The wave of gene advance under diverse systems of mating

Xin-Xin Zhang^{1,2} · Xiang Cheng^{1,2} · Ling-Ling Li^{1,2} · Xi Wang^{1,2} · Wei Zhou^{1,2} · Xiao-Yang Chen^{1,2} · Xin-Sheng Hu^{1,2}

Received: 17 December 2019 / Revised: 9 June 2020 / Accepted: 9 June 2020 / Published online: 30 June 2020

© The Author(s), under exclusive licence to The Genetics Society 2020

Abstract

Mating systems will influence gene spread across the natural distribution of a plant species. Existing theories have not fully explored the role of mating systems on the wave of advance of an advantageous gene. Here, we develop a theory to account for the rate of spread of both advantageous and neutral genes under different mating systems, based on migration-selection processes. We show that a complex relationship exists between selfing rate and the speed of gene spread. The interaction of selfing with gametophytic selection shapes the traveling wave of the advantageous gene. Selfing can impede (or enhance) the spread of an advantageous gene in the presence (or absence) of gametophytic selection. The interaction of selfing with recombination shapes the spread of a neutral gene. Linkage disequilibrium, mainly generated by selfing, enhances the traveling wave of the neutral gene that is tightly linked with the selective gene. Recombination gradually breaks down the genetic hitchhiking effects along the direction of advantageous gene spread, yielding decreasing waves of advance of neutral genes. The stochastic process does not alter the pattern of selfing effects except for increasing the uncertainty of the waves of advance of both advantageous and neutral genes. This theory helps us to explain how mating systems act as a barrier to spread of adaptive and neutral genes, and to interpret species cohesion maintained by a low level of adaptive gene flow.

Introduction

Gene spread in space is affected by many factors, including the mode of gene inheritance (uni- vs. bi-parental inheritance), physical barriers, gene neutrality, irregular dispersal of pollen and seeds for plants (or gametes and zygotes for animals), variation in population density, and variation of selection (intrinsic vs. extrinsic) in different habitats. Fisher (1937) first studied the wave of advance of advantageous nuclear genes in one-dimensional homogeneous space, showing that the velocity is proportional to the square root of the product of selection advantage and diffusion

coefficient. When applied to biological invasion, this conclusion generally holds in theory under the selection-migration processes but is modified in various situations (Shigesada and Kawasaki 1997), such as biological invasion within heterogeneous habitats or in the presence of competing species. In flowering plants, both gene spread and biological invasion (or species range expansion) are potentially affected by mating systems (Grossenbacher et al. 2015; Razanajatovo et al. 2016; Pickup et al. 2019). For instance, pollen limitation can lead to Allee effects and slow down species invasion (Groom 1998; Davis et al. 2004). Mating systems may act as a barrier to gene flow or interact with selection, which naturally influences the wave of advance of advantageous genes, or the species range expansion that could co-evolve with the wave of advance of advantageous genes (Hu et al. 2019). Nevertheless, the significance of mating systems in regulating the wave of advance of advantageous genes remains to be elaborated.

The interaction of mating systems with selection in shaping gene spread can occur in both the gametophyte and sporophyte stages of the life cycle. We recently showed that selfing as a barrier to adaptive gene flow can limit a species range under gametophytic selection (Hu et al. 2019). Differences between outcrossing rates could provide a biological basis for pre-mating barriers to both neutral and selective gene flow between genetically related species (Pickup et al. 2019).

Associate editor: Jinliang Wang

Supplementary information The online version of this article (<https://doi.org/10.1038/s41437-020-0333-9>) contains supplementary material, which is available to authorized users.

✉ Xin-Sheng Hu
xinsheng@scau.edu.cn

¹ College of Forestry and Landscape Architecture, South China Agricultural University, Guangdong 510642, China

² Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, South China Agricultural University, Guangdong 510642, China

Genes may be expressed in either the gametophyte or sporophyte stage, or in both stages (Tanksley et al. 1981; Shaw et al. 2011), and are likely involved in gametophytic and sporophytic selection. Peters and Weis (2018) show the influences of selfing on the spread and maintenance of pollen-expressed genes that affect pollen competitiveness in the model of a single locus controlling pleiotropic antagonism between pollen competitiveness and diploid fitness. Selfing interacts with selection in different ways in the gametophyte and sporophyte stages of a flowering plant species (Charlesworth and Charlesworth 1992; Damgaard et al. 1994; Hu and Li 2003). For instance, when both gametic and zygotic selection are present, the interaction between selfing and zygotic selection increases allele frequency, but the interaction between selfing and gametic selection reduces allele frequency (Hu 2015, p 1177). Also, mating systems interact with alternation of generations in plant life cycle. Early theories show that selfing enhances the predominance of the gametophytic generation of the life cycle, while outcrossing favors the sporophyte generation of the life cycle (Otto and Marks 1996). This potentially influences the wave of advance of advantageous genes among species with different rates of selfing. Thus, it is necessary to bring together the joint effects of selfing with gametophytic and/or sporophytic selection to examine their influence on the rate of gene spread.

The interaction of mating systems with migration is realized through changing pollination, an essential phase in the life cycle of flowering plants. There are diverse systems of mating ranging from outcrossing to selfing in plants, with many species exhibiting a mixed mating system (Igic and Kohn 2006). At one extreme, complete self-fertilization assures reproductive success and maintains population growth (Darwin 1876), even in the absence of pollinators. Here the wave of advance of advantageous genes is carried out by seed flow only. At another extreme, outcrossing or self-incompatible species maximize pollen transfer from other individuals while minimizing inbreeding depression (Charlesworth and Willis 2009). Here the wave of advance of advantageous genes can be carried out by pollen or seed flow alone, or both. The relative contribution to gene spread is often associated with the presence or absence of physical barriers and the type of vectors for seed and pollen dispersal. This varies with species, as implied from the estimates of the rates of pollen to seed flow (m_p/m_s , Table 1) derived from molecular markers (Ennos 1994). An intermediate system is mixed mating where both selfing and outcrossing modes of transmission of parental gene copies to progeny occur with certain probabilities. Although the evolutionary stability of this system remains under debate (Charlesworth 2006; Johnston et al. 2009; Barrett 2014), it does add more complexity to analyzing the role of mating systems in regulating gene spread (Hu 2011). Thus, it is important to consider the interaction of mating systems with pollen flow in addressing the wave of gene advance.

In addition, when a nuclear neutral gene is considered, a potential wave generated by seed and pollen flow dissipates with time in homogeneous habitats, irrespective of presence or absence of a genetic drift process. When the neutral gene is tightly linked with a selective gene, the transient wave of advance of the neutral gene is maintained by migration and linkage disequilibrium (LD) processes. This is analogous to the transient clines of neutral genes in the cytonuclear system (Hu 2008). LD is generated by gene flow but eroded by recombination between the two loci. This differs from genetic hitchhiking or background selection effects where the effective population size for the neutral gene is decreased owing to selection at its linked selective site (Nordborg et al. 1996a, b; Barton 2000; Hu and He 2005). A further effect of the selfing process is strengthening LD (Nordborg et al. 1996a, b; Nordborg 2000), and the transient wave of neutral gene advance is anticipated to follow the wave of advance of advantageous genes. Selfing alters neutral gene spread and could produce a wave different in shape from that under a random mating system or in a subdivided population (Slatkin and Charlesworth 1978; Slatkin and Wiehe 1998; Barton 2000).

Here, we bring together the processes of mating systems, selection and migration to examine the wave of advance of both advantageous and neutral genes in flowering plants. We begin by describing the basic theory for modeling the dynamics of selective and neutral nuclear genes in one-dimensional homogeneous space, followed by both numerical and Monte Carlo (MC) simulations to examine the effects of mating systems on gene spread. Inferences on the role of mating systems in shaping the wave of advance of advantageous and neutral genes are drawn from both the analytical model and simulation examples.

Basic theory

A single selective gene

The theory deals with a hermaphrodite plant species distributed in one-dimensional homogeneous space. Consider an initial population that is stable in demography and genetic composition. The influence on the initial population from reverse dispersal is assumed to be negligible. A mutant allele has an advantage over preexisting alleles and has attained a high frequency due to natural selection in the starting population. Adaptive alleles spread from the initial population and move along one-dimensional space. Figure 1 shows the life cycle for each local population that follows the sequence of events: generation of pollen and ovules, pollen flow, selection at the gametophyte stage, mixed mating, seed flow, and selection at the sporophyte stage (Hu et al. 2019). The difference from a previous study

Table 1 Symbols and their meanings used in the model.

Symbol	Definition
σ_s^2, σ_p^2	Dispersal variances of seeds and pollen, respectively
α	Selfing rate
$\tilde{\sigma}^2 (= (1 - \alpha)\sigma_p^2/2 + \sigma_s^2)$	Composite variance of seed and pollen dispersal
Δt	A time interval
s_h, s_d	Selection coefficients of advantageous allele A at gametophytic and sporophytic stages, respectively
w_A, w_a	Fitness of alleles A and a , respectively
W_{AA}, W_{Aa}, W_{aa}	Fitness of genotypes AA , Aa , and aa , respectively
$w_P, w_O, \bar{W}$	Average fitness in pollen, ovules, and sporophyte stage, respectively
$p_i, p_i^* (i = A, a)$	Allele frequency before and after pollen dispersal, respectively
$p_{ij}, p_{ij}^* (i, j = A, a)$	Genotype frequency before and after seed dispersal, respectively
n	Population density at position x
D_A	Hardy–Weinberg disequilibrium coefficient
K	Environmental capacity
ϵ	Per-capital growth rate
$p_{ik}, p_{ijkl} (I, j = A, a; k, l = B, b)$	Frequency of gamete ij and genotype $ijkl$, respectively
D_{AB}	Linkage disequilibrium (LD) between alleles A and B on the same gametes
$D_{A/B}$	LD between alleles A and B on different gametes
$p_{A/B}$	Frequency of alleles A and B on different gametes
Δ_{AB}	Composite digenic disequilibrium
δ_{AB}	Twice the difference between the frequencies of double heterozygotes at two loci in two linkage phases
$I_{ijkl} (I, j = A, a; k, l = B, b)$	Composite genotype frequency after a mixed mating system
c_0, c_α	The rates of gene spread under outcrossing ($\alpha = 0$) and mixed mating ($\alpha \neq 0$) systems, respectively
m_p, m_s	Migration rates of pollen and seeds between adjacent populations, respectively
$\dot{f}, f', f'' (f = p_A, p_B, D_A, D_{AB}, D_{A/B}, \dots)$	$\partial f / \partial t, \partial f / \partial x, \partial^2 f / \partial x^2$

Fig. 1 The life cycle in the model shows a sequence of events, including the interactions of mating systems with pollen flow and gametophytic selection, seed flow and sporophytic selection. These

different events are biologically connected within one life cycle, with arrow lines showing the occurrence of sequential events.

(Hu et al. 2019) lies in that we focus here on the spread of advantageous mutant alleles where traveling waves occur (Fisher 1937). A traveling wave is the term given to the processes in which an allele advances at a constant speed while maintaining a constant shape of the distribution of allele frequencies (Shigesada and Kawasaki 1997, p 47). Weak selection is considered in both gametophyte and sporophyte stages so that the terms with selection coefficients of the second or higher orders are neglected in mathematic deduction. For the rigorous theoretical analysis,

the life cycle is assumed to occur within a time interval Δt so that partial differentiation equations are derived when Δt becomes infinitely small (Nagylaki 1975). The diffusion process is applied to approximate seed and pollen flow in space. Dispersal variances of pollen (σ_p^2) and seeds (σ_s^2) are assumed to be of the same order as the selection coefficients so that terms with the product of gene flow and selection coefficients are neglected. Mutation rate is assumed to be small and its effects are excluded. Genetic drift effects are only considered in MC simulations.

In the Supporting Information (SI) Appendix the derivations of recursive gene and genotype frequencies at two loci are detailed. Here we briefly describe a selective nuclear locus, with alleles A and a . In the gametophyte stage, let $w_A = 1 + s_h$ be the fitness for gamete A in pollen and ovules, where s_h is the selection coefficient, and $w_a = 1$ for gamete a . Allele A is advantageous over allele a in adaptation to the habitat. The average fitness in pollen and ovules, denoted by w_p and w_o , respectively, is calculated using the conventional method $w_p = \sum_{i=A,a} w_i p_i^*$, where p_i^* is the allele frequency after pollen flow, and $w_o = \sum_{i=A,a} w_i p_i$ where p_i is the allele frequency in adults before pollen and ovule formation because ovules do not migrate between populations. Selection in pollen increases the frequency of the advantageous allele, which removes maladaptive alleles from both resident and migrant pollen, resulting in pollen discounting. For simplicity but without neglecting the role of selfing in shaping gene spread, we consider an additive selective model for genotype fitness in the sporophyte stage. Let $W_{AA} = 1 + 2s_d$ be the fitness of genotype AA , where s_d is the selection coefficient; $W_{Aa} = 1 + s_d$ for the fitness of heterozygotes Aa ; and $W_{aa} = 1$ for genotype aa . The average fitness in the sporophyte stage is $\bar{W} = W_{AA}p_{AA}^* + W_{Aa}p_{Aa}^* + W_{aa}p_{aa}^*$, where p_{ij}^* is the frequency of genotype ij ($i, j = A, a$) after seed dispersal.

Let α be the selfing rate in each local population, $\alpha = 1$ for complete self-fertilization and $\alpha = 0$ for outcrossing. Let n be population density at position x and time t . For simplicity of mathematical expression, we use notation $\dot{f}$ for $\partial f / \partial t$, f' for $\partial f / \partial x$, and f'' for $\partial^2 f / \partial x^2$ for a function f . With these settings, the dynamics for the frequency of the advantageous allele is obtained from Eq. (A25) in SI Appendix:

$$\begin{aligned} \dot{p}_A = & \tilde{\sigma}^2 (\ln(n))' p_A' + \frac{\tilde{\sigma}^2}{2} p_A'' \\ & + s_h \left(\left(1 - \frac{1}{2} \alpha \right) p_A p_a - \frac{1}{2} \alpha D_A \right), \\ & + s_d \left(\left(1 + \frac{1}{2} \alpha \right) p_A p_a + \frac{1}{2} \alpha D_A \right) \end{aligned} \quad (1)$$

where $D_A = (P_{Aa} - 2p_A p_a)/2$ is the Hardy–Weinberg disequilibrium (HWD) coefficient and $\tilde{\sigma}^2 = (1 - \alpha)\sigma_p^2/2 + \sigma_s^2$ is a composite rate of seed and pollen flow. The first term on the right side of Eq. (1) represents the effects of asymmetric dispersal that carries genes from dense populations to less dense populations when $(\ln(n))' > 0$, or vice versa when $(\ln(n))' < 0$. The second term represents the contribution from seeds or pollen flow. The proportion $(1 - \alpha)\sigma_p^2/2$ in $\tilde{\sigma}^2$ is involved in pollination, and only $\alpha\sigma_p^2/2$ of cross pollen is not used, which is one type of pollen discounting (reduced outcross pollen due to selfing). A high selfing rate causes a large proportion of

pollen discounting (Kohn and Barrett 1994). The third and fourth components are effects of selection, showing that selfing differently interacts with gametic and zygotic selection, respectively (Hu and Li 2003; Hu 2015).

From Eq. (A27) in SI Appendix, the dynamics of HWD coefficient D_A with space x and time t is derived as follows:

$$\begin{aligned} \dot{D}_A = & \left(-1 + \frac{1}{2} \alpha + \left(s_d - \frac{1}{2} s_h \right) \alpha (p_a - p_A) p_a p_A \right) D_A \\ & + \left(\frac{1}{2} + \left(s_d + \frac{1}{2} s_h \right) (p_a - p_A) \right) \alpha p_A p_a + \sigma_s^2 (p_A')^2 \\ & + \frac{1}{2} \alpha \sigma_s^2 \left((\ln(n))' (p_A p_a + D_A)' + \frac{1}{2} (p_A p_a + D_A)'' \right) \end{aligned} \quad (2)$$

Note that previous studies address the steady-state distribution of gene frequency using the approximation of $D_A = F p_A p_a$ where inbreeding coefficient F is equal to $\alpha/(2 - \alpha)$ (Caballero and Hill 1992). This is reasonable because other components in D_A have the second or higher orders of selection coefficients or the dispersal variances and are neglected.

Two scenarios are examined. One is that population density is uniformly distributed, and hence the terms with $(\ln(n))'$ in Eqs. (1) and (2) are equal to 0. The advantageous allele starts from the initial population and spreads by replacing the pre-existing alleles (Fig. 2a), like the case addressed by Fisher (1937). The second scenario is that the initial population colonizes new habitats along one-dimensional space (Fig. 2b). The advantageous gene spreads along with population colonization where variation in density is present $(\ln(n))' \neq 0$.

In the second scenario, we assume that there is no effect of density-dependent regulation or environmental capacity on the number of pollen and seeds in the gametophyte stage, but there is in the sporophyte stage. We assume that each population follows a logistical growth model. From Hu and He (2005), the change in population density with time and space, $n(x, t)$, can be described as follows:

$$\dot{n} = \frac{\sigma_s^2}{2} n'' + \epsilon n \left(1 - \frac{n}{K} \right), \quad (3)$$

where the first term represents the direct contribution from seed dispersal, the second term represents the logistic growth contribution in which K is the environmental capacity and is the per-capital growth rate derived as $\epsilon = \bar{W} - 1 = 2s_d p_A$ at position x and time t at a single selective locus. The maximum ϵ is equal to $2s_d$ when $p_A = 1$, which is much smaller than 1.0 under weak selection. Oscillation in population size that requires a large ϵ (say, $\epsilon > 1$) does not occur in the traveling waves simulated in the next section. A more complex model is needed when capacities in the sporophyte stage differ among genotypes (Roughgarden 1979; Hu and He 2005).

Fig. 2 Two hypothetical scenarios considered in the theory. a Uniform density where the advantageous allele in the initial population spreads along one-dimensional space; **b** the initial population colonizes new space and adaptive genes spread accompanying the colonization process. In (a), black solid line represents uniform density

distribution in space. Black dashed and solid lines represent population density at earlier and later times in (b), respectively; while red dashed and solid lines represent frequencies of adaptive genes at earlier and later times in (a) and (b), respectively.

A neutral gene linked with the selective gene

Now, we describe a neutral locus B linked with the selective locus A, with alleles B and b . The recombination rate between the two loci is designated by r ($0 \leq r \leq 0.5$). Let p_{ik} and p_{ijkl} be the frequencies of gamete ik ($i = A, a; k = B, b$) and genotype $ijkl$ ($i, j = A, a; k, l = B, b$) in the current adult population at position x and time t . Let D_{AB} ($= p_{AB} - p_A p_B$) be the LD in current adults between alleles A and B on the same gametes, and $D_{A/B}$ ($= p_{A/B} - p_A p_B$) be the LD in current adults between alleles A and B on different gametes. Note that the forward slash “/” stands for alleles from separate gametes. The two digenic frequencies p_{AB} and $p_{A/B}$ in current adults are calculated by:

$$p_{AB} = p_{AABB} + (p_{AABb} + p_{AaBB} + p_{AB/ab})/2, \quad (4a)$$

$$p_{A/B} = p_{AABB} + (p_{AABb} + p_{AaBB} + p_{Ab/aB})/2, \quad (4b)$$

denote $\Delta_{AB} = D_{AB} + D_{A/B}$, a composite digenic disequilibrium in current adults (Weir 1996, p 126), and $\delta_{AB} = 2(D_{AB} - D_{A/B})$, which equals twice the difference between the frequencies of double heterozygotes at two loci in two linkage phases ($p_{AB/ab} - p_{Ab/aB}$).

Genotypic frequencies at two loci can be derived based on Wright's (1969) approach that combines the selfing and random mating parts in the mixed mating system. From Eq. (A36) in SI Appendix, the change of allele frequency at the neutral locus is given by:

$$\dot{p}_B = \tilde{\sigma}^2 (\ln(n))' p_B' + \frac{\tilde{\sigma}^2}{2} p_B'' + (s_h + s_d) \left((1 - \alpha)(1 - r)D_{AB} + \frac{1 - 2r}{4} \alpha \delta_{AB} \right) + \alpha s_d \Delta_{AB} \quad (5)$$

Strong gametic LD (e.g., close linkage) enhances the trend toward concordant spread of selective and neutral

genes in traveling waves. Composite LD (Δ_{AB}) interacts with selection in the sporophyte stage and imposes additional effects on neutral gene spread.

To evaluate the spread of neutral genes, we need calculations of two digenic LDs (D_{AB} and $D_{A/B}$). From Eq. (A44) in SI Appendix, the change of gametic LD (D_{AB}) per generation is as follows:

$$\begin{aligned} \dot{D}_{AB} = & -rD_{AB} - \frac{1}{2} \alpha r \delta_{AB} + (1 - r) \tilde{\sigma}^2 (\ln(n))' D_{AB}' \\ & + \frac{\tilde{\sigma}^2}{2} (2p_A' p_B' + (1 - r) D_{AB}'') - \frac{1}{2} \alpha r \sigma_s^2 \left((\ln(n))' \delta_{AB}' + \frac{1}{2} \delta_{AB}'' \right) \\ & + s_h \left(\frac{(1 - \alpha)(p_a - p_A)(1 - r)D_{AB}}{4} + \frac{1}{4} \alpha (p_{AaBB} + p_{AB/ab} - (p_A + (p_a - p_A)r) \delta_{AB} - p_B p_{Aa}) \right) \\ & + s_d \left((p_a - p_A)(1 - r)D_{AB} + \alpha ((1 - r)D_{AB} - r p_a \delta_{AB} - p_A \Delta_{AB}) \right. \\ & \left. - \frac{1}{4} \alpha (p_{AaBB} + p_{AB/ab} - r \delta_{AB} + (1 - 2r)p_A \delta_{AB} - p_B p_{Aa}) \right) \end{aligned} \quad (6)$$

From Eq. (A45), the change of the digenic LD for the two genes on different gametes ($D_{A/B}$) is derived as follows:

$$\begin{aligned} \dot{D}_{A/B} = & \frac{1}{2} \alpha \Delta_{AB} - D_{A/B} + \frac{1}{2} \alpha \sigma_s^2 (\ln(n))' \Delta_{AB}' + \sigma_s^2 (p_A' p_B' + \frac{1}{4} \alpha \delta_{AB}'') \\ & + \frac{1}{4} \alpha s_h (p_{AABb} + p_{AB/ab} - (p_A + (p_a - p_A)r) \sigma_{AB} - p_B p_{Aa}) \\ & + s_d (-2 \alpha p_A p_a p_B + \alpha (p_a - p_A) \Delta_{AB} - \frac{1}{4} \alpha (p_{AaBB} + p_{Ab/aB} \\ & + (p_A + (p_a - p_A)r) \delta_{AB} - p_B p_{Aa})) \end{aligned} \quad (7)$$

Pollen flow does not directly change $D_{A/B}$ because each pollen grain (haploid) only carries one allele at one locus, but directly affects D_{AB} because each pollen grain carries one of four types of gametes (AB , Ab , aB , and ab).

To evaluate two digenic LDs in the mixed mating system, we need to calculate three genotype frequencies (p_{AaBB} , p_{AABb} , and the double heterozygotes with two

linkage phases, $p_{AB/ab}$ and $p_{Ab/aB}$). Note that these three frequencies together with the frequency of any genotype of double homozygotes (e.g., p_{AABB}) are independent among all nine two-locus genotypes. For simplicity, let $l_{AaBB} = 2(1 - \alpha)p_{AB}p_{aB} + \frac{1}{2}\alpha(p_{AaBB} + r(1 - r)(p_{AB/ab} + p_{Ab/aB}))$ where $p_{ik}(i = A, a; k = B, b)$ is the gametic frequency in pollen and ovules (next generation). l_{AaBB} represents the frequency of genotype $AaBB$ that is derived under both random mating and selfing but without effects of selection and gene flow. From Eq. (A38) of SI Appendix, the change of genotype frequency p_{AaBB} is derived as follows:

$$\begin{aligned} \dot{p}_{AaBB} = & l_{AaBB} - p_{AaBB} + (1 - \alpha)\sigma_p^2(\ln(n))'(p_{AB}p_{aB})' \\ & + \frac{(1 - \alpha)\sigma_p^2}{2}(p_{AB}p_{aB}'' + p_{AB}''p_{aB}) \\ & + \sigma_s^2(\ln(n))'l_{AaBB}' + \frac{\sigma_s^2}{2}l_{AaBB}'' \\ & + 2s_h(1 - \alpha)(p_a - p_A)p_{AB}p_{aB} + s_d(p_a - p_A)l_{AaBB} \end{aligned} \quad (8)$$

which can be numerically calculated.

Similarly, let $l_{AABb} = 2(1 - \alpha)p_{AB}p_{Ab} + \frac{1}{2}\alpha(p_{AABb} + r(1 - r)(p_{AB/ab} + p_{Ab/aB}))$. From Eq. (A39) of SI Appendix, the recursive equation for genotypic frequency p_{AABb} is derived as follows:

$$\begin{aligned} \dot{p}_{AABb} = & l_{AABb} - p_{AABb} + (1 - \alpha)\sigma_p^2(\ln(n))'(p_{AB}p_{Ab})' \\ & + \frac{(1 - \alpha)\sigma_p^2}{2}(p_{AB}p_{Ab}'' + p_{AB}''p_{Ab}) \\ & + \sigma_s^2(\ln(n))'l_{AABb}' + \frac{\sigma_s^2}{2}l_{AABb}'' \\ & + 2s_h(1 - \alpha)(p_a - p_A)p_{AB}p_{Ab} + 2s_d p_A l_{AABb} \end{aligned} \quad (9)$$

The frequency of genotype $AaBb$ is separately considered in distinct linkage phases. Let $l_{AB/ab} = 2(1 - \alpha)p_{AB}p_{ab} + \frac{1}{2}\alpha((1 - r)^2 p_{AB/ab} + r^2 p_{Ab/aB})$. From Eq. (A40) of SI Appendix, the recursive equation for genotypic frequency $p_{AB/ab}$ is derived as follows:

$$\begin{aligned} \dot{p}_{AB/ab} = & l_{AB/ab} - p_{AB/ab} + (1 - \alpha)\sigma_p^2(\ln(n))'(p_{AB}p_{ab})' \\ & + \frac{(1 - \alpha)\sigma_p^2}{2}(p_{AB}p_{ab}'' + p_{AB}''p_{ab}) \\ & + \sigma_s^2(\ln(n))'l_{AB/ab}' + \frac{\sigma_s^2}{2}l_{AB/ab}'' \\ & + 2s_h(1 - \alpha)(p_a - p_A)p_{AB}p_{ab} + s_d(p_a - p_A)l_{AB/ab} \end{aligned} \quad (10)$$

Let $l_{Ab/aB} = 2(1 - \alpha)p_{Ab}p_{aB} + \frac{1}{2}\alpha((1 - r)^2 p_{Ab/aB} + r^2 p_{AB/ab})$. From Eq. (A41) of SI Appendix, the recursive

equation for genotypic frequency $p_{Ab/aB}$ is derived as follows:

$$\begin{aligned} \dot{p}_{Ab/aB} = & l_{Ab/aB} - p_{Ab/aB} + (1 - \alpha)\sigma_p^2(\ln(n))'(p_{Ab}p_{aB})' \\ & + \frac{(1 - \alpha)\sigma_p^2}{2}(p_{Ab}p_{aB}'' + p_{Ab}''p_{aB}) + \sigma_s^2(\ln(n))'l_{Ab/aB}' \\ & + \frac{\sigma_s^2}{2}l_{Ab/aB}'' + 2s_h(1 - \alpha)(p_a - p_A)p_{Ab}p_{aB} + s_d(p_a - p_A)l_{Ab/aB} \end{aligned} \quad (11)$$

Thus, the spread of neutral genes can be thoroughly assessed from Eqs. (5) to (11). The dynamics of composite digenic LD is calculated according to the relationships of $\Delta_{AB} = \dot{D}_{A/B} + \dot{D}_{AB}$ and $\dot{\delta}_{AB} = 2(\dot{D}_{AB} - \dot{D}_{A/B})$.

Simulations

Methods

Two approaches are separately applied to assess the wave of advance of both advantageous and neutral genes. The first approach is based on analytical calculations described above and uses the forward-finite difference method to investigate the waves of gene advance. The initial population is fixed in density (n_0) and genetic composition. The step size is set as $dx = 1$ for the space and $dt = 1$ for the time (generation). The dispersal variances are set so that the values for $(\sigma_s^2/2)\partial t/(\partial x)^2$ and $(\sigma_p^2/2)\partial t/(\partial x)^2$ are smaller than 1, leading to stable simulations (DuChateau and Zachmann 1989, p 352–365).

The second approach is based on MC simulation where a one-dimensional stepping-stone model is used to simulate gene flow among populations (Kimura 1953). The initial population is stable in density and genetic composition. Each population follows the same life cycle (Fig. 1) together with additional genetic drift after sporophytic selection. In the scenario of uniform density (Fig. 2a), the MC simulation considers all populations are initially occupied by genotype $aabb$. A mutant allele A of adaptive advantage over allele a and its linked neutral allele B spread along the one-dimensional space. We assume that allele A has attained a high frequency in the initial population due to selection, and so has the neutral allele B linked with allele A due to genetic hitchhiking effects (Barton 2000). This assumption ensures that the advantageous A allele has a very small probability of going extinct in the initial population. Fixation of allele A in the initial population is an extreme case. Exchanges of migrants between two nearest-neighbor demes are allowed, each with migration rates of $m_s/2$ and $m_p/2$ for seeds and pollen, respectively. Genotypic frequencies after selection in the sporophyte stage are randomly sampled according to the multinomial distribution of genotypic frequencies, given the population size.

In the scenario of nonuniform density (Fig. 2b), the MC simulation considers that the initial population colonizes new habitats along one-dimensional homogeneous space. Dispersal of seeds and pollen (m_s and m_p) occurs between the nearest-neighbor populations. Joint changes in population density and gene frequency are examined. Each derived local population originates from immigrant seeds from its nearest-neighbor population that was established earlier. When a derived population starts (for simulation purpose, say, $n(x, t) \geq 1$), we consider both seed and pollen flow between nearest-neighbor populations. The logistical growth model (the second term in Eq. (3)) is used in each population. Gametic and allelic frequencies in migrants of pollen and seeds are assumed to be the same as in the donor populations. Genotypic frequencies after selection in the sporophyte stage are randomly sampled according to the multinomial distribution of genotypic frequencies with the obtained population size.

One thousand independent data sets are generated for a particular set of simulation conditions, and each data set is used to estimate mean and standard deviations of different variables within each population. Results are extracted at 50 generation intervals for analysis. Scripts in C have been submitted to Dryad.

Spread of advantageous genes

In the scenario of uniform population density (Fisher 1937), i.e. $(\ln(n))' = 0$, the advantageous allele A advances in theory at a constant speed to replace the previous allele a and maintains the same shape of distribution (Shigesada and Kawasaki 1997). Let c_0 be the rate at which allele A spreads under outcrossing ($\alpha = 0$). The speed c_0 is calculated according to Fisher (1937) and Kolmogorov et al. (1937):

$$c_0 = (2(s_h + s_d)(\sigma_s^2 + \sigma_p^2/2))^{1/2}. \quad (12)$$

The speed c_0 can be derived by application of the method provided by Shigesada and Kawasaki (1997, p 57–60). The speed positively correlates with selection coefficients in both gametophyte and sporophyte stages.

When selfing is the major process generating HWD in comparison with gene flow and selection, a stable HWD coefficient can be achieved and is approximated by $D_A = \frac{\alpha}{2-\alpha} p_A p_a$ from Eq. (2). According to Shigesada and Kawasaki (1997, p 57–60), substitution of D_A into Eq. (1) yields the speed c_α as follows:

$$c_\alpha = \left(2 \left(s_d + s_h + (s_d - s_h) \frac{\alpha}{2 - \alpha} \right) \left(\sigma_s^2 + \frac{1 - \alpha}{2} \sigma_p^2 \right) \right)^{1/2} \quad (13)$$

indicating a very complicated relationship between the selfing rate and speed under different levels of pollen and seed dispersal.

To look at the effects of selfing rate on the speed, we take the derivative of c_α with respect to α (suggestion of one anonymous reviewer):

$$\frac{\partial c_\alpha}{\partial \alpha} = \frac{1}{(2 - \alpha)^2 c_\alpha} (2\sigma_s^2(s_d - s_h) - \sigma_p^2(s_d + s_h(3 - \alpha)(1 - \alpha))). \quad (14)$$

It can be inferred that self-fertilization impedes gene flow ($\frac{\partial c_\alpha}{\partial \alpha} < 0$) when the condition of $\frac{\sigma_p^2}{\sigma_s^2} > \frac{2(1 - s_h/s_d)}{1 + (3 - \alpha)(1 - \alpha)s_h/s_d}$ is met, but enhances gene flow ($\frac{\partial c_\alpha}{\partial \alpha} > 0$) when $\frac{\sigma_p^2}{\sigma_s^2} < \frac{2(1 - s_h/s_d)}{1 + (3 - \alpha)(1 - \alpha)s_h/s_d}$. Self-fertilization always impedes gene flow when $s_d < s_h$. The maximum c_α is related to the relative values of selection coefficients and pollen/seed dispersal variances.

In the scenario of uniform population density, SI Fig. 1a shows the traveling wave of advantageous alleles based on Eqs. (1) and (2). The HWD coefficient exhibits one unimodal distribution (SI Fig. 1b), with the maximum occurring at around the middle of the wave front. Figure 3a, b shows that gene spread decreases, while the maximum HWD coefficient increases as the selfing rate increases from $\alpha = 0$ to 1.0, indicating that selfing impedes the spread of advantageous genes.

Alternatively, when selection coefficients are negative, allele A is maladaptive to the habitat compared with allele a . If the initial population is fixed by allele A , gene spread is counteracted by selection, eventually resulting in a limited range of distribution. SI Fig. 2 shows the limitation of gene spread at $t = 500$ under uniform population density and maladaptive allele A ($s_h = -0.01$ and $s_d = -0.02$). The rate of gene spread tends to decrease as the selfing rate increases.

When only gametophytic selection occurs ($s_h \neq 0$, $s_d = 0$), gene spread decreases as the selfing rate increases under a given rate of seed flow (Fig. 4a for the case of uniform population density), indicating that a greater proportion of pollen discounting due to the joint effects of selfing and gametic selection can substantially impede gene spread. This also confirms the relation of $\partial c_\alpha / \partial \alpha < 0$ in the presence of gametic selection only. The maximum HWD increases as the selfing rate increases (Fig. 4b). When only sporophytic selection occurs ($s_h = 0$, $s_d \neq 0$), selfing slightly regulates the wave of gene advance. For instance, Fig. 4c shows that selfing enhances gene spread to some extent under certain levels of seed flow. The pattern is consistent with $\partial c_\alpha / \partial \alpha > 0$ under this parameter setting. The maximum HWD increases as the selfing rate increases (Fig. 4d). Note allele A does not

Fig. 3 Effects of selfing on the wave of gene advance. a, c Traveling wave of allele frequency; **b, d** HWD coefficients. Results are derived at generation $t = 100$. **a, b** Refer to the results under uniform population density. **c, d** Refer to the results under nonuniform population density, $(\ln(n))' \neq 0$. Common parameters are the gametic selection coefficient $s_h = 0.02$, zygotic selection coefficient $s_d = 0.04$, pollen dispersal variance = 0.05, and seed dispersal variance = 0.01. The initial population is set as $p_A = 1.0$, $D_A = 0.0$, and $n_0 = 100$ for the case of $(\ln(n))' \neq 0$.

spread without seed dispersal in the selfing system. In the absence of pollen flow ($\sigma_p^2 = 0.0$) but presence of seed dispersal (e.g., $\sigma_s^2 = 0.04$ in Fig. 4e, f), selfing discounts outcross pollen from resident populations and facilitates gene spread, which is consistent with the result of $\partial c_d / \partial \alpha > 0$ under $\sigma_p^2 = 0$ and $s_d > s_h$.

MC simulations indicate that patterns for selfing effects are qualitatively similar between the stochastic and deterministic processes except that their traveling waves fluctuate under genetic drift. Selfing impedes the wave of advance of advantageous genes. For instance, the wave of advance of the advantageous allele has the largest rate of speed under random mating ($\alpha = 0$) but the smallest rate under selfing ($\alpha = 1$) (Fig. 5a). The maximum value of standard deviation occurs at around $p_A = 0.5$ since the variance of allele frequency is proportional to $p_A(1 - p_A)$ (Weir 1996, p 42). Also, the position with the largest standard deviations of allele frequency moves forward along the spread of the advantageous allele as the selfing rate decreases (Fig. 5b).

MC simulations indicate that genetic drift effects shape the waves of advance of the advantageous allele. SI Fig. 3a shows that the wave of gene advance speeds up as population size increases. In addition, as the population size

increases, the maximum standard deviation of the position of gene advance decreases (SI Fig. 3b).

In the scenario of nonuniform population density, i.e. $(\ln(n))' \neq 0$, both population density and gene frequency exhibit traveling waves (SI Fig. 1a–c). Patterns regarding the effects of selfing rate are qualitatively similar between the two scenarios except that variation in population density enhances the rate of gene spread. For instance, gene spread decreases as the selfing rate increases (Fig. 3c). The HWD distribution shrinks slightly as the selfing rate increases from 0 to 1.0 (Fig. 3d). This is because pollen dispersal affects population density and HWD by changing the spatial distribution of the frequencies of adaptive alleles.

MC simulations also confirm that selfing impedes the wave of advance of advantageous alleles. Selfing restricts colonization in the presence of gametophytic selection (Hu et al. 2019). For instance, population density exhibits a traveling wave with the largest rate of spread under random mating (Fig. 6a), and so does the frequency of advantageous alleles (Fig. 6c). As the selfing rate increases, the standard deviations increase for the waves of advance of population density and advantageous allele frequency (Fig. 6b, d). This is because the uncertainty

Fig. 4 Effects of gametophytic or sporophytic selection on the wave of gene advance. **a, c, e** Traveling wave of allele frequency; **b, d, f** HWD coefficients. Results are derived at generation $t = 100$ under uniform population density. **a, b** Refer to the results with gametophytic selection only $s_h = 0.02$ and $s_d = 0.0$. **c, d** Refer to the results with sporophytic selection only $s_d = 0.04$ and $s_h = 0.0$. Common parameters in (**a–d**) are the pollen dispersal variance = 0.05, and seed dispersal variance = 0.04. **e, f** Refer to the results with selection $s_d = 0.04$ and $s_h = 0.02$, the pollen dispersal variance = 0.0, and seed dispersal variance = 0.04. The initial population in each case is set as $p_A = 1.0$ and $D_A = 0.0$.

increases when genetic drift effects are high (a small population density) and the allele frequency p_A is low ($<<0.5$), but is reduced when density is large and p_A is high ($>>0.5$).

Spread of neutral genes

Numerical simulations for neutral genes are based on Eqs. (5)–(11). Though accompanying the traveling wave of

advantageous alleles, the wave of advance of neutral genes gradually declines under both the scenario of uniform ($(\ln(n))' = 0$) and nonuniform ($(\ln(n))' \neq 0$) population density (Fig. 7a, b). This is associated with the wave of advance of the two digenic LDs (D_{AB} and $D_{A/B}$) that are gradually weakened with time owing to recombination (Fig. 7c, d). The genetic hitchhiking effects are broken down by recombination as the advantageous allele spreads. D_{AB} is greater in value than $D_{A/B}$ although both LDs exhibit an unimodal

Fig. 5 Effects of selfing rates on traveling waves under the stochastic process and uniform population density. **a** Average allele frequency p_A ; **b** standard deviations $Sd(p_A)$; **c** average neutral gene frequency p_B ; and **d** standard deviations $Sd(p_B)$. All results are derived at the 100th generation. One thousand independent runs are conducted. Initial population settings are $n = 50$, $p_{AABB} = 0.5$, $p_{AABb} = 0.3$, $p_{AAAb} = 0.03$, $p_{AaBB} = 0.1$, $p_{AaBb} = 0.04$, $p_{Aabb} = 0.01$, $p_{aaBB} = 0.01$, $p_{aaBb} = 0.01$, and $p_{aabb} = 0.0$. Other common parameters are the recombination rate = 0.05, gametic selection coefficient $s_h = 0.01$, zygotic selection coefficient $s_d = 0.02$, pollen dispersal variance = 0.05, and seed dispersal variance = 0.01.

distribution. Like the wave of advance of the advantageous allele, variation in population density promotes wave advance of the neutral genes through strengthening LDs.

The recombination rate between selective and neutral loci is crucial in determining the wave of advance of the neutral gene. For instance, spread of the neutral gene quickly slows down as the recombination rate increases from 0.01 to 0.5 under both the scenario of uniform and nonuniform population density (SI Fig. 4a, d). The two digenic LDs also quickly decrease (SI Fig. 4b, c, e, f), and the spread of neutral genes mainly relies on the diffusion process under loose linkage. Thus, only those neutral loci that are tightly linked with the selective locus tend to follow the wave of advantageous genes due to strong gametic LDs (Fig. 7). The difference between the two scenarios lies in the fact that variation in population density enhances both digenic LDs and weakens the effects of recombination.

Selfing impedes the spread of neutral genes under both the scenario of uniform (Fig. 8b, c) and nonuniform (Fig. 8e, f) population density. Such effects are substantial for the neutral locus that is tightly linked with a selective locus when pollen dispersal is much stronger than seed dispersal ($\sigma_p^2 \geq \sigma_s^2$). Like the spread of advantageous alleles, random mating ($\alpha = 0$) maintains the largest rate of spread of neutral genes.

In the scenario of uniform population density, MC simulations confirm that the average traveling wave of neutral genes is similar under genetic drift to that under deterministic processes. Selfing impedes the rate of neutral gene spread. For instance, the average wave of advance of neutral genes slows down as the selfing rate increases from $\alpha = 0$ to 1 (Fig. 5c). The maximum values of standard deviation occur at around $p_B = 0.5$. Positions with the largest standard deviation of allele frequency move forward along the spread of the advantageous allele as the selfing rate decreases (Fig. 5d). Also, MC simulations indicate that genetic drift effects shape the average wave of advance of neutral genes. The average wave of neutral genes speeds up as the population size increases (SI Fig. 3c). The maximum standard deviation of the traveling waves decreases as the population size increases (SI Fig. 3d).

In the scenario of nonuniform population density, MC simulations confirm that selfing impedes the wave of advance of neutral genes. Although selfing initially enhances the spread of neutral genes owing to LD reinforced by selfing, the spread of neutral gene quickly slows down as the spread of the advantageous allele decreases with the selfing rate (Fig. 6e). As the selfing rate increases from $\alpha =$

Fig. 6 Effects of selfing rates on traveling waves under the stochastic process and nonuniform population density. **a** Average population density; **b** standard deviations of population density; **c** average allele frequency p_A ; **d** standard deviations of allele frequency $Sd(p_A)$; **e** average neutral gene frequency p_B ; and **f** standard deviations of allele frequency $Sd(p_B)$. All results are derived at the 500th generation. One thousand independent runs are conducted. Initial population settings are $n_0 = 50$,

environmental capacity $K = 50$ for all genotypes, $p_{AABB} = 0.5$, $p_{AABb} = 0.3$, $p_{AAbb} = 0.03$, $p_{AaBB} = 0.1$, $p_{AaBb} = 0.04$, $p_{Aabb} = 0.01$, $p_{aaBB} = 0.01$, $p_{aaBb} = 0.01$, and $p_{aabb} = 0.0$. Other common parameters are the recombination rate = 0.05, gametic selection coefficient $s_h = 0.01$, zygotic selection coefficient $s_d = 0.02$, pollen dispersal variance = 0.1, and seed dispersal variance = 0.1.

0 to 1, the standard deviation increases for the wave of advance of neutral genes (Fig. 6f). Similarly, the uncertainty is high when population density is low (large genetic drift effects) and allele frequency p_B is low, but is reduced when density is large and p_B is high (say, $\gg 0.5$).

MC simulations indicate that recombination slightly affects the wave of neutral gene spread when a certain level of seed dispersal is involved in population colonization. This is similar to the results in the scenario of

uniform density. For instance, the neutral gene generally exhibits a pattern similar to that of the advantageous allele under different recombination rates (SI Fig. 6a, c). Nevertheless, neutral genes spread more extensively when they are more closely linked with the selective locus (SI Fig. 6c). Large fluctuations occur in frontal waves of gene frequencies for both advantageous and neutral genes due to genetic drift effects (SI Fig. 6b, d).

Fig. 7 Traveling waves of both advantageous and neutral genes. **a** Allele frequency of an advantageous gene; **b** neutral gene frequency; **c** gametic LD D_{AB} ; and **d** gametic LD $D_{A/B}$.

Lines from left to right in each figure represent the results at generation 50, 100, 150, ..., 500, respectively. The Y-axis values are higher for D_{AB} than for $D_{A/B}$. Lines in blue refer to the results under uniform density $(\ln(n))' = 0$. Lines in red refer to the results under nonuniform density $(\ln(n))' \neq 0$. Initial population settings are $n_0 = 100$ for nonuniform density, $p_{AABB} = 0.5$, $p_{AABb} = 0.3$, $p_{AAbb} = 0.03$, $p_{AaBB} = 0.1$, $p_{AaBb} = 0.04$, $p_{Aabb} = 0.01$, $p_{aaBB} = 0.01$, $p_{aaBb} = 0.01$, and $p_{aabb} = 0.0$. Other common parameters are the selfing rate $\alpha = 0.0$, the recombination rate $= 0.01$, gametic selection coefficient $s_h = 0.02$, zygotic selection coefficient $s_d = 0.04$, pollen dispersal variance $= 0.05$, and seed dispersal variance $= 0.04$.

Discussion

Since Fisher's (1937) pioneering work, there have been extensive studies on the rate of spread of advantageous and neutral genes under random mating. The present study examines the role of mating systems in affecting the spread of advantageous and neutral genes, and provides an addition to the extant theories where effects of selfing are not considered. Selfing reduces outcross pollen from both immigrants and resident populations. The automatic transmission advantage of selfers over outcrossers (Fisher 1941) ensures that selfing is more effective than outcrossing in impeding gene dispersal. Two essential processes in our theory remain the same as Fisher (1937) and Haldane (1948), i.e., the migration (seed and pollen dispersal) and selection (gametophytic and sporophytic selection). An additional stochastic process is considered in MC simulations. Since mating systems play a critical role in structuring genetic variation within and between populations, it is important in both practice and theory to elucidate how selfing regulates the spread of advantageous and neutral genes within a species.

A few specific conclusions are summarized as follows. One is that a complex relationship exists between selfing rate and the speed of gene spread. The interaction of selfing with gametophytic selection is crucial in determining the role of

mating systems in regulating the wave of advance of advantageous genes. Selfing can impede (or promote) the spread of advantageous genes in the presence (or absence) of gametophytic selection. This is similar to the role of selfing in regulating a species' range (Hu et al. 2019). The second conclusion is that the stochastic process (genetic drift) does not alter the pattern of selfing effects except for increasing the uncertainty of the wave of advance for both advantageous and neutral genes. The third conclusion is that the interaction of selfing with recombination affects the rate of neutral gene spread, given the migration-selection processes for spreading advantageous genes. When selfing impedes the spread of advantageous genes, it also restricts the spread of the tightly linked neutral gene due to LD generated by both selfing and pollen/seed dispersal. The waves of advance of neutral genes gradually go down along the direction of spread of advantageous genes as the genetic hitchhiking effects are broken down by recombination with time. The fourth conclusion is that, given a particular mating system, variation in population density enhances the spread of advantageous genes and their tightly linked neutral genes. This also tends to weaken the recombination effects. The above results are applicable to interpreting how selfing impedes or enhances the spread of both advantageous and neutral genes in the natural distribution of a plant species.

Fig. 8 Effects of selfing rates on traveling waves of neutral genes. **a, d** Allele frequency of the advantageous gene; **b, e** neutral gene frequency; **c, f** gametic LD D_{AB} . All results are derived at the 100th generation. **a–c** are the results under uniform density $(\ln(n))' = 0$; **d–f** are the results under nonuniform density $(\ln(n))' \neq 0$. Initial population settings are $n_0 = 100$ for $(\ln(n))' \neq 0$, $p_{AABB} = 0.5$, $p_{AABb} = 0.3$, $p_{AAbb} = 0.03$, $p_{AaBB} = 0.1$, $p_{AaBb} = 0.04$, $p_{Aabb} = 0.01$, $p_{aaBB} = 0.01$, $p_{aaBb} = 0.01$, and $p_{aabb} = 0.0$. Other common parameters are the recombination rate = 0.01, gametic selection coefficient $s_h = 0.02$, zygotic selection coefficient $s_d = 0.04$, pollen dispersal variance = 0.05, and seed dispersal variance = 0.01.

Concerned with the spread of advantageous genes, previous studies have examined the effects of different barriers, including population subdivision (Slatkin 1976; Slatkin and Charlesworth 1978), physical obstacles, and random genetic drift effects (Pialek and Barton 1997; Barton 2000). Using the simulation approach, Slatkin and Charlesworth (1978) showed that the wave of advantageous alleles going to fixation can be quickly attained in a subdivided population, but stochastic effects reduce the wave speed. Pialek and Barton (1997) also confirmed the effects of random drift on delaying the spread of an advantageous allele in a separate model, as does the effect of selection against heterozygotes. The present study further expands our understanding of the role of mating systems in impeding the spread of advantageous genes. Although the interaction of selfing with physical barriers (Hu 2015) or with isolation by distance via reducing neighbor size (Wright 1946) is not examined in this study, this interaction could reinforce the impediment to gene flow.

A physical barrier may delay the spread of neutral genes (Barton 1979). With genetic hitchhiking effects, Slatkin and Charlesworth (1978) showed that the probability (p_i) of new alleles occurring in i populations is similar for detrimental and neutral genes under interaction of gene flow with selection and drift. When migration rate is smaller than selection strength, Slatkin and Wiehe (1998) showed that genetic hitchhiking enhances population differences at neutral loci that are closely linked to advantageous genes. Under similar condition, Kim and Maruki (2011) showed that population subdivision diminishes genetic hitchhiking

effects because the breakdown by recombination is enhanced in the case of restricted spread of advantageous genes. We here demonstrate new insights into the spread of neutral genes: (1) selfing impedes the spread of neutral genes; (2) variation in population density enhances neutral gene flow for a given mating system.

The difference between scenarios of uniform vs. non-uniform population density is that the latter has an additional contribution to gene flow due to unequal population sizes. In the scenario of uniform population density (Fisher 1937), i.e. $(\ln(n))' = 0$, each population is already established in space. Advantageous genes spread through seed and pollen dispersal and replace the preexisting disadvantageous alleles through natural selection in any local population. In the scenario of nonuniform density, i.e. $(\ln(n))' \neq 0$, spread of adaptive genes and species colonization in new habitats co-evolve because fitness at multiple loci is the genetic basis of population growth and adaptation to local environments. Asymmetric dispersal from dense to less dense populations ($(\ln(n))' > 0$) also tends to increase frequency of adaptive genes ($p_A' > 0$) or the frequency of linked neutral genes ($p_B' > 0$), or vice versa (i.e., if $(\ln(n))' < 0$, then $p_A' < 0$ and $p_B' < 0$) (Kirkpatrick and Barton 1997). High frequencies of advantageous genes enhance population growth. Thus, spread of adaptive alleles follows or is concordant with the expansion of species range (Hu et al. 2019). Thus, variation in population density produces a larger gradient for adaptive allele frequencies, which enhances gene spread, than does uniform population density.

With our theory, we assume that the mutant allele with advantage over the preexisting allele at a diallelic locus has attained a high frequency or become fixed before spreading out to neighboring populations (Fisher 1937). This is plausible for a large population where an adaptive mutant allele becomes extinct with a small probability (Wright 1969). For a small population, achievement of high frequencies of the advantageous allele is conditional on the ultimate fixation probability where the mutant allele is not lost in the genetic drift process (Haldane 1927; Kimura 1962; Barton 2000). Thus, the assumption that the copies of the advantageous allele exceed a certain level is needed so that the extinction of this allele is rare. Also, in the presence of the mixed mating system, selfing reduces the efficacy of selection (Glemin 2007) and hence could modify the fixation probability of the mutant allele in a population. This ultimate fixation probability could change with the selfing rate. Consequently, the waiting time for the mutant allele to reach fixation should be different between the systems with and without selfing besides the effects due to genetic drift.

Several implications can be derived from the present theory. The first implication concerns the role of selfing in advancing or impeding the spread of an advantageous gene in the distribution of a species. Selfing regulates the rate of spread of advantageous genes through two approaches. One is through its interaction with gametophytic selection. The antagonistic relationship between two stages of selection in one life cycle leads to their distinct interactions with selfing rate. This is an indirect form of regulation, like the role of selfing in regulating a species' range since only selective alleles are associated with population growth and adaption to habitats (Hu et al. 2019). The second effect is to regulate the proportion of migrant pollen that combines with ovules, which is a form of regulation and could be more efficient than the indirect way.

The two approaches regulate the speed of gene spread in complex ways. A nonlinear relationship exists between the speed and the selfing rate, as implied from Eq. (13). Given a constant rate of gene spread, a lower level of gene flow can effectively meet the requirement for the spread of alleles with a higher selective advantage than the existing alleles. Alternatively, a high level of gene flow is needed for the spread of an allele with a lower selective advantage than those already established. Thus, evolution of mating systems from outcrossing to selfing implicates that different rates of selfing are applied to regulate the rate of spreading adaptive genes in different species (Barrett 2014). This is crucial to maintain species integrity or cohesion since the spread of selective genes is more effective in preventing reproductive isolation between genetically different pools of populations (Slatkin 1976). This is different from neutral gene flow that is often estimated and applied to infer population history. If a certain level of exchange of genes

responsible for reproductive isolation is needed to maintain species integrity, the theory could provide a basis for explaining how a low level of such gene flow is sufficient for alleles to spread across the entire species (Morjan and Rieseberg 2004; Lowe and Allendorf 2010).

The second implication concerns a better understanding of the spread of neutral genes. Previous studies imply that a neutral gene can eventually spread across the whole species, regardless of the presence or absence of physical barriers (Barton 1979; Barton and Bengtsson 1986). Genetic hitchhiking can generate transient traveling waves or clines of neutral genes in both nuclear and cytonuclear systems (Slatkin and Wiehe 1998; Barton 2000; Hu 2008). The present theory shows that selfing reinforces the maintenance of the transient traveling waves of neutral genes although the pattern of the frontal waves is reduced as the genetic hitchhiking effects break down with time. It is predicted that transient waves of neutral genes could last longer than that under random mating where LD is generated by gene flow only. Such a difference could be more substantial for species with selfing or mixed mating systems than for species with random mating. Practical observations of multilocus waves or clines of gene frequencies among populations at both selective and neutral marker loci could likely signal the reinforcement of genetic hitchhiking from selfing.

The third result to emerge from our models is that they predict potential irregular waves of advance of advantageous genes under heterogeneous habitats (Shigesada and Kawasaki 1997) or under variation in mating systems. This could arise from variation in diffusion coefficients (e.g., environment-dependent diffusion for pollen $\sigma_p^2(x)$ or seeds $\sigma_s^2(x)$ dispersal) and/or selection coefficients (e.g., environment-dependent selection in gametophytic $s_h(x)$ or sporophytic $s_d(x)$ stage) under heterogeneous habitats. It could also arise from the influence of some factors (e.g., population density and environmental factors) that cause variation of mating systems among populations (Whitehead et al. 2018; Zhang et al. 2019). The selfing rate likely changes at different geographic positions (e.g., $\alpha(x)$), and hence could unequally affect the rate of gene spread in different populations. Nevertheless, the function of selfing should remain the same in interacting with selection and gene flow except that the wave becomes irregular under diverse selfing rates or the wave of gene advance stops in unfavorable environments.

The fourth implication of our models concerns the relationship between population density (n) and gene flow that has been investigated in empirical studies and exhibits mixed patterns. Although the direction of gene flow from less dense to dense populations is recorded under heterogeneous conditions (Kennington et al. 2003; Palstra et al. 2007), more reports are of asymmetrical gene flow from dense to less dense populations ($(\ln(n))' > 0$ and $p_A' > 0$).

This is supported by studies in both animal (Fraser et al. 2004; Manier and Arnold 2005) and plant species (Nagamitsu et al. 2014). The pattern is also conceptually related to an extreme case of classical mainland-island model or the source-sink model in metapopulation dynamics (Hanski and Gaggiotti 2004) where unidirectional gene flow takes place from mainland to island populations. A common phenomenon in tropical forest is the habitat fragmentation that creates many isolating patches with various population densities of a tree species. Although there are controversial issues about fragmentation and gene flow (Stephanie et al. 1992; White et al. 2002; Sebbenn 2011), the asymmetric population sizes could be an important factor, given a physical distance or presence/absence of physical barriers between fragments. Our theory implies that asymmetric population densities enhance gene flow among fragments, provided that long-distance dispersal is neglected. So far, this perspective has rarely been emphasized in the literature in explaining genetic structure and gene flow among fragments.

Data availability

The codes used for simulations have been submitted to Dryad (<https://doi.org/10.5061/dryad.83bk3j9p0>).

Acknowledgements We appreciate editors and three anonymous reviewers for very helpful comments that significantly improved the presentation of this article. We are grateful to Richard A. Ennos for proof reading this manuscript. This work is supported by forestry reform and development fund of China central finance, forestry science and technology promotion and demonstration (2018-GDTP-08), and funding from the South China Agricultural University (4400-K16013).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Barrett SCH (2014) Evolution of mating systems: outcrossing versus selfing. In: Losos J (ed.) The Princeton guide to evolution. Princeton University Press, Princeton, NJ, p 356–362
- Barton NH (1979) Gene flow past a cline. *Heredity* 43:333–340
- Barton NH (2000) Genetic hitchhiking. *Philos Trans R Soc Lond Ser B Biol Sci* 355:1553–1562
- Barton NH, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Heredity* 57:357–376
- Caballero A, Hill WG (1992) Effective size of non-random mating populations. *Genetics* 130:909–916
- Charlesworth D (2006) Evolution of plant breeding systems. *Curr Biol* 16:R726–R735
- Charlesworth D, Charlesworth B (1992) The effects of selection in the gametophytic stage on mutational load. *Evolution* 46:703–720
- Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. *Nat Rev Genet* 10:783–796
- Darwin C (1876) The effects of cross- and self-fertilization in vegetable kingdom. Murray, London
- Damgaard C, Guldbrandsen B, Christiansen FB (1994) Male gametophytic selection against a deleterious allele in a mixed mating model. *Heredity* 120:13–18
- Davis HG, Taylor CM, Lambrinos JG, Stong DR (2004) Pollen limitation causes an Allee effect in a wind-pollinated invasive grass (*Spartina alterniflora*). *Proc Natl Acad Sci USA* 101:13804–13807
- DuChateau P, Zachmann D (1989) Applied partial differential equations. Harper & Row Publisher, New York, NY
- Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* 72:250–259
- Fisher RA (1937) The wave of advance of advantageous genes. *Ann Eugen* 7:355–369
- Fisher RA (1941) Average excess and average effect of a gene substitution. *Ann Eugen* 11:53–63
- Fraser DJ, Lippe C, Bernatchez L (2004) Consequences of unequal population size, asymmetric gene flow and sex-biased dispersal on population structure in brook charr (*Salvelinus fontinalis*). *Mol Ecol* 13:67–80
- Glemin S (2007) Mating systems and the efficacy of selection at the molecular level. *Genetics* 177:905–916
- Groom MJ (1998) Allee effects limit population viability of an annual plant. *Am Nat* 151:487–496
- Grossenbacher DL, Runquist RB, Goldberg EE, Brandvain Y (2015) Geographic range size is predicted by plant mating system. *Ecol Lett* 18:706–713
- Haldane JBS (1927) A mathematical theory of natural and artificial selection. Part V: selection and mutation. *Math Proc Camb Philos Soc* 23:838–844
- Haldane JBS (1948) Theory of a cline. *J Genet* 28:227–284
- Hanski I, Gaggiotti O (2004) Ecology, genetics and evolution of metapopulations. Elsevier Academic Press, London
- Hu XS (2008) Barriers to the spread of neutral alleles in the cyto-nuclear system. *Evolution* 62:2260–2278
- Hu XS (2011) Mating system and the critical migration rate for swamping selection. *Genet Res* 93:233–254
- Hu XS (2015) Mating system as a barrier to gene flow. *Evolution* 69:1158–1177
- Hu XS, He FL (2005) Background selection and population differentiation. *J Theor Biol* 235:207–219
- Hu XS, Li B (2003) On migration load of seeds and pollen grains in a local population. *Heredity* 90:162–168
- Hu XS, Zhang XX, Zhou W, Hu Y, Wang X, Chen XY (2019) Mating system shifts a species' range. *Evolution* 73:158–174
- Igic B, Kohn JR (2006) The distribution of plant mating systems: study bias against obligately outcrossing species. *Evolution* 60:1098–1103
- Johnston MO, Porcher E, Cheptou PO, Eckert CG, Elle E, Geber MA et al. (2009) Correlations among fertility components can maintain mixed mating in plants. *Am Nat* 173:1–11
- Kennington WJ, Gockel J, Partridge L (2003) Testing for asymmetrical gene flow in a *Drosophila melanogaster* body-size cline. *Genetics* 165:667–673
- Kim Y, Maruki T (2011) Hitchhiking effect of a beneficial mutation spreading in a subdivided population. *Genetics* 189:213–226
- Kimura M (1953) "Stepping-stone" model of population. *Annu Rep Natl Inst Genet* 3:62–63
- Kimura M (1962) On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719
- Kirkpatrick M, Barton NH (1997) Evolution of a species' range. *Am Nat* 150:1–23
- Kolmogorov A, Petrovsky N, Picoonov NS (1937) A study of the equation of diffusion with increase in the quantity of matter, and

- its application to a biological problem. *Mosc Univ Bull Math* 1:1–25
- Kohn JR, Barrett SCH (1994) Pollen discounting and the spread of selfing variant in tristylous *Eichhornia paniculata*: evidence from experimental populations. *Evolution* 48:1576–1594
- Lowe WH, Allendorf FW (2010) What can genetics tell us about connectivity? *Mol Ecol* 19:30038–30051
- Manier MK, Arnold SJ (2005) Population genetic analysis identifies source-sink dynamics for two sympatric garter snake species (*Thamnophis elegans* and *Thamnophis sirtalis*). *Mol Ecol* 14:3965–3976
- Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Mol Ecol* 13:1341–1356
- Nagamitsu T, Kikuchi S, Hotta M, Kenta T, Hiura T (2014) Effects of population size, forest fragmentation, and urbanization on seed production and gene flow in an endangered maple (*Acer miyabei*). *Am Naturalist* 172:303–316
- Nagylaki T (1975) Conditions for existence of clines. *Genetics* 80:595–615
- Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154:923–929
- Nordborg M, Charlesworth B, Charlesworth D (1996a) The effect of recombination on background selection. *Genet Res* 67:159–174
- Nordborg M, Charlesworth B, Charlesworth D (1996b) Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proc Biol Sci* 263(1373):1033–1039
- Otto SP, Marks JC (1996) Mating systems and the evolutionary transition between haploidy and diploidy. *Biol J Linn Soc* 57:197–218
- Palstra FP, O'Connell MF, Ruzzante DE (2007) Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. *Mol Ecol* 16:4504–4522
- Peters MAE, Weis AE (2018) Selection for pollen competitive ability in mixed-mating systems. *Evolution* 72:2513–2536
- Pialek J, Barton NH (1997) The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. *Genetics* 145:493–504
- Pickup M, Brandvain Y, Fraise C, Yakimowski S, Barton NH, Dixit T et al. (2019) Mating system variation in hybrid zones: facilitation, barriers and asymmetries to gene flow. *N Phytol* 224:1035–1047
- Razanajatovo M, Maurel N, Dawson W, Essl F, Kreft H, Pergl J et al. (2016) Plants capable of selfing are more likely to become naturalized. *Nat Commun* 7:13313
- Roughgarden J (1979) Theory of population genetics and evolutionary ecology: an introduction. Macmillan Publishing Co., Inc, New York, NY
- Sebbenn A (2011) Effects of forest fragmentation on the effective and realized gene flow of Neotropical tree species: implications for genetic conservation. *BMC Proc* 5(7):1–2
- Shaw AJ, Szovenyi P, Shaw B (2011) Bryophyte diversity and evolution: windows into the early evolution of land plants. *Am J Bot* 98(2):352–369
- Shigesada N, Kawasaki K (1997) Biological Invasions: theory and Practice. Oxford University Press, Oxford
- Slatkin M (1976) The rate of spread of an advantageous allele in a subdivided population. In: Karlin S, Nevo E (eds) Population genetics and ecology. Academic Press, Inc, New York, NY, p 767–780
- Slatkin M, Charlesworth D (1978) The spatial distribution of transient alleles in a subdivided population: a simulation study. *Genetics* 89:793–810
- Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genet Res Camb* 71:155–160
- Stephanie AF, Hickey RJ, Vankat JL, Guttman SI, Schaefer RL (1992) Genetic structure after forest fragmentation: a landscape ecology perspective on *Acer saccharum*. *Can J Bot* 70:1659–1668
- Tanksley S, Zamir D, Rick CM (1981) Evidence for extensive overlap of sporophytic and gametophytic gene expression in *Lycopersicon esculentum*. *Science* 213:453–455
- Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates, Inc Publishers, Sunderland, MA
- White GM, Boshier DH, Powell W (2002) Increased pollen flow counteracts fragmentation in a tropical dry forest: an example from *Swietenia humilis Zuccarini*. *Proc Natl Acad Sci USA* 99:2038–2042
- Whitehead MR, Lanfear R, Mitchell RJ, Karron JD (2018) Plant mating systems often vary widely among populations. *Front Ecol Evolution* 6:38
- Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics* 31:39–59
- Wright S (1969) Evolution and the genetics of populations. The theory of gene frequencies, vol. 2. University Chicago Press, Chicago
- Zhang XX, Wang X, Hu Y, Zhou W, Chen XY, Hu XS (2019) Advances in the study of population genetic diversity at plant species' margins. *Chin J Plant Ecol* 43(5):383–395