

Electronic Supplementary Materials for Establishment of New Mutations - Feder et al.

Parameter combinations used in the simulations

The simulations were run varying the parameter values for migration probability (m), recombination rate (r), divergent selection strength on habitat-related survivorship (s), and number of loci ($nloci$). 100,000 trials were performed for each combination of parameter values separately introducing the new A mutant allele into population 1 (where it was favoured) and population 2 (where it was selected against). We performed most of the stochastic simulation runs considering only one additional gene (locus 2) other than the new mutation under divergent selection, but also ran some simulations with three and five unlinked loci also each experiencing strong selection ($s = 0.5$). We examined five different rates of recombination between the new mutation and pre-diverged locus ranging from very tight linkage to random assortment ($r = 0.001, 0.01, 0.05, 0.10, \text{ and } 0.50$). Three different levels of migration probability were considered ($m = 0.001$ [low], 0.01 [moderate], and 0.1 [high]). We considered these rates of migration to be appropriate for most cases of divergence-with-gene-flow. When divergence-with-gene-flow occurs in nature, it is likely that some form of spatial structuring is often associated with the distribution of habitats, and hence populations, across the landscape (e.g., insects associated with different host plants that may remain on their natal plant during their lifetimes). As a consequence, migration rates are often not expected to be random ($m = 0.5$) between populations, but reflect to varying degrees the structure of the different habitats and the dispersal capacities of individuals across space. Our values of $m = 0.001, 0.01, \text{ and } 0.1$ were chosen to reflect this structuring from greater to lesser significance. We leave it to future work to examine the probabilities of establishment for new habitat choice and assortative mating

mutations. It is important to note in this regard that in the current simulations, as was the case in our previous work (1), mating is occurring in a habitat specific manner. Thus, our models differ from a Levene (2) model in which surviving adults in the two habitats contribute gametes to a single, common gene pool. Instead, individuals are mating randomly only with the other surviving adults from the same habitat and not with adults from the alternate habitat (i.e., mating is habitat specific, as may often be the case during divergence-with-gene flow). In addition, following mating, offspring are not distributing themselves randomly but tending to remain in their natal habitats with probability $1-m$, so there is a degree of habitat fidelity (innate habitat choice) in the models.

Several different intensities of divergent selection affecting the already established locus or loci (designated s_o) and the new mutation (s_n) were analyzed ($s = 0.01$ [weak], 0.1 [moderate], 0.2 [moderately strong], 0.333 [strong], and 0.50 [very strong]). Fitness interactions between loci were multiplicative, while segregating alleles under selection at a locus were considered to interact in a partially dominant manner ($h = 0.5$), such that the relative fitnesses of alternate homozygotes and the heterozygote at a locus in population 1 were 1 , $1-s_1$ and $1-s_1/2$, respectively, and were 1 , $1-s_2$ and $1-s_2/2$ in population 2 (i.e., divergent selection generated performance tradeoffs between the habitats occupied by populations 1 and 2). For all of the simulation runs in the current study, all parameter values were symmetric between populations. Finally, we also varied the number of loci ($nloci$) under divergent selection prior to the new mutation from zero (a new mutation occurring in an undifferentiated genome) to five (a new mutation occurring at a recombination distance r to a previously differentiated locus in a genome containing four other unlinked, differentiated loci). In effect, adding a locus to the simulations was equivalent to adding an additional chromosome to the genome of an organism that contained

a gene under divergent selection. In so doing, we could examine the effects of increased population differentiation for divergence and genome hitchhiking during the speciation process.

Analytical methods

We compared our simulation results to analytical predictions for the probability of establishment of a new mutation based on the approach described in Yeaman and Otto (3) and Yeaman and Whitlock (4). The approach uses equations developed by Barton and Bengtsson (5) to calculate the effective migration (m_e) rate experienced by a new mutation that is linked by some rate of recombination (r) to an already diverged locus, or loci (Barton and Bengtsson equation A4). This calculated m_e value is then used to predict the probability of establishment of a mutation introduced in the patch where it is favoured based on equations 3 and 9 from Yeaman and Otto (3). The Barton and Bengtsson (5) model assumes that the alleles at the already diverged locus are initially fixed in each population, such that the new mutation is always introduced into a genetic background with a favoured allele. As such, the analytical results do not provide exact predictions of the true establishment probability of a randomly occurring mutation and only consider mutations arising in the population in which they are favoured. As discussed in the body of the paper and below, these assumptions cause the analytical estimates to overestimate establishment probabilities, particularly under conditions of higher migration rate and weaker selection, key considerations for speciation-initiated-with-gene-flow.

Comparison of analytical approximations to simulations

Analytical approximations for the probability of establishment of new mutations arising in the favoured habitat using the approach of Yeaman and Otto (3) and Yeaman and Whitlock (4)

were generally in good agreement with the simulation results, especially for low migration rates (see Figs. S2A and B; Table S1). However, for higher migration rates ($m = 0.1$), the analytical approximation overestimated the effects of divergence hitchhiking for new mutations closely linked to the pre-diverged locus (Fig. S2C). The reason for the overestimation was that the analytical approach assumes that the new mutation arises in populations differentially fixed for locally adapted alleles. With low migration rates and very strong divergent selection, this is a reasonable assumption. However, with high migration rates and weaker selection, populations will possess a significant frequency of locally maladapted genotypes. In this case, new mutations will often arise linked to a disfavoured allele(s) at other loci. In the absence of recombination, these new mutations will be eliminated from the population. Our simulations took this into account by randomly introducing new mutations into genotypes initially present in selection/migration equilibrium between populations. The analytical approximations did not, and therefore increasingly overestimated establishment probabilities for new mutations in ever closer linkage to a selected locus (Fig. S2C), as new mutations would always benefit from increasingly tighter association with other locally favoured alleles of like kind in the genome. The overestimation by the analytical approximation under conditions of high migration and weaker selection underscores the need for caution in interpreting the role of divergence hitchhiking in speciation-with-gene-flow, as the conditions where the process should be most important are those in which analytical estimates may be most inaccurate, perhaps contributing to an enhanced historical appreciation for the potential role of linkage in speciation-with-gene-flow. Full spreadsheets of simulation and analytical results are available from the authors upon request.

Supplementary Literature cited

1. Feder JL, Nosil P. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*. 2010. 64:1729-1747.
2. Levene H. Genetic equilibrium when more than one niche is available. *Am. Nat.* 1953; 87: 331-333.
3. Yeaman S, Otto SP. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution*. 2011; 65:2123-2129.
4. Yeaman S, Whitlock MC. The genetic architecture of adaptation under migration-selection balance. *Evolution*. 2011; 65:1897-1911.
5. Barton N, Bengtsson BO. The barrier to genetic exchange between hybridizing populations. *Heredity*. 1986; 57:357-376.

Table S1. For a subset of the parameters explored, a comparison of the results for the probability of establishment ($P[e]$) of a new mutation derived from simulations (sims) versus analytical approximations (analytic). Details of how the analytical approximations were derived, and a full comparison of the simulation and analytical approaches, can be found in the text of the electronic supplementary materials. Likewise, a spreadsheet of the full results from both approaches is available upon request from the authors.

m	s_o	s_n	r	(P[e])sims	(P[e])analytic
0.001	0.01	0.01	0.5	0.00835	0.00826
0.001	0.01	0.01	0.1	0.00847	0.00833
0.001	0.01	0.01	0.05	0.00848	0.00840
0.001	0.01	0.01	0.01	0.00869	0.00881
0.001	0.01	0.01	0.001	0.00884	0.00972
0.001	0.1	0.01	0.5	0.00878	0.00840
0.001	0.1	0.01	0.1	0.00897	0.00882
0.001	0.1	0.01	0.05	0.00892	0.00912
0.001	0.1	0.01	0.01	0.00972	0.00973
0.001	0.1	0.01	0.001	0.00954	0.01001
0.001	0.333	0.333	0.5	0.36753	0.39193
0.001	0.333	0.333	0.1	0.36775	0.39251
0.001	0.333	0.333	0.05	0.36792	0.39271
0.001	0.333	0.333	0.01	0.36941	0.39294

0.001	0.333	0.333	0.001	0.36925	0.39301
0.001	0.333	0.1	0.5	0.10065	0.10382
0.001	0.333	0.1	0.1	0.10290	0.10453
0.001	0.333	0.1	0.05	0.10210	0.10478
0.001	0.333	0.1	0.01	0.10132	0.10507
0.001	0.333	0.1	0.001	0.10162	0.10515
0.001	0.333	0.01	0.5	0.00894	0.00873
0.001	0.333	0.01	0.1	0.00927	0.00941
0.001	0.333	0.01	0.05	0.00921	0.00966
0.001	0.333	0.01	0.01	0.00910	0.00996
0.001	0.333	0.01	0.001	0.00897	0.01004
0.01	0.333	0.333	0.5	0.35815	0.38223
0.01	0.333	0.333	0.1	0.35376	0.38797
0.01	0.333	0.333	0.05	0.35473	0.38998
0.01	0.333	0.333	0.01	0.35468	0.39229
0.01	0.333	0.333	0.001	0.35206	0.39294
0.01	0.333	0.1	0.5	0.09098	0.09244
0.01	0.333	0.1	0.1	0.09529	0.09903
0.01	0.333	0.1	0.05	0.09607	0.10144
0.01	0.333	0.1	0.01	0.09656	0.10426
0.01	0.333	0.1	0.001	0.09793	0.10507
0.01	0.333	0.01	0.5	0.00470	0.00318

0.01	0.333	0.01	0.1	0.00674	0.00539
0.01	0.333	0.01	0.05	0.00839	0.00682
0.01	0.333	0.01	0.01	0.00915	0.00914
0.01	0.333	0.01	0.001	0.00933	0.00995
0.1	0.333	0.333	0.5	0.25466	0.29339
0.1	0.333	0.333	0.1	0.25559	0.34002
0.1	0.333	0.333	0.05	0.25201	0.36317
0.1	0.333	0.333	0.01	0.24127	0.38581
0.1	0.333	0.333	0.001	0.23717	0.39226
0.1	0.333	0.1	0.5	0.03498	0.03507
0.1	0.333	0.1	0.1	0.04521	0.05881
0.1	0.333	0.1	0.05	0.05031	0.07350
0.1	0.333	0.1	0.01	0.05664	0.09650
0.1	0.333	0.1	0.001	0.05817	0.10422

Supplementary Figure Legends

Figure S1. The effect of varying selection strength from $s_o = 0.5$ to $s_o = 0.01$ on an already existing locus on the mean probability of establishment of a new mutation under moderate divergent selection ($s_n = 0.1$) arising in the favoured or disfavoured population with high migration rate ($m = 0.1$). Also shown is the baseline probability of establishment for a new mutation on its own not influenced by selection on any other loci in the genome (noH).

Figure S2. Comparison of the analytical approximations and simulation results for the probabilities of establishment of a new mutation under moderate divergent selection ($s_n = 0.1$) linked at varying recombination distance to an already diverged locus under strong selection ($s_o = 0.33$). Results are given for (A) low migration rate ($m = 0.001$), (B) moderate migration rate ($m = 0.01$), and (C) high migration rate ($m = 0.1$). Light stippled line represents analytical approximations for the probabilities of establishment for new beneficial mutations arising in the favoured population 1, heavy dashed line represents computer simulation estimates for the probabilities of establishment for new beneficial mutations arising in the favoured population 1, light dashed line represents the computer simulation estimates for the probabilities of establishment for new beneficial mutations arising in the disfavoured population 2, and the solid line represents the computer simulation estimates for the mean probabilities of establishment for new beneficial mutations arising in the favoured population 1 and disfavoured population 2.

Fig S1.

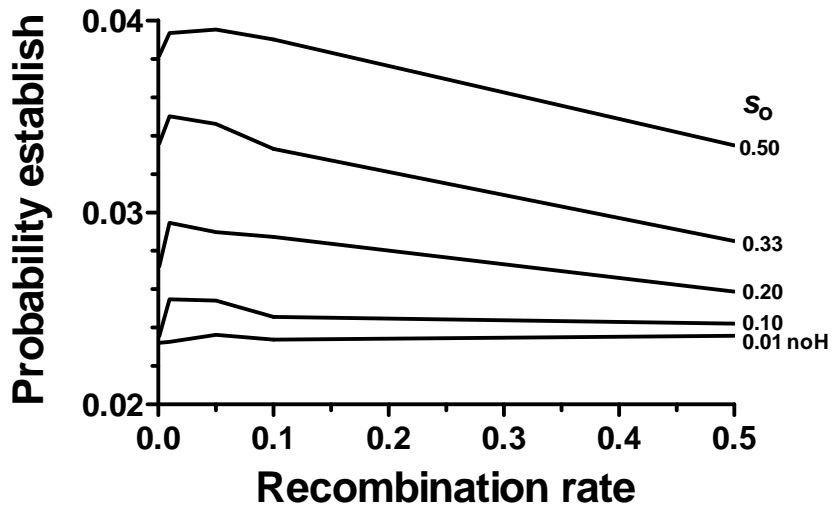


Fig S2.

