Spatial gene drives and pushed genetic waves

Hidenori Tanaka\textsuperscript{a,b,1}, Howard A. Stone\textsuperscript{c}, and David R. Nelson\textsuperscript{d,e,1}

\textsuperscript{a}School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; \textsuperscript{b}Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, MA 02138; \textsuperscript{c}Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544; \textsuperscript{d}Department of Physics, Harvard University, Cambridge, MA 02138; and \textsuperscript{e}Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138

Edited by Andrea J. Liu, University of Pennsylvania, Philadelphia, PA, and approved June 20, 2017 (received for review April 9, 2017)

Gene drives have the potential to rapidly replace a harmful wild-type allele with a gene drive allele engineered to have desired functionalities. However, an accidental or premature release of a gene drive construct to the natural environment could damage an ecosystem irreversibly. Thus, it is important to understand the spatiotemporal consequences of the super-Mendelian population genetics before potential applications. Here, we use a reaction–diffusion model for sexually reproducing diploid organisms to study how a locally introduced gene drive allele spreads to replace the wild-type allele, although it possesses a selective disadvantage \(s > 0\). Using methods developed by Barton and collaborators, we show that socially responsible gene drives require \(0.5 < s < 0.697\), a rather narrow range. In this “pushed wave” regime, the spatial spreading of gene drives will be initiated only when the initial frequency distribution is above a threshold profile called “critical propagule,” which acts as a safeguard against accidental release. We also study how the spatial spread of the pushed wave can be stopped by making gene drives uniquely vulnerable (“sensitizing drive”) in a way that is harmless for a wild-type allele. Finally, we show that appropriately sensitized drives in two dimensions can be stopped, even by imperfect barriers perforated by a series of gaps.

Significance

Gene constructs introduced into natural environments have been proposed to solve various ecological problems. The CRISPR-Cas9 technology greatly facilitates construction of gene drives that allow desired traits to rapidly replace wild types, even if these convey a selective growth rate disadvantage \(s > 0\). However, accidental release of a gene drive could damage ecosystems irreversibly. We have modeled the spatial spread of gene drives and find a preferred range of selective disadvantages, \(0.5 < s < 0.697\). In this regime, gene drives spread but only when a nucleus exceeds a critical size and intensity. By making gene drives uniquely susceptible to a compound, their advance can be stopped in two dimensions by finite-width barriers, even when interrupted by gaps.

Author contributions: H.T., H.A.S., and D.R.N. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper. The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

\textsuperscript{1}To whom correspondence may be addressed. Email: tanaka@g.harvard.edu or drnelson@fas.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/doi/10.1073/pnas.1705868114/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1705868114
and diffusive dispersal at the leading edge, and they are difficult to slow down and stop.

However, for somewhat larger selective disadvantages (0.5 < s < 0.097), we find that propagation proceeds instead via a pushed genetic wave (33–35), where the genetic wave advances via accented growth from populations somewhere behind the front that spill over the leading edge. These waves, characterized by a strong Allee effect (36, 37), are more socially responsible than the pulled Fisher waves, because (i) only inoculations with spatial size and density that exceed a critical nucleus or “critical propagule” (28) are able to spread spatially, thus providing protection against a premature or accidental release of a gene drive; (ii) the gene drive pushed waves can be stopped by making them uniquely vulnerable to a specific compound (“sensitizing drive” (9)), which is harmless for a WT allele; and (iii) appropriately sensitized gene drives can be stopped even by barriers punctuated by defects, analogous to regularly spaced fire breaks used to contain forest fires. Similar pushed or “excitable” waves also arise, for example, in neuroscience in simplified versions of the Hodgkin–Huxley model of action potentials (38). When the selective disadvantage associated with the gene drive is too large (s > 0.697 in our model), the excitable wave reverses direction, and the region occupied by the gene drive homozygotes collapses to zero.

The same mathematical analyses apply to spatial evolutionary games of two competing species in one dimension, which are governed by a class of reaction–diffusion equations that resemble the gene drive system. The fitness levels of the two interacting red and green species (wR, wG) are related to their frequencies [f(x, t), 1 − f(x, t)] by wR(x, t) = g + α[1 − f(x, t)], wG(x, t) = q + βf(x, t), where g is a background fitness, assumed identical for the two alleles for simplicity. The mutualistic regime α > 0, β > 0 in the first quadrant of Fig. 2 has been studied already (40), including the effect of genetic drift, with two lines of directed “percolation” transitions out of a mutualistic phase. Here, we apply the methods from ref. 28 to study the evolutionary dynamics near the line of first-order transitions that characterize the competitive exclusion regime in the third quadrant of Fig. 2. Because the mathematics parallels the analysis inspired by gene drive systems in the text, we relegate discussion of this topic to SI Appendix, which also has discussion of conversion efficiencies c < 1, an analogy with nucleation theory, and other matters.

Mathematical Model of the CRISPR Gene Drives

We start with a Hardy–Weinberg model (42) and incorporate an MCR with 100% conversion rate to construct a model for a well-mixed system. This model is the limiting case of “c = 1” in the work by Unckless et al. (29). Conversion efficiencies c < 1 can be handled by similar techniques. We consider a well-mixed diploid system with a WT allele a and a gene drive allele A with frequencies p = p(t) and q = q(t), respectively, at time t, with p(t) + q(t) = 1. Within a random mating model, the allele frequencies after one generation time τg are given by

\[
(pa + qA)^2 = p^2(a, A) + 2pq(a, A) + q^2(A, A),
\]

and the ratios of fertilized eggs with diploid types (a, a), (a, A), and (A, A) are \(p^2 : 2pq : q^2\). In a heterozygous (a, A) egg, the CRISPR-Cas9 machinery encoded on a gene drive allele A converts the WT allele \(a\) into a gene drive allele \(A\). Here, we assume a perfect conversion rate \((a, A) \leftrightarrow (A, A)\) in the embryo, which has been approximated already for yeast (19) and fruit flies (20). Genetic engineering will typically reduce the fitness of individuals carrying the gene drive alleles compared with WT organisms, which have already gone through natural evolution and may be near a fitness maximum.

The selective disadvantage of a gene drive allele s is defined by the ratio of the fitness \(w_{\text{wild}}\) of WT organisms (\(a, a\)) to the fitness \(w_{\text{drive}}\) of (\(A, A\)) individuals carrying the gene drive:

\[
\frac{w_{\text{drive}}}{w_{\text{wild}}} \equiv 1 - s, 0 \leq s.
\]

[In the limit c → 1, no heterozygous (\(a, A\)) individuals are born (29).] Taking the fitness into account, the allele frequencies after one generation time \(\tau_g\) are

\[
p' : q' = w_{\text{wild}}p^2 : w_{\text{wild}}(1 - s)(q^2 + 2pq),
\]

where \(p' \equiv p(t + \tau_g)\) and \(q' \equiv q(t + \tau_g)\). On approximating \(q' - q = q(t + \tau_g) - q(t)\) by \(\tau_g dq/dt\), we obtain a differential equation
which governs population dynamics of the MCR with 100% conversion efficiency in a well-mixed system. To take spatial dynamics into account, we add a diffusion term (28) and obtain a deterministic reaction–diffusion equation for the MCR model, namely

\[
\tau_g \frac{dq}{dt} = \frac{(1-s)(q^2 + 2pq)}{p^2 + (1-s)(q^2 + 2pq)} - q
\]

\[
= \frac{sq(1-q)(q - q^*)}{1 - sq(2-q)}, \quad \text{where} \quad q^* = \frac{2s - 1}{s},
\]

which will be the main focus of this article. For later discussions, we name the reaction term of the reaction–diffusion equation

\[
f_{\text{MCR}}(q, s) = \frac{sq(1-q)(q - q^*)}{1 - sq(2-q)}.
\]

The reaction term reduces to a simpler cubic expression

\[
f_{\text{cubic}}(q, s) = sq(1-q)(q - q^*)
\]

by ignoring \(-sq(2-q)\) in the denominator, which is a reasonable approximation if the selective disadvantage \(s\) is small. This form of the reaction–diffusion equation has been well-studied, as reviewed in ref. 28.

Although population genetics is often studied in the limit of small \(s\), \(s\) is, in fact, fairly large in the regime of pushed excitable waves of most interest to us here, \(0.5 < s < 1.0\). Hence, we will keep the denominator of the reaction term, which was also done in ref. 28 with a different reaction term. Comparison of results for the full nonlinear reaction term with those for the cubic approximation will give us a sense of the robustness of the cubic approximation. Although it might also be of interest to study corrections to the continuous time approximation arising from higher-order time derivatives in \((q - q^*)/\tau_g = \partial q/\partial t + 1/2\tau_g q^2/\partial t^2 + \ldots\) \(q\) \((\tau_g^2q/\partial t^2\) are formally of order \(s^2\)), this complicated problem will be neglected here; however, ref. 43 is a study of the robustness of the continuous time approximation motivated by a model of dengue suppressing Wolbachia in mosquitoes.

**Initiation of the Pushed Waves**

The reaction terms \(f_{\text{MCR}}(q, s)\) and \(f_{\text{cubic}}(q, s)\) have three identical fixed points, \(q = 0, 1\) and \(q^* = (2s - 1)/s\). As discussed in SI Appendix in connection with classical nucleation theory in physics and following ref. 26, we can define the potential energy function

\[
U(q) = -\frac{1}{\tau_g} \int_0^q \frac{sq'(1-q')(q' - q^*)}{1 - sq'(2-q')} dq'
\]

to identify qualitatively different parameter regimes. In a well-mixed system, without spatial structure, the gene drive frequency \(q(t)\) obeys Eq. 4 and evolves in time, so that it arrives at a local minimum of \(U(q)\). For the spatial model of interest here, \(q(x, t)\) shows qualitatively distinct behaviors in three parameter regimes depending on the selective disadvantage \(s\) (Fig. 3A). We plot the potential energy functions \(U(q)\) in these parameter regimes in Fig. 3B.

\(i\) First, when \(s < s_{\text{min}} = 0.5\), fixation of a gene drive allele \(q(x) = 1\) for all \(x\) is the unique stable state, and there is no energy barrier to reach the ground state starting from \(q = 0\). In this regime, any finite frequency of gene drive allele locally introduced in space (provided it overcomes genetic drift) will spread and replace the WT allele. The frequency profile will evolve as a pulled traveling wave \(q(x, t) = Q(x - vt)\) with wave velocity \(v\). Such a wave was first found by Fisher (31) and Kolmogorov et al. (32) in the 1930s in studies of how locally introduced organisms with advantageous genes spatially spread and replace inferior genes. However, the thresholdless initiation of population waves of engineered gene drives with relatively small selective disadvantages seems highly undesirable, because the accidental escape of a single gene drive construct can establish a population wave that spreads freely into the extended environment.

\(ii\) There is a second regime for \(0.5 < s < 0.697\), in which the potential energy function \(U(q)\) exhibits an energy barrier between \(q = 0\) and \(q = 1\). In this regime, a pushed traveling wave can be excited only when a threshold gene drive allele frequency is introduced over a sufficiently broad region of

![Fig. 3](image-url)

(A) Spatial dynamics of gene drives can be determined by both the selective disadvantage \(s\) and (when \(0.5 < s < 0.697\)), the size and intensity of the initial condition. (B) The energy landscapes \(U(q)\) with various selective disadvantages \(s\). (i) Pulled Fisher wave regime. When \(s\) is small, \(s \leq s_{\text{min}} = 0.5\) (red and yellow curves), fixation of the gene drive allele \((q = 1)\) is the unique stable state, and there is no energy barrier between \(q = 0\) and \(1\). Any finite introduction of a gene drive allele is sufficient to initiate a pulled Fisher population wave that spreads through space to saturate the system. (ii) Pushed excitable wave regime. When \(s\) is slightly larger (green curve) and satisfies \(s_{\text{min}} = 0.5 < s < 0.697\), \(q = 1\) is still the preferred stable state, but an energy barrier at \(q = q^*\) appears between \(q = 0\) and \(1\). In this regime, the introduction of the gene drive allele at sufficient concentration and over a sufficiently large spatial extent is required for a pushed wave to spread to global fixation. (iii) Wave reverses direction. When \(s\) is large, \(s > s_{\text{min}} = 0.697\) (blue and purple curves), \(q = 0\) is the unique ground state, and the gene drive species cannot establish a traveling population wave and therefore dies out.
space that exceeds the size of a critical nucleus, which we investigate in the next section. The existence of this threshold acts as a safeguard against accidental release. In addition, such excitable waves are easier to stop, which we will discuss later. It seems that gene drives in this relatively narrow intermediate regime are the most desirable from a biosafety perspective.

iii) When \( s > s_{\text{max}} = 0.697 \), the fixation of a gene drive allele throughout space is no longer absolutely stable (Fig. 3B), and a gene drive population wave cannot be established. Indeed, the excitable wave reverses direction for \( s > s_{\text{max}} \). An implicit equation for \( s_{\text{max}} \) results from equating \( U(0) = U(1) = 0 \), which yields

\[
0 = \int_0^1 \frac{sq(1-q)(q-q^*)}{1-sq(2-q)} dq,
\]

or

\[
0 = -2 + s_{\text{max}} + 2 \sqrt{1 + \frac{1}{s_{\text{max}}} \arcsin(\sqrt{s_{\text{max}}})}
\]

\[\Rightarrow s_{\text{max}} \approx 0.697,\]

where we used \( q^* = (2s-1)/s \). When \( s > s_{\text{max}} \), the locally introduced gene drive allele contracts rather than expands relative to the WT allele and simply dies out. SI Appendix has the analogous results with an arbitrary conversion rate \( 0 < c < 1 \).

**Critical Nucleus in the Pushed Wave Regime**

When the selective disadvantage \( s \) is in the intermediate regime, \( s_{\text{min}} = 1/2 < s < s_{\text{max}} = 0.697 \), we can control initiation of the pushed excitable wave by the initial frequency profile of the gene drive allele \( q(x,0) \) just sufficient to excite a pushed genetic wave in 1D (critical propagule). Numerically calculated critical propagules for the MCR model of Eq. 5 (solid lines) are compared with analytical results available for the cubic model of Eq. 7 (dashed lines) (28). When \( s = 0.51 \), the two equations give almost identical results, but as \( s \) increases, the critical propagule shape of the MCR model deviates significantly from that of the cubic model. The critical propagule of the cubic equation consistently overestimates the height of \( q(x,0) \), because the \( sq(2-q) > 0 \) term in the denominator of the MCR model always increases the growth rate.

Fig. 5. Initial critical frequency profiles of the MCR allele \( q_i(x) \) just sufficient to excite a pushed genetic wave in 1D (critical propagule). Numerically calculated critical propagules for the MCR model of Eq. 5 (solid lines) are compared with analytical results available for the cubic model of Eq. 7 (dashed lines) (28). When \( s = 0.51 \), the two equations give almost identical results, but as \( s \) increases, the critical propagule shape of the MCR model deviates significantly from that of the cubic model. The critical propagule of the cubic equation consistently overestimates the height of \( q(x,0) \), because the \( sq(2-q) > 0 \) term in the denominator of the MCR model always increases the growth rate.

We show the spatial concentration profile \( q(x,t) \) that constitutes that (Gaussian) critical nucleus is just sufficient to initiate an excitable wave in Fig. 5. The solid lines in Fig. 5 represent numerically obtained critical nuclei of the MCR model. Note the consistency for \( s = 0.58 \) with the pushed excitable waves shown in Fig. 4. The dashed lines in Fig. 5 represent analytically derived critical propagules of the cubic model as a reference (details are in SI Appendix). Fig. 5 shows that the cubic model overestimates the height of critical propagule, particularly for larger \( s \). The difference between the reaction terms of the MCR model \( f_{\text{MCR}}(q) \) (Eq. 6) and those of its cubic approximation \( f_{\text{cubic}}(q) \) (Eq. 7) arises from the term \(-sq(2-q)\) in the denominator of Eq. 5. In the biologically relevant regime \( 0 < s < 1 \), \( sq(2-q) \) is always positive, and \( f_{\text{MCR}}(q) > f_{\text{cubic}}(q) \) is satisfied, which explains why there is a larger critical propagule in the cubic approximation and why the discrepancy is larger for larger \( s \). The critical nucleus with a step function-like circular boundary is studied both numerically and analytically in two dimensions in SI Appendix.

**Stopping of Pushed, Excitable Waves by a Selective Disadvantage Barrier**

Thus far, we have found that (i) we can control initiation of the spatial spread of a gene drive provided \( s_{\text{min}} = 0.5 < s < s_{\text{max}} = 0.697 \) and that (ii) the pushed population waves in this...
Numerical simulations of pushed, excitable waves generated by Eq. 5 in one dimension with a barrier of strength \( s_0 = 0.958 \) placed in a region \( 25 \sqrt{\tau_D} < x < 27 \sqrt{\tau_D} \). When the selective disadvantage outside the barrier is small (\( s < 0.5 \)) and the population wave travels as the pulled Fisher wave, even a tiny fraction of the MCR allele diffusing through the insecticide region can easily reestablish the population wave, which is shown in Fig. 6a. However, when the system is in the pushed wave regime \( 0.5 < s < 0.697 \), the wave can be stopped provided that the spatial profile of the gene drive allele that leaks through does not constitute a critical nucleus, which is illustrated in Fig. 6b. SI Appendix has numerically calculated plots of the critical width and barrier selective disadvantage needed to stop pushed waves for various values of \( s \).

**Fig. 6.** Numerical simulations of pushed, excitable waves generated by Eq. 5 with barriers in one dimension, with time increments \( \Delta t = 0.07g \). As the waves advance from left to right, the early time response is shown in red, with later times shown in blue. The fitness disadvantage inside the barrier is set to \( s_0 = 0.958 \) within a region \( 25 \sqrt{\tau_D} < x < 27 \sqrt{\tau_D} \) (shown as purple bars). The initial conditions are step function-like, \( q(x, 0) = q_0/(1 + e^{10(x-x_0)}/\sqrt{\tau_D}) \), with \( q_0 = 1.0 \) and \( x_0 = 5.0 \sqrt{\tau_D} \), similar to the initial condition SI Appendix, Eq. S30 that we used in two dimensions (SI Appendix). (A) In the case of a Fisher wave with \( s = 0.479 \) \( < s_{\text{max}} = 0.5 \), a small number of individuals diffuse through the barrier, which is sufficient to reestablish a robust traveling wave. (B) In the case of the excitable wave \( s = 0.542 > s_{\text{max}} = 0.5 \), a small number of individuals also diffuse through the barrier. However, because the tail of the penetrating wave front is insufficient to create a critical nucleus, the barrier causes the excitable wave to die out.

**Fig. 7.** Population waves impeded by a selective disadvantage barrier of strength \( s_0 = 1.0 \) (purple) with a gap. This imperfect barrier has a region without insecticide in the middle of width \( 6 \sqrt{\tau_D} \). (A) The pulled Fisher wave with \( s = 0.48 < 0.5 \) always leaks through the gap and reestablishes the gene drive wave (red and yellow). (B) The pushed wave that arises when \( s = 0.62 > 0.5 \) is deexcited by a gapped barrier, provided that the gap width is comparable to or smaller than the width of the gene drive wave.

**Excitable Wave Dynamics with Gapped Barriers in Two Dimensions**

In the previous section, we showed that pushed excitable waves can be stopped by a selective disadvantage barrier in one dimension. However, in two dimensions, it may be difficult to make barriers without defects. Hence, we have also studied the effect of a gap in a 2D selective disadvantage barrier. We find that, although the gene drive population wave in the Fisher wave regime \( s < 0.5 \) always leaks through the gaps, the excitable wave with \( 0.5 < s < 0.697 \) can be stopped, provided that the gap is comparable with or smaller than the width of the traveling wave front. In Fig. 7, we illustrate the gene drive dynamics for two different parameter choices. In Fig. 7, the strength of the selective disadvantage barrier is set to be \( s_0 = 1.0 \), and the width of the gap in the barrier is set to be \( 6 \sqrt{\tau_D} \). The engineered selective disadvantage in the nonbarrier region \( s \) differs in the two plots. In Fig. 7A, \( s = 0.48 < 0.5 \); therefore, the gene drive wave propagates as a pulled Fisher wave, and the wave easily leaks through the gap. If genetic drift can be neglected, we expect that Fisher wave excitations will leak through any gap, regardless of how small. However, when the selective disadvantage barrier is in the pushed wave regime \( 0.5 < s < 0.697 \), the population wave can be stopped by a gapped selective disadvantage barrier as shown in Fig. 7B. To stop a pushed excitable wave, the gap dimensions must be smaller than the front width; alternatively, we can say that the gap must be smaller than size of the critical nucleus.

**Discussion**

The CRISPR-Cas9 system has greatly expanded the design space for genome editing and construction of MCRs with non-Mendelian inheritance. We analyzed the spatial spreading of gene drive constructs, applying reaction–diffusion formulations that have been developed to understand spatial genetic waves with bistable dynamics (26–28). For a continuous time and space version of the model by Unckless et al. (29), in the limit of 100% conversion efficiency, we found that a critical nucleus or propagule is required to establish a gene drive population wave when the selective disadvantage satisfies \( 0.5 < s < 0.697 \). This range is even narrower when we relax the assumption of 100% conversion efficiency, so that \( c < 1 \) (SI Appendix). Our model led us to study termination of pushed gene drive waves using a barrier that acts only on gene drive homozygotes, corresponding to an insecticide in the case of mosquitoes. In this parameter...
regime, the properties of pushed waves allow safeguards against the accidental release and spreading of the gene drives. One can, in effect, construct switches that initiate and terminate the gene drive wave. In the future, it would be interesting to study the stochasticity caused by finite population size (genetic drift), which is known to play a role in the first quadrant in Fig. 2 (40, 41). We expect that genetic drift can be neglected provided that \( N_e \gg 1 \), where \( N_e \) is an effective population size, say, the number of organisms in a well-mixed critical propagule. SI Appendix has a brief discussion of genetic drift. It could also be important to study the effect of additional mutations on an excitable gene drive wave, particularly those that move the organism outside the preferred range \( 0.5 < s < 0.607 \). Finally, we address possible experimental tests of the theoretical predictions. Because it seems inadvisable to conduct field tests without thorough understanding of the system, laboratory experiments with microbes would be a good starting point. Recently, the transition from pulled to pushed waves was qualitatively investigated with haploid microbial populations (35). Because the MCR has already been realized in Saccharomyces cerevisiae (19), it may also be possible to test the theory in the context of range expansions on a petri dish, which has already been done for haploid mutational yeast strains in ref. 44. Here, the frontier approximates a 1D stepping stone model, and jostling of daughter cells at the frontier leads to an effective diffusion constant in one dimension (45, 46). Finally, as illustrated in SI Appendix, Fig. S2, the mathematics of the spatial evolutionary games in one dimension parallels the dynamics of diploid gene drives in the pushed wave regime, providing another arena for experimental tests, including the effects of genetic drift.

Materials and Methods

To simulate the dynamics governed by Eq. 5 in Figs. 4, 6, and 7 and SI Appendix, Fig. S6, we used the method of lines and discretized spatial variables to map the partial differential equation to a system of coupled ordinary equations (ODEs). Then, we solved the coupled ODEs with a standard ODE solver. The widths of the spatial grids were varied from \( 1/200 < 1/\sqrt{D} < 1/20 \), always making sure that the mesh size was much smaller than the width of the fronts of the pushed and pulled genetic waves that we studied.

ACKNOWLEDGMENTS. We thank N. Barton, S. Block, S. Sawyer, T. Stearns, and M. Turelli for helpful discussions and two anonymous reviewers for useful suggestions. N. Barton also provided a critical reading of our manuscript. V.W. T. T. and D. R. N. was supported by National Science Foundation Grant DMR1608501 and Harvard Materials Science Research and Engineer- ing Center Grant DMR1435999. H. A. S. acknowledges support from National Science Foundation Grants MCB1344191 and DMS1614907.