection. Among all the selection
332 coefficients for both the Conflict model and Buffering model, the true α's stayed high or
333 increased in the phases where positive selection is expected and stayed constant or decreased in
334 the phases of neutral evolution and purifying selection.
335 In the Conflict Base model, the average of iMKT estimates of α were almost all positive
336 for selection coefficients with $N_e|s|$>1 and showed clear periodic changes as *Wolbachia* comes in
337 and out of the population across all three phase lengths (fig. 3A, row 2). Surprisingly, the
338 magnitude of the iMKT α's increased in the phase without the imposed positive selection. This
339 unexpected increase is explained by the change of nonsynonymous polymorphisms (P_n) in the
340 population. In the *Wolbachia*-infection phase, both D_n and P_n accumulated due to the positive
341 selection of nonsynonymous mutations, as expected; however, after the sudden change to the
342 *Wolbachia*-absence phase, D_n was largely unchanged while P_n experienced a sudden decrease as
343 nonsynonymous mutations were all selected against (fig 3A, row 4). Given the equation for
344 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
345 with the implemented purifying selection that followed the positive selection. For the effectively
346 neutral case of $N_e|s| = 1$, the iMKT $\alpha's$ in the Conflict Base model fluctuated around 0.
347 In the Buffering Base model across all selection coefficients, the iMKT α's were mostly
348 negative across the whole simulation, but changes in magnitude are evident in different selection
349 phases, e.g., iMKT α decreased during neutral phases (*Wolbachia* present) but increased in
350 phases with selection (*Wolbachia* absent) (fig. 3B, row 2). These observed negative iMKT
351 estimates of α were due to the contributions from both D_n and P_n. In the initial *Wolbachia*-
352 infection phase of the Buffering model, nonsynonymous polymorphisms were negatively
353 selected in the constrained codons and neutrally buffered by *Wolbachia* in the codons under

354 selection, with few such mutations in the latter category going to fixation, explaining the
355 negative iMKT α's in the initial phases (fig 3B, row 4). Following this "buffering" period, a
356 subset of nonsynonymous mutations was selected for. However, the number of nonsynonymous
357 mutations that could be positively selected in the Buffering Base model was much less than those
358 in the Conflict Base model, leading to a smaller D_n and thus a smaller (possibly < 0) iMKT α,
359 even when positive selection was present. While a periodicity of iMKT estimated α's was
360 observed in both models, the small number of polymorphisms and fixed differences made these
361 estimates only vaguely reflective of the true α's.

362 The boxplots of differences between the true and iMKT α's were used to evaluate the
363 accuracy of iMKT α's. In general, iMKT α's systematically underestimate the true α's due to
364 the presence of deleterious polymorphisms (Fay et al. 2001, Eyre-Walker and Keightley 2009,
365 Messer and Petrov 2013). For the Conflict Base model with effectively neutral evolution
366 ($N_e|s|$=1), iMKT α's usually underestimated the true α's (fig 3A, row 3). However, under
367 stronger selection ($N_e|s|$=10 or 100), iMKT α's underestimated the true α's only during the
368 *Wolbachia*-infection phase; there was good accuracy in iMKT α's estimation when *Wolbachia*
369 was lost, which reflected the delay in detecting selection based on changing P_n as previously
370 explained. In the Buffering Base model, iMKT α's also tended to underestimate the true α's
371 (especially with $N_e|s|$>1 in the later *Wolbachia*-infection phases), with the boxplots distributed
372 above 0.

373 For the Conflict Base model, the pattern of the true α's and the iMKT α's was not
374 dramatically influenced by the magnitude of the selection coefficient ($N_e|s|$=10 or 100) and the
375 varying lengths of the infection/absence periods that we examined. In contrast, for the Buffering
376 Base model, longer *Wolbachia* infection periods resulted in larger true α's, presumably due to
377 the greater time to accumulate buffered deleterious mutations in the presence of *Wolbachia*.
378 Nevertheless, the length of *Wolbachia* infection and absence periods had only a minor impact on
379 the final magnitude of D_n, D_s, P_n, and P_s observed at the end of the simulations (fig. 3A&B, row
380 4). Only P_n show dramatic periodic fluctuations due to the cyclic infection and absence periods.

381 Additionally, we looked at the distributions of p-values in iMKT (FWW correction, SNPs
382 frequency > 5%) and the correlation between iMKT α's and their corresponding p-values at the
383 end of the simulation. In general, a statistically significant rejection of neutrality in the direction
384 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. 5A). On the other hand, 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 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).

All together, these results demonstrated that α estimated from iMKT identified departures from neutrality in the direction of positive selection in the Conflict Base model but not reliably in the Buffering Base model. The phase without imposed positive selection in the Conflict model introduced a cyclical pattern of the iMKT α (but not the true α). The iMKT α was basically unreliable in all phases of the Buffering model. Lastly, the MK test parameter D_n reached a higher magnitude in the Conflict model, which was reflected in the smaller p-value statistics, suggesting greater power to detect positive selection with the MK test in the Conflict model than in the Buffering model.

Conflict and Buffering Complex Models

We next implemented the models with a more complex parameterization of selection coefficients based on Miyata scores. The MK test results for these Conflict and Buffering 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 for both the Conflict and Buffering Complex models compared to their Base models (fig 4A&B, row 1). For the Buffering Complex model, this difference was particularly pronounced, with a barely perceptible increase in true α for even 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 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. However, D_n had a slight increase in the Buffering Complex model compared with the Buffering Base model,

416 potentially due to the introduction of the neutral region leading to a small number of additional
417 nonsynonymous fixations by genetic drift alone (fig 4A&B, row 4). The distributions of p-values
418 were also similar between the Complex and Base cases (fig 5, C & D).
419
420 **Distributions of Miyata scores**
421 We expect amino acid substitutions to be more diversified in the Conflict model than in
422 the Buffering model in both the Base and Complex cases, as the former is based on the premise
423 of a sequence evolving *away* from the ancestral sequence and the latter is based on the premise
424 of a sequence evolving *toward* the ancestral sequence. To assess this, we calculated Miyata
425 scores between each amino acid substitution and its ancestral amino acid throughout the
426 simulations.
427 For both the Base and Complex cases, the distributions of Miyata scores per amino acid
428 from the Conflict and Buffering models were most distinguishable from each other in the
429 strongest selection scenario at $N_e|s| = 100$. Here, the interquartile ranges of Miyata score
430 distributions of the two models were completely separated at the end of simulations (fig6A, 6B;
431 row 1), but they were basically indistinguishable from each other throughout the simulations
432 when the selection is the weakest at $N_e|s| = 1$ (fig6A, 6B; row 3).
433 For $N_e|s| = 10$, the distributions of Miyata scores overlap more in the Base case
434 compared with the Complex case (fig6A, 6B; row 2) because the positively selected mutations
435 had a higher concentration of Miyata scores between 1 and 3 in the Complex case, which made
436 the differences between Miyata scores more prominent. The same results can be observed in the
437 strongest selection simulations, where the distance between interquartile ranges was also larger
438 in the Complex cases than in the Base cases.
439 In summary, we are better able to distinguish between the Conflict and Buffering models
440 using Miyata score distributions when there is strong selection, and when there is a more
441 complicated parameterization of selection scheme as in the Complex models. Different
442 infection/absence phase lengths did not have a large impact on the average Miyata scores across
443 the simulations.
444
445 **Comparison with the empirical data**

446 To evaluate which model in our analysis better captures *bam's* observed patterns of
447 sequence evolution within and between natural populations of *Drosophila*, we first performed
448 our MK test on a population sample (n=89) of *D. melanogaster* (Lack et al. 2015), using
449 divergence to the predicted common ancestral sequence with *D. simulans* as the outgroup and a
450 randomly sampled sequence as the reference sequence used in the iMKT estimate. Analysis of
451 these data reject neutrality in the direction of positive selection using the MK test with a p =
452 0.00015 and α estimated to be 0.91 (FWW correction, SNPs > 5% only). We used the number of
453 nonsynonymous substitutions per nonsynonymous site (dN) calculated from iMKT results as the
454 summary statistic to tune selection parameters of the two simulation models with only one
455 *Wolbachia* infection-loss cycle.
456 While examining polymorphism levels would seem important to distinguish between
457 Conflict and Buffering models, these levels are very sensitive to the length of *Wolbachia*
458 infection and absence as we have modeled it, for example, due to the strong purifying selection
459 occurring in the Conflict model when *Wolbachia* is lost. Thus, determining the time point for
460 sampling is problematic for the empirical data as we do not know for a species that is infected
461 with *Wolbachia* how long it has been infected, nor do we know for uninfected species when the
462 last time they were infected (or even if they were). The problematic effect of this timing choice
463 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
464 and represents the number of amino acid changes in *bam*, we chose to only use this parameter to
465 evaluate how well our models fit to empirical data.
466 Applying our custom iMKT script to our empirical sequence data, we found that iMKT
467 D_n = 34 and dN = 0.033 for the empirical *D. melanogaster* population. We initially found that
468 Conflict models always predicted a much higher D_n than the empirical observation, while
469 Buffering models often exhibited a much lower D_n. Such results showed the initially assumed
470 ratio of codons under selection (RS=15%) and ratio of codons under constraints (RC=75%)
471 could not reproduce similar results for the evolution of amino acids for either model. Thus, we
472 chose to tune these two ratios of selected and constrained codons (RS and RC) under different
473 strengths of positive selection ($N_e|s|$ = 100, 10, 1) and explore under which parameter settings
474 could we fit the empirical dN in each of our proposed models. When we achieved a matching
475 dN, we then compared the Miyata scores per amino acid change in the observed data and our

476 simulation results and see whether a Conflict or a Buffering model is more similar to our
477 observations.
478 For each selection coefficient s, we first ran simulations using RS and RC both sampled
479 from a uniformly distributed grid of nine points ranging from 0 to 0.8, since the maximum
480 proportion of conserved codons is 0.85, and assessed the resulting dN. Preliminary simulations
481 revealed that for the Conflict models, dN was consistently more than two-fold overestimated for
482 any proportion of selected sites greater than 0 (e.g., RS > 0, data not shown). Therefore, we
483 refined the Conflict model grid search for RS to a uniform grid of 6 points from [0, 0.1], while
484 keeping the full grid range for RC. For the Buffering models, we kept the full range of the RS
485 grid as we did find parameters that fit the observed dN. For each pair of parameters for all
486 models, we ran 50 simulations and calculated the mean of dN ($\bar{dN}$) across the runs. We then
487 compared the difference between the empirical dN and $\bar{dN}$. The best pair of RS and RC was the
488 one that led to the smallest difference between $\bar{dN}$ and the empirical dN under each selection
489 coefficient s.
490 For the models with selection coefficient $N_e|s| = 1$, all combinations of the two ratios
491 reproduced similar results consistent with effective neutrality (fig 7). For moderate or strong
492 selection, the best-fit parameters are shown in Table 2.
493
494 **Analyses of Conflict and Buffering Models Best Fitting the Empirical Data**
495 To evaluate how well the Conflict and Buffering models implemented with the best-fit
496 pairs of RS and RC recapitulate the empirical data for *D. melanogaster*, we performed the same
497 iMKT analysis and Miyata score analysis for the resulting simulations. Positive true α's were
498 observed in the Conflict Base, Buffering Base and Conflict Complex models across different
499 phase lengths, indicating that positive selection was present under these scenarios. However,
500 iMKT could only identify positive selection by α and statistically significant p-values in the two
501 Conflict models with strong selection at $N_e|s| = 100$. Moderate selection at $N_e|s| = 10$ in the
502 Conflict models or any levels of selection in the two Buffering models was not detected by
503 iMKT p-values or inferred α (fig. 8).
504 In addition, we calculated the empirical per-site Miyata scores between the current *D.*
505 *melanogaster* sequences and their predicted common ancestral sequence shared with *D. simulans*
506 (supplemental file S2), and compared it with the distributions of per-site Miyata scores simulated

507 from the best-fitted RS and RC at different timepoints. The end of the simulations at 45,000
508 scaled generations represents the actual divergence time between the ancestral sequence and the
509 extant *D. melanogaster* and *D. simulans* species. At this time point, the interquartile ranges of
510 Miyata scores of the Conflict and Buffering models have separated from each other, with fully
511 non-overlapping interquartile distributions in the Complex models. In all models, the per-site
512 Miyata score of *D. melanogaster* are located closer to the center of the distributions from the
513 Buffering models than to the center of distributions from the Conflict models (Figure 9).
514 In summary, the best-fit Conflict models with strong selection reproduced the most
515 significant iMKT p-values and high estimates of α like that observed in the *D. melanogaster*
516 sample, but the Miyata-score analysis indicated the Buffering models as a better fit for the
517 evolution of the amino acids' biochemical properties. It is important to note that while the
518 average iMKT α is extremely close to zero for all Buffering models, the lower whiskers on the
519 box plots in fig 8B and 8C show that high iMKT inferred α's can, although infrequently, occur
520 under the Buffering models as well.
521
522 **Discussion**
523 The increasing availability of DNA sequence datasets for diverse genes, genomes and
524 organisms has led experimentalists to scan genes and genomes for footprints of natural selection.
525 Using tests like the MK test, a striking number of cases of departures from neutrality have been
526 revealed (e.g., Eyre-Walker 2006). In some cases, statistical evidence of strong positive selection
527 can be easily associated with a proposed causal factor (e.g., genes involved in antiviral immunity
528 or in mating behavior; e.g., McLaughlin and Malik 2017). In other cases, the driving factor is
529 less clear.
530 In this study, we proposed two different models, the Conflict model and the Buffering
531 model, to investigate different types of driving forces behind the signatures of positive selection
532 at *bam*, motivated by its biological interactions with *Wolbachia*. The Conflict model is based on
533 an arms race dynamics previously proposed to model the interactions between competing
534 symbiotic species, which positively selected diversified amino acids. The alternative Buffering
535 model we newly propose in this paper is based on another possible interaction between *bam* and
536 *Wolbachia*, in which *Wolbachia* protects *bam* from the effects of deleterious mutations that can
537 therefore accumulate during the *Wolbachia* infection phase by drift (equivalent to the relaxation

538 of functional constraints for amino acid mutations). When *Wolbachia* is lost, the constraints are
539 reimposed and amino acids similar to *bam*'s ancestral state are selected for.
540 We used simulations to study the evolutionary process involved in each model and found
541 that both models can generate positive selection as measured by the true α. A positive true α in
542 the Conflict Base and Complex models aligns with our expectations of the represented
543 interaction. More interestingly, the positive true α in the Buffering models reveals that
544 *Wolbachia* need not function as a reproductive parasite in conflict with *bam* to drive positive
545 selection in the host gene.
546 However, we must highlight the difference in the Buffering model's ability to generate
547 positive selection and our ability to detect it. We found that in all simulations, the iMKT α
548 generally underestimated the true α. This underestimation had a minimal effect on our
549 interpretation of evolution in the Conflict models, as the true α was very large in all simulations
550 outside of those with the weakest selection ($N_e|s| = 1$). On the other hand, with a maximum true
551 α (~ 0.25) in the Buffering models' simulations, an underestimation led to a weak or absent
552 signal of positive selection detectable by iMKT, which could further be confounded by statistical
553 noise. Such findings highlight some limitations of the MK test that are consistent with the
554 findings of others (Akashi 1999, Fay et al. 2001, Eyre-Walker and Keightley 2009, Zhai et al.
555 2009, Messer and Petrov 2013). Nevertheless, even with these limitations, the iMKT could still
556 infer high α's, representing detection of positive selection, in some simulation runs under the
557 Buffering models.
558 The Buffering models require the fixation of *Wolbachia*-buffered deleterious
559 nonsynonymous mutations by drift for there to be resulting positive selection during a
560 subsequent phase without *Wolbachia*. This effect is seen across the three different infection
561 lengths that we simulated in the Buffering Base model. Thus, longer *Wolbachia* infection phases
562 will increase the chance of detecting positive selection in a subsequent *Wolbachia* absence phase,
563 though never to the level resulting from Conflict models. The average length of *Wolbachia*
564 infection time is unknown for *Drosophila*, but a few studies have documented the minimum
565 length of current *Wolbachia* infections. These include two independent studies that found the
566 *w*Mel *Wolbachia* variant to have been in *D. melanogaster* for 79,000 and 80,000 *Drosophila*
567 generations thus far (Richardson et al. 2012; Choi and Aquadro, 2014). These time periods are
568 shorter than what we have simulated, but there is evidence to suggest turnover of *Wolbachia*

569 variants that could act as a longer standing infection period than currently documented (Riegler
570 et al. 2005, Kriesner et al. 2013). Thus, *Wolbachia* infection of the length we have simulated, and
571 with it a potential for subsequent positive selection, is not out of question.

572 In addition to phase lengths affecting the Buffering model results, population size could
573 also play a large role. Because drift during the *Wolbachia* infection phase is what allows the
574 buffered deleterious nonsynonymous mutations to fix, this model will likely lead to stronger
575 signatures of positive selection for species with smaller N_e than this large population size
576 *Drosophila* species since the time to fixation of neutral mutations is approximately $4N_e$
577 generations (Kimura and Ohta 1969).

578 To better evaluate the fit of the observed data from *D. melanogaster* to the predictions of
579 the Conflict and Buffering models, we tuned the simulation selection parameters of both models
580 to fit the observed nonsynonymous sequence divergence per nonsynonymous site (dN) between
581 *D. melanogaster* and the inferred common ancestor with *D. simulans*. Explorative simulations
582 are reflected in the empirically tuned simulations. Only the tuned Conflict model recapitulated
583 the statistically significant positive iMKT α's that we observed for the *D. melanogaster*
584 population. As in the general Buffering results discussed above, the tuned Buffering model can
585 result in evidence of positive selection as indicated by a true α under certain conditions, but we
586 can rarely detect it with the iMKT in *bam* with statistical significance. Interestingly, for the
587 Miyata score analysis, we found that the Buffering models better fit our empirical data, as the
588 Conflict models predicts greater amino acid diversity (assessed by the Miyata score) than we
589 observe. Thus, combining these two results, we suggest that the Buffering model is a possible
590 explanation behind the observed evolution in the *D. melanogaster bam* gene on the rare occasion
591 that the iMKT is significant in the direction of positive selection. Nevertheless, a p-value less
592 than or equal to 0.05 for the empirical MKT result is the typical criteria used by experimentalists
593 to infer a departure from an equilibrium neutral model. Thus, with the current assumptions of our
594 models, the Conflict model of an arms race between *Wolbachia* and *bam* is the better explanation
595 for the signature of selection that we observe at *bam.*

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

600 and deleterious mutations, while they could actually be drawn from some distribution. The
601 cutoffs of Miyata scores to determine the characteristics of each mutation were also fixed and
602 thus could also involve more customization based on the properties and functions of actual
603 amino acids as well. It is also possible that a mixture of Conflict and Buffering models may be
604 operating, with each driving evolution at a subset of sites. Importantly, while the underlying
605 simplified assumptions as implemented in our models give them only a limited ability to capture
606 the full details of the evolutionary processes, our simulations do demonstrate that the Conflict
607 models have enough power to generate statistically significant signatures of positive selection at
608 *bam*.

609 With regard to resolving the evolutionary interactions between *bam* and *Wolbachia* in
610 *Drosophila*, it will now be important to explore other experimental evidence with respect to
611 potential conflict, change in function or buffering. For example, we could test for evidence of
612 positive selection in *Wolbachia* genes, as positive selection in *Wolbachia* is expected under the
613 Conflict model where *Wolbachia* would co-evolve with *bam* to continue its impact on
614 *Drosophila* fertility. There is already some evidence of positive selection at a few genes across
615 different *Wolbachia* strains of arthropods and nematodes (Baldo et al. 2002; Baldo et al. 2010)
616 but a much more thorough analysis of closely related *Wolbachia* strains infecting *D.*
617 *melanogaster* is needed. To test for a conflict-like interaction between germline stem cell genes
618 and *Wolbachia*, signals of selection in *Wolbachia* need to be examined solely within the
619 *Drosophila* genus, since *Wolbachia* has a very different relationship with its nematode hosts
620 (Taylor et al. 2005).

621 While our modeling study was motivated by the bursts of positive selection at *bam* and
622 *Sxl* in *Drosophila* and the experimental interactions between *Wolbachia* and hypomorphs at these
623 genes, both the Conflict model and the Buffering model should be investigated when we try to
624 understand the signals of positive selection at other genes in *Drosophila*. Considering that
625 *Wolbachia* infects some 50% or more of all arthropods (Hilgenboecker et al. 2008; Zug et al.
626 2012; Weinert et al. 2015), this means that there are potentially many yet undiscovered cases of
627 strong episodic positive selection in other species. Genes of a greater length than *bam* are of
628 particular interest as there is potentially more statistical power to detect positive selection
629 resulting from the Buffering model.

630 We also want to emphasize that a fit to the Conflict compared to the Buffering model
631 does not by itself imply a conflict drives the positive selection observed. A change in function
632 that favors diversification of the protein-coding gene would also give similar results because, like
633 the Conflict model, selection to refine a new function would likely favor positive selection for
634 physicochemically different amino acids, which is selection for increasingly diverse Miyata
635 scores. A recent analysis of CRISPR/Cas-9 generated nulls in five *Drosophila* species raises this
636 possibility for *bam* (Bubnell et al. 2021). Whether the observed changes in function are
637 associated with conflict with *Wolbachia* remains an open question as the two are not mutually
638 exclusive.
639 Ultimately, we suggest that the Buffering model is a new entry to the suite of models that
640 need to be considered in cases where molecular population genetic evidence is found for
641 departures from selective neutrality consistent with positive selection. The Buffering model is a
642 framework that could apply to populations that experience cycles of higher mutational loads,
643 followed by positive selection. This is observed in seasonally small populations, where a drop in
644 population size allows the fixation of some deleterious alleles that are subsequently purged from
645 the population. Additionally, populations in changing environments may experience higher
646 mutational loads at the onset of the change. This phenomenon would be similar to that of
647 antagonistic pleiotropy, in which, for our case, one environment is more tolerant of various
648 alleles, allowing some alleles to fix that would be considered deleterious in the subsequent
649 environment. In the subsequent environment, positive selection favors new mutations that return
650 the gene to its optimum sequence (Chen and Zhang 2020).
651
652 **Supplemental Materials**:
653 Supplemental materials are available at *Molecular Biology and Evolution online*.
654
655 **Acknowledgements**
656 This work was supported by the National Institute of General Medical Sciences at the National
657 Institutes of Health grant number R01GM095793 to C.F.A. We thank Jesús Murga Moreno,
658 Antoni Barbadilla Prados, and Sònia Casillas Viladerrams for discussing and sharing their iMKT
659 script. We also thank Benjamin C. Haller for constructive discussion on implementing our

models in SLiM and feedback to improve the clarity of this manuscript and Jeffrey D. Jensen and Andrew G. Clark for thoughtful suggestions on an earlier draft of the manuscript.

Author Contributions

RS - coding, investigation, analysis, and writing. MW – conceptualization, investigation, analysis, and writing. PWM – writing and supervision. CFA - conceptualization, funding acquisition, and writing and supervision.

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.

675 **References**

676

677 Akashi H. 1999. Inferring the fitness effects of DNA mutations from polymorphism and
678 divergence data: statistical power to detect directional selection under stationarity and free
679 recombination. *Genetics* 151:221-238.
680 Baldo L, Bartos JD, Werren JH, Bazzocchi C, Casiraghi M, Panelli S. 2002. Different rates of
681 nucleotide substitutions in Wolbachia endosymbionts of arthropods and nematodes: arms race or
682 host shifts? *Parassitologia* 44:179-187.
683 Baldo L, Desjardins CA, Russell JA, Stahlhut JK, Werren JH. 2010. Accelerated microevolution
684 in an outer membrane protein (OMP) of the intracellular bacteria Wolbachia. *BMC Evol Biol*.
685 10:48.
686 Bauer DuMont VL, Flores HA, Wright MH, Aquadro CF. 2007. Recurrent positive selection at
687 Bgcn, a key determinant of germ line differentiation, does not appear to be driven by simple
688 coevolution with its partner protein bam. *Mol Biol Evol.* 24:182-191.
689 Bauer DuMont VL, White SL, Zinshteyn D, Aquadro CF. 2021. Molecular population genetics
690 of Sex-lethal (Sxl) in the Drosophila melanogaster species group: a locus that genetically
691 interacts with Wolbachia pipientis in Drosophila melanogaster. *G3(Bethesda)* 11(8):jkab197.
692 Bubnell JE, Fernandez-Begne P, Ulbing CKS, Aquadro CF. 2021. Diverse wMel variants of
693 Wolbachia pipientis differentially rescue fertility and cytological defects of the bag of marbles
694 partial loss of function mutation in Drosophila melanogaster. *G3 (Bethesda)* 11(12): jkab312.
695 Campos JL, Zhao L, Charlesworth B. 2017. Estimating the parameters of background selection
696 and selective sweeps in Drosophila in the presence of gene conversion. *Proc Natl Acad Sci U S*
697 *A.* 114(24):E4762-E4771.
698 Charron C, Nicolaï M, Gallois JL, Robaglia C, Moury B, Palloix A, Caranta C. 2008. Natural
699 variation and functional analyses provide evidence for co-evolution between plant eIF4E and
700 potyviral VPg. *Plant J* 54:56-68.
701 Chen P, Zhang J. 2020. Antagonistic pleiotropy conceals molecular adaptations in changing
702 environments. *Nat Ecol Evol*. 4:461-469.
703 Choi, J. Y., and Aquadro, C. F. 2014. The coevolutionary period of Wolbachia pipientis infecting
704 Drosophila ananassae and its impact on the evolution of the host germline stem cell regulating
705 genes. *Mol Biol and Evol*, *31*(9), 2457–2471.

706 Comeron JM, Ratnappan R, Bailin S. 2012. The Many Landscapes of Recombination in
707 Drosophila melanogaster. *PLOS Genetics* 8:33-35.
708 Demogines A, Abraham J, Choe H, Farzan M, Sawyer SL. 2013. Dual Host-Virus Arms Races
709 Shape an Essential Housekeeping Protein. *PLOS Biology* 11:e1001571.
710 Eyre-Walker A. 2006. The genomic rate of adaptive evolution. *Trends Ecol Evol*. 21(10): 569-
711 575.
712 Eyre-Walker A and Keightley PD. 2009. Estimating the rate of adaptive molecular evolution of
713 slightly deleterious mutations and population size change. *Mol Biol and Evol*. 26:2097-2108.
714 Fay JC, Wyckoff GJ, Wu C-I. 2001. Positive and Negative Selection on the Human Genome.
715 *Genetics* 158:1227.
716 Flores HA, Bauer DuMont VLB, Fatoo A, Hubbard D, Hijji M, Barbash DA, Aquadro CF.
717 2015a. Adaptive evolution of genes involved in the regulation of germline stem cells in
718 Drosophila melanogaster and D. simulans. *G3 (Bethesda)* 5(4):583-592.
719 Flores HA, Bubnell JE, Aquadro CF, Barbash DA. 2015b. The Drosophila bag of marbles Gene
720 Interacts Genetically with Wolbachia and Shows Female-Specific Effects of Divergence. *PLOS*
721 *Genetics* 11:1-31.
722 Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science*
723 185:862-864.
724 Haller BC and Messer PW. 2019. SLiM 3: Forward genetic simulations beyond the Wright-
725 Fisher model. *Mol Biol and Evol*. 36(3):632-637.
726 Haller BC, Galloway J, Kelleher J, Messer PW, Ralph PL. 2019. Tree-sequence recording in
727 SLiM opens new horizons for forward-time simulation of whole genomes. *Mol Ecol Resour.*
728 19(2):552-566.
729 Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl*
730 *Acad Sci U S A* 89:10915-10919.
731 Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. 2008. How many
732 species are infected with Wolbachia?--A statistical analysis of current data. *FEMS Microbiol Lett*
733 281:215-220.
734 Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous
735 mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics* 196:313-
736 320.

737 Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the
738 genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines.
739 Genome Res. 19:1195-1201.
740 Kimura M, Ohta T. 1969. The average number of generations until fixation of a mutant gene in a
741 finite population. *Genetics* 61:763-771.
742 Kriesner P, Hoffman AA, Lee SF, Turelli M, Weeks AR. 2013. Rapid sequential spread of two
743 Wolbachia variants in Drosophila simulans. *PLoS Pathog* 9(9): e1003607.
744 Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH,
745 Pool JE. 2015. The Drosophila genome nexus: A population genomic resource of 623 Drosophila
746 melanogaster genomes, including 197 from a single ancestral range population. *Genetics*
747 199:1229-1241.
748 Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. *Methods Mol Biol*. 1079:155-
749 170.
750 McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila.
751 *Nature* 351:652-654.
752 McLaughlin RN, Malik HS. 2017. Genetic conflicts: The usual suspects and beyond. *J Exp Biol*.
753 220:6-17.
754 Meany MK, Conner WR, Richter SV, Bailey JA, Turelli M, Cooper BS. 2019. Loss of
755 cytoplasmic incompatibility and minimal fecundity effects explain relatively low Wolbachia
756 frequencies in Drosophila mauritiana. *Evolution* 73:1278-1295.
757 Messer PW and Petrov DA. 2013. Frequent adaptation and the McDonald Kreitman test. *PNAS*
758 110:8615-8620.
759 Miyata T, Miyazawa S, Yasunaga T. 1979. Two types of amino acid substitutions in protein
760 evolution. *J Mol Evol*. 12:219-236.
761 Murga-Moreno J, Coronado-Zamora M, Hervas S, Casillas S, Barbadilla A. 2019. iMKT: The
762 integrative McDonald and Kreitman test. *Nucleic Acids Res* 47:W283-W288.
763 Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM.
764 2012. Population genomics of the Wolbachia endosymbiont in Drosophila melanogaster. *PLOS*
765 *Genetics* 8:e1003129.
766 Riegler M, Sidhu M, Miller WJ, O'Neill SL. 2005. Evidence for a global Wolbachia replacement
767 in Drosophila melanogaster. *Curr Biol.* 15:1428-1433

768 Russo CA, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid
769 species. *Mol Biol and Evol*. 12:391-404.
770 Signor SA, New FN, Nuzhdin S. 2018. A Large Panel of Drosophila simulans Reveals an
771 Abundance of Common Variants. *Genome Biol and Evol*. 10:189-206.
772 Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in Drosophila. *Nature* 415:1022-
773 1024.
774 Starr DJ, Cline TW. 2002. A host-parasite interaction rescues Drosophila oogenesis defects.
775 *Nature* 418:76-79.
776 Taylor MJ, Bandi C, Hoerauf A. 2005. Wolbachia bacterial endosymbionts of filarial nematodes.
777 *Adv Parasitol* 60:245-284.
778 Turelli M, Cooper BS, Richardson KM, Ginsberg PS, Peckenpaugh B, Antelope CX, Kim KJ,
779 May MR, Abrieux A, Wilson DA, et al. 2018. Rapid global spread of wRi-like Wolbachia across
780 multiple Drosophila. *Curr Biol*. 28:963-971.e968.
781 Weinert LA, Araujo-Jnr EV, Ahmed MZ, Welch JJ. 2015. The incidence of bacterial
782 endosymbionts in terrestrial arthropods. *Proc R Soc B Biol Sci*. 282:20150249.
783 Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol and Evol*.
784 24:1586-1591.
785 Zhai W, Nielsen R, Slatkin M. 2009. An investigation of the statistical power of neutrality tests
786 based on comparative and population genetic data. *Mol Biol and Evol.* 26:273-283.
787 Zug R, Hammerstein P. 2012. Still a host of hosts for Wolbachia: Analysis of recent data
788 suggests that 40% of terrestrial arthropod species are infected. *PLoS One* 7:e38544.
789
790

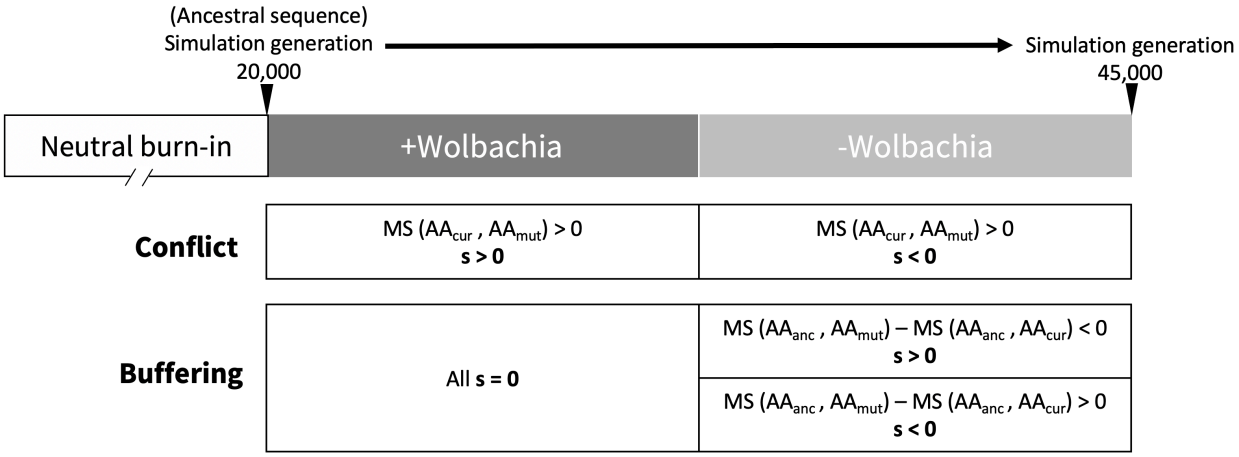


Fig. 1. Simulation setup for Conflict and Buffering Base models. Selection on new nonsynonymous mutations (mutated amino acid, AA_{mut}) is determined by their Miyata score (MS) to the appropriate reference amino acid (the current amino acid, AA_{cur} , or the ancestral amino acid, AA_{anc}).

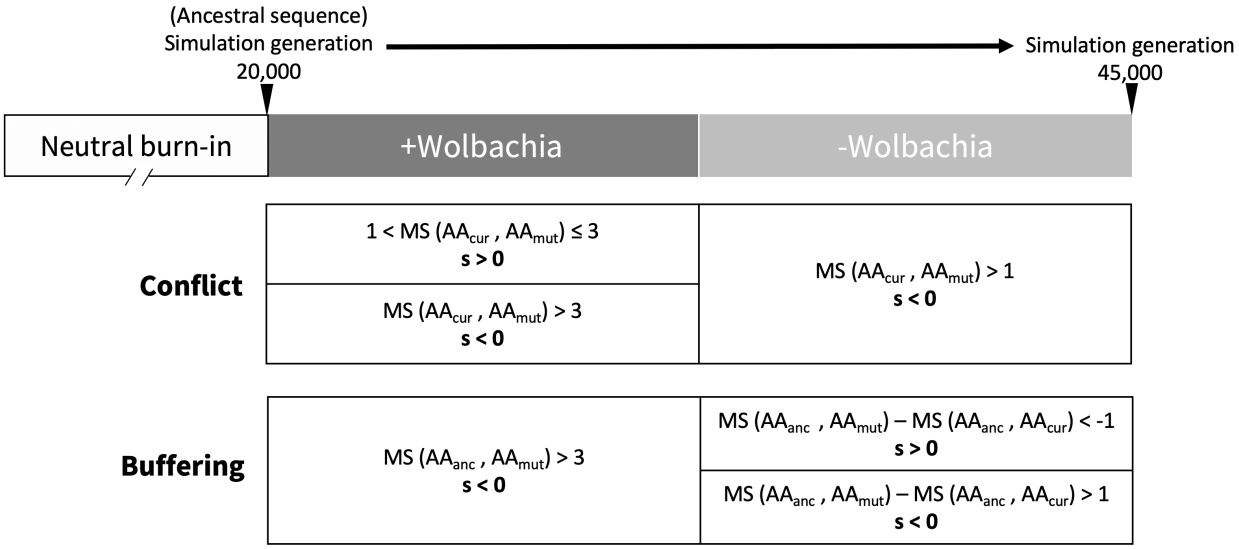
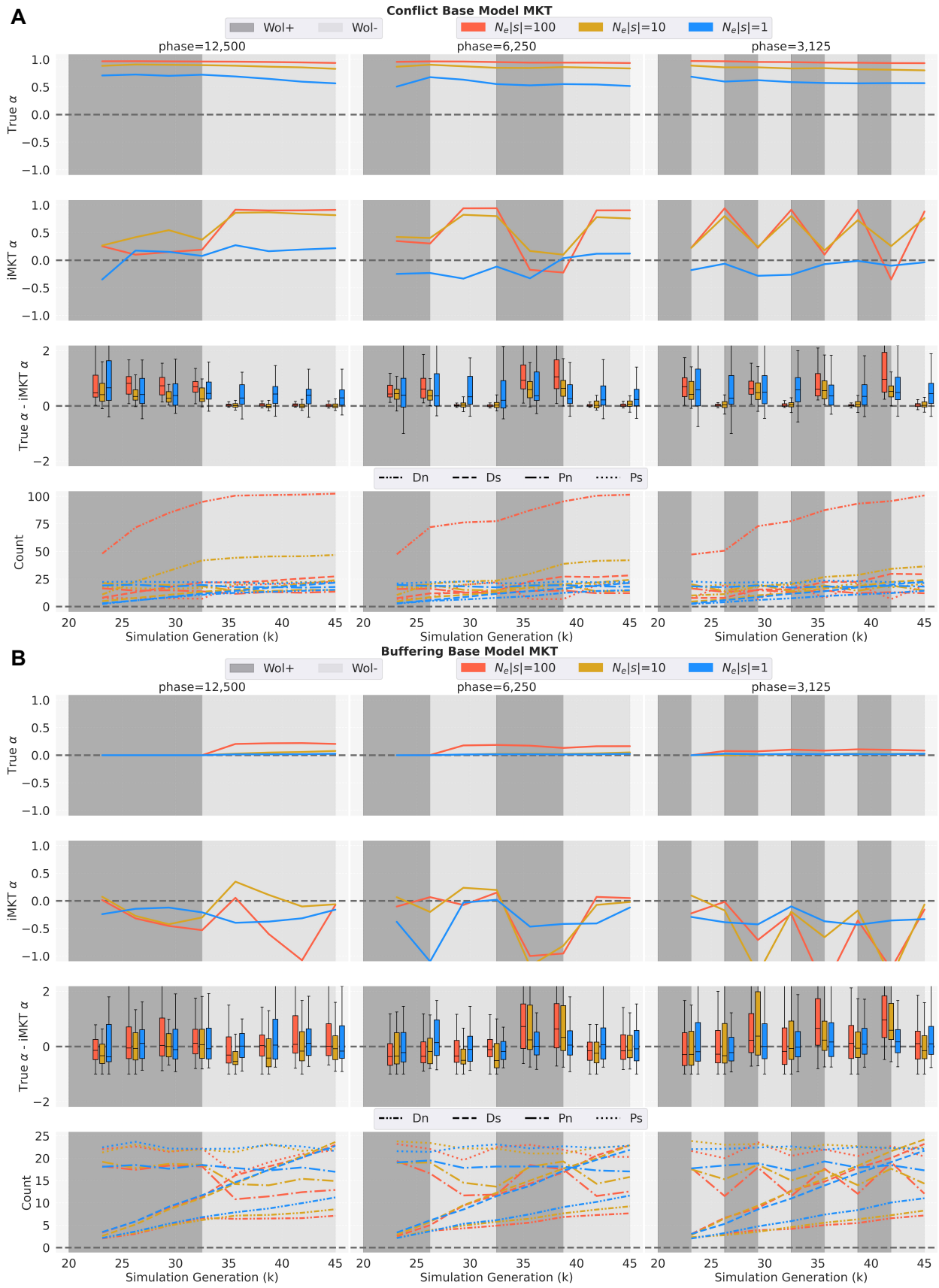
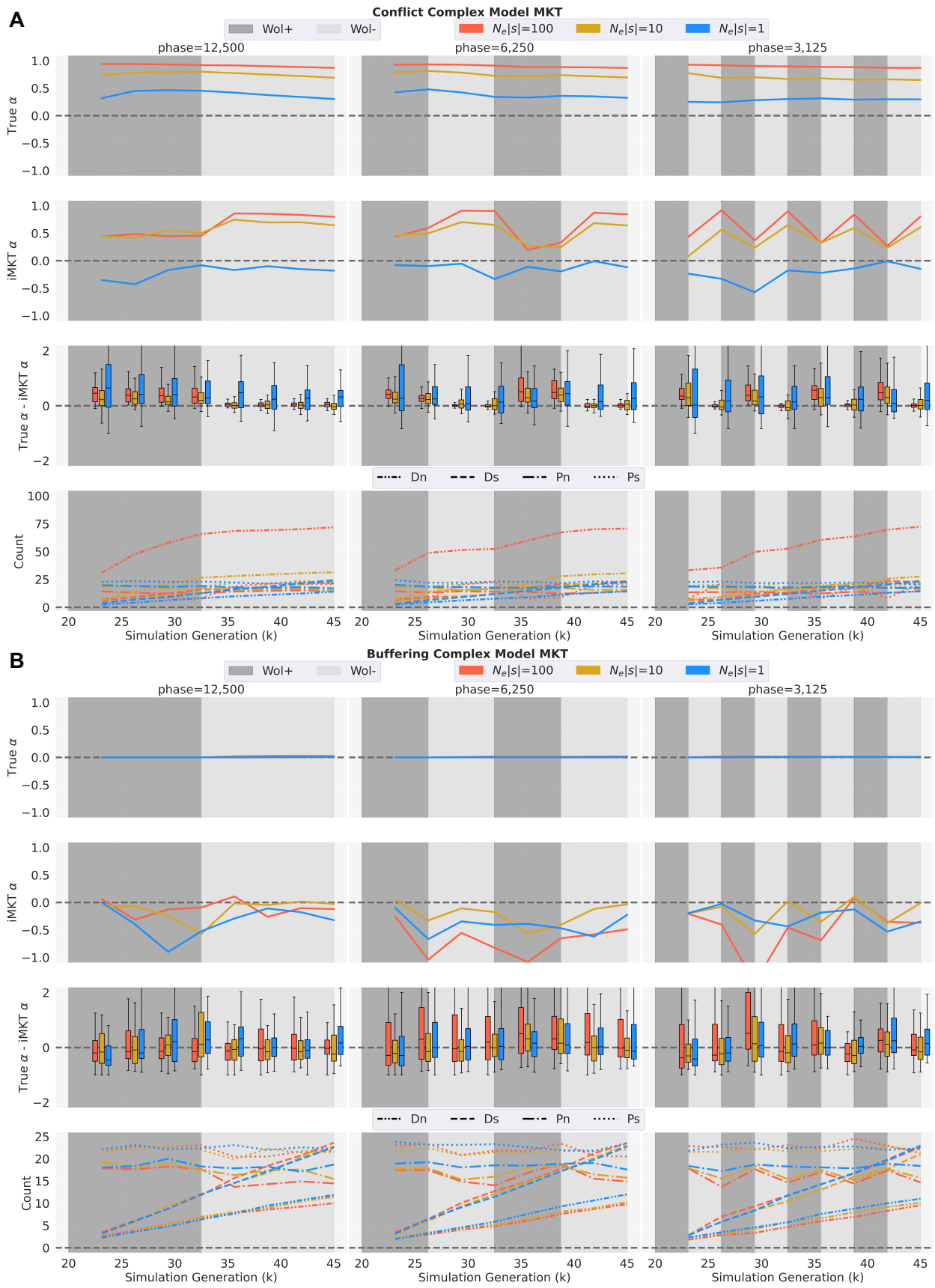


Fig. 2. Simulation setup for Conflict and Buffering Complex models. Selection on new nonsynonymous mutations (mutated amino acid, AA_{mut}) is determined by their Miyata score (MS) to the appropriate reference amino acid (the current amino acid, AA_{cur}, or the ancestral amino acid, AA_{anc}).


A
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
Buffering Base Model MKT

806 **Fig. 3. MK test results of simulations for two base models.** (A) Conflict model; (B) Buffering
807 model. Each panel shows MKT analyses with different selection coefficients of $N_e|s|$=100,
808 $N_e|s|$=10, and $N_e|s|$=1 graphed across alternating phases (phase length=12,500, 6,250, 3,125 and
809 simulation generations) of *Wolbachia* infection (Wol+, dark grey) and *Wolbachia* absence (Wol-,
810 light grey) post-burn-in period. In each panel, row 1: the average true α in the simulations; row
811 2: the average iMKT α in the simulations (FWW correction, SNPs frequency > 5% only); row 3:
812 the distributions of differences between the true and iMKT α every 3,125 simulation
813 generations; row 4: The average of each iMKT component (D_n, D_s, P_n, P_s).
814


A
Conflict 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)
B
Buffering Complex Model MKT

816 **Fig. 4. MK test results of simulations for two complex models:** (A) Conflict model, (B)
817 Buffering model. Each panel shows MKT analyses with different selection coefficients of
818 $N_e|s|$=100, $N_e|s|$=10, and $N_e|s|$=1 graphed across alternating phases (phase length=12,500, 6,250,
819 3,125 and simulation generations) of *Wolbachia* infection (Wol+, dark grey) and *Wolbachia*
820 absence (Wol-, light grey) post-burn-in period. In each panel, row 1: the average true α in the
821 simulations; row 2: the average iMKT α in the simulations (FWW correction, SNPs frequency >
822 5% only); row 3: the distributions of differences between the true and iMKT α every 3,125
823 simulation generations; row 4: The average of each iMKT component (D_n, D_s, P_n, P_s).

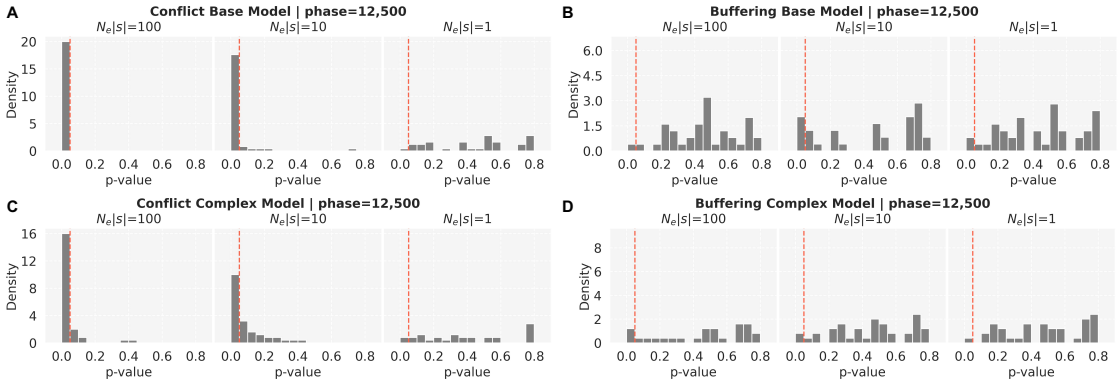
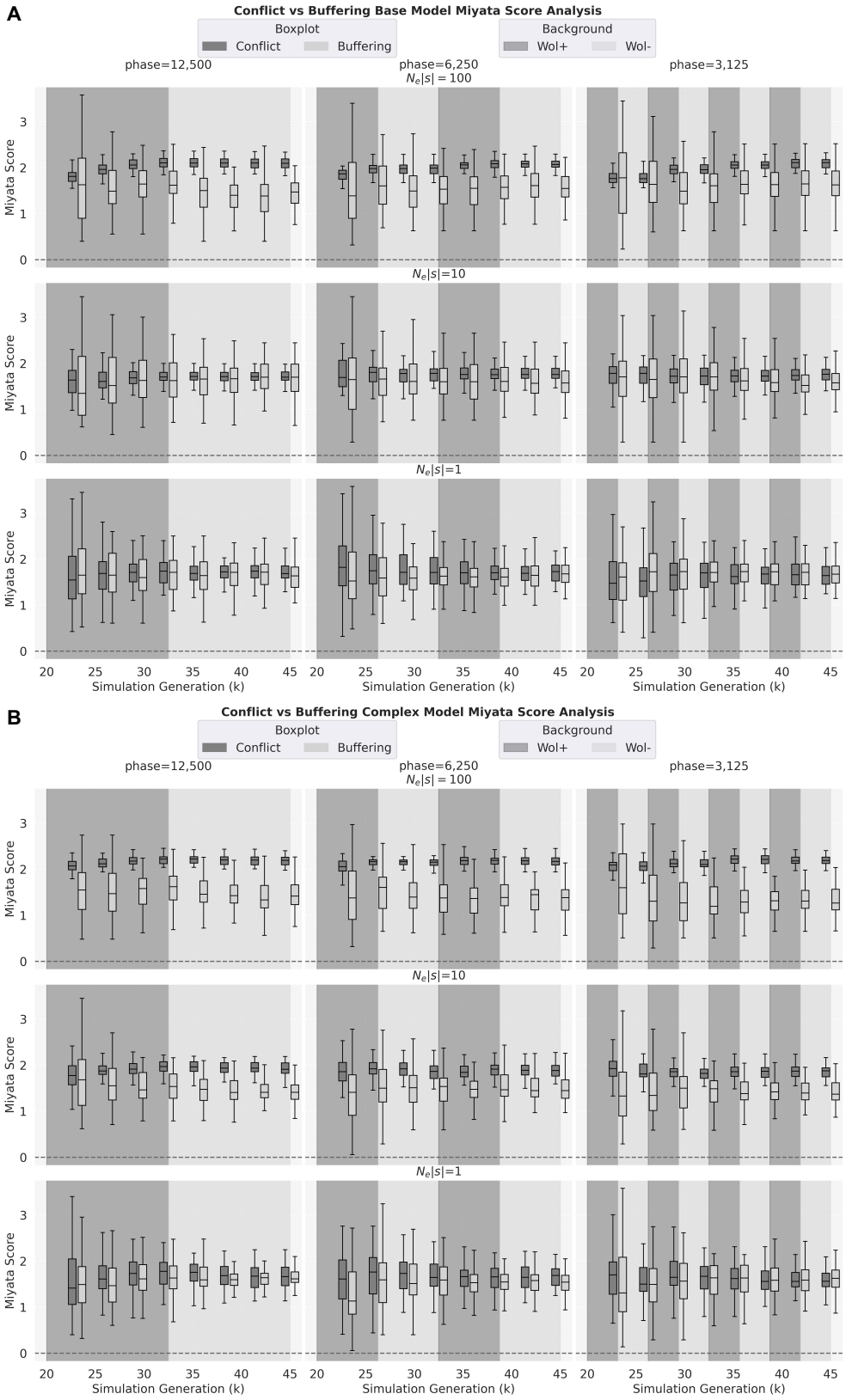


Fig. 5. Distributions of iMKT p-values. iMKT p-values (FWW correction, SNPs frequency > 5% only) for simulation runs with *Wolbachia* phase=12,500 simulation generations and $N_e|s|$=100, 10, 1 for the simulated models at 45,000 simulation generation. The vertical red line denotes p-value=0.05. Note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

831

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

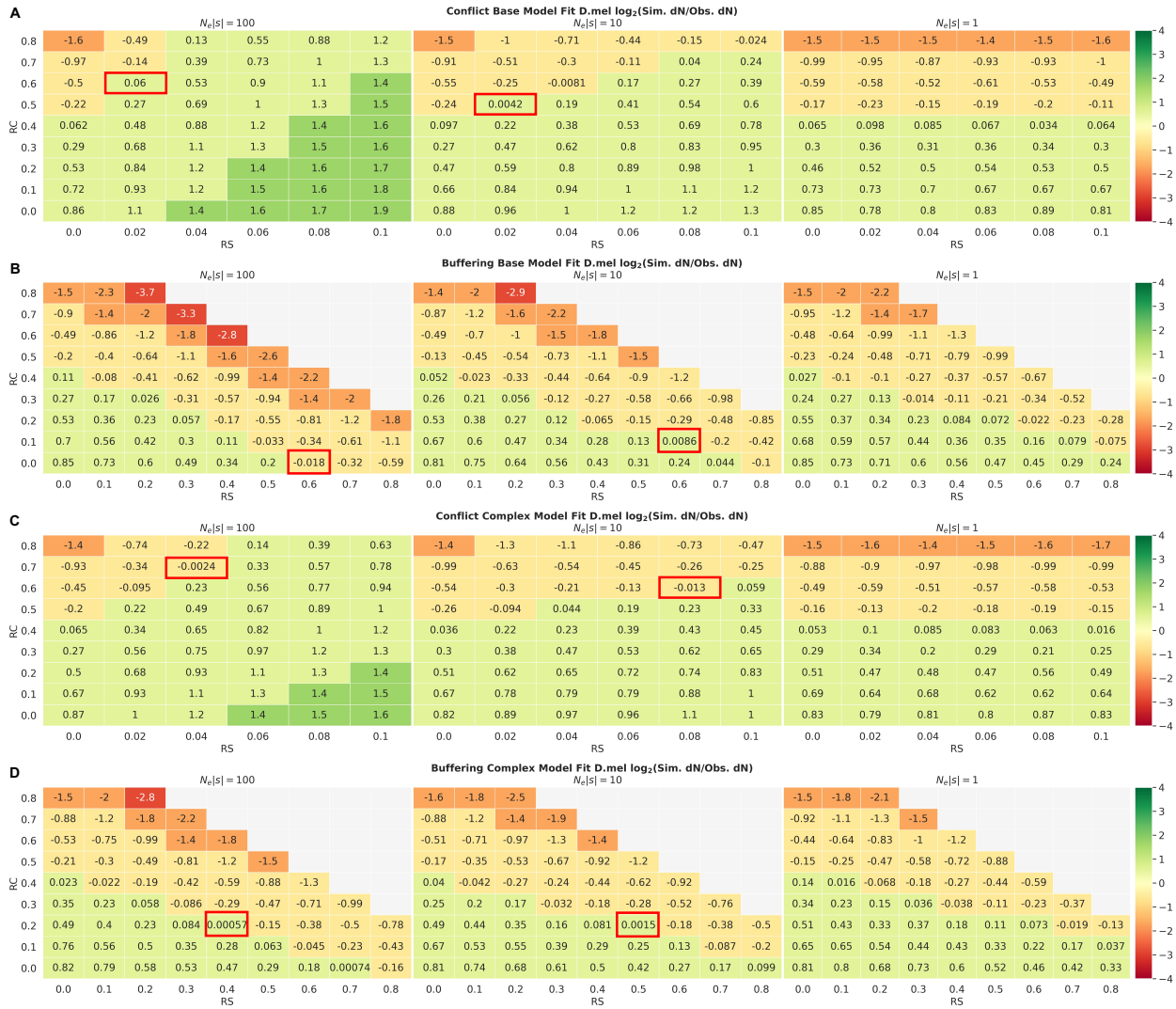
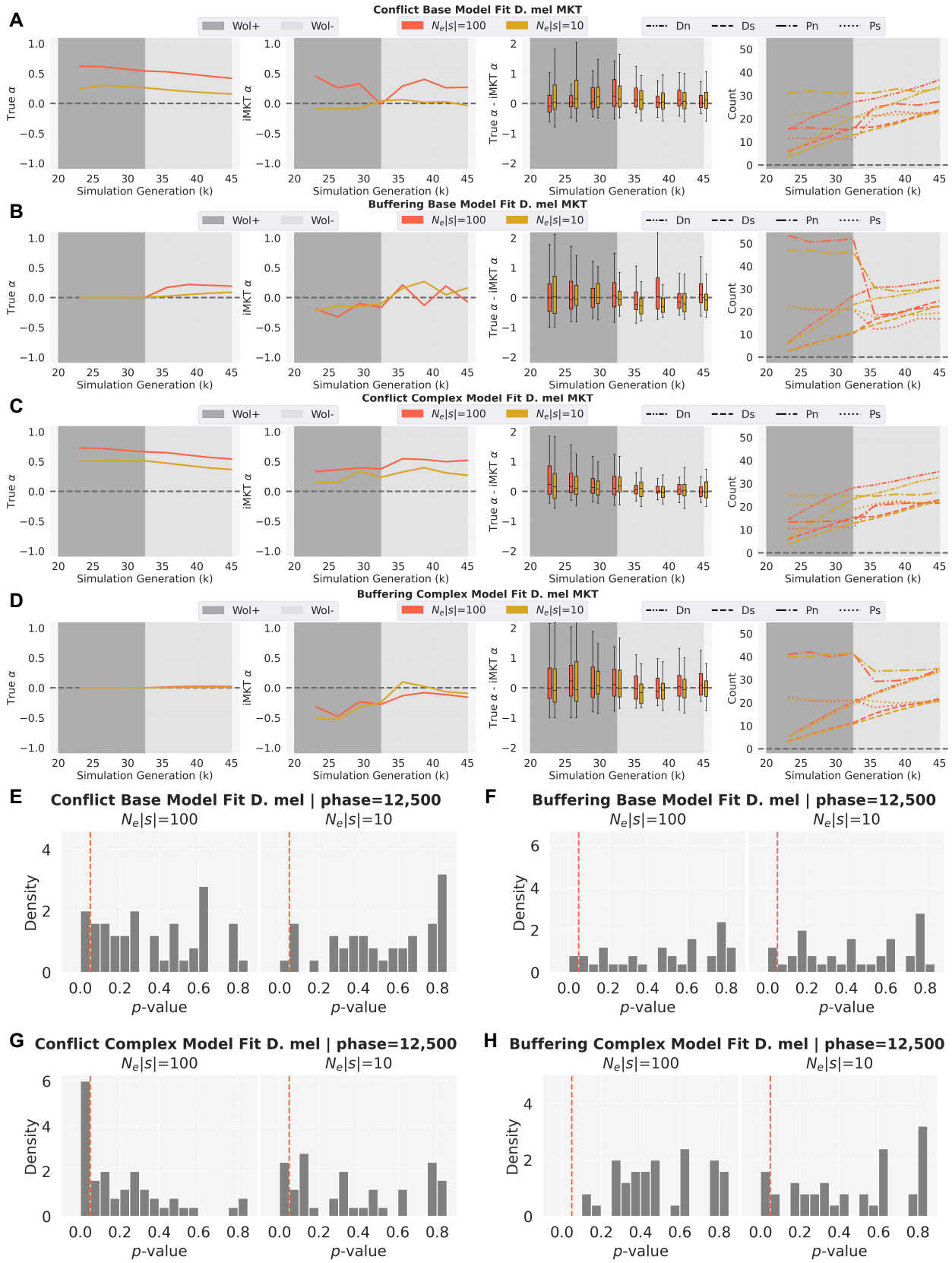


Fig. 7. Heatmap of $log_2 \frac{Simulation\ dN}{Observed\ dN}$ for four models to fit the empirical data. For each model, average dN was calculated across 50 simulation runs per each pair of conserved-site ratio and selection-site ratio to find the combination of parameters best fitting the empirical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


A
Conflict Base Model Fit D. mel MKT
B
Buffering Base Model Fit D. mel MKT
C
Conflict Complex Model Fit D. mel MKT
D
Buffering 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
Conflict Base Model Fit D. mel | phase=12,500
F
Buffering Base Model Fit D. mel | phase=12,500
G
Conflict Complex Model Fit D. mel | phase=12,500
H
Buffering Complex Model Fit D. mel | phase=12,500
Density
p-value

846 **Fig. 8. MK test results of simulations for best-fit models for *D. melanogaster*.** (A) MK test α
847 analysis of each model with different selection coefficients of $N_e|s|$=100, $N_e|s|$=10 graphed at
848 phase length=12,500 simulation generations of *Wolbachia* infection (Wol+, dark grey) and
849 *Wolbachia* absence (Wol-, light grey) post-burn-in period. In each panel, row 1: the average true
850 α in the simulations; row 2: the average iMKT α in the simulations (FWW correction, SNPs
851 frequency > 5% only); row 3: the distributions of differences between the true and iMKT α
852 every 3,125 simulation generations; row 4: The average of each iMKT component (D_n, D_s, P_n,
853 P_s). (B) Distributions of iMKT p-values. iMKT p-values (FWW correction, SNPs frequency >
854 5% only) for simulation runs with *Wolbachia* phase=12,500 simulation generations and
855 $N_e|s|$=100, 10 for the simulated models at 45,000 simulation generation. The vertical red line
856 denotes p-value=0.05. Note that distributions are normalized to have an area of 1 under the
857 histograms.
858

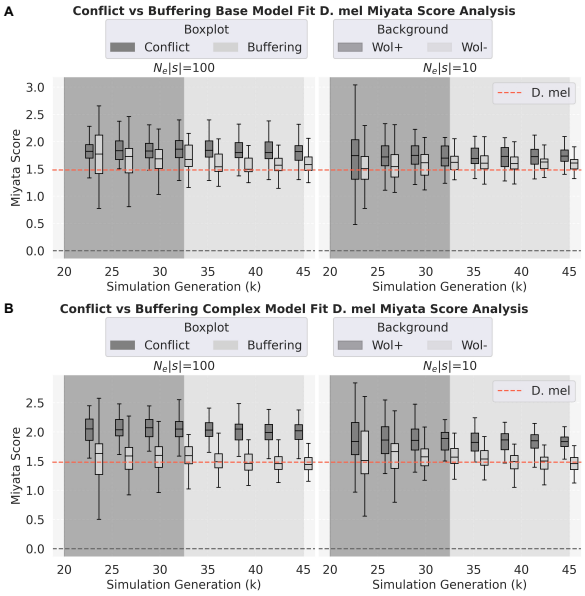


Fig. 9. Distributions of Miyata score for the *D. melanogaster* samples from the simulations with best-fit parameters. The boxplots are the distributions of Miyata scores for each model at every 3,125 generation. (A) Conflict and Base models; (B) Conflict and Buffering Complex model. The distributions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

866 **Table 1.** Empirical (biological) estimates for evolutionary parameters and scaled estimates used
867 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

868 N_e: effective population size; ρ: recombination rate; μ: mutation rate; t: time in unit of
869 generation
870

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

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

872