
Article:
Shafiey, H., Gossmann, T.I. and Waxman, D. (2017) Evolutionary control: Targeted change of allele frequencies in natural populations using externally directed evolution. Journal of Theoretical Biology, 419. pp. 362-374. ISSN 0022-5193

https://doi.org/10.1016/j.jtbi.2017.01.023

www.elsevier.com/locate/yjtbi

Evolutionary control: Targeted change of allele frequencies in natural populations using externally directed evolution

Hassan Shafiey, Toni I. Gossmann, David Waxman

PII: S0022-5193(17)30023-1
DOI: http://dx.doi.org/10.1016/j.jtbi.2017.01.023
Reference: YJTBI8933

To appear in: *Journal of Theoretical Biology*

Received date: 23 May 2016
Revised date: 13 November 2016
Accepted date: 16 January 2017

Cite this article as: Hassan Shafiey, Toni I. Gossmann and David Waxman, Evolutionary control: Targeted change of allele frequencies in natural populations using externally directed evolution, *Journal of Theoretical Biology*, http://dx.doi.org/10.1016/j.jtbi.2017.01.023

17/1/2017

Evolutionary control: Targeted change of allele frequencies in natural populations using externally directed evolution

Hassan Shafiey[1], Toni I. Gossmann[2] and David Waxman[1]

[1]Centre for Computational Systems Biology,

Fudan University, Shanghai 20433, PRC.

[2]Department of Animal and Plant Sciences,

University of Sheffield, Sheffield S102TN, UK

Running Title: Evolutionary control by externally directed evolution

Correspondence to: Professor D. Waxman, Centre for Computational Systems Biology, Fudan University, 220 Handan Road, Shanghai 200433, People's Republic of China.

E-mail: davidwaxman@fudan.edu.cn

Keywords: evolutionary theory, random genetic drift, Markov chain, diffusion approximation, genetic control

ABSTRACT

Random processes, in particular random genetic drift, often make it difficult to predict the fate of a particular mutation in a biological population. In this work we show, using principles of theoretical population genetics, a form of biological control that ensures that a focal allele's frequency, at a given locus, achieves a prescribed probability distribution at a given time. This control is in the form of an additional evolutionary force that acts on a population. We provide the mathematical framework that determines the additional force. Our analysis indicates that, generally, the additional force depends on the frequency of the focal allele, and may also depend on the time. We argue that translating this additional force into an externally controlled process, which has the possibility of being implemented in a number of different ways, corresponding to selection, migration, mutation, or a combination of these, may provide a flexible instrument for targeted change of traits of interest in natural populations. This framework may be applied, or used as an informed form of guidance, in a variety of different biological scenarios including: yield and pesticide optimisation in crop production, biofermentation, the local regulation of human-associated natural populations, such as parasitic animals, or bacterial communities in hospitals.

1 Introduction

Since the beginning of the 20th century, it has become clear that controlled intervention in many biological phenomena is crucial to avoid or reduce costs that are potentially substantial. Examples of phenomena, with large consequences for economics and health, include pest-induced reduction of crop yields and insect-borne human diseases (McFadyen 1998; Kamareddine 2012). While such interventions may be described as 'biological control', we note that in classical terms this phrase means "the use of all natural organic checks, bacterial and fungous diseases as well as parasitic and predaceous insects" to deal with insect pests (Smith 1919). Here we adopt a broader viewpoint of biological control, as *any* steering influence of a biological population. This then makes the subject particularly relevant in the era of large genetic data, which arises, for example, from next-generation-sequencing (Buermans and den Dunnen 2014). Such genetic information, combined with phenotypic information, may be exploited for the biological control of natural populations. As an example, genomic selection uses markers that are in linkage with particular traits, and these can ultimately be used to predict yield gains or to increase breeding values of successive generations (de Los Campos et al. 2013; Daetwyler et al. 2013). This method performs better than previous methods, even when traits are controlled by many loci of

small effect (Spindel et al. 2015a). However, usually, the method can only be applied over short time-scales and it shows little consistency between different populations with limited relatedness (Daetwyler et al. 2013).

Application of biological control theory has been driven by straightforward practical reasoning, e.g., the extinction of a parasite. This, however, may either be difficult or impossible to achieve. For example, an excessive use of pathogens to control the insects pests of crops and forests may be harmful to the environment (Lacey et al. 2015), and may also lead to resistance in the pest population against the pathogen or pesticides in general (Owen and Zelaya 2005). Moreover, using natural enemies typically involves perturbing a complex network of interacting biological species, making it difficult to predict long-term outcomes and damaging ecological side-effects (Reilly and Elderd 2014). Consequently, when it is necessary or desirable to steer a biological population in a predictable manner, the external forces which are applied, e.g., the amount of pathogen/pesticide used on a population, need to be effectively controlled (for this case, minimised). We argue that taking genetic information, in a population-genetics framework, into account, and intervening in a population's natural behaviour, may overcome or ameliorate these issues.

In this work we provide a mathematical framework for an additional

evolutionary force (*additional force*, for short) which, when imposed on a natural population of finite size, controls the population in a predictable manner, to the extent that a specified *final probability distribution* of the frequency of a focal allele is obtained.

The additional force of this work is given in terms of quantities that describe the original population (where only pre-existing forces act). This form of the additional force is not unique: other additional forces can also can also produce the final distribution. Thus what we present is a particular solution for the additional force that turns out to have the simplest form. We devote some of the Discussion to the issue of the uniqueness of the additional force, and ways this feature may be usefully exploited.

The additional evolutionary force always acts in concert with other (pre-existing) forces. In finite populations, random genetic drift is always a pre-existing force, and other forces may include selection, mutation, migration, and although we do not consider it here, also recombination. In spite of the final (specified) distribution being the joint outcome of multiple forces, we will adopt the convenient shorthand of simply saying that the additional force *produces* (or is chosen to produce) the final distribution.

The additional force introduced in this work has a characteristic feature that it generally depends on the frequency of the focal allele. The additional

force may also depend on the time that it acts. Under some circumstances, the additional force may be viewed as a form of additional selection that acts on the population; such additional selection will generally be frequency dependent.

One possible application of the additional force is to a number of population *lines*[1] where a focal allele at a biallelic locus has, in the simplest case, the same initial frequency in all lines. An intervention, in the form of an additional evolutionary force, can lead to the production of a specific probability distribution of the allele frequency, over all lines, at a specified final time[2].

Pursuing this idea, we note that the classic experiment of Buri (1956), as reanalysed by Hartl and Clark (1977, Chapter 7), illustrated the action of random genetic drift over a number of population lines which were followed over time. In Figure 1 we have presented related simulation results that illustrate the action of genetic drift on a neutral focal allele in a haploid Wright-Fisher population. For all three panels of Figure 1, a focal allele started with a frequency of 0.5 in a large number of lines that each main-

[1] A set of lines is a number of identical (or near identical) biological subpopulations of the same size, that have no migration between them. They may exhibit different allele frequencies due to the operation of random genetic drift.

[2] Given a finite number of lines, the final frequency distribution of the focal allele, achieved over all lines, will be a finite sample from the prescribed final distribution. As a consequence, the final distribution will not precisely coincide with the prescribed distribution.

Figure 1A

tained the finite number of 100 adults in every generation. The distribution of the frequency of the focal allele has been plotted over 60 generations. Figure 1A contains a plot of the distribution in the absence of any additional force, while Figures 1B and 1C show results when two different forms for the additional force were acting. In particular, in Figure 1B the additional force was chosen so that fixation of the focal allele occurs by the final time (generation 60), while in Figure 1C the additional force was chosen so that the final distribution corresponds to the occurrence of fixation with probability 0.7, and the occurrence of loss with probability of 0.3. The additional forces required for both Figure 1B and Figure 1C both depend on the frequency of the focal allele and the time.

Figure 1B

The additional evolutionary forces that produced the final distributions in Figures 1B and 1C are special cases of a general result, we present later, for the additional force. In this general result, the frequency of the focal allele does not start from a definite value, but from an initial probability distribution, and the additional force ensures production of a given final distribution. The general result applies for an ensemble of populations that, collectively, have the appropriate initial and final distributions. An example of such an ensemble is a set of population lines/subpopulations described above.

The additional evolutionary force can also be applied to a single population (not, e.g., to a set of lines). If the final distribution of the focal allele

Figure 1C

Figure 1 Caption: This figure shows the distribution of the allele frequency over time of uncontrolled and controlled populations. The figure is based on simulations of a neutral biallelic haploid population in a Wright-Fisher model. The parameter-values were arbitrarily chosen as follows. The simulations involved 10^5 lines (or replicate populations). Each line started in generation 0 with a focal allele at a frequency of 0.5. Number regulation was implemented, so that in every generation, 100 adults were maintained in each line. In Panel A, the probability distribution of the focal allele's frequency is plotted over 60 generations when there was no additional evolutionary force. Panel B shows what occurs under the action of an additional evolutionary force that was chosen to ensure fixation occurs by generation 60. Panel C shows what occurs under the action of an additional evolutionary force that was chosen so that the frequency does not have the probability distribution of a neutral population. Rather, the additional force was chosen so that by generation 60 there is a probability of 0.7 of the allele achieving fixation, a probability of 0.3 of the allele achieving loss, and zero probability of the allele segregating.

involves more than a single allele frequency, then a single population cannot, in any sense, achieve the final distribution (in a single population there is only a single value of the frequency at any time, including the final time)[3]. The interpretation is that while the additional force does not guarantee a particular outcome, its application could significantly bias matters, to the extent that some outcomes could become much more probable than would be the case in the absence of the additional force. Thus, in this case, the additional force can be viewed as modifying the *probabilities* of final outcomes. As an example, application of an additional force, to a single population, can change the probability of fixation from its value in the absence of the additional force (its 'natural' value), to a quite different (prescribed) value.

We shall consider examples of the additional force acting in plant populations, but also discuss other examples where the framework of the present work may be applied. These include parasitic animals that are immune to pesticides, bacterial communities in hospitals that exhibit antibiotic resistance and to personalised/precision medicine. Generally, we argue that the CRISPR/Cas9 system (Gantz and Bier 2015) provides the ultimate tool to conduct evolutionary control on any population of interest, and for a recent analysis of the effects of the mutagenic chain reaction associated with this,

[3]In the Discussion, we address the question of a single population which has uncertainty in the value of the initial frequency.

see the paper by Unckless et al. (2015).

2 Theoretical background

Prior to presenting the results of this work, we give some necessary theoretical background.

Consider a focal allele at a biallelic locus where, in what follows, we shall often refer to the frequency of this allele as just the *frequency*.

With $X(t)$ the frequency at time t, we assume that in an effectively infinite population, the frequency obeys[4]

$$\frac{dX}{dt} = F_0(X). \tag{1}$$

In this equation we have suppressed the time dependence of X so $X \equiv X(t)$, and the quantity $F_0(x)$ is the 'original force' acting when the focal allele has frequency x. When the force $F_0(x)$ is non-zero it systematically drives changes in the allele's frequency, and causes evolution to occur. For simplicity, we have assumed that the force $F_0(x)$ depends only on the frequency, and not on the time, but this is not an essential restriction.

[4]Throughout this work, we shall need to refer to 'the frequency of the allele at time t' for a number of different cases. In each case, we shall write this frequency as $X(t)$, avoiding a more cumbersome notation where each case is notationally distinguished. The context should make it clear which frequency is being referred to.

Equation (1) naturally arises in an overlapping generation (i.e., continuous time) model. However, we shall primarily envisage it as arising as an approximation of a discrete time model, where time is measured in generations ($t = 0, 1, 2, ...$) and all changes in the allele frequency over adjacent generations are small.

As an example of the systematic force $F_0(x)$ which drives evolution in an effectively infinite population, consider an unlinked locus in a randomly mating diploid population. If each copy of the focal allele additively changes the relative fitness of its carriers by s then to leading order in s we have $F_0(x) = sx(1-x)$.

In a finite population the behaviour of the frequency is more complex than that described by Eq. (1) because random genetic drift has also to be included. For a finite population, under a diffusion approximation of a discrete time model, Eq. (1) becomes

$$dX = F_0(X)dt + \sqrt{V(X)}dW \tag{2}$$

where[5] the quantity $\sqrt{V(X)}dW$ is the change in X that arises from random

[5]Using Eq. (2) for a finite population with discrete generations is equivalent to making a diffusion approximation: Eq. (2) treats frequency and time as continuous quantities and directly leads to a diffusion equation for the probability density of $X(t)$ (Tuckwell 1995). All analytical results we give in this work are obtained under a diffusion approximation.

genetic drift, when acting over a time interval of dt. This 'drift force' is a source of time-dependent randomness, and consists of the product of two factors. One factor involves a function $V(x)$, which is the rate at which genetic drift introduces variance across replicate populations, when the frequency is x. Assuming the population behaves as a randomly mating diploid population of effective size N_e we have $V(x) = x(1-x)/(2N_e)$ (Wright 1931; Gale 1990). The other factor in the 'drift force' involves the quantity $W \equiv W(t)$ which is a random function of time, namely a Wiener process (Tuckwell 1995).

Equation (2) is a stochastic differential equation[6] where changes in $X(t)$ arise from two terms, one which acts systematically ($F_0(X)dt$), the other which acts stochastically ($\sqrt{V(X)}dW$).

3 Aim of this work

In this work we focus solely on finite populations. Broadly speaking, we wish to determine the necessary change in the dynamics, so that a specific outcome is achieved. To this end, we find it sufficient to change only the systematic part of the force in Eq. (2), (i.e., to only change the coefficient

[6]Equation (2) is the mathematically preferred way of writing the equation for X, rather than $dX/dt = F_0(X) + \sqrt{V(X)}dW/dt$. Deriving Eq. (2) from a discrete generation model, leads naturally to an Ito stochastic differential equation (for Ito stochastic differential equations, see e.g., Tuckwell 1995).

of dt, in Eq. (2)).

To go further, let us consider a population where the focal allele has frequency x at time t. We assume that, apart from the original force, $F_0(x)$, which naturally arises within the population, an additional systematic force also acts, which arises from sources *external* to the population. We write the additional force as $F_{add}(x,t)$. The original force in Eq. (2) must then be replaced by $F_0(x) + F_{add}(x,t)$ and the equation that governs the frequency becomes $dX = [F_0(X) + F_{add}(X,t)]\,dt + \sqrt{V(X)}dW$. We take the additional force, $F_{add}(x,t)$, to arise from human intervention into the behaviour of the population.

We can now precisely state the aim of this work. This is to determine the form of the additional force, $F_{add}(x,t)$, which results in the achievement of a specified probability distribution (of the frequency of the focal allele) at a specific time T. In the general case considered in this work, there is an arbitrary probability distribution at an initial time, and the desired outcome, at the final time T, is also an arbitrary distribution. We shall call the distribution of the frequency at the final time the *target distribution*. Thus the additional force, $F_{add}(x,t)$, ensures that the desired target distribution is produced at the desired time[7].

[7] As we shall show, the additional force has to generally be a function of both frequency and time. In other words, when the additional force acts on a population, the value of the

This work is related to a previous work, where knowledge of the state of a population at a final time, led to a time-dependent conditioned distribution for the population. However, that problem was shown to be fully equivalent to a different problem, where there was no knowledge or restriction on the final state of the population, but there was an additional evolutionary force acting (Zhao et al. 2013). Here, we use related ideas, in a significant generalisation with a different logical order: we specify an arbitrary final distribution, and infer a form of the additional force that drives the population to this final distribution, from an arbitrary initial distribution.

4 General result

We shall give the form of an additional force, $F_{add}(x,t)$, which produces a target distribution at a final time, given the distribution at an initial time of 0; the final distribution, the final time, and the initial distribution are all arbitrary.

We first need to introduce three ingredients from which this force is constructed. These are: (i) the probability distribution of the allele frequency in the *absence* of the additional force, (ii) the initial distribution of the allele frequency, and (iii) the target (i.e., final) distribution of the allele frequency.

force depends on the composition of the population (frequency of the focal allele) and the time the force acts.

4.1 Probability distribution in the absence of the additional force

Equation (2) is the stochastic differential equation for the frequency at time t, in the absence of any additional force, and hence only involves the original force, $F_0(x)$. Let $X(t)$ denote the frequency described by Eq. (2) which, at an initial time of u, takes the value y (i.e., $X(u) = y$). We write the probability density of $X(t)$, as a function of x, as $K_0(x, t|y, u)$. This distribution is the solution of a diffusion equation (see Eq. (A1)).

4.2 Initial and target distributions

We write the arbitrary initial allele frequency distribution as $\alpha(x)$.

We write the target distribution as $\beta(x)$. There appears to be a more general way of writing the target distribution, but we shall not make use of this in the present work[8].

[8] A more general way of writing the target distribution is in terms of the function $\beta(x|q)$, which denotes the target distribution, as a function of x, conditional on an initial frequency of q. The target distribution is then given by $\beta(x) = \int_0^1 \beta(x|q)\alpha(q)dq$.

4.3 Additional force

We can now give the full form of the additional force. Defining a function $B(x,t;p)$ by

$$B(x,t;q) = \int_0^1 \beta(p) \frac{K_0(p,T|x,t)}{K_0(p,T|q,0)} dp \tag{3}$$

and given an initial probability distribution of $\alpha(x)$ at time 0, additional force, which ensures that the target probability distribution $\beta(x)$ is achieved at time T, is

$$F_{add}(x,t) = V(x) \frac{\int_0^1 \frac{\partial B(x,t;q)}{\partial x} K_0(x,t|q,0)\alpha(q)dq}{\int_0^1 B(x,t;q) K_0(x,t|q,0)\alpha(q)dq}. \tag{4}$$

A derivation of the additional force in Eq. (4) is given in Appendix A.

The additional force in Eq. (4) does not have a simple form. This is perhaps understandable, since the additional force has to modify the shape of a probability distribution over time and the frequency and time-dependent form it has is a way of achieving this non-trivial outcome.

4.4 Interpretation and implementation of the additional force

Equation (4) represents an explicit formula for the additional force, in terms of the ingredients given in Sections 4.1 and 4.2. We note that we have not

given an explicit interpretation of the additional force, for example we have not said it definitely represents selection or, indeed, another evolutionary force. This is because any external agency that produces this force will cause the final distribution to coincide with the target distribution. The external force is thus arbitrary in this sense, and it may be thought of (and implemented) in different ways. If, for example, the additional force is viewed as additional additive selection that acts on the focal allele, then with such an interpretation, we would write $F_{add}(x,t) = s_{add}(x,t)x(1-x)$ where $s_{add}(x,t)$ is a frequency and time dependent selection coefficient[9] whose form is given by dividing the right hand side of Eq. (4) by $x(1-x)$. However, other interpretations of the additional force are possible and may be more useful or more natural. We could, for example, write $F_{add}(x,t) = u(x,t)(1-x)$ where $u(x,t)$ is a frequency and time dependent mutation rate. Generally, we can interpret the additional force as arising from a mixture of different forces (such as selection, mutation and migration), with the interpretation driven by considerations of external interventions with a population that are feasible.

[9]This interpretation works if the typical (most probable) values of the selection coefficient are small ($|s_{add}(x,t)| \ll 1$). If this is not the case then the picture of additive selection breaks down (it is based on small selection coefficients). An alternative interpretation of the additional force may then be possible, such as it being mutational in nature - see later.

5 Illustrative examples of the additional force

We now give some illustrative *analytical* examples of the additional force. There are many possibilities, because of the range of choices possible in the additional force given in Eq. (4). To reduce the range of options, we shall assume the target distribution corresponds to the occurrence of fixation and loss, with no segregating frequencies, and in Appendix B we give a number of special cases for this target distribution. In the main text, we assume further that the requirement on the target distribution is that it only be achieved at a very large time, i.e., by a final time of $T = \infty$. We thus consider the additional force that controls the probabilities of the ultimate occurrence of fixation and loss. We shall assume fixation and loss occur with the probabilities of β_1 and $\beta_0 = 1 - \beta_1$, respectively.

5.1 Time-dependent additional force

Consider the case where there is a non-trivial distribution of the initial allele frequency, $\alpha(x)$. This would occur, for example, when there is an initial distribution of frequencies across a set of population lines (as described in the Introduction), where all lines do not start at the same allele frequency.

In Appendix B we show that the additional force is given by

$$F_{add}(x,t) = V(x)\frac{\int_0^1 \left[\frac{\partial}{\partial x}\left(\beta_1 \frac{P_{fix}(x)}{P_{fix}(q)} + \beta_0 \frac{P_{loss}(x)}{P_{loss}(q)}\right)\right] K_0(x,t|q)\alpha(q)dq}{\int_0^1 \left(\beta_1 \frac{P_{fix}(x)}{P_{fix}(q)} + \beta_0 \frac{P_{loss}(x)}{P_{loss}(q)}\right) K_0(x,t|q)\alpha(q)dq} \tag{5}$$

where $P_{fix}(x)$ is the probability of ultimate fixation of the focal allele when its initial frequency is x and only the original force, $F_0(x)$, is acting[10].

Note that for Eq. (5), despite the target distribution being achieved at the final time of $T = \infty$, the additional force generally has dependence on the time, t.

5.2 Time-independent examples of the additional force

The time dependence in Eq. (5) follows because the initial distribution, $\alpha(x)$ contains a range of frequencies of the focal allele, not a single frequency. We shall now assume that only a single frequency is present in the initial distribution. Taking this initial frequency to be y, we find from Eq. (5) that the additional force is independent of time and given by

$$F_{add}(x) = V(x)\frac{[\beta_1 - P_{fix}(y)]\, P'_{fix}(x)}{\beta_1 P_{fix}(x) + \beta_0 P_{fix}(y) - P_{fix}(y)P_{fix}(x)} \tag{6}$$

where $P'_{fix}(x) = dP_{fix}(x)/dx$ (see Appendix B for details).

[10] We shall sometimes refer to $P_{fix}(x)$ as the 'natural fixation probability'.

We again assume the population behaves as a randomly mating diploid population with effective size N_e, and so take $V(x) = x(1-x)/(2N_e)$ (Wright 1931; Gale 1990). We also assume an original force that corresponds to additive selection, where each copy of the focal allele additively changes the relative fitness of its carrier by s. Then the original force is

$$F_0(x) = sx(1-x) \tag{7}$$

to leading order in s.

Using the scaled strength of selection

$$S = 4N_e s \tag{8}$$

the natural probability of fixation is

$$P_{fix}(x) = \frac{1 - e^{-Sx}}{1 - e^{-S}} \tag{9}$$

(Kimura 1962).

Explicit, time-independent results for the additional force for four cases are as follows[11] (see Appendix B for derivations).

[11] Where, below, we speak of selection, we are referring purely to the selection on the focal allele from the original force, Eq. (7).

1. Fixation ultimately occurs ($T = \infty$), and selection is non-zero. Thus $\beta_1 = 1$, $\beta_0 = 0$, $S \neq 0$, and

$$F_{add}(x) = \frac{x(1-x)}{2N_e} \frac{S}{e^{Sx} - 1}. \tag{10}$$

2. By time $T = \infty$, fixation is *enhanced* from its natural value, $P_{fix}(y)$, by a factor r (with $0 \leq r \leq 1/P_{fix}(y)$) and selection is non-zero. Thus $\beta_1 = rP_{fix}(y)$, $\beta_0 = 1 - rP_{fix}(y)$, $S \neq 0$, and

$$F_{add}(x) = -\frac{x(1-x)}{2N_e} \frac{S\,(1-r)}{1 - e^{-S(1-x)} + r(e^{-S(y-x)} - 1)}. \tag{11}$$

3. By time $T = \infty$, fixation and loss can generally both occur, and there is no selection. Thus $\beta_1 \neq 0$ or 1, $\beta_0 \neq 0$ or 1, $S = 0$, and

$$F_{add}(x) = \frac{x(1-x)}{2N_e} \frac{\beta_1 - y}{\beta_1\,(x-y) + y(1-x)}. \tag{12}$$

4. Loss ultimately occurs ($T = \infty$), and there is no selection. Thus $\beta_1 = 0$, $\beta_0 = 1$, $S = 0$, and

$$F_{add}(x) = -\frac{x}{2N_e}. \tag{13}$$

For an initial frequency y that lies in the range $0 < y < 1$, we note that the additional forces in Eqs. (10) and (13) are independent of y (see Appendix B). This independence illustrates a general feature of Eq. (4), that when only a *single frequency* is present in the target distribution, the additional force is independent of the initial distribution, $\alpha(x)$. A direct consequence is that the additional forces in Eqs. (10) and (13) apply for any initial distribution. Later we shall present results that are based on the additional forces in Eqs. (10) and (13), when the initial distribution is uniform over a range of frequencies.

In Figure 2 we plot the additional forces given in Eqs. (10)-(12) as functions of the frequency, x.

In Figure 2 we have not shown the additional force given in Eq. (13), which applies when there is no selection on the focal allele and loss ultimately occurs, because the corresponding additional force is mathematically simple, namely a linear function of frequency[12]. The additional force in this case does not vanish at $x = 1$, and cannot be taken to correspond to additive selection, since this would entail a selection coefficient of $-[2N_e(1-x)]^{-1}$, which can become arbitrarily negative. Instead, the additional force given in Eq.(13) is best interpreted/implemented as another systematic force, for

[12]The additional force is also a linear function of frequency when there is no selection, and fixation ultimately occurs.

Figure 2 Caption: The time-independent additional forces given in Eqs. (10) - (12) are plotted as functions of the frequency, x, for an initial frequency of $y = 0.1$ and an effective population size of $N_e = 100$. These results apply when the target distribution is achieved by $T = \infty$. The thick solid curve denotes the additional force when selection is of scaled strength $S = 4$ and the target distribution corresponds solely to fixation ($\beta_1 = 1$). The thick dashed curve denotes the additional force when selection is of scaled strength $S = 4$ and the target distribution corresponds to *doubling* the natural probability of fixation, i.e., $\beta_1 = 2P_fix(y = 0.1) \simeq 0.67$ and $\beta_0 \simeq 0.33$. The thin dashed curve denotes the additional force under neutrality ($S = 0$) when the target distribution corresponds to the occurrence of both fixation and loss, with probabilities of $\beta_1 = 0.3$ and $\beta_0 = 0.7$, respectively.

example as recurrent mutation of the focal allele, at a rate of $(2N_e)^{-1}$. Such an additional force systematically pushes the population away from high frequencies, so loss ultimately occurs. In contrast, when the target distribution corresponds to non-zero probabilities of both fixation and loss at long times, the additional force vanishes at the boundaries $x = 0$ and $x = 1$, and can, if desired, be interpreted as arising from a form of (frequency-dependent) additive selection.

6 Simulations

We illustrate this work further, with a presentation of results from simulations.

All simulations used for this work were of a hybrid nature, in the sense that the additional force was derived under a diffusion approximation, which involves continuous time and continuous frequencies, but the additional force was employed within a stochastic equation equivalent to Wright-Fisher model, where time and frequencies are discrete, (details of the simulations are given in Appendix C).

Figure 3A

6.1 Results for time-independent additional forces

In Figure 3 we illustrate results of simulations where we have used the time-independent additional forces of Eqs. (10) and (13). For each panel of Figure 3, properties of a large set of simulated trajectories are shown when the target distribution corresponds to either fixation or loss being ultimately achieved ($T = \infty$).

In Figure 4, we illustrate results of simulations where we employed the time-independent additional forces of Eqs. (11) and (12). The target distribution contains two states of the population, namely fixation and loss. In Figure 4A, the additional force was used to *enhance* the natural fixation probability by a factor r. The level of enhancement (fixation probability

Figure 3B

Figure 3 Caption: Results are presented for simulated trajectories where the target distribution corresponds to the ultimate achievement of a single frequency. The additional forces of Eqs. (10) and (13) were employed within the stochastic dynamics of a Wright-Fisher model (see Appendix C for details of the simulations). For both panels, 10^5 simulated trajectories were used, and all trajectories were initially distributed uniformly over frequencies in the range 0.2 to 0.4 (i.e., the initial frequency distribution, $\alpha(x)$, was uniform over this range). In each generation, all trajectories where sorted according to the value of their frequency and percentiles were determined. In both panels, the shaded area shows the middle 50% of all trajectories (i.e., lying between 25th and 75th percentiles), and the thick solid lines denote median trajectories. For Panel A, selection was additive in the number of focal alleles (Eq. (7)) and of scaled strength $S = 4$, and all trajectories were directed to ultimately achieve fixation ($\beta_1 = 1$) using the force of Eq. (10). For Panel B, there was no selection ($S = 0$) and the force given in Eq. (13) was imposed, so loss was ultimately achieved.

divided by natural fixation probability) observed in the simulations is very close to the linear value that the force was designed to produce. In Figure 4B, a neutral focal allele was directed to achieve fixation with probability 0.7, using the additional force of Eq. (12). The dashed and shaded areas in Figure 4B show the lower 30% and upper 70% of all trajectories, respectively[13]. The lines in each area show the *conditional medians* of the respective areas. From the conditional median in the dashed area, it is apparent that by approximately generation 100, 15% of all trajectories have been lost, and similarly, by approximately generation 300, 35% of all trajectories have fixed. For a comparison, Figure 4C shows the dynamics of the same population, in the absence of any additional force.

6.2 Results for time-dependent additional forces

We have numerically determined the additional force, Eq. (4), in some situations where the final time, T, is finite ($< \infty$) and the additional force, apart from being frequency dependent, also has time dependence.

We again assume *additive selection*, as given in Eq. (7), and the form of $V(x)$ adopted is appropriate to an ideal randomly mating diploid population of size N, namely $V(x) = x(1-x)/(2N)$. We take the target distribution to

[13]To be precise, the indicated areas of Figure 4B are calculated from the 29.5th and 30.5th percentiles of all trajectories, so the two areas indicated contain 99% of all trajectories.

Figure 4A

Figure 4B

Figure 4C

Figure 4 Caption: In this figure, we show the outcomes of a large set of simulated trajectories for a target distribution involving both fixation and loss by the final time $T = \infty$. For Panel A, the natural fixation probability of Eq. (9), with a scaled strength of selection of $S = 4$ and an initial frequency of $y = 0.1$, was enhanced by a factor of r, by application of the additional force of Eq. (11). For each value of r, a data point (black dot) was plotted, representing the simulated value of 'the probability of fixation, divided by $P_fix(0.1)$'. A straight line was fitted to the data, and a slope and intercept of approximately 0.98 and 0.01 were obtained (theoretically, the slope and intercept are 1 and 0, respectively). For Panel B, the simulated trajectories started from an initial allele frequency of $y = 0.2$. The strength of selection was $S = 0$ and the focal allele was, by time $T = \infty$, directed to achieve fixation with probability $\beta_1 = 0.7$, and loss with probability $\beta_0 = 0.3$, using the additional force of Eq. (12). We found that in each generation, 70% of all trajectories lay within the shaded area, and the bold solid line in this area shows the *conditional* median of this 70% of all trajectories. Similarly, 30% of all trajectories lay in the dashed area, each generation, and the bold dashed line gives the conditional median of these trajectories. For comparison, in Panel C we have plotted the outcomes of a large set of simulated trajectories, for the same parameters as Panel B, when no additional force acts.

correspond only to fixation. The additional force in Eq. (4) then reduces to $F_{add}(x,t) = V(x)\partial \ln\left[P_{fix}\left(T|x,t\right)\right]/\partial x$ where $P_{fix}(T|x,t)$ is the probability of fixation of the focal allele (under the dynamics of Eq. (2)) by time T given an initial frequency of x at time t (see Appendix B for details). Since the form of $P_{fix}(T|x,t)$, as a function of x, is not explicitly known, we replaced it by the corresponding quantity of a Wright Fisher model, with the derivative replaced by a finite difference. In the Wright-Fisher model, $P_{fix}(T|x,t)$ corresponds to an element of the $(T-t)$'th power of the transition matrix, hence when t is small, the transition matrix is raised to higher power than when t is large. This allows the quantitative conclusion (borne out numerically and evident in Figure 5) that when t is small, the force is smaller than when t is large. In Figure 5A we plot this force as a function of x for four different values of the time, when the target distribution is fixation by time $T = 100$. Figure 5B shows simulation results of the dynamics of such a population. The results can be viewed as the additional force 'pushing' the trajectories to a restricted region of frequency, by the final time.

At the final time ($T = 100$) in Figure 5B, all simulated trajectories achieved a frequency of unity (i.e., fixed).

Figure 6 gives another example for a time-dependent additional force, where the target distribution is achieved by a finite time. This is for a *neutral*

Figure 5A

focal allele, when the target distribution involves both fixation and loss, but no segregating frequencies. The force is given in Eq. (B4). The target distribution involves more than one frequency, and the form of the additional force strongly depends on time (cf. Figure 5A, where the target distribution involves only a single frequency and there is much weaker time dependence in the additional force). In Figure 6B, simulation of the dynamics is shown.

The time and frequency dependence of the additional force shown in Figure 6A is complicated, and despite a relatively simple outcome (fixation and loss, by a finite time, of a neutral allele) it would be very hard to guess the additional force that achieves this.

We have carried out simulations involving both time-dependent and

Figure 5B

Figure 5 Caption: In this figure we plot the time-dependent additional force and results of simulations, assuming additive selection, when the target distribution corresponds solely to fixation. In Panel A, the additional force is plotted as a function of allele frequency, x, for four different values of the time. The target distribution was required to be achieved by time $T = 100$. The population had equal census and effective sizes: $N = N_e = 100$ and selection was of scaled strength $S = 4Ns = 4$. The additional force was numerically approximated (using probabilities from a Wright-Fisher model, with derivatives replaced by finite differences). In Panel B, results of simulations for 10^5 trajectories are plotted, using the above parameters (see Appendix C for details of simulations). All trajectories were initially distributed uniformly over allele frequencies in the range 0.2 to 0.4 (i.e., the initial distribution, $\alpha(x)$, was uniform over this range of frequencies). The 25th, 50th and 75th percentiles of the allele frequency are shown, hence half of all trajectories lie in the shaded area at any time.

Figure 6A

time-independent additional forces. When the target distribution corresponds purely to fixation, or purely to loss, we have observed that by the final time the target distribution is fully achieved. However, when the target distribution corresponds to the possibility of both fixation and loss (i.e., both have non-zero probabilities of occurrence in the target distribution), we have not observed any segregating trajectories, but have observed errors in the prescribed probabilities of fixation and loss that are smaller than 1%. These errors do not primarily arise because of finiteness of the simulations (we used a very large number of trajectories), but rather because the additional force was calculated under a diffusion approximation, but was employed within a Wright-Fisher model, and the calculated force is an *approximation* of the

Figure 6B

Figure 6 Caption: In this figure we plot time-dependent additional forces and the results of simulations for target distributions with multiple frequencies, assuming no selection. In Panel A, the additional force of Eq. (B4) is plotted as a function of allele frequency, x, at four different times, when there was a single initial frequency of $y = 0.2$. The target distribution corresponds, by time $T = 100$, to achieving fixation with probability 0.7 and loss with probability 0.3. The population was of equal census and effective size: $N = N_e = 100$, and there was no selection ($S = 0$). The additional force of Eq.(B4) was numerically approximated (see Figure 5 Caption). In Panel B, results of simulations for 10^5 trajectories are plotted, using the above parameters (see Appendix C for details of simulations). We found that in each generation, 70% of all trajectories lay in the shaded area, and the bold solid line in this area shows the conditional median of these trajectories. Similarly, 30% of all trajectories lay in the dashed area, each generation, and the bold dashed line gives the conditional median of these trajectories.

additional force that is required within this model. Errors of this magnitude are present in Figures 1C, 4B and 6B.

7 Discussion

In this work we have provided a mathematical framework for a form of biological control. We have introduced and described an additional evolutionary force that controls the frequency of a focal allele, in the sense that from an arbitrary initial distribution of the focal allele, the additional force produces an arbitrary target (or final) distribution at a specific time.

An issue that has been noted in the Introduction and Appendix A, but not addressed in the main text is whether the additional force we have found is unique, or whether there can be other forms of the additional force with the same outcome, namely producing the target distribution at the final time. The additional force was calculated from a probability distribution of the frequency, that was determined by properties of the original population (i.e., it depended on pre-existing forces that acted in the population). The probability distribution of the frequency written down was the simplest that went from the initial distribution to the target distribution, at the appropriate times. There are various ways to show that other forces can also produce the target distribution from the initial distribution. We give

a somewhat general method in Appendix A (see the paragraph following Eq. (A6)). Here we give two other ways, in particular examples, to see this. One way is to put further requirements on the probability distribution of the frequency (apart from achieving the initial and target distributions), for example, requiring that a particular distribution is achieved at an intermediate time. This further requirement generally leads to an additional force that is different to that presented in Eq. (4). Another way of showing the non-uniqueness of the additional force, applies when the additional force $F_{add}(x,t)$ produces just fixation by a finite time T. In this case, the force $k \times F_{add}(x,t)$ with any $k > 1$ will produce fixation *before* time T, and hence also *by* time T. This last example is instructive; we can use results like this to introduce the idea that we can 'tune' the additional force to modify the time taken or in other ways, to produce a modified outcome.

There remains the question whether the specific additional force we have presented, Eq. (4), has a particular property that distinguishes it from alternative forms of the additional force (that also produce the target distribution). As we argue in Appendix A, it has the *simplest form*, and it may be conjectured that in a general sense, it is the minimal force that produces the target distribution. But the notion of 'minimal' requires an explicit measure of an additional force that depends on both time and frequency.

At the present time, we have no such measure and we leave open the issue of what makes the additional force in Eq. (4) genuinely 'special'.

We note that the only method we have of analytically calculating the additional force is via the diffusion approximation. In principle, this can lead to additional forces of large magnitude. However, in the time independent examples we have considered, the additional force typically takes small values. For example, from Figure 2, the additional force is of the order 10^{-3}, while for Figure 3, the magnitude of the additional force, when calculated along the median trajectory, is less than 0.005. In time dependent cases, the force is a function of both frequency and time. In Figures 5A and 6A the additional force is of order 10^{-2} but for times close to termination time T, the magnitude of F_{add} increases. Since we have a finite final time, T, those trajectories that are 'far' from the target distribution at times close to T, can suffer a large additional force, to push them to the target by the final time. However, the additional force is applied to all trajectories from the initial time onwards, and as shown in Figures 5B and 6B, the distributions of the resulting trajectories are indicative of being directed to the target from earlier times. A consequence is that only a tiny fraction of all trajectories lie in a frequency-time region where there is a large additional force. In other words, the probability of requiring a large additional force is small,

and typically, we do not need to concern ourselves with additional forces of large magnitude. The good performance of the additional force within the Wright-Fisher model where a bounded version of the additional force is used (see Appendix C), is an illustration of this.

Consider now the initial distribution of the frequency. With real data about a single population, the initial frequency is generally not precisely known, but is an estimate that is derived by sampling the population. It might be thought that incorporating uncertainty about the initial frequency, in the initial distribution, would be a way to deal with such uncertainty. However, it seems to us that there is a real difference between an initial distribution that reflects genuinely different allele frequencies that are initially present (as could occur for a set of population lines, that do not all have the same initial frequency), and a distribution that reflects uncertainty about the initial frequency. To take one example of this, consider with a single population, which certainly just has a single initial frequency. In the case where the target distribution corresponds to fixation and loss by time $T = \infty$, the additional force will, according to Eq. (6), be time independent. If, however, we incorporate uncertainty in the initial frequency, by using an initial distribution to encapsulate this, then the additional force will be time dependent (see Eq. (5)). It seems to us that this time dependence is

an artefact of using the initial distribution to model uncertainty, and that a preferable procedure to deal with such uncertainty would be to average the time independent additional force (for a single initial frequency), over different initial frequencies. This would still lead to a time independent additional force. This argumentation leads us to believe that generally, uncertainty in the initial frequency should not be reflected in the initial distribution, $\alpha(x)$.

We have already pointed out that the additional force can be interpreted in different ways, i.e., as selection, migration, mutation or some combination of these (with some non-trivial frequency dependence in most cases). Therefore implementation of the additional force in a biological problem is highly flexible and may be customised according to available techniques. Although we derive the mathematical framework for a single locus, the additional force can, as it stands, be applied to an arbitrary number of unlinked loci of a population, in the absence of epistasis.

Intuitively, certain desirable target distributions (e.g. immediate extinction of a resistant parasite) may require very large interventionist forces. This may turn out to be impracticable, and thus indicates an effectively inaccessible choice of the target distribution. Within our mathematical framework we can investigate this problem, by theoretically exploring ways of reducing the extent of the additional force applied. We can determine,

for example, how allowing additional time before the target distribution has to be achieved, or choosing more reasonable target distributions, affect the practicality of an intervention.

The framework we present also provides results for the distributions that are achieved between the initial and final times (see Eq. (A7)). Such distributions are generally different to the initial and target distributions, as is illustrated in, e.g., Figure 1C. During application of an additional force, the distributions that are observed at intermediate times may be compared to the theoretical distribution (Eq. (A7)). If significant discrepancies are detected, this may suggest modifications of the additional force, in the form of fine scale adjustments of the experimental setup. This may be useful since quantifying the additional force from real data requires estimation of parameters of the population, which may not necessarily be easy (Barbosa et al. 2011).

The additional force can be applied in a repeated experimental setup where the same lines are exposed to the same evolutionary force, and this is similar to 'evolve and resequence experiments' (Schlötterer et al. 2015). However, potential applications are not restricted to an artificial experimental setup. Repeated exposure of the same additional force to identical populations may be feasible for human monocultured plantations, such as

crop fields and monocultured forests, or in animal breeding such as cattle. In such a framework, evolutionary control can be used to reduce the frequency of detrimental alleles or to increase the genetic diversity for particular loci, e.g. regions of the genome with a reduced local effective population size (Gossmann et al. 2011) as observed in highly inbred individuals.

We note that different humans constitute different lines, as far as somatic cell lineages, diseases and parasites are concerned. Hence an individualised treatment of different individuals may yield a desirable population-wide distribution. A drug that thus targets a specific disease/parasite can be administered at an individual level that is guided by the mathematical framework of the present work. The same framework can used to quantify prophylactic steering of human populations, e.g. a population wide treatment, such as nutrition add-on of iodine (Zimmermann and Andersson 2012).

As noted at the beginning of the paper, the additional evolutionary force can also be applied to a single population. In general, the target distribution involves more than a single allele frequency, but a single population is characterised by a single frequency at any time, and the additional force is perhaps best not thought of as producing a particular distribution, but rather of biasing matters, so some outcomes are made more probable (or less) than they would be in the absence of the additional force. With a

single population in mind, the additional force could be used as a method of population control. It could be applied to urban animal populations, such as pest bird populations in cities (Fuller et al. 2009) or antimicrobial resistance populations in hospitals (Ferri et al. 2015). Allele invasions into populations may only be stably introduced at certain allele frequencies (Unckless et al. 2015), therefore shifting allele frequencies for a certain limited period of time, by applying an additional evolutionary force, may be crucial for the successful establishment of a trait into a population.

The results we have presented are based on the stochastic differential equation given in Eq. (2). Related equations arise in different contexts, with different interpretations. It may be that, with the necessary changes made, the results we have derived have applications to a variety of systems other than the genetic ones considered here.

To sum up, we believe the theory presented in this work provides a useful and flexible method for the control of biological populations. The additional force we have introduced can be applied to a diverse set of biological problems, and it may have interesting extensions/generalisations/applications to a wider set of problems than those considered in this work.

Acknowledgements

It is a pleasure to thank Yu Shi and Lei Zhao for helpful conversations on this work. We also thank two anonymous reviewers for comments that have significantly improved this paper. One of us (TIG) was supported by a Leverhulme Early Career Fellowship (Grant ECF-2015-453) and NERC grant N013832/1.

Appendices

Appendix A: Determining the additional force

In this appendix we motivate and determine expressions for the form of an additional systematic force that acts on the frequency of an allele in a population. This force, in concert with the original force and random genetic drift, produces an arbitrary target distribution at a given time, from an arbitrary initial distribution.

As we shall point out, the form of the additional force is not unique. However, we believe that the additional force we present has the simplest form, as we shall reason later in this Appendix. We establish the form of the additional force as follows. (i) We write down a time-dependent probability density in terms of quantities describing the original population (i.e., in the absence of any additional force). (ii) The probability density is explicitly constructed so that it starts at the initial distribution and achieves the target distribution at the final time. (iii) This time-dependent probability density is shown to obey a diffusion equation. (iv) Inspection of this diffusion equation allows us to infer the form of the additional force required to achieve the target distribution.

Background

We begin with Eq. (2) of the main text,namely $dX = F_0(X)dt + \sqrt{V(X)}dW$, which is the stochastic differential equation for the allele frequency when the additional force is absent. Assuming the frequency, $X(t)$, takes the value y at initial time u (i.e., $X(u) = y$), we write the probability density of $X(t)$, when evaluated at frequency x, as $K_0(x, t|y, u)$. This obeys the forward and backward diffusion equations

$$-\frac{\partial}{\partial t}K_0(x,t|y,u) = -\frac{1}{2}\frac{\partial^2}{\partial x^2}\left[V(x)K_0(x,t|y,u)\right] + \frac{\partial}{\partial x}\left[F_0(x)K_0(x,t|y,u)\right] \tag{A1}$$

$$\frac{\partial}{\partial u}K_0(x,t|y,u) = -\frac{V(y)}{2}\frac{\partial^2}{\partial y^2}K_0(x,t|y,u) - F_0(y)\frac{\partial}{\partial y}K_0(x,t|y,u) \tag{A2}$$

(Kimura 1964) where $V(x)$ is the genetic variance introduced per unit time by genetic drift when the allele frequency is x. The distribution $K_0(x, t|y, u)$ also obeys the initial condition $K_0(x, u|y, u) = \delta(x - y)$, where $\delta(x)$ denotes a Dirac delta function of argument x.

The distribution $K_0(x, t|y, u)$ we use is *complete* in the sense that it describes all possible states of a population, that is to say where the focal allele is segregating, lost or fixed (McKane and Waxman 2007; Waxman 2011; Zhao et al. 2013). Since $K_0(x, t|y, u)$ is complete, it has the property

that for all $t \geq u$, it is normalised to unity: $\int_0^1 K_0(x,t|y,u)dx = 1$. Generally $K_0(x,t|y,u)$ is singular, in that it contains Dirac delta functions located at $x = 0$ and $x = 1$, which represent loss and fixation, while the non-singular part represents populations where the allele is segregating (McKane and Waxman 2007; Waxman 2011; Zhao et al. 2013).

We treat the distribution $K_0(x,t|y,u)$ as known, since all of its properties follow from the known functions $V(x)$ and $F_0(x)$, and analytical or numerical approaches may be used to elicit these properties.

We now consider a *conditioned* probability density of the frequency of the focal allele. We assume the frequency of the focal starts at the value y at initial time 0 and achieves (or 'is conditioned upon') the value z at final time T. The conditioned probability density of the frequency at an intermediate time t $(0 \leq t \leq T)$, as a function of x, follows by an application of Bayes' Theorem

$$K^{[z,T]}(x,t|y,0) = C(z,T,x,t,y)K_0(x,t|y,0) \tag{A3}$$

where

$$C(z,T,x,t,y) = \frac{K_0(z,T|x,t)}{K_0(z,T|y,0)} \tag{A4}$$

(see Zhao et al. 2013).

Assuming loss and fixation can occur, the result in Eq. (A3) takes dif-

ferent forms, depending on whether the final frequency corresponds to the allele segregating ($0 < z < 1$) or fixed ($z = 1$) or lost ($z = 0$).

When $0 < z < 1$ only the non-singular parts of the K_0's in the ratio $C(z, T, x, t, y)$ of Eq. (A4) contribute (the non-singular parts describe segregating alleles). When $z = 1$, $C(z, T, x, t, y)$ takes the value $P_{fix}(T|x, t)/P_{fix}(T|y, 0)$ where $P_{fix}(T|x, t)$ is the probability of fixation of the focal allele by time T (under the dynamics of Eq. (2) of the main text) given that it had an initial frequency of x at time t. When $z = 0$, $C(z, T, x, t; y)$ takes the value $P_{loss}(T|x, t)/P_{loss}(T|y, 0)$ where $P_{loss}(T|x, t)$ is the corresponding probability of loss of the focal allele (for further details, see Appendix A of Zhao et al. 2013)[14].

It may be directly verified, using the forward and backward diffusion equations, Eqs. (A1) and (A2), that $K^{[z,T]}(x, t|y, 0)$ obeys the diffusion

[14]A simple way to understand the value of $C(z, T, x, t; y)$ when e.g., $z = 1$, is to argue that the term in $K_0(z, T|x, t)$ involving $\delta(1 - z)$ dominates the numerator, while the corresponding term in $K_0(z, T|y, 0)$ involving $\delta(1 - z)$ dominates the denominator, and the two Dirac delta functions cancel, leaving the ratio of their coefficients, which is $P_{fix}(T|x, t)/P_{fix}(T|y, 0)$. This argumentation can be made more formal by first 'weakening' the Dirac delta functions so they are non-singular distributions. After setting, e.g., $z = 1$, the weakened Dirac delta functions can be allowed to become singular. The result obtained, namely $P_{fix}(T|x, t)/P_{fix}(T|y, 0)$, is in full accord with the corresponding result in the Wright-Fisher model.

equation

$$-\frac{\partial}{\partial t}K^{[z,T]}(x,t|y,0) = -\frac{1}{2}\frac{\partial^2}{\partial x^2}\left[V(x)K^{[z,T]}(x,t|y,0)\right]$$

$$+\frac{\partial}{\partial x}\left\{\left[F_0(x)+f_{add}(x,t)\right]K^{[z,T]}(x,t|y,0)\right\} \tag{A5}$$

where

$$f_{add}(x,t) = V(x)\frac{\partial}{\partial x}\ln\left[C(z,T,x,t,y)\right] \tag{A6}$$

(Zhao et al. 2013).

Note that Eq. (A5) has the form of diffusion equation for an *unconditioned problem* that is subject to a total force of $F_0(x) + f_{add}(x,t)$, i.e., the original force $F_0(x)$ plus the 'additional force' $f_{add}(x,t)$. Thus although Eq. (A5) originally arose in a conditioned problem, we can ignore this connection, and simply note that for the force $F_0(x) + f_{add}(x,t)$, the solution of the diffusion equation, Eq. (A5), is given by the distribution $K^{[z,T]}(x,t|y,0)$ of Eq. (A3). This solution has the properties $K^{[z,T]}(x,0|y,0) = \delta(x-y)$ and $K^{[z,T]}(x,T|y,0) = \delta(x-z)$ which shows that, from a definite starting value of y at time 0, the force $F_0(x) + f_{add}(x,t)$ systematically 'pushes' the frequency to a definite final value of z at time T. The form of the additional

force $f_{add}(x,t)$ is not unique but has the simplest form[15].

Generalisation

Continuing, we generalise the above results, so that $f_{add}(x,t)$ is replaced by a force $F_{add}(x,t)$ which, starting from an arbitrary initial distribution, $\alpha(x)$, drives the allele frequency to the arbitrary target distribution $\beta(x)$. Both $\alpha(x)$ and $\beta(x)$ are normalised to unity ($\int_0^1 \alpha(x)dx = 1$ and $\int_0^1 \beta(x)dx = 1$) as befits probability densities.

We begin the generalisation by constructing

$$P(x,t) = \int_0^1 dp \int_0^1 dq \beta(p) K^{[p,T]}(x,t|q,0)\alpha(q). \tag{A7}$$

This is a probability density defined over the times $0 \leq t \leq T$ with u the initial time and T the final time. The distribution in Eq. (A7) has the following properties: (i) $\int_0^1 P(x,t)dx = 1$ (because $\int_0^1 K^{[p,T]}(x,t|q,0)dx = 1$); (ii) when $t = 0$ we have $P(x,0) = \alpha(x)$ (because $K^{[p,T]}(x,0|q,0) =$

[15] A somewhat general way to see that the additional force $f_{add}(x,t)$ is not unique is as follows. Suppose that in Eq. (A3), instead of using $K^{[z,T]}(x,t|y,0)$ we used the conditioned distribution $\widetilde{K}^{[z,T]}(x,t|y,0) = \frac{K_1(z,T|x,t)K_1(x,t|y,0)}{K_1(z,T|y,0)}$ where $K_1(x,t|y,u)$ is the distribution of the frequency for the force $F_1(x,t)$. The distribution $\widetilde{K}^{[z,T]}(x,t|y,0)$ has the properties $\widetilde{K}^{[z,T]}(x,0|y,0) = \delta(x-y)$ and $\widetilde{K}^{[z,T]}(x,T|y,0) = \delta(x-z)$, thus (like $K^{[z,T]}(x,t|y,0)$), the frequency also 'goes' from y to z. However, the total force driving $\widetilde{K}^{[z,T]}(x,t|y,0)$ is $F_1(x,t) + V(x)\frac{\partial}{\partial x}\ln\left[\frac{K_1(z,T|x,t)}{K_1(z,T|y,0)}\right]$, and the additional force, *relative* to the force acting in the original problem ($F_0(x)$), is $F_1(x,t) + V(x)\frac{\partial}{\partial x}\ln\left[\frac{K_1(z,T|x,t)}{K_1(z,T|y,0)}\right] - F_0(x)$. This additional force generally differs from the additional force in Eq. (A6). The additional force has the simplest form when $F_1(x,t)$ *coincides* with $F_0(x)$.

$\delta(x-q)$); (iii) when $t = T$ we have $P(x,T) = \int_0^1 \beta(x|q)\alpha(q)dq = \beta(x)$ (because $K^{[p,T]}(x,T|q,0) = \delta(x-p)$). Thus the form of the distribution in Eq. (A7) satisfies the requirements that it starts at an initial distribution of $\alpha(x)$ at time 0 and becomes the target distribution $\beta(x)$ at time T. We shall show that $P(x,t)$ obeys a diffusion equation where, apart from the original force $F_0(x)$, an additional force acts. This additional force thus serves to produce $\beta(x)$ from $\alpha(x)$, but because of reasons advanced above, does not have a unique form.

Proceeding, we note that using Eq. (A3) we can write $P(x,t)$ as

$$P(x,t) = \int_0^1 B(x,t;q)K_0(x,t|q,0)\alpha(q)dq \tag{A8}$$

where

$$B(x,t;q) = \int_0^1 \beta(p)\frac{K_0(p,T|x,t)}{K_0(p,T|q,0)}dp. \tag{A9}$$

As a function of x and t, $B(x,t;q)$ obeys the *backward* diffusion equation, Eq. (A2). Because of this, the quantity

$$Q(x,t;q) = B(x,t;q)K_0(x,t|q,0) \tag{A10}$$

obeys, as a function of x and t, the equation

$$-\frac{\partial}{\partial t}Q(x,t;q) = -\frac{1}{2}\frac{\partial^2}{\partial x^2}\left[V(x)Q(x,t;q)\right] + \frac{\partial}{\partial x}\left[F_0(x,t)Q(x,t;q)\right]$$

$$+\frac{\partial}{\partial x}\left[V(x)\left(\frac{\partial}{\partial x}\ln\left[B(x,t;q)\right]\right)Q(x,t;q)\right]. \tag{A11}$$

On multiplying this equation by $\alpha(q)$ and integrating over all q yields (using $P(x,t) = \int_0^1 Q(x,t;q)\alpha(q)dq$, which follows from Eqs. (A7) and (A10))

$$-\frac{\partial}{\partial t}P(x,t) = -\frac{1}{2}\frac{\partial^2}{\partial x^2}\left[V(x)P(x,t)\right] + \frac{\partial}{\partial x}\left[F_0(x,t)P(x,t)\right]$$

$$+\frac{\partial}{\partial x}\left[V(x)\int_0^1\left(\frac{\partial}{\partial x}\ln\left[B(x,t;q)\right]\right)Q(x,t;q)\alpha(q)dq\right]. \tag{A12}$$

We write the last term as

$$\frac{\partial}{\partial x}\left[\frac{V(x)\int_0^1\left(\frac{\partial}{\partial x}\ln[B(x,t;q)]\right)Q(x,t;q)\alpha(q)dq}{\int_0^1 Q(x,t;q)\alpha(q)dq}\times\int_0^1 Q(x,t;q)\alpha(q)dq\right]$$

$$=\frac{\partial}{\partial x}\left[\frac{V(x)\int_0^1\left(\frac{\partial}{\partial x}\ln[B(x,t;q)]\right)Q(x,t;q)\alpha(q)dq}{\int_0^1 Q(x,t;q)\alpha(q)dq}\times P(x,t)\right] \tag{A13}$$

This is of the form $\frac{\partial}{\partial x}\left[F_{add}(x,t)P(x,t)\right]$ where

$$F_{add}(x,t) = V(x)\frac{\int_0^1 \left(\frac{\partial}{\partial x}\ln[B(x,t;q)]\right)Q(x,t;q)\alpha(q)dq}{\int_0^1 Q(x,t;q)\alpha(q)dq}. \tag{A14}$$

Using Eq. (A10) allows this to be written as

$$F_{add}(x,t) = V(x)\frac{\int_0^1 \frac{\partial B(x,t;q)}{\partial x}K_0(x,t|q)\alpha(q)dq}{\int_0^1 B(x,t;q)K_0(x,t|q)\alpha(q)dq} \tag{A15}$$

and $F_{add}(x,t)$ is the generalisation of the force appearing in Eq. (A6).

Appendix B: Special cases and examples of the additional force

In this appendix we consider some special cases of the additional force including some time-independent examples. Throughout this appendix, we take the target distribution to correspond to just the occurrence of fixation and loss, and take these to occur with probabilities of β_1 and $\beta_0 = 1 - \beta_1$, respectively. That is

$$\beta(x) = \beta_1\delta(1-x) + \beta_0\delta(x) \tag{B1}$$

where $\delta(x)$ is a Dirac delta function of argument x.

The function $B(x,t;q)$ of Eq. (3), that appears in Eq. (A15) for the additional force, is given by $B(x,t;q) = \int_0^1 \beta(p)\frac{K_0(p,T|x,t)}{K_0(p,T|q,0)}dp$. Due to Eq. (B1), the ratio $\frac{K_0(p,T|x,t)}{K_0(p,T|y,0)}$ within $B(x,t;q)$ is evaluated at the frequencies $p=1$ and $p=0$ (fixation and loss) and has (as stated in Appendix A) to be replaced by the appropriate ratio of fixation or loss probabilities. For $p=1$ the ratio becomes $\frac{P_{fix}(T|x,t)}{P_{fix}(T|y,0)}$ where $P_{fix}(T|x,t)$ is the probability of fixation of the focal allele (under the dynamics of Eq. (2)) by time T given an initial frequency of x at time t; similarly, for $p=0$ the ratio becomes $\frac{P_{loss}(T|x,t)}{P_{loss}(T|y,0)}$ where $P_{loss}(T|x,t)$ is the corresponding probability of loss by time T. We thus have

$$B(x,t;q) = \int_0^1 \beta(p)\frac{K_0(p,T|x,t)}{K_0(p,T|q,0)}dp$$

$$= \beta_1\frac{P_{fix}(T|x,t)}{P_{fix}(T|q,0)} + \beta_0\frac{P_{loss}(T|x,t)}{P_{loss}(T|q,0)}. \tag{B2}$$

Using this result in Eq. (A15) for the force gives

$$F_{add}(x,t) = V(x)\frac{\int_0^1 \left[\frac{\partial}{\partial x}\left(\beta_1\frac{P_{fix}(T|x,t)}{P_{fix}(T|q,0)} + \beta_0\frac{P_{loss}(T|x,t)}{P_{loss}(T|q,0)}\right)\right] K_0(x,t|q)\alpha(q)dq}{\int_0^1 \left(\beta_1\frac{P_{fix}(T|x,t)}{P_{fix}(T|q,0)} + \beta_0\frac{P_{loss}(T|x,t)}{P_{loss}(T|q,0)}\right) K_0(x,t|q)\alpha(q)dq}. \tag{B3}$$

Special case: only a single initial frequency

A special case of the additional force in Eq. (B3) occurs when the initial distribution corresponds to only the presence of the single frequency y (that is $\alpha(q) = \delta(q - y)$). We then obtain the simpler result

$$F_{add}(x,t) = V(x)\frac{\frac{\partial}{\partial x}\left(\beta_1 \frac{P_{fix}(T|x,t)}{P_{fix}(T|y,0)} + \beta_0 \frac{P_{loss}(T|x,t)}{P_{loss}(T|y,0)}\right)}{\beta_1 \frac{P_{fix}(T|x,t)}{P_{fix}(T|y,0)} + \beta_0 \frac{P_{loss}(T|x,t)}{P_{loss}(T|y,0)}}. \tag{B4}$$

Assuming the time of attaining the target distribution, T, is finite, this additional force generally has time dependence.

Special case: only fixation occurs

A special case of the additional force in Eq. (B3) occurs when the final distribution corresponds to only the occurrence of fixation (that is $\beta_1 = 1$ and $\beta_0 = 0$). We then obtain the simpler result

$$F_{add}(x,t) = V(x)\frac{\partial}{\partial x}\ln\left[P_{fix}(T|x,t)\right]. \tag{B5}$$

Assuming the time of attaining the target distribution, T, is finite, this additional force will generally depend on the time, t.

Special case: $T \to \infty$

A different special case of the additional force of Eq. (B3) is the limit of

large T, corresponding to the target distribution being achieved at very long times. We shall assume the original force and the effective population size are independent of time. Then when $T \to \infty$, $P_{fix}(T|x,t)$ becomes $P_{fix}(x)$, the probability of ultimate fixation of the focal allele when the initial frequency is x and similarly, $P_{loss}(T|x,t)$ becomes $P_{loss}(x) = 1 - P_{fix}(x)$. We thus obtain

$$F_{add}(x,t) = V(x)\frac{\int_0^1 \left[\frac{\partial}{\partial x}\left(\beta_1 \frac{P_{fix}(x)}{P_{fix}(q)} + \beta_0 \frac{P_{loss}(x)}{P_{loss}(q)}\right)\right] K_0(x,t|q)\alpha(q)dq}{\int_0^1 \left(\beta_1 \frac{P_{fix}(x)}{P_{fix}(q)} + \beta_0 \frac{P_{loss}(x)}{P_{loss}(q)}\right) K_0(x,t|q)\alpha(q)dq}. \tag{B6}$$

Assuming an initial distribution, $\alpha(q)$, that covers a range of frequencies (not just a single frequency) we note that despite taking the large T limit ($T \to \infty$) the additional force in Eq. (B6) generally depends on the value of the time, t. At large values of the time, t, however, this additional force becomes independent of time (because when $t \to \infty$, the distribution $K_0(x,t|q)$ becomes independent of time: $K_0(x,t|q) \to \delta(1-x)P_{fix}(q) + \delta(x)P_{loss}(q)$).

Special case: $T \to \infty$, only a single initial frequency

We can specialise this last result further, to the case where, additionally, the initial distribution corresponds to only the single frequency y being present (that is $\alpha(q) = \delta(q-y)$). We then obtain the time independent

result

$$F_{add}(x,t) = V(x)\frac{\frac{\partial}{\partial x}\left(\beta_1\frac{P_{fix}(x)}{P_{fix}(y)} + \beta_0\frac{P_{loss}(x)}{P_{loss}(y)}\right)}{\beta_1\frac{P_{fix}(x)}{P_{fix}(y)} + \beta_0\frac{P_{loss}(x)}{P_{loss}(y)}},\ T\to\infty,\ \text{initial freq. is } y$$

$$= V(x)\frac{\left[\beta_1 - P_{fix}(y)\right]P'_{fix}(x)}{\beta_1 P_{fix}(x) + \beta_0 P_{fix}(y) - P_{fix}(x)P_{fix}(y)} \tag{B7}$$

where $P'_{fix}(x) = dP_{fix}(x)/dx$.

Deriving examples of the special case: $T \to \infty$, only a single initial frequency

We shall now derive some time-independent examples for the additional force, based on Eq. (B7).

Assuming a diploid population that behaves as though it undergoes random mating, with effective size N_e, we have $V(x) = x(1-x)/(2N_e)$ (Wright 1931, Gale 1990) and for additive selection, where each copy of the focal allele changes the relative fitness of its carrier by s, the original force is $F_0(x) = sx(1-x)$, with assumed negligible corrections of order s^2. The probability of ultimate fixation is then given by Kimura's (1962) result

$$P_{fix}(x) = \frac{1-e^{-Sx}}{1-e^{-S}} \tag{B8}$$

where $S = 4N_e s$.

To obtain the results in Eqs. (10) - (13) of the main text we have the following.

For $(\beta_1 = 1, \beta_0 = 0)$, Eq. (B7) yields $F_{add}(x) = \frac{V(x)P'_{fix}(x)}{P_{fix}(x)}$. Use of Eq. (B8) then leads to Eq. (10) of the main text.

For $(\beta_1 = rP_{fix}(y), \beta_0 = 1 - rP_{fix}(y))$, Eq. (B7) yields $F_{add}(x) = -\frac{V(x)(1-r)P'_{fix}(x)}{1-P_{fix}(x)+r\left[P_{fix}(x)-P_{fix}(y)\right]}$. Use of Eq. (B8) then leads to Eq. (11) of the main text.

For $S = 0$, Eq. (B8) reduces to $P_{fix}(x) = x$ and Eq. (B7) yields $F_{add}(x) = \frac{V(x)(\beta_1 - y)}{\beta_1(x-y)+y(1-x)}$ which is Eq. (12) of the main text.

For $(S = 0,\ \beta_1 = 0, \beta_0 = 1)$, Eq. (B7) yields $F_{add}(x) = -\frac{V(x)P'_{fix}(x)}{1-P_{fix}(x)}$. Since $P_{fix}(x) = x$, we obtain Eq. (13) of the main text.

Appendix C: Wright-Fisher simulations

In this appendix we give details of the simulations we used in this work, that were based on a Wright-Fisher model.

Starting with an effectively infinite population, where the frequency of a focal allele at time t is X_t, we assume that X_t changes according to $X_{t+1} = X_t + F_0(X_t)$ (thus $F_0(x)$ is the evolutionary force acting when the frequency is x).

In a finite diploid population of census size N the Wright-Fisher model

corresponds to the equation

$$X_{t+1} = \frac{\text{Bin}(2N, X_t + F_0(X_t))}{2N} \tag{C1}$$

where we use $\text{Bin}(n, p)$ to denote a random variable (*not* a distribution) that is drawn from a binomial distribution with parameters n and p, such that $\text{Bin}(n, p)$ takes the values $m = 0, 1, 2, ..., n$ with probabilities $\frac{n!}{(n-m)!m!}p^m(1-p)^{n-m}$.

Generating a large number of copies of trajectories (X_0, X_1, X_2, ...), from Eq. (C1), is the most basic form of the simulations adopted in this work.

When we incorporate the additional force into the simulations, we replace $F_0(X_t)$ in Eq. (C1) by $F_{total}(X_t, t) = F_0(X_t) + F_{add}(X_t, t)$, with $F_{add}(X_t, t)$ obtained from a diffusion approximation. This leads to simulations of a *hybrid nature* since the Wright-Fisher model has discrete generations and discrete frequencies, but the diffusion approximation treats both frequency and time as continuous quantities.

In principle, the diffusion approximation can lead to additional forces of large magnitude, but when a force as a function of frequency x, say $F(x)$, appears in a Wright-Fisher model, it always occurs in the combination

$x + F(x)$. This combination has the interpretation as a probability, which means that for all x we must have $0 \leq x + F(x) \leq 1$. When $0 \leq X_t + F_{total}(X_t, t) \leq 1$ we can directly use $F_{total}(X_t, t)$ in the simulations. When $X_t + F_{total}(X_t, t) < 0$ we reset it to 0 while when $X_t + F_{total}(X_t, t) > 1$ we reset it to 1.

Typically resetting of $X_t + F_{total}(X_t, t)$ is not necessary in the examples we have considered where the additional force is time independent. When the target distribution is to be achieved at a finite time T, the diffusion approximation of the additional force is time dependent and can become large. This occurs for trajectories that are far from probable target frequencies at times close to T. Because the additional force acts from early times, such trajectories are highly improbable, and resetting needs rarely to be performed.

References

Barbosa, S., W. C. Black and I. Hastings, 2011 Challenges in estimating insecticide selection pressures from mosquito field data. PLoS Neglected Tropical Diseases 5(11), e1387.

Buermans, H. P. J. and J. T. den Dunnen, J. T., 2014 Next generation sequencing technology: Advances and applications. Biochimca Biophysica Acta 1842: 1932–1941.

Buri, P., 1956 Gene Frequency in Small Populations of Mutant Drosophila. Evolution 10: 367-402.

Daetwyler, H. D., M. P. L. Calus, R. Pong-Wong, G. de Los Campos and J. M. Hickey, 2013 Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193: 347–365.

Ferri, M., E. Ranucci, P. Romagnoli and V. Giaccone, 2015 Antimicrobial Resistance: A Global Emerging Threat to Public Health Systems. Critical Reviews in Food Science and Nutrition, 0. [Epub ahead of print]

Fuller, R. A., J. Tratalos and K. J. Gaston, 2009 How many birds are there in a city of half a million people? Diversity and Distributions 15: 328–337.

Gale, J. S., 1990 Theoretical Population Genetics. Unwin Hyman, London.

Gantz, V. M., and E. Bier, 2015 Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations.

Science 348: 442–444.

Gossmann, T. I., M. Woolfit,and A. Eyre-Walker, 2011 Quantifying the variation in the effective population size within a genome. Genetics 189: 1389–1402.

Hartl, D. L., and A. G. Clark 1997 Principles of Population Genetics, 3rd Edition. Sinauer Associates, Sunderland, MA.

Kamareddine, L., 2012 The biological control of the malaria vector. Toxins 4: 748–767.

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

Kimura, M., 1964 Diffusion Models in Population Genetics. Journal of Applied Probability 1: 177-232.

Lacey, L. A. D. Grzywacz, D. I. Shapiro-Ilan, R. Frutos, M. Brownbridge and M. S. Goettel 2015 Insect pathogens as biological control agents: Back to the future. Journal of Invertebrate Pathology 132: 1–41.

de Los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler and M. P. L. Calus 2013 Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193: 327–345.

McFadyen, R. E., 1998 Biological control of weeds. Annual Review of Entomology 43: 369–393.

McKane, A. J., and D. Waxman, 2007 Singular solutions of the diffusion equation of population genetics. Journal of Theoretical Biology 247: 849–858.

Owen, M. D. K., and I. A. Zelaya 2005 Herbicide-resistant crops and weed resistance to herbicides. Pest Management Science 61: 301–311.

Reilly, J. R. and B. D. Elderd, 2014 Effects of biological control on long-term population dynamics: identifying unexpected outcomes. Journal of Applied Ecology 51: 90–101.

Schlötterer, C., R. Kofler, E. Versace, R. Tobler and S. U. Franssen, 2015 Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity (Edinb) 114: 431–440.

Smith, H. S., 1919 On Some Phases of Insect Control by the Biological Method. Journal of Economic Entomology 12: 288–292.

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard, E. Redoña, G. Atlin, J.-L. Jannink and S. R. McCouch 2015 Genomic selection and association mapping in rice (Oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genetics 11(2), e1004982.

Tuckwell, H. C., 1995 Elementary Applications of Probability Theory. 2nd Edition. Chapman and Hall, London.

Unckless, R. L., P. W. Messer, T. Connallon and A. G. Clark, 2015 Modeling the Manipulation of Natural Populations by the Mutagenic Chain Reaction. Genetics 201: 425–431.

Waxman, D., 2011 Comparison and content of the Wright–Fisher model of random genetic drift, the diffusion approximation, and an intermediate model. Journal of Theoretical Biology 269: 79–87.

Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97–159.

Zhao, L., M. Lascoux, A. D. J. Overall and D. Waxman, 2013 The characteristic trajectory of a fixing Allele: A consequence of fictitious selection that arises from conditioning. Genetics 195: 993-1006.

Zhao, L., X. Yue and D. Waxman, 2013 Complete numerical solution of the diffusion equation of random genetic drift. Genetics 194: 973–985.

Zimmermann, M. B. and M. Andersson, 2012 Assessment of iodine nutrition in populations: past, present, and future. Nutrition Reviews 70: 553–570.