Theoretical Expression for the Evolution of Genomic Island of Speciation: From Its Birth to Preservation

T. Sakamoto* **and H. Innan*,[1]**

*SOKENDAI, The Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan.

ABSTRACT Ecological speciation could be driven by divergent selection that works to maintain phenotypes that are adaptive to each niche. In its early stages, genetic divergence (or F_{ST}) can be maintained around the target sites of divergent selection, while in other regions, genetic variation can be mixed by gene flow or migration. Such regions of elevated genetic divergence are called genomic islands of speciation. In this work, we theoretically consider the evolutionary process of a genomic island of speciation, from its birth to stable preservation. Under a simple two-population model, we use a diffusion approach to obtain analytical expressions for the probability of initial establishment of a locally adaptive allele, the reduction of genetic variation due to the spread of the adaptive allele, and the process to the development of a sharp peak of divergence. Our result would be useful to understand how genomes evolve through ecological speciation with gene flow.

KEYWORDS speciation; population genetics; diffusion theory; migration; gene flow; divergent selection

A genomic island of speciation arises in the earlier stages of ecological speciation with gene flow (Wu 2001; Turner *et al.* 2005; Nosil 2012). Speciation can be initiated by the initial establishment of a locally adapted allele to a certain subpopulation. This local establishment could be stably maintained by divergent selection when the allele confers sufficient benefit in the adaptive subpopulation(s), but not (or even deleterious) in others. Due to recombination, the genomic regions that are affected by divergent selection is limited, thereby creating a peak of divergence along chromosome, that is, a genomic island of speciation. We are here interested in the evolutionary behavior of a genomic island of speciation, from its initial establishment to stable preservation.

Theoretically, it would be convenient to consider the process by dividing into three phases, the establishment, erosion and equilibrium phases, as illustrated in Figure 1. We consider a simple situation with two subpopulations, I and II. Assuming a relatively high migration rate between them, the levels of polymorphism within the two subpopulations are similar to each other (measured by the average numbers of pairwise nucleotide differences, π_{w1} and π_{w2}, for subpopulations I and II). In the meantime, the population divergence (measured by π_b, the average numbers of pairwise nucleotide differences between the two subpopulations) is very low (Figure 1A). Then, a *de novo* mutation (the star in Figure 1A) arises in subpopulation I, in which the mutation is advantageous but maladaptive (or deleterious) in subpopulation II. In the establishment phase, the mutation spreads in subpopulation I and nearly fixes (Figure 1B), but its frequency in subpopulation II is low because it should be selected against if migrated into subpopulation II. In a strict sense, this is not a fixation that can be mathematically treated as an absorbing state, because migration keeps providing maladaptive alleles. Therefore, after Kimura (1954), we hereafter use the terminology of "quasi-fixation" for this nearly fixed state. The quasi-fixation should occur quickly, and a selective sweep occurs in

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Thursday 7th February, 2019
[1] *Corresponding author:* Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan. E-mail: innan_hideki@soken.ac.jp

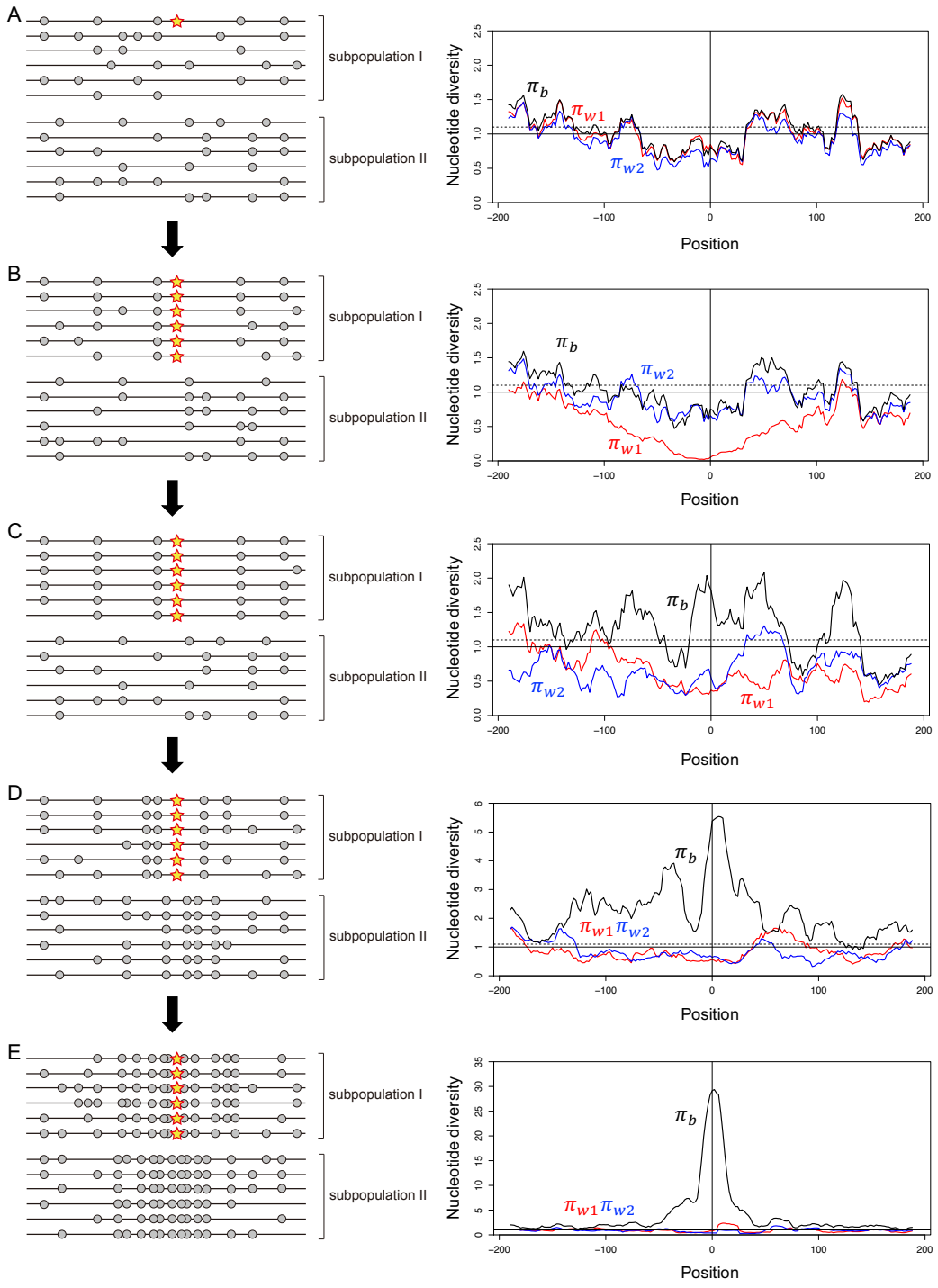


Figure 1 Illustrating the evolution of a genomic island of speciation in a simple two-population model with fairly high migration between them. (A) A locally adaptive de novo mutation arises in subpopulation I at position 0. A typical pattern of polymorphism is shown in left. The star is the locally adaptive mutation and gray circles are neutral polymorphism in the surrounding region. The right panel shows the spacial distributions of nucleotide diversity obtained by a simulation. The polymorphism levels within the two populations (π_{w1} and π_{w2}) are in red and blue, and divergence between the two populations (π_b) is in black. The y-axis is adjusted such that $E(\pi_{w1}) = E(\pi_{w2}) = 1$ under neutrality (the solid line), and the broken line exhibits $E(\pi_b)$. The entire simulated region is 400 kb if the population recombination rate = 0.001 per site is assumed. (B) The mutation quasi-fixes in subpopulation I, causing a drastic reduction in π_{w1}. (C) Migration shuffles polymorphisms in the two subpopulations, while selection works to maintain the quasi-fixation of the mutation. (D) The divergence gradually increases around the mutation, and (E) a clear peak of divergence arises.

subpopulation I (Figure 1B), thereby creating an initial island. The initial genomic island should be as large as the region where genetic variation within the genomic island should be very low in subpopulation I, whereas genetic variation in subpopulation II may not be much affected by the selective sweep (Figure 1B). The erosion phase starts after the initial establishment, during which the genomic island gradually shrinks over time by recombination and migration (Figure 1C). Then, at the end, the genomic island appears as a stable sharp peak of divergence in the equilibrium phase (Figure 1D). The equilibrium size of the genomic island is mainly determined by the balance between selection intensity and the rates of recombination and migration.

The scope of this work is to provide a unified and comprehensive theoretical understanding of the evolution of a new genomic island of speciation, from its birth to stable preservation in equilibrium. We use a simple two-population model, where migration is allowed between subpopulations I and II. Suppose a de novo mutation arises that confers a selective advantage specific to subpopulation I, which is the initial state of our system. Under this model, we derived the following:

for the establishment phase,

(i) The establishment probability of the de novo mutation, that is, the probability that the mutation quasi-fixes in subpopulation I.

(ii) The expected reduction of genetic variation within subpopulations I and II after the quasi-fixation (i.e., local sweep).

for the erosion phase,

(iii) The expected erosion of the initial island as a function of time since the quasi-fixation.

and for the equilibrium phase,

(iv) The expected shape of the genomic island at equilibrium.

There have been several theoretical works that focused on a specific part of these aspects. For (i) the established probability, perhaps the most flexible, useful theoretical framework was introduced by Barton (1987) in a general multiple-island-model. By using a diffusion approximation, Barton (1987) derived a partial differential equation for the establishment probability. Essentially the same result was obtained by Pollak (1966), who used a branching process and the establishment probability was derived from the probability generating function. Barton's differential equation was solved and closed forms of the establishment probability have been available only in several specific situations in *continuous* habitat models. In a one-dimensional continuous habitat model, Barton (1987) solved his partial differential equation analytically assuming two forms of fitness gradient (linear and pocket). Kirkpatrick and Peischl (2013) used a branching process, from which they obtained a partial differential equation that is similar to that of Barton (1987). Then, the authors successfully incorporated changes in fitness gradient along time, and derived an approximate establishment probability.

In *discrete* population models, Barton's general formula (and also Pollak's one) is difficult to handle and have not been fully explored even in a simple two-population model with symmetric migration. Therefore, the currently available theoretical results are not based on Barton's differential equation, and have some limitations. In a continent-island model with *unidirectional* migration, Aeschbacher and Bürger (2014) solved the establishment probability of a locally beneficial mutation linked to another locally beneficial mutation that was already established, where mathematical treatment is

quite straightforward because of unidirectional migration (see also Yeaman *et al.* 2016). Yeaman and Otto (2011) obtained an approximate establishment probability by using a heuristic approach that is a combination of the leading eigenvalue of the transition matrix of deterministic process and Kimura's formula of fixation probability (Kimura 1962). As shown in their paper, this formula well describes the establishment probability when a de novo mutation arises in the adapted subpopulation (i.e., subpopulation I in our model), but it does not work when it arises in the maladapted subpopulation (i.e., subpopulation II in our model). Recently, Tomasini and Peischl (2018) provided an approximate established probability by assuming a slightly supercritical branching process. Their formula works well under the assumption of slightly supercritical approximation, namely, the leading eigenvalue of the transition matrix of deterministic model is not large, but it may not work well when the selection intensity in the adapted subpopulation is very large.

In this work, we derive a closed form formula of the establishment probability in a two-population model with bidirectonal migration along the formulation of Barton (1987). We extended Barton's derivation with simultaneous quadratic equations and solved them allowing unequal subpopulation sizes. Our formula is more general than previous ones (Yeaman and Otto 2011; Tomasini and Peischl 2018); it works with strong selection and it allows that a de novo mutation can arise either subpopulation I or II.

To our best knowledge, there is no theoretical work on the hitch-hiking process of a local sweep in a two-population model. With regard to a single population model, many studies theoretically investigated the reduction of polymorphism due to a selective sweep. These studies considered a selected site and a linked neutral site, and assumed that a very advantageous mutation arises and goes to fixation in the population. Along this fixation, they derived how much polymorphism can be reduced at the linked site. Maynard Smith and Haigh (1974) first obtained the reduction of polymorphism, where the stochastic effect of genetic drift at the linked site was ignored. The model was extended to include the stochastic effect by using a coalescent approach (Kaplan *et al.* 1989) and by using a diffusion method (Stephan *et al.* 1992). It is worthy to note that Stephan *et al.* (1992) derived a nice analytical approximate formula (see also Barton 1998; Etheridge *et al.* 2006). Durrett and Schweinsberg (2004) used a different approach for a faster approximate simulation of a selective sweep and derived some analytical expressions (see also Schweinsberg and Durrett 2005).

There are several theoretical studies on a sweep in multi-population models available, but these considered a fixation across multiple subpopulations, not a local fixation. In a model with multiple subpopulations, Slatkin and Wiehe (1998) considered the process where a beneficial mutation fixes in the entire population through weak migration. Santiago and Caballero (2005) considered a two-population model with a more general initial state and derived analytical expressions under the assumptions of weak migration. Kim and Maruki (2011) allowed stronger migration and derived analytical expression in a two-population model. Our interest is different from these studies in that we consider a locally beneficial mutation that can quasi-fix only in the adaptive subpopulation (not the entire population). We here extended the theory of Stephan's diffusion model (Stephan *et al.* 1992) to a two-population model, and considered how much polymorphism can be reduced at a linked site after a local sweep.

The common interest in the erosion phase is how a genomic island decays along time. We here consider this process after a local sweep as described in Figure 1. A local sweep creates a "block" of a fairly long region with almost no genetic variation in the adaptive population (i.e., subpopulation I in our model). In this work, given an arbitral configuration of genetic variation after a local sweep, we analytically obtain the moments of allele frequency at a linked site, with which we describe how a genomic island decays. Yeaman *et al.* (2016) investigated a similar problem in a different situation, where a secondary contact occurred between already diverged populations. In their model, erosion

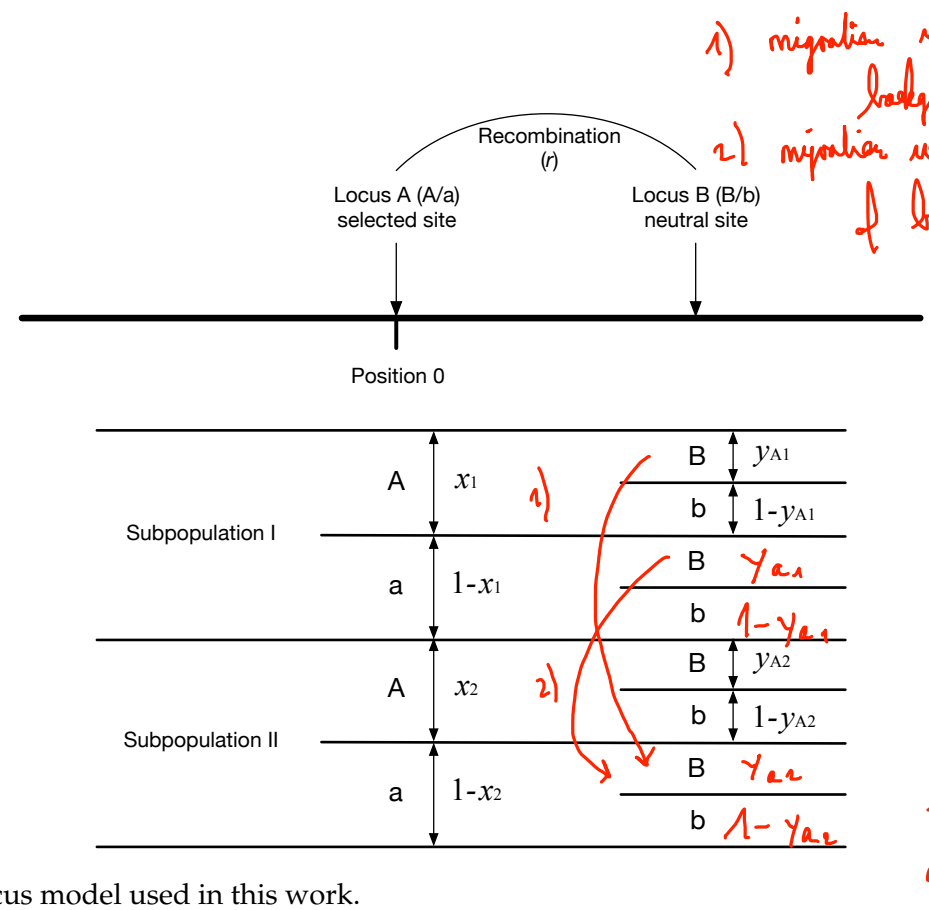


Figure 2 Two-locus model used in this work.

starts when there already are a large number of fixed sites that spread over the genome, and islands appear because selection works to maintain divergence at selected site(s), while losing divergence in other regions through homogenization by migration. By using the structured coalescent, they obtained the expected spacial distribution of F_{ST} (in terms of relative coalescent time) around a selected site as a function of the time since the secondary contact. They also considered the scenario where a de novo mutation creates a genomic island, but their derivation did not consider the effect of selective sweep of the de novo mutation, which may be slightly unrealistic. It should be noted that, because our derivation accepts any arbitral initial allele frequency at a linked site, it can be applied to any situation, not only that after a secondary contact but also that after a local sweep.

In the equilibrium phase, the balance between selection, migration, recombination and mutation holds. Theoretical treatment at equilibrium is relatively straightforward, and there are several theoretical studies on the spacial distribution of F_{ST} (Charlesworth *et al.* 1997; Akerman and Bürger 2014; Yeaman *et al.* 2016). Under our framework for the erosion phase, essentially the same result can be provided as a special case with time going to infinity.

MODEL

We consider a random mating two-population model with discrete generation, which follows the Wright-Fisher reproduction. The diploid population sizes of subpopulations I and II are assumed to be constant at N_1 and N_2, respectively. As illustrated in Figure 1, we are specifically interested in selection for local adaptation in subpopulation I. We consider a genomic region encompassing a selected site at position 0, which is referred to as locus A (Figure 2). At locus A, two alleles (A/a) are allowed with no recurrent mutation between them. Allele A confers a selection coefficient s_1 in subpopulation I and s_2 in subpopulation II (we assume $s_1 > 0$ and $s_2 < 0$). Additive selection is assumed so that the fitness of individuals with AA, Aa and aa are given by $1 + 2s_1$, $1 + s_1$ and 1 in subpopulation I, and $1 + 2s_2$, $1 + s_2$ and 1 in subpopulation II. Selection works only at this selected

site, and all remaining sites are assumed to be neutral. For the following derivation under a two-locus model, we consider a secondary neutral site (locus B), at which two alleles (B/b) are allowed with recurrent mutation between them (Figure 2). The mutation rate from B to b is u and that from b to B is v. r is the recombination rate between the two loci, A and B.

The system starts when a de novo mutation (allele A) arises in a single individual either in population I or II, where allele a is fixed in both subpopulations. Therefore, the initial state is $(x_1, x_2) = (1/2N_1, 0)$ or $(0, 1/2N_2)$, where x_1 and x_2 are frequencies of the new allele A in subpopulations I and II, respectively. Throughout this article, we assume strong selection and weak migration so that maladapted individuals are rare in each subpopulation once the initial establishment is achieved.

Establishment probability

We derive the establishment probability of a new de novo allele using the general framework of Barton (1987), who derived a simultaneous quadratic equation from the diffusion theory. This section focuses only on the selected locus A (see Figure 2), at which we are interested in the probability that allele A quasi-fixes in subpopulation I. Following previous studies (Haldane 1927), we consider that the establishment probability can be essentially obtained as the probability that the new mutation increases in frequency and escapes from immediate extinction. This is because, with the assumption of strong selection, the behavior of such an allele is almost deterministic once it escapes from extinction by genetic drift.

Let $u(x_1, x_2)$ be the establishment probability when the frequencies of allele A are x_1 and x_2 in the two subpopulations. By using an analogous procedure to Barton (1987), we derive $p_1 = u(1/2N_1, 0)$ and $p_2 = u(0, 1/2N_2)$, the establishment probability when the new allele arises in subpopulations I and II, respectively. According to the diffusion theory, u satisfies the Kolmogorov backward equation:

$$0 = \frac{x_1}{4N_1}\frac{\partial^2 u}{\partial x_1^2} + \frac{x_2}{4N_2}\frac{\partial^2 u}{\partial x_2^2} + \{s_1 x_1 + m_1(x_2 - x_1)\}\frac{\partial u}{\partial x_1} + \{s_2 x_2 + m_2(x_1 - x_2)\}\frac{\partial u}{\partial x_2}, \tag{1}$$

where $m_1(m_2)$ is the proportion of immigrant individuals just after migration in subpopulation I (II). To keep the subpopulation sizes constant, we assume $N_1 m_1 = N_2 m_2$, and we ignore higher order terms of $o(x_i)$ (i.e., x_1^2, x_2^2). This is reasonable because of the assumption that the establishment probability is mainly determined at low frequencies. Because the extinction probability of each individual mutation is independent to each other, we can write u as

$$u(x_1, x_2) = 1 - \exp(-2N_1 x_1 \psi_1 - 2N_2 x_2 \psi_2) \tag{2}$$

where $\exp(-\psi_i)$ is the extinction probability of a new mutant in subpopulation i, therefore, $p_i = 1 - \exp(-\psi_i)$. By substituting Equation 2 into Equation 1, we have

$$\begin{aligned} \psi_1^2 &= 2(s_1 - m_1)\psi_1 + 2\frac{N_2}{N_1} m_2 \psi_2 \\ \psi_2^2 &= 2\frac{N_1}{N_2} m_1 \psi_1 + 2(s_2 - m_2)\psi_2, \end{aligned} \tag{3}$$

which corresponds to Equation 4b in Barton (1987). Then, the above equations deduce

$$\psi_1(\psi_1^3 - 2a\psi_1^2 + (a^2 - bd)\psi_1 + b(ad - bc)) = 0 \tag{4}$$

$$\psi_2 = \frac{\psi_1^2 - a\psi_1}{b}, \tag{5}$$

where $a = 2(s_1 - m_1)$, $b = 2\frac{N_2}{N_1}m_2 = 2m_1$, $c = 2\frac{N_1}{N_2}m_1 = 2m_2$, and $d = 2(s_2 - m_2)$. Equation 4 can be solved by using the solution of cubic equation. Equations 4 and 5 have at most one solution which fulfills $p_1 > 0$ and $p_2 > 0$. The condition where Equations 4 and 5 have such a solution is $a + d > 0$ or $ad - bc < 0$, which corresponds to the situation where the deterministic growth rate of the mutant allele is positive (see APPENDIX A for details).

Figure 3 shows the establishment probability from Equations 4 and 5 as a function of migration rate. We first consider a symmetric model ($N_1 = N_2 = 1000$), and two selection intensities ($s_1 = 0.02$ and $s_1 = 0.1$) are assumed, while $s_2 = -0.01$ is fixed (Figures 3A and B). The establishment probability can be computed when a locally adaptive mutation arises either in subpopulation I or II, represented as $u(1/2N_1, 0)$ and $u(0, 1/2N_2)$, respectively. We performed a forward simulation to check the performance of our analytical result. For each parameter set, we ran 1,000,000 independent replications of simulation and counted the number of replications where the new allele A was preserved in 10,000 generations. The establishment probability was then obtained as the proportion of such replications. Therefore, it includes replications where alleles A and a coexisted (case C) and those where A is completely fixed in both subpopulations (case F). The proportion of case C in the established replications (Pc) decreases with increasing the migration rate (see below).

Our result (red in Figure 3) is in an excellent agreement with the simulation result: $u(1/2N_1, 0)$ is approximately $u_{m=0} = \frac{1-\exp(-2s_1)}{1-\exp(-4N_1 s_1)}$ when the migration rate is very low, consistent with the prediction in a single population model (Kimura 1957). As the migration rate increases, $u(1/2N_1, 0)$ decreases and $u(0, 1/2N_2)$ increases, and they become similar to each other. With a very high migration rate ($m \sim 0.5$), the two subpopulations can be considered as a single random-mating population, and the fixation probability of a single mutation is mainly determined by the average selection coefficient, $\bar{s} = \frac{s_1 N_1 + s_2 N_2}{N_1 + N_2}$, namely, $u_{m=0.5} = \frac{1-\exp(-2\bar{s})}{1-\exp(-4N_T\bar{s})}$ where $N_T = N_1 + N_2$ (Nagylaki 1980). Indeed, in our simulations, allele A was fixed in both populations in almost all established cases ($Pc = 1$). In each panel in Figure 3, a gray region is placed such that $Pc > 0.9$ in the left, while $Pc < 0.1$ in the right. It is indicated that the pattern dramatically changes in a short range of m_1, and the left side is the scope of this article. Similar results were also obtained in asymmetric models ($N_1 = 3N_2$ in Figure 3C and $N_1 = N_2/3$ in Figure 3D).

Figure 3 quantitatively compares our analytical results with those of previous studies (Tomasini and Peischl 2018; Yeaman and Otto 2011). It is found that $u(1/2N_1, 0)$ from Yeaman and Otto (2011) is almost as good as ours, but unfortunately $u(0, 1/2N_2)$ was not provided by Yeaman and Otto (2011). It seems that Tomasini and Peischl (2018) overestimates $u(1/2N_1, 0)$ and underestimates $u(0, 1/2N_2)$.

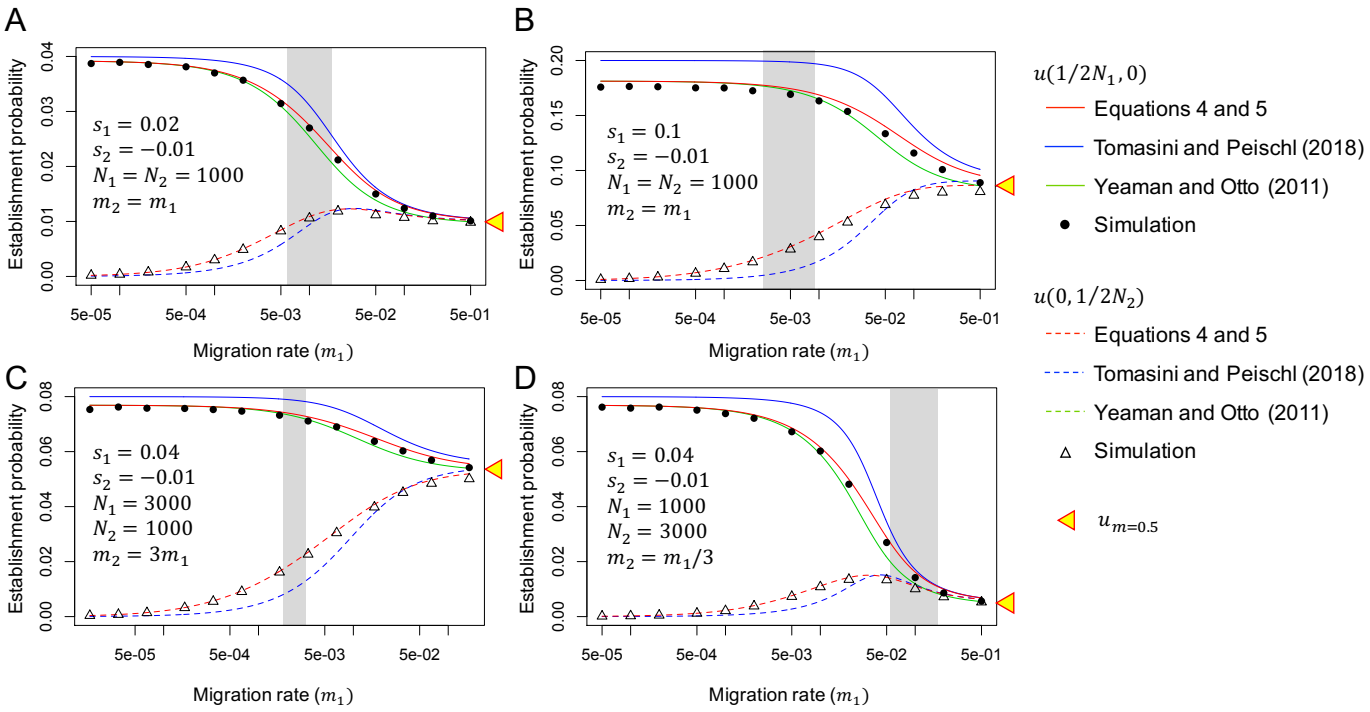


Figure 3 Establishment probability as a function of migration rate. (A) Weak selection ($s_1 = 0.02$ and $s_2 = -0.01$) and strong selection ($s_1 = 0.1$ and $s_2 = -0.01$) are assumed in a symmetric model ($N_1 = N_2$). (C, D) Asymmetric population settings are considered ($N_1 = 3N_2$ in C and $N_1 = N_2/3$ in D). Our result in red is compared with those of Tomasini and Peischl (2018) and Yeaman and Otto (2011), together with the result of our forward simulation. The establishment probability for a mutation that arises in subpopulation I ($u(1/2N_1, 0)$) is shown by solid lines and closed circles, and that for a mutation that arises in subpopulation II ($u(0, 1/2N_2)$) is shown by broken lines and open triangles. The establishment probability at the high migration limit ($m = 0.5$) is shown by a yellow triangle. In each panel in Figure 3, a gray region is placed such that $Pc > 0.9$ in the left, while $Pc < 0.1$ in the right.

as the reader reads this the first time before reading the def. of P_c, spell out what $P_c > 0.9$ & $P_c < 0.1$ means

Reduction of genetic variation due to a selective sweep

When a new locally adaptive mutation (a→A) arises and quasi-fixes in subpopulation I, genetic variation in the surrounding region in subpopulation I should be dramatically reduced due to the hitch-hiking effect. In this section, we consider a two-locus model as defined in Figure 2. We derive the degree of reduction in heterozygosity at a linked neutral site (locus B) in subpopulation I, $D_{local\ sweep}$, by extending the diffusion approach of Stephan *et al.* (1992), who investigated the effect of hitch-hiking in a single population model with no population structure.

Overview of Stephan et al. (1992): We first introduce the approach of Stephan *et al.* (1992) briefly, which provides the basis of our derivation below. The expected reduction of heterozygosity at locus B for a single population model with diploid size N is denoted by $D_{\text{Stephan et al.}}$. With the assumption of strong selection, Stephan *et al.* (1992) assumed that the behavior of the frequency (x) of the beneficial allele A with selection coefficient, s, follow a deterministic function:

$$\frac{dx}{dt} = sx(1-x), \tag{6}$$

where selection is additive. It should be noted that x with no subscript denotes the frequency of A in the single population model, whereas in our two-population model, the frequencies of A in subpopulations I and II are denoted by x_1 and x_2, respectively (see Figure 2). We consider another biallelic neutral locus (B/b), and the recombination rate between between this neutral locus and the selected locus is assumed to be r. y_A is the frequency of B among A-chromosomes and y_a is the frequency of B among a-chromosomes. Then, the expected changes of an arbitrary function $f(y_A, y_a)$ is described as the following ordinary differential equation:

$$\frac{d}{dt}E(f) = E(L(f)), \tag{7}$$

where L is a differential operator of the Kolmogorov backward equation:

$$L = \frac{y_A(1-y_A)}{4Nx}\frac{\partial^2}{\partial y_A^2} + r(1-x)(y_a - y_A)\frac{\partial}{\partial y_A} + \frac{y_a(1-y_a)}{4N(1-x)}\frac{\partial^2}{\partial y_a^2} + rx(y_A - y_a)\frac{\partial}{\partial y_a}. \tag{8}$$

By using this formula, Stephan *et al.* (1992) solved the first and second moments of y_A and y_a after a sweep, from which the expected reduction of heterozygosity at the linked site can be computed numerically. With some approximation, Stephan *et al.* (1992) further obtained a nice closed form of the solution:

$$D_{\text{Stephan et al.}} = \frac{2r}{s}(2Ns)^{-2r/s}\Gamma\left(-\frac{2r}{s}, \frac{1}{2Ns}\right). \tag{9}$$

In this work, we found that this equation somehow undervalues the effect of random genetic drift perhaps due to the approximation of Stephan *et al.* (1992). It is known that heterozygosity decreases by genetic drift by a factor of $1/2N$ per generation. To correct for this factor, we obtain the expected reduction of heterozygosity along the fixation at the selected site as $\exp(-\log(2N)/Ns)$, because the fixation time is approximately given by

$$\begin{aligned} T &= \int_{1/2N}^{1-1/2N} \frac{dx}{sx(1-x)} \\ &\approx \frac{2\log(2N)}{s} \end{aligned}$$

see previous page

Then, we add this factor into Equation 9:

$$D_{\text{modified Stephan et al.}} = \frac{2r}{s}(2Ns)^{-2r/s}\Gamma\left(-\frac{2r}{s}, \frac{1}{2Ns}\right)\exp\left(-\frac{\log(2N)}{Ns}\right). \quad (10)$$

We found that this heuristic approach is in a very good agreement with the numerical solution obtained by directly computing Equation 8.

Local sweep in the two-population model: In this work, we extend Stephan *et al.*'s derivation (1992) to the two-population model defined above (Figure 2). We first consider the dynamics of the new mutant allele frequency (x_1) at the selected locus (position 0) in the subpopulation I. The major difference from the corresponding formula in Stephan *et al.* (1992) (i.e., Equation 6) is that the effect of migration should be considered in the two-population model. Because ~~maladaptive~~ allele A is very rare in subpopulation II under the assumption of strong selection and low migration, we can ignore migrants with A allele from subpopulations II to I. Then, the dynamics of x_1 could be approximated by a deterministic function:

$$\frac{dx_1}{dt} = s_1 x_1(1-x_1) - m_1 x_1. \quad (11)$$

We set the time such that $t = 0$ when the mutation arises and $t = \tau$ when the mutation quasi-fixes. We next consider the neutral locus B (B/b). As illustrated in Figure 2, y_{A1} (y_{A2}) is the frequency of haplotype A-B among A-chromosomes in subpopulation I (II), and y_{a1} (y_{a2}) is the frequency of haplotype a-B among a-chromosomes in subpopulation I (II). We assume that y_{A2} is very small throughout the sweep process. Then, the expected changes of an arbitrary function $f(y_{A1}, y_{a1}, y_{a2})$ is described as the following ordinary differential equation:

$$\frac{d}{dt}E(f) = E(L(f)), \quad (12)$$

where L is a differential operator of the Kolmogorov backward equation. Following Ohta and Kimura (1969), we obtain L for our model as

change in y_{A_1} due to recomb.
change in y_{a_1} due to recomb.
immigration

$$\begin{aligned} L =& \frac{y_{A1}(1-y_{A1})}{4N_1 x_1(t)}\frac{\partial^2}{\partial y_{A1}^2} + r(1-x_1(t))(y_{a1}-y_{A1})\frac{\partial}{\partial y_{A1}} \\ &+ \frac{y_{a1}(1-y_{a1})}{4N_1(1-x_1(t))}\frac{\partial^2}{\partial y_{a1}^2} + \{r x_1(t)(y_{A1}-y_{a1}) + \frac{m_1}{(1-x_1(t))(1-m_1)+m_1}(y_{a2}-y_{a1})\}\frac{\partial}{\partial y_{a1}} \\ &+ \frac{y_{a2}(1-y_{a2})}{4N_2}\frac{\partial^2}{\partial y_{a2}^2} + \{x_1(t)m_{e,1\to 2}(y_{A1}-y_{a2}) + (1-x_1(t))m_2(y_{a1}-y_{a2})\}\frac{\partial}{\partial y_{a2}}. \end{aligned} \quad (13)$$

In this equation, the effect of selection on locus B is incorporated such that A-chromosomes will be selected out immediately if migrated from subpopulations I to II. In other words, with the linkage effect, the migration rate of A-chromosomes at the B locus is *effectively* reduced to $m_{e,1\to 2}$:

$$m_{e,1\to 2} = \frac{(1+s_2)r}{1-(1+s_2)(1-r)}m_2 \quad (14)$$

(Bengtsson 1985). Then, we can compute the first and second moments of y_{A1} and y_{a2} after the quasi-fixation of allele A (i.e., $y_{A1}(\tau)$ and $y_{a2}(\tau)$), from which we can obtain heterozygosity within each subpopulations (h_{w1} and h_{w2}) and between them (h_b) at $t = \tau$ as

$$\begin{aligned} h_{w1}(\tau) &= 2E(y_{A1}(\tau)) - 2E(y_{A1}(\tau)^2), \\ h_{w2}(\tau) &= 2E(y_{a2}(\tau)) - 2E(y_{a2}(\tau)^2), \\ h_b(\tau) &= E(y_{A1}(\tau)) + E(y_{a2}(\tau)) - 2E(y_{A1}(\tau)y_{a2}(\tau)). \end{aligned} \quad (15)$$

Note: $Het_w = 2\,Var(p)$; $Het_b = E[p] + E[q] - 2E(pq)$

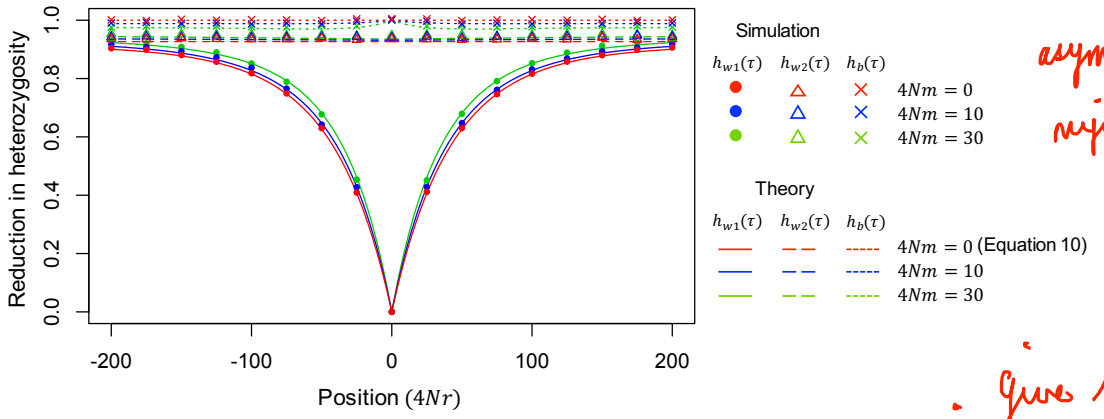


Figure 4 The expected reduction of heterozygosity after a sweep in the two-population model. Position is shown in $4Nr$ from the selected site. $N_1 = N_2 = 1000$ and $m = m_1$ are assumed. Theoretical results for $h_{w1}(\tau)$ $h_{w2}(\tau)$ and $h_b(\tau)$ computed from 13-16 by assuming $y_{a1}(0) = y_{a2}(0) = 0.3$ for convenience, but very similar results were obtained for other values of $y_{a1}(0)$ and $y_{a2}(0)$. In the case of no migration (red), our results is identical to Stephan *et al.* (1992) (i.e., Equation 10)

Then, the expected reduction of heterozygosity is obtained as

$$D_{local\ sweep} = h_{w1}(\tau)/h_{w1}(0). \tag{16}$$

Generally, $D_{local\ sweep}$ involves the initial frequencies, $y_{a1}(0)$ and $y_{a2}(0)$. However, it should be noted that their quantitative effect on $D_{local\ sweep}$ is not large unless $y_{a1}(0)$ and $y_{a2}(0)$ are not very similar.

Figure 4 shows the effect of migration on the reduction in heterozygosity. The plot in red is the case of no migration, where our result is essentially identical to Stephan *et al.* (1992). For the cases with migration, we assumed $N_1 = N_2 = 1000$, $m_1 = m_2 = m$. For each parameter set, filled circles represent the average over 100,000 replications of forward simulation. In Figure 4, $h_{w1}(\tau)$, $h_{w2}(\tau)$ and $h_b(\tau)$ are plotted such that $h_{w1}(0) = h_{w2}(0) = 1$ before the sweep, so that $h_{w1}(\tau)$ directly corresponds to $D_{local\ sweep}$. In all cases, our theoretical result from Equation 13 is in excellent agreement with the simulation results. It is found that the effect of a sweep seems to be only on subpopulation I, and there is almost no effect on the variation in subpopulation II. As going further from the selected site at position 0, $D_{local\ sweep}$ is larger for a higher migration rate. It is indicated that migration brings standing variation maintained in subpopulation II into subpopulation I, thereby increasing the polymorphism level in subpopulation I. We can observed a slight increase of $h_b(\tau)$ around the selected site at position 0. If we assume $1 - h_w(\tau)/h_{all}(\tau)$ roughly approximates F_{ST} where h_{all} is heterozygosity when the two subpopulations are merged together, it can be said that a local sweep creates a relatively wide region of elevated F_{ST}, which can be considered as an initial genomic island of speciation.

Erosion and growth of a genomic island

When a new locally adaptive mutation (a→A) quasi-fixes in subpopulation I, a block of region in which genetic variation in subpopulation I is dramatically reduced arises (Figure 1B), which is referred to as an initial genomic island. In this section, by using the two-locus model defined in Figure 2, we consider the process after this state, but our derivation is flexible enough to plug in any initial state.

We use a similar diffusion approach to the previous section but we focus on the behavior of y_{A1} and y_{a2}. The expected changes of an arbitrary function $f(y_{A1}, y_{a2})$ is described as the following ordinary differential equation:

$$\frac{d}{dt}E(f) = E(L(f)), \tag{17}$$

where L is a differential operator of the Kolmogorov backward equation, which is given by

$$\begin{aligned} L =& \frac{y_{A1}(1-y_{A1})}{4N_1}\frac{\partial^2}{\partial y_{A1}^2} + \frac{y_{a2}(1-y_{a2})}{4N_2}\frac{\partial^2}{\partial y_{a2}^2} \\ &+ [v-(u+v)y_{A1} + m_{e,2\to1}(y_{a2}-y_{A1})]\frac{\partial}{\partial y_{A1}} + [v-(u+v)y_{a2} + m_{e,1\to2}(y_{A1}-y_{a2})]\frac{\partial}{\partial y_{a2}}. \end{aligned} \tag{18}$$

As well as the previous section, we use the effective migration rate (Bengtsson 1985):

$$m_{e,2\to1} = \frac{(1+\tilde{s_1})r}{1-(1+\tilde{s_1})(1-r)}m_1, \tag{19}$$

where $\tilde{s_1} = 1/(1+s_1) - 1$ is the relative selection coefficient of maladapted individuals in subpopulation I. $m_{e,1\to2}$ is defined by Equation 14. We consider the dynamics of the first and second order moments, and put $\boldsymbol{y} = (E(y_{A1}), E(y_{a2}), E(y_{A1}^2), E(y_{A1}y_{a2}), E(y_{a2}^2))^T$. By using Equation 17, we derive a differential equation for $\boldsymbol{y}$ as follows:

$$\frac{d\boldsymbol{y}}{dt} = \boldsymbol{Q}\boldsymbol{y} + \boldsymbol{a}, \tag{20}$$

where $\boldsymbol{Q}$ is the 5×5 matrix given by

$$\boldsymbol{Q} = \begin{pmatrix} -(u+v+m_{e,2\to1}) & m_{e,2\to1} & 0 & 0 & 0 \\ m_{e,1\to2} & -(u+v+m_{e,1\to2}) & 0 & 0 & 0 \\ 2v+\frac{1}{2N_1} & 0 & -2(u+v+m_{e,2\to1}+\frac{1}{4N_1}) & 2m_{e,2\to1} & 0 \\ v & v & m_{e,1\to2} & -(2u+2v+m_{e,2\to1}+m_{e,1\to2}) & m_{e,2\to1} \\ 0 & 2v+\frac{1}{2N_2} & 0 & 2m_{e,1\to2} & -2(u+v+m_{e,1\to2}+\frac{1}{4N_2}) \end{pmatrix} \tag{21}$$

and $\boldsymbol{a} = (v, v, 0, 0, 0)^T$. By solving Equation 20, $\boldsymbol{y}$ is given by

$$\boldsymbol{y}(t) = \exp(t\boldsymbol{Q})\boldsymbol{y}(0) + \boldsymbol{Q}^{-1}(\exp(t\boldsymbol{Q}) - \boldsymbol{I})\boldsymbol{a} \tag{22}$$

where $\boldsymbol{I}$ is the identity matrix of size 5. $\boldsymbol{y}$ at equlibrium is given by $\tilde{\boldsymbol{y}} = -\boldsymbol{Q}^{-1}\boldsymbol{a}$. Our solution at equilibrium is well consistent with previous studies (Charlesworth *et al.* 1997; Yeaman *et al.* 2016) that used the coalescent approach (see APPENDIX B).

Figure 5 compares our theoretical results from Equation 22 (broken lines) with simulation results. $N_1 = N_2 = 1000, s_1 = -s_2 = 0.05, u = v = 2.5\times10^{-6}, m_1 = m_2 = 1.25\times10^{-3}$ are assumed. As the initial condition ($t = 0$), we set $h_{w1} = 0$, $h_{w2} = 0.18$ and $h_b = 0.1$, representing a situation after

a local sweep in subpopulation I. Equation 22 describes how a sharp peak of divergence grows along time. As time goes, h_{w1} and h_{w2} become closer to each other, and eventually reaches their equilibrium values (t$\gg$10,000). h_b also decreases except for a short region surrounding the selected site. The rate of erosion (decrease of h_b) is high as going apart from the selected site. At the selected site, h_b gradually increases and eventually develops a sharp peak. It reaches an equilibrium after a significant amount of time, where the selection-migration balance holds so that the shape of the peak does not change much.

Figure 5 shows that Equation 22 is well consistent with the simulation results (broken lines), but they could be further improved if we include the effect of maladapted alleles, which were completely ignored in Equation 22. In practice, although their effect on the long-term dynamics is ignorable, they stay in the population, thereby constituting a certain proportion; their expected frequencies in subpopulations I and II are, respectively, $1 - x_1 \approx -m_1/\tilde{s_1}$ and $x_2 \approx -m_2/s_2$. Let us focus on the trajectory of the frequency of a single maladaptive migrant. We ask how long such a maladaptive migrant can survive (as a maladaptive allele). It could be eliminated by selection, or recombine with an adaptive allele and escape from the maladaptive state. Because the expected time until a migrant dies or recombines in subpopulations I and II are, respectively, given by

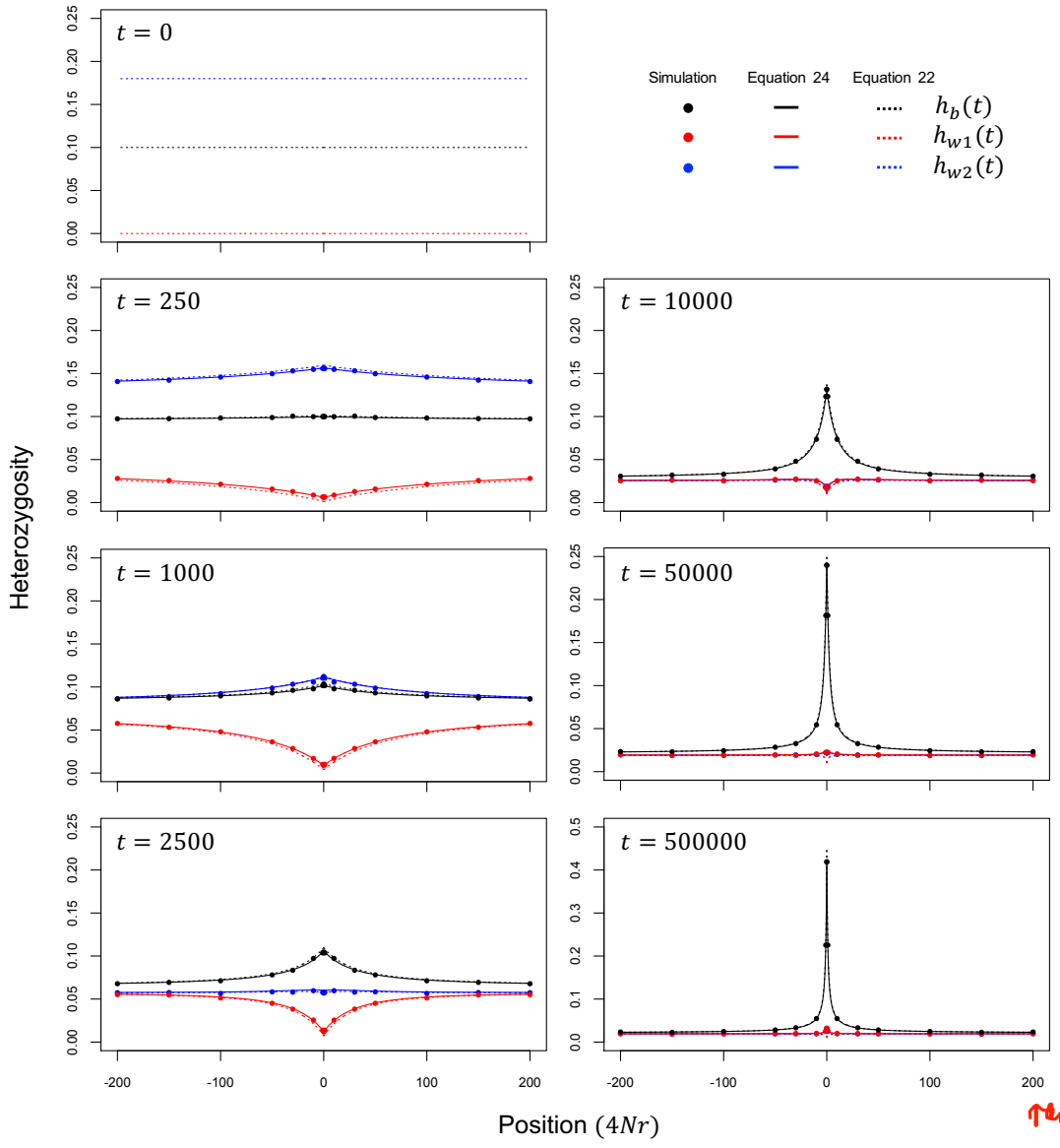
$$\begin{cases} t_{2\to1} = \sum_{i=0}^{\infty}\{(1+\tilde{s_1})(1-r)\}^i = \dfrac{1}{1-(1+\tilde{s1})(1-r)}, \\ t_{1\to2} = \sum_{i=0}^{\infty}\{(1+s_2)(1-r)\}^i = \dfrac{1}{1-(1+s_2)(1-r)}. \end{cases} \tag{23}$$

Therefore, the expected numbers of neutral alleles from the other subpopulation with the maladapted allele is $N_2 m_2 t_{1\to2}$ and $N_1 m_1 t_{2\to1}$ in subpopulations I and II, respectively.

Let the frequencies of B in subpopulations I and II including maladapted ones are denoted by $\tilde{y_1}$ and $\tilde{y_2}$. Together with this effect of maladaptive alleles, the first and second-order moments of $\tilde{y}_i$ are given as follows,

$$\begin{aligned}
E(\tilde{y_1}) &= (1 - m_1 t_{2\to1})E(y_{A1}) + m_1 t_{2\to1} E(y_{a2}) \\
E(\tilde{y_2}) &= m_2 t_{1\to2} E(y_{A1}) + (1 - m_2 t_{1\to2})E(y_{a2}) \\
E(\tilde{y_1^2}) &= (1 - m_1 t_{2\to1})^2 E(y_{A1}^2) + m_1^2 t_{2\to1}^2 E(y_{a2}^2) + 2m_1 t_{2\to1}(1 - m_1 t_{2\to1})E(y_{A1}y_{a2}) \\
E(\tilde{y_1}\tilde{y_2}) &= (1 - m_1 t_{2\to1}) m_2 t_{1\to2} E(y_{A1}^2) + m_1 t_{2\to1}(1 - m_2 t_{1\to2})E(y_{a2}^2) \\
&\quad + \left\{(1 - m_1 t_{2\to1})(1 - m_2 t_{1\to2}) + m_1 t_{2\to1} m_2 t_{1\to2}\right\} E(y_{A1}y_{a2}) \\
E(\tilde{y_2^2}) &= m_2^2 t_{1\to2}^2 E(y_{A1}^2) + (1 - m_2 t_{1\to2})^2 E(y_{a2}^2) + 2m_2 t_{1\to2}(1 - m_2 t_{1\to2})E(y_{A1}y_{a2}).
\end{aligned} \tag{24}$$

Figure 5 shows that Equation 24 fits to the simulation results better than Equation 22.



highlight the change of the range of y-values

Figure 5 Temporal change of heterozygosity (h_{w1}, h_{w2}, h_b) after a local sweep in subpopulation I. Position is shown in $4Nr$ from the selected site. $N_1 = N_2 = 1000, s_1 = -s_2 = 0.05, u = v = 2.5 \times 10^{-6}, m_1 = m_2 = 1.25 \times 10^{-3}, y_1(0) = 0.0$ and $y_2(0) = 0.1$ are assumed. Theoretical results from Equations 22 and 24 are shown by broken and solid lines, respectively. Simulation results (closed circles) are the averages over 50,000 replications of forward simulation.

DISCUSSION

In the earlier stages of ecological speciation with gene flow, divergent selection should work to maintain phenotypes that are adaptive to each niche (Wu 2001; Turner *et al.* 2005; Nosil 2012). Therefore, it is predicted that genetic variations responsible to those adaptive phenotypes should appear as genomic islands of speciation. This article theoretically considers the evolutionary behavior of a genomic island of speciation, from its initial establishment to stable preservation. The process was divided into three phases, the establishment, erosion and equilibrium phases (Figure 1). We obtained (i) the establishment probability of a locally adaptive mutation, (ii) the expected reduction of genetic variation within subpopulations I and II after a local sweep that creates an initial genomic island, (iii) the expected erosion of the initial island as a function of time since the sweep, and (iv) the expected shape of the peak of divergence in the island in equilibrium.

For (i), we have successfully derived a close-form formula of the establishment probability along the formulation of Barton (1987). Our simulations showed that our theoretical results for $u(1/2N_1, 0)$ and $u(0, 1/2N_2)$ outperform the previous studies, although Yeaman and Otto (2011)'s heuristic approach is almost as good as ours. It would be intriguing to discuss the analogy between our result and those of Gavrilets and Gibson (2002) and Whitlock and Gomulkiewicz (2005). Because this work focuses on divergent selection so that allele A is quasi-fixed in subpopulation I whereas allele a is quasi-fixed in subpopulation II, we assume $s_1 > 0$ and $s_2 < 0$. However, as showed in Figure 3, it is possible that either A or a could fix in the entire population even if $s_1 > 0$ and $s_2 < 0$ hold, although it might take an extremely long time. In contrast, Gavrilets and Gibson (2002) and Whitlock and Gomulkiewicz (2005) obtained the probability of such eventual fixation in the entire population. These studies and ours can be understood in a single framework as follows. Assuming $s_1 > 0$ and $s_2 < 0$, the establishment of A first occurs and maintained quite stably for a long time, but with time going infinity, allele A could fix in the entire population most likely when the average selection coefficient $\bar{s}$ is positive, while allele a could likely fix when $\bar{s}$ is negative. This is why our formula of the establishment probability (Equation 2) is the same as the numerator of the fixation probability when $\bar{s}$ is positive (Equations 7 and 8 in Gavrilets and Gibson 2002 and Equation 6 in Whitlock and Gomulkiewicz 2005). On the other hand, the establishment probability significantly differs from the fixation probability of Gavrilets and Gibson (2002) and Whitlock and Gomulkiewicz (2005) when $\bar{s}$ is negative because such a mutation hardly goes to eventual fixation, although it can be maintained as a quasi-fixed state for a sufficiently long time.

For (ii), we extended the diffusion method of Stephan *et al.* (1992) to our two-population model. Because the beneficial allele A fixes only in one subpopulation, the process is very similar to that of a single population model (Stephan *et al.* 1992), except that migration between two subpopulations has some effect. Our theoretical result (see Figure 4) demonstrated a relatively minor effect of migration; with an increasing migration rate, the level of polymorphism in subpopulation I increases because migration brings genetic variation from subpopulation II.

For (iii) and (iv), we considered the erosion of an initial island created by a local sweep, followed by the development of a stable island at equilibrium. This process from erosion to equilibrium can be described by a single formula 22. Furthermore, Equation 22 is flexible enough to plug in any initial state, such as a secondary contact of already diverged subpopulation. To demonstrate this, in Figure A1, we compare the pattern after a local sweep (left panels) and that after a secondary contact (right panels). After a secondary contact, h_b is already high across the genome, and h_b gradually decreases but selection works to keep divergence around the selected site, thereby creating a peak of divergence (i.e., island). After a very long time (i.e., in equilibrium), the shape of the peak becomes identical to that after a sweep.

We have thus developed analytical expressions for the evolutionary behavior of genomic island

an overstatement

of speciation, from the emergence of an initial island by a local sweep to stable maintenance of the island in equilibrium. Genomic islands of speciation can arise in the earlier stages of ecological speciation, but it does not necessarily mean that the emergence of genomic islands of speciation always results in speciation. It is possible that genomic islands of speciation could disappear by environmental changes or by chance, and no speciation occurs. To achieve speciation, there would be many other forces necessary, including emergence of additional islands (Feder *et al.* 2012a; Via 2012; Feder *et al.* 2012b; Aeschbacher and Bürger 2014; Yeaman *et al.* 2016), further divergence on a genomic-scale possible due to a reduction in migration rate, and environmental changes. More theoretical works are needed to fully understand the process to ecological speciation.

Literature Cited

Aeschbacher, S. and R. Bürger, 2014 The effect of linkage on establishment and survival of locally beneficial mutations. Genetics **197**: 317–336.

Akerman, A. and R. Bürger, 2014 The consequences of gene flow for local adaptation and differentiation: a two-locus two-deme model. J. Math. Biol. **68**: 1135–1198.

Barton, N. H., 1987 The probability of establishment of an advantageous mutant in a subdivided population. Genet. Res. **50**: 35–40.

Barton, N. H., 1998 The effect of hitch-hiking on neutral genealogies. Genet. Res. **72**: 123–133.

Bengtsson, B. O., 1985 The flow of genes through a genetic barrier. In *Evolution: Essays in Honor of John Maynard Smith.*, edited by J. J. Greenwood, P. H. Harvey, and M. Slatkin, pp. 31–42, Cambridge University Press, New York.

Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. **70**: 155–174.

Durrett, R. and J. Schweinsberg, 2004 Approximating selective sweeps. Theor. Popul. Biol. **66**: 129–138.

Etheridge, A., P. Pfaffelhuber, and A. Wakolbinger, 2006 An approximate sampling formula under genetic hitchhiking. Ann. Appl. Probab. **16**: 685–729.

Feder, J. L., S. P. Egan, and P. Nosil, 2012a The genomics of speciation-with-gene-flow. Trends Genet. **28**: 342–350.

Feder, J. L., R. Gejji, S. Yeaman, and P. Nosil, 2012b Establishment of new mutations under divergence and genome hitchhiking. Philos. Trans. R. Soc. Lond. B Biol. Sci. **367**: 461–474.

Gavrilets, S. and N. Gibson, 2002 Fixation probabilities in a spatially heterogeneous environment. Popul. Ecol. **44**: 51–58.

Haldane, J. B. S., 1927 A mathematical theory of natural and artificial selection, part v: selection and mutation. Proc. Camb. Philos. Soc. **23**: 838–844.

Kaplan, N. L., R. R. Hudson, and C. H. Langley, 1989 The" hitchhiking effect" revisited. Genetics **123**: 887–899.

Kim, Y. and T. Maruki, 2011 Hitchhiking effect of a beneficial mutation spreading in a subdivided population. Genetics **189**: 213–226.

Kimura, M., 1954 Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. Genetics **39**: 280–295.

Kimura, M., 1957 Some problems of stochastic processes in genetics. Ann. Math. Statist. **28**: 882–901.

Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Genetics **47**: 713–719.

Kirkpatrick, M. and S. Peischl, 2013 Evolutionary rescue by beneficial mutations in environments that change in space and time. Philos. Trans. R. Soc. Lond. B Biol. Sci. **368**: 20120082.

Maynard Smith, J. and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. **23**: 23–35.
Nagylaki, T., 1980 The strong-migration limit in geographically structured populations. J. Math. Biol. **9**: 101–114.
Nosil, P., 2012 *Ecological speciation*. Oxford University Press.
Ohta, T. and M. Kimura, 1969 Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. Genetics **63**: 229–238.
Pollak, E., 1966 On the survival of a gene in a subdivided population. J. Appl. Prob. **3**: 142–155.
Santiago, E. and A. Caballero, 2005 Variation after a selective sweep in a subdivided population. Genetics **169**: 475–483.
Schweinsberg, J. and R. Durrett, 2005 Random partitions approximating the coalescence of lineages during a selective sweep. Ann. Appl. Probab. **15**: 1591–1651.
Slatkin, M. and T. Wiehe, 1998 Genetic hitch-hiking in a subdivided population. Genet. Res. **71**: 155–160.
Stephan, W., T. H. Wiehe, and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Popul. Biol. **41**: 237–254.
Tomasini, M. and S. Peischl, 2018 Establishment of locally adapted mutations under divergent selection. Genetics **209**: 885–895.
Turner, T. L., M. W. Hahn, and S. V. Nuzhdin, 2005 Genomic islands of speciation in anopheles gambiae. PLoS Biol. **3**: e285.
Via, S., 2012 Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos. Trans. R. Soc. Lond. B Biol. Sci. **367**: 451–460.
Wakeley, J., 2009 *Coalescent Theory: An Introduction*, Ed. 1. macmillan learning.
Whitlock, M. C. and R. Gomulkiewicz, 2005 Probability of fixation in a heterogeneous environment. Genetics **171**: 1407–1417.
Wu, C.-I., 2001 The genic view of the process of speciation. J. Evol. Biol. **14**: 851–865.
Yeaman, S., S. Aeschbacher, and R. Bürger, 2016 The evolution of genomic islands by increased establishment probability of linked alleles. Mol. Ecol. **25**: 2542–2558.
Yeaman, S. and S. P. Otto, 2011 Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evolution **65**: 2123–2129.

vertex : point of maximum curvature ; for a parabola, it is the point of the local extrema

APPENDIX

Appendix A: The solution of Equations 4 and 5

First, we present a proof that there is at most one solution which fulfills $p_1 > 0$ and $p_2 > 0$, and the condition on which such a solution exists is $a + d > 0$ or $ad - bc < 0$. Then, we give a closed expression of the solution.

For ψ_1 and ψ_2 to satisfy $p_1 > 0$ and $p_2 > 0$, $\psi_1 > 0$ and $\psi_2 > 0$ are needed. Notice that $b, c > 0$ because migration rate and population size are always positive. We put $f(x) = x^3 - 2ax^2 + (a^2 - bd)x + (abd - b^2c)$ and $f'(x) = 3x^2 - 4ax + (a^2 - bd)$.

1. $a \geq 0$

   From Equation 5, $\psi_1 > a$ is needed. Because x-cordinate of vertex of $f'(x)$, $\frac{2}{3}a$, is not greater than a, $f'(x)$ monotonically increases when $x > a$. Noting $f(a) = -b^2c < 0$, there is only one solution.

2. $a < 0$ and $d \leq 0$

   From Equation 5, $\psi_1 > 0$ is needed. Because $f'(0) = a^2 - bd \geq 0$ and the x-coodinate of vertex of $f'(x)$, $\frac{2}{3}a$, is smaller than 0, $f'(x) > 0$ when $x > 0$. Therefore, whether $f(x) = 0$ has a solution or not in $(0, \infty)$ depends on the sign of $f(0)$. If $f(0) \geq 0$, i.e. $b(ad - bc) \geq 0$, there is no solution. Otherwise, there is only one solution.

3. $a < 0$ and $d > 0$

   From Equation 5, $\psi_1 > 0$ is needed. Because x-coodinate of vertex of $f'(x)$, $\frac{2}{3}a$, is smaller than 0, $f'(x)$ monotonically increases when $x > 0$. Noting $f(0) = abd - b^2c < 0$, there is only one solution.

Noting that $ad - bc$ is negative when $ad \leq 0$, the condition on which one solution exists is reduced to $a + d > 0$ or $ad - bc < 0$. This is the same as the condition where a deterministic model,

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \tag{A1}$$

has a positive growth rate, in other words, the matrix in Equation A1 has at least one positive eigenvalue.

Next, we present a closed form of ψ_1. From the above proof, if there is a nonzero real root of $f(\psi_1) = 0$ which fulfills $p_1 > 0$ and $p_2 > 0$, the root is the largest real root of $f(\psi_1) = 0$. Therefore, by using the solution of cubic equation, ψ_1 can be expressed as

$$\psi_1 = \begin{cases} 0 & \text{when } a + d \leq 0 \text{ and } ad - bc \geq 0 \\ \sqrt[3]{-\frac{Q}{2} + \sqrt{R}} + \sqrt[3]{-\frac{Q}{2} - \sqrt{R}} - \frac{A_2}{3} & \text{when } R > 0 \text{ and } (a + d > 0 \text{ or } ad - bc < 0) \\ 2S\cos\left(\frac{1}{3}\arccos\left(\frac{T}{2S}\right)\right) - \frac{A_2}{3} & \text{when } R \leq 0 \text{ and } (a + d > 0 \text{ or } ad - bc < 0) \end{cases}, \tag{A2}$$

where $A_0 = abd - b^2c, A_1 = a^2 - bd, A_2 = -2a, P = A_1 - \frac{A_2^2}{3}, Q = A_0 - \frac{A_1A_2}{3} + \frac{2}{27}A_2^3, R = (\frac{P}{3})^3 + (\frac{Q}{2})^2, S = \sqrt{-\frac{P}{3}}, T = -\frac{Q}{S^2}$. In the above expression, we assume the range of principal value of $y = \arccos(x)$ as $0 \leq y \leq \pi$.

Appendix B: Comparison between diffusion and coalescent at equilibrium phase

In the main text, we show that replacing the migration rate in the neutral diffusion equation by the effective migration rate well approximates the effect of linkage with the locus under divergent selection. In a neutral model, heterozygosity at equilibrium in a structured population is already well studied by the coalescent theory under the infinite-site model (reviewed in Wakeley 2009). In this work, we alternatively used the forward diffusion approach because the diffusion approach can be applied to more general conditions. In this Appendix, we show our diffusion result at equilibrium is consistent with that of the coalescent theory.

We attempt to derive the expected heterozygosity under the infinite-site setting along our diffusion-based derivation. In practice, we first consider a K-allele model, and then the results will be transformed to the infinite-site model. Let B allele be one of the alleles at the locus. We put y_1 and y_2 as frequency of allele B in subpopulation I and II, respectively. In the following derivation, we assume $N_1 = N_2 = N$ and $m_1 = m_2 = m$. The differential operator of the Kolmogorov backward equation is as follows,

$$\begin{aligned} L =& \frac{y_1(1-y_1)}{4N}\frac{\partial^2}{\partial y_1^2} + \frac{y_2(1-y_2)}{4N}\frac{\partial^2}{\partial y_2^2} \\ &+ [v-(u+v)y_1 + m(y_2-y_1)]\frac{\partial}{\partial y_1} + [v-(u+v)y_2 + m(y_1-y_2)]\frac{\partial}{\partial y_2}, \end{aligned} \tag{B1}$$

At the equilibrium, we derive the moments up to the second order as

$$\begin{aligned} E(y_1) = E(y_2) &= \frac{V}{U+V}, \\ E(y_1^2) = E(y_2^2) &= \frac{V(V+1)(U+V+M) + V^2M}{(U+V)(U+V+M+1)(U+V+M) - M^2(U+V)}, \\ E(y_1y_2) &= \frac{MV(V+1) + V^2(U+V+M+1)}{(U+V)(U+V+M+1)(U+V+M) - M^2(U+V)}, \end{aligned} \tag{B2}$$

where $U = 4Nu$, $V = 4Nv$ and $M = 4Nm$. In the limit to the infinite-allele model, that is, $v = \frac{u}{K-1}$ and $K \to \infty$, the expected heterozygousity within and between subpopulation goes to

$$\begin{aligned} h_w = 1 - KE(y_1^2) &\to \frac{U^2 + 2UM}{(U+M+1)(U+M) - M^2}, \\ h_b = 1 - KE(y_1y_2) &\to \frac{U(U+2M+1)}{(U+M+1)(U+M) - M^2}. \end{aligned} \tag{B3}$$

This result udner the infinite allele-setting can be transformed to the infinite-site mode: If we put $U = \frac{\theta}{n}$ and n goes to ∞, π_w and π_b are described as

$$\begin{aligned} \pi_w &= nh_w \to 2\theta \\ \pi_b &= nh_b \to \theta(2 + \frac{1}{M}), \end{aligned} \tag{B4}$$

which is identical with the result from the coalescent theory (Charlesworth *et al.* 1997; Yeaman *et al.* 2016).

Appendix C: Equation 22 for a scenario of secondary contact

We compute Equation 22 for a scenario of secondary contact, where we assume that already diverged two subpopulation have merged so that there are a number of fixed sites between the two subpopulations. To make a realization of this situation, we set $y_1(0) = 0.1$ and $y_2(0) = 0.9$, and the other parameters are identical to those used for Figure 5. Figure A1 compares the patterns after a local sweep (left panels) and after a secondary contact (right panels). After a secondary contact, h_b is already high across the genome, and h_b gradually decreases but selection works to keep divergence around the selected site, thereby creating a peak of divergence. In equilibrium, the shape of the peak becomes identical to that after a sweep.

This is analogous to what Yeaman et al. demonstrated for a two-peak island w. unidirectional gene flow.

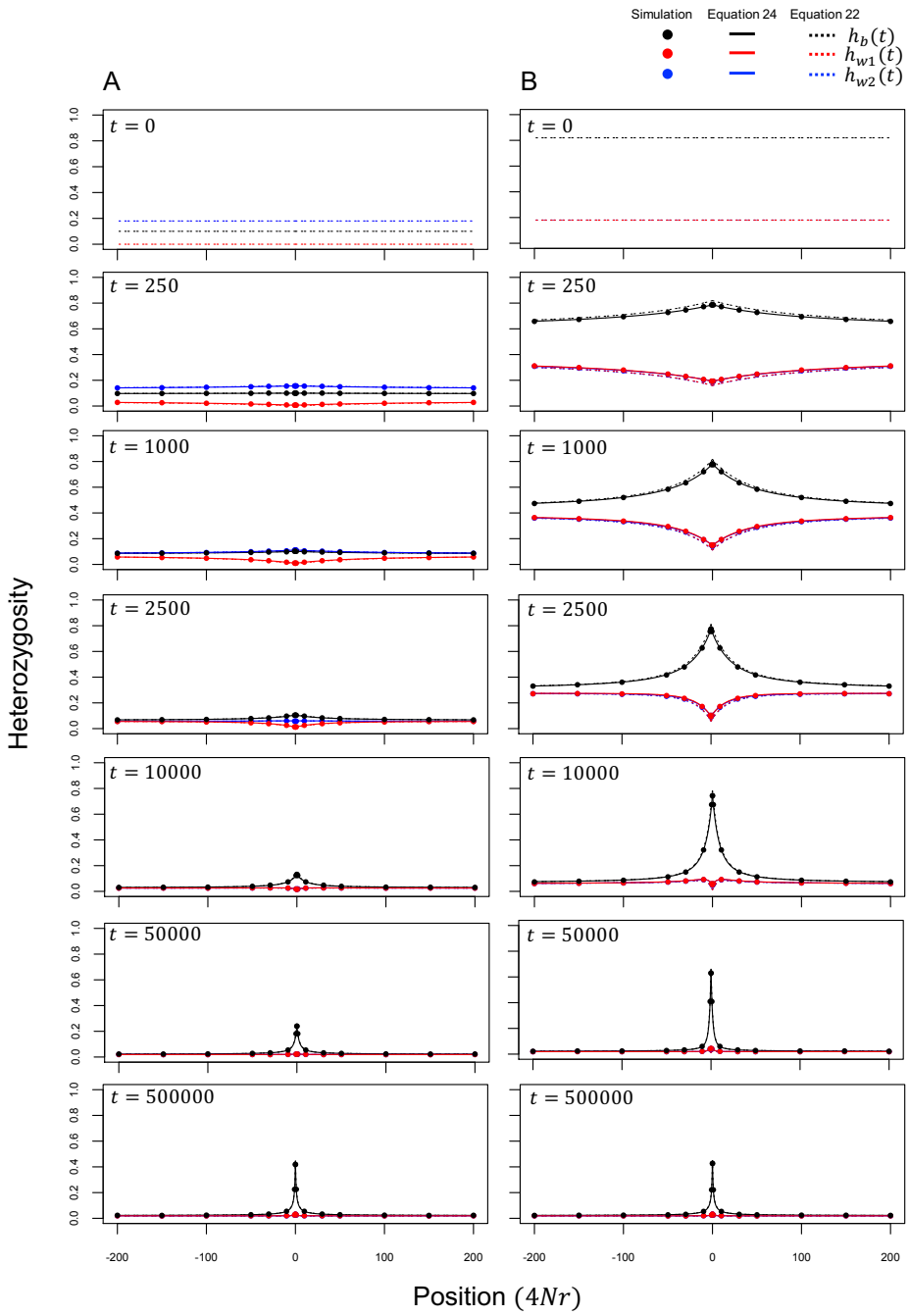


Figure A1 Temporal change of heterozygosity (h_{w1}, h_{w2}, h_b) as a function of recombination rate (A) after a local sweep in subpopulation I and (B) after a secondary contact. $y_1(0) = 0.0$ and $y_2(0) = 0.1$ are assumed in (A), whereas $y_1(0) = 0.1$ and $y_2(0) = 0.9$ in (B). Theoretical results from Equations 22 and 24 are shown by broken and solid lines, respectively. Simulation results (closed circles) are the averages over 50,000 replications of forward simulation. The left panel is identical to Figure 5.