



Universiteit
Leiden
The Netherlands

Calculating hazard rates of introgression with branching processes

Ghosh, A.

Citation

Ghosh, A. (2012, May 22). *Calculating hazard rates of introgression with branching processes*. Retrieved from <https://hdl.handle.net/1887/18976>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/18976>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/18976> holds various files of this Leiden University dissertation.

Author: Ghosh, Atiyo

Title: Calculating hazard rates of introgression with branching processes

Date: 2012-05-22

CHAPTER 3: QUANTIFYING TIME-INHOMOGENEOUS STOCHASTIC INTROGRESSION PROCESSES WITH HAZARD RATES

Reprinted with minor edits from Ghosh, Haccou, 2012. *Theor. Popul. Biol.*, 81, 253-263

ABSTRACT

Introgression is the permanent incorporation of genes from one population into another through hybridization and backcrossing. It is currently of particular concern as a possible mechanism for the spread of modified crop genes to wild populations. The hazard rate is the probability per time unit that such an escape takes place, given that it has not happened before. It is a quantitative measure of introgression risk that takes the stochastic elements inherent in introgression processes into account. We present a methodology to calculate the hazard rate for situations with time-varying gene flow from a crop to a large recipient wild population. As an illustration, several types of time-inhomogeneity are examined, including deterministic periodicity as well as random variation. Furthermore, we examine the effects of an extended fitness bottleneck of hybrids and backcrosses in combination with time-varying gene flow. It is found that bottlenecks decrease the hazard rate, but also slow down and delay its changes in reaction to changes in gene flow. Furthermore, we find that random variation in gene flow generates a lower hazard rate than analogous deterministic variation. We discuss the implications of our findings for crop management and introgression risk assessment.

1. INTRODUCTION

Through backcrossing and hybridization, genes from one population can become permanently incorporated into the genome of another population. This process is called introgression (Riesberg and Wendel, 1993; Ellstrand et al., 1999; Hails and Morley, 2005). Introgression of crop genes into wild relatives may have severe negative environmental effects, such as the spread of insecticide or herbicide resistance genes. In particular, there are strong concerns about transgene escape and its consequences, e.g. the production of superweeds (Maan, 1987; Snow et al., 1999; Thompson et al., 2003; Kelly et al., 2005).

The likelihood of such scenarios, given environmental conditions, crop management, and characteristics of the species involved can be studied with mathematical models. Such models allow us to perform thought experiments, and identify factors that crucially determine introgression risk. Introgression usually involves many random components, such as hybridization and backcross events, and demographic stochasticity in hybrid populations. In a previous paper (Ghosh and Haccou, 2010) we showed that it is important to take this stochasticity into account, since stochastic models may give very different predictions from deterministic ones. We considered a situation where foreign genes invade repeatedly

into a resident wild population, and each invasion has a small probability of establishing a permanent lineage (see also Haygood et al., 2004). We showed that there can be an extensive period of failed invasions, and that the length of this period largely determines introgression risk. Furthermore, we derived a measure, the hazard rate, that quantifies the distribution of such periods. In the context of introgression, the hazard rate is defined as the probability per time unit that a permanent lineage is initiated, given that this has not happened before. It is derived from a multitype branching process model of hybrid population dynamics (Demon et al., 2007; Serra and Haccou, 2007).

In our previous paper we assumed that the distribution of numbers of newly created hybrids is the same in each time period. We considered a model with an initial fitness bottleneck (i.e. F1 hybrids have a lower fitness than the wild type) and showed that in such a situation, the hazard rate increases monotonically from zero to a constant asymptotic value. As a consequence, the distribution of the initial period before establishment of a permanent lineage can be approximated by a time-lagged geometric distribution. In many applications, however, the hybridization probability will vary in time, due to, for example, crop rotation or termination, or random variation, such as weather-dependent pollinator activity. In the current paper we generalize the method to include such time-inhomogeneity. We calculate the hazard rate for general time-inhomogeneous hybridization schemes and examine the effects of crop management schemes such as (gradually) stopping or increasing crop cultivation, or rotating crops. We show that, in the latter case, periods in which the hazard rate increases alternate with periods of decrease, and that, in the long run, it converges to a periodic function. We also examine how stochastic fluctuations in hybridization rates affect the hazard rate.

As an example we consider a model for a monocarpic species (it dies after flowering), that is monoecious (flowers have both male and female functions), and non-selfing. We first consider a situation where F1 hybrids have a reduced fitness when compared to the wild-type, and all backcrosses have the same life history parameters, and superior fitness. Then the model is generalized to examine the effects of an extended fitness bottleneck, where several initial backcross generations have a reduced fitness.

There are many other contexts in which repeated invasions with low initial fitness occur, such as tumor spread and growth, where usually several mutations must occur before cells proliferate (as in Michor et al., 2006), or pathogen host switching, where adjustments to new hosts imply an initial fitness bottleneck (as in Reluga et al., 2007). Time-inhomogeneity of invasions may play a role in such contexts too. For instance, there may be time-varying risks of exposure to carcinogenic environments (e.g. Bos et al., 2004). Furthermore, many epidemics show time-varying infection patterns (as in Welliver, 2009). Our methods and results therefore have implications for research in such contexts too.

2. THE MODEL

We consider a plant species that dies after flowering once. For simplicity, we assume that there is no age-dependence. Furthermore, it is assumed that there is a large, stable wild population, and random numbers of hybrid seeds are produced by pollen flow from a nearby crop. We consider time periods of one year. Seeds

may germinate at the beginning of the year, and plants grow up to be adults and may flower later in the same year. We denote the probability that a seed germinates and that the seedling survives to become an adult plant by p_0 . In this paper we will consider p_0 as a given parameter. Its value is determined by the population dynamics of the wild population, and is such that this population is stable (see Ghosh and Haccou, 2010, for an example of its calculation).

Hybrid formation can be followed by repeated backcrossing with wild plants. F1 hybrids are assumed to be less fit than wild individuals, but backcrossed individuals have a positive probability of producing a permanent introgressed lineage. We assume that all backcross generations are equivalent with respect to their life history parameters, and therefore they do not need to be distinguished as separate types (this assumption is relaxed in section 6). As a consequence, there are two types of plants in the model: F1 hybrids (labelled type-1) and backcrossed individuals (labelled type- E).

Since the population of wild plants is large and the numbers of individuals containing crop genes are initially small, it can be assumed that these individuals do not interact with each other, but only with wild plants. This has several implications. Firstly, since we consider a non-selfing species, reproduction can only occur through outcrossing with wild plants. Secondly, competition occurs only with the wild population. This is quantified through the probability p_0 . For convenience, we assume that there are no other factors apart from this competition that affect germination probability of hybrids and backcrosses. The model can be easily generalized to account for e.g. effects of spatial variation.

Because hybrid and backcrossed plants do not affect each other's reproduction and survival initially, their invasion dynamics can be modeled as a branching process. The production of hybrid seeds is modeled by means of an artificial type, which we will call type-0. There is one permanently present individual of this type, that produces a stochastic number of hybrid seeds in each year. Fig. 1 shows a schematic summary of the invasion dynamics.

The model thus involves three different types of individuals: type-0, type-1 and type- E . Each year, a type-0 individual produces one individual of type-0 and a random number of F1 hybrid seeds. In our previous paper we assumed that the probability distribution of these random numbers was the same over time. In this paper, we let it vary over years. The number of hybrid seeds produced in year k is a random variable denoted by $\xi_{0,k}$. Each one of these seeds germinates and produces a type-1 individual with probability p_0 . Type-1 individuals flower with probability r_1 , and produce a random number, ξ_1 , of backcrossed seeds, either by male or female functions. In the case that a type-1 individual does not flower (with a probability $(1 - r_1)$), it may then survive to become a type-1 individual in the next year with probability p_1 , or it will die with a probability $1 - p_1$. Each backcrossed seed germinates and survives with probability p_0 , to produce a type- E individual. Type- E individuals produce only type- E offspring in their lineage. We denote the probability that a lineage started by one type- E individual goes extinct by q . This value can be calculated straightforwardly from the life history parameters of type- E individuals, by standard methods (see e.g. Haccou et al., 2005; Ghosh and Haccou, 2010). Here, we will treat it as a parameter in the model, taking values between zero and one.

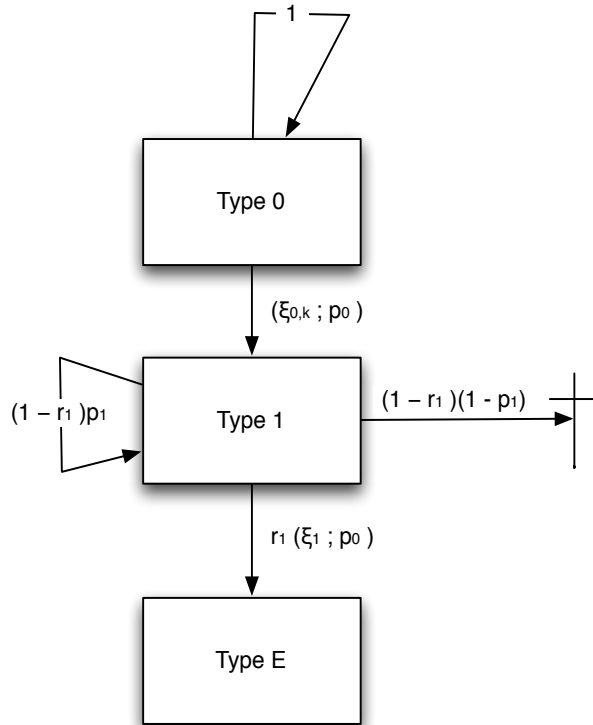


FIGURE 1. Schematic representation of the model. $(\xi_{0,k}; p_0)$ and $(\xi_1; p_0)$ represent the production of $\xi_{0,k}$ and ξ_1 seeds respectively, where each seed has a germination probability p_0 . Each type- E individual initiates a lineage which eventually becomes extinct with probability q .

3. DERIVATION OF THE HAZARD RATE

Probability generating functions are important tools in deriving the hazard rate. Let X be a non-negative discrete random variable, then its probability generating function (p.g.f.) is a function from $[0, 1]$ to $[0, 1]$ which is defined as $E[s^X]$, where $E[\cdot]$ denotes expectation. The p.g.f. of $\xi_{0,k}$ is denoted by $G_0(k; s)$, and that of ξ_1 by $G_1(s)$.

Define the random variable $I_i(k, n)$ ($n, k \in \mathbb{N}_0$, $i = 0, 1$) to be the total number of type- E individuals with non type- E parents, appearing up to and including year n , in the *line of descent* of a single individual of type- i that was produced in year k . The expression *line of descent* refers to the population process stemming from the referred individual. For a general scenario where individuals can have

offspring of any type, this definition leads to the following equalities:

$$I_i(k, n) = \begin{cases} 0 & \text{if } k \geq n \\ Z_E^{(i)}(k+1) + \sum_{m=0}^1 Z_m^{(i)}(k+1) \sum_{j=1}^1 I_m^{(j)}(k+1, n) & \text{if } k < n \end{cases} \quad (1)$$

where $Z_m^{(i)}(k+1)$ represents the number of type- m individuals that the type- i individual (born in year k) produced in year $k+1$. The $I_m^{(j)}(k+1, n)$ terms represent the total number of type- E individuals that have non type- E parents, appearing up to year n in the line of descent of the j^{th} individual of type- m that was born in year $k+1$ from the initial type- i individual.

In the specific scenario described in Fig. 1, we find the following recursive relationships in k for the different p.g.f.'s of the $I_i(k, n)$'s, where $f_{I_i(k, n)}(s)$ denotes the p.g.f. of $I_i(k, n)$ (see Appendix A.1):

$$\begin{aligned} f_{I_0(k, n)}(s) &= f_{I_0(k+1, n)}(s)G_0(k; p_0 f_{I_1(k+1, n)}(s) + 1 - p_0) \\ f_{I_1(k, n)}(s) &= (1 - r_1)(1 - p_1) + (1 - r_1)p_1 f_{I_1(k+1, n)}(s) + r_1 G_1(p_0 s + 1 - p_0) \end{aligned} \quad (2)$$

with the initial conditions $f_{I_1(n, n)}(s) = f_{I_0(n, n)}(s) = 1$. Note that, since the seed production of type-1 individuals is homogeneous,

$$f_{I_1(k, n)}(s) = f_{I_1(0, n-k)}(s). \quad (3)$$

The time of an introgression event, T , is defined as the time that the first type- E individual is produced whose lineage never becomes extinct. The population starts with a single type-0 individual, therefore:

$$P(T > n) = f_{I_0(0, n)}(q), \quad (4)$$

since the probability that an introgression event occurs after a time n is the probability that all type- E individuals produced at or before year n have become extinct.

The hazard rate of introgression is defined as the probability per time unit that an introgression event occurs given that it has not occurred before. With time units of one year, this gives:

$$H_n(q) = P(T = n | T > n - 1) / \text{year} = \left(1 - \frac{f_{I_0(0, n)}(q)}{f_{I_0(0, n-1)}(q)} \right) \text{year}^{-1} \quad (5)$$

with $n \in \mathbb{N}_0$.

The second equation of (2) can be solved to yield (see Appendix A.2):

$$f_{I_1(0, n)}(s) = 1 - \beta_1(s) + \beta_1(s) b_1^n, \quad (6)$$

where, in order to simplify future expressions, we have introduced the quantities

$$b_1 = (1 - r_1) p_1 \quad \text{and} \quad \beta_1(s) = \frac{r_1 (1 - G_1(p_0 s + 1 - p_0))}{1 - b_1} \quad (7)$$

Putting (2), (3), (4) and (5) together gives us the following expression for the hazard rate (see Appendix A.3):

$$H_n(q) = \begin{cases} 0 & \text{if } n \in \{0, 1\} \\ 1 - \frac{\prod_{j=1}^{n-1} G_0(j-1; p_0 f_{I_1(0, n-j)}(q) + 1 - p_0)}{n-2} & \text{if } n \geq 2 \\ \prod_{j=1}^{n-2} G_0(j-1; p_0 f_{I_1(0, n-1-j)}(q) + 1 - p_0) & \end{cases} \quad (8)$$

which can be computed by using (6). This result provides us with a general method for calculating the hazard rate with time-inhomogeneous hybridization. In the next sections we examine several situations.

4. DETERMINISTICALLY VARYING HYBRIDIZATION

For mathematical convenience we assume that hybrids are generated according to a Poisson distribution with a time-dependent mean, i.e.:

$$G_0(k; s) = e^{-m_0(k)(1-s)}, \quad s \in [0, 1]. \quad (9)$$

We also take ξ_1 as Poisson-distributed with mean m_1 in presented numerical work. Combining (6) to (9) gives:

$$H_n(q) = \begin{cases} 0 & \text{if } n \in \{0, 1\} \\ 1 - e^{-p_0 \beta_1(q)(1-b_1) b_1^{n-2} \sum_{j=0}^{n-2} m_0(j) b_1^{-j}} & \text{if } n \geq 2. \end{cases} \quad (10)$$

From (10) it follows that the long term behaviour of the hazard rate depends on the limit behaviour, as $k \rightarrow \infty$, of:

$$b_1^k \sum_{j=0}^k \frac{m_0(j)}{b_1^j}.$$

For example, if $m_0(j) = m_0^j$, the hazard rate converges to zero when $0 < m_0 < 1$ and it converges to one when $m_0 > 1$. If there is constant hybridization, i.e. $m_0(j) = m_0$, the hazard rate tends to a constant value between zero and one (as was also derived in Ghosh and Haccou, 2010). It can easily be shown that, for the current model, this value equals

$$1 - \exp\{-p_0 \beta_1(q) m_0\}. \quad (11)$$

In the next subsections we will examine the effects of specific frequently used crop-management schemes.

4.1. Temporary crops. Crop cultivation may be stopped for a variety of reasons. In the case of transgene crops, e.g., legislation may change, or termination of cultivation may be used as a management strategy to lower the chance of introgression. In this sub-section we examine the case where hybridization occurs at a constant rate, and is then stopped at a fixed time S , i.e.:

$$m_0(j) = \begin{cases} m_0 & \text{if } 0 \leq j < S \\ 0 & \text{if } j \geq S, \end{cases} \quad (12)$$

with $m_0 > 0$.

Substituting this into (10) gives:

$$H_n(q) = \begin{cases} 0 & \text{if } n \in \{0, 1\} \\ 1 - e^{-m_0 p_0 \beta_1(q) (1-b_1^{n-1})} & \text{if } 2 \leq n \leq S+1 \\ 1 - e^{-m_0 p_0 \beta_1(q) b_1^{n-(S+1)} (1-b_1^S)} & \text{if } n \geq S+2 \end{cases} \quad (13)$$

Thus, the hazard rate increases monotonically to a maximum level of $1 - e^{-m_0 p_0 \beta_1(q) (1-b_1^S)}$ at time $S+1$ and decays monotonically afterwards. The decay is only seen to start at time $S+2$ because stopping hybridization at year S will only affect the population of type-1 individuals at time $S+1$, and the population of type- E individuals at time $S+2$. The rate of increase as well as that of decay is mainly governed by b_1 , which represents the probability that individuals do not flower but do survive (see (7)). A larger value of b_1 makes the hazard rate increase and decrease more slowly. When b_1 tends to zero (i.e. when the probability of flowering in the first year is high and/or the survival probability of non-flowering adults is low), the maximum level is reached quickly and, unless S is very small, it is therefore virtually independent of S . Furthermore, after stopping cultivation, the hazard rate returns rapidly to zero. As b_1 tends to zero or S tends to infinity, the maximum level approaches the asymptotic level of the hazard rate in the situation without stopping. The effect of the life history parameters on this asymptotic level can be inferred from (11).

With temporary crops, there is a positive probability that introgression never occurs. From (4), (9), (12) and the derivation in Appendix A.3 it is apparent that this probability equals:

$$\lim_{n \rightarrow \infty} P(T > n) = \lim_{n \rightarrow \infty} f_{I_0(0,n)}(q) = e^{-m_0 p_0 \beta_1(q) S} \quad (14)$$

Thus, it decreases exponentially with the stopping time S , at a rate determined by the hybridization rate and the life history parameters.

A numerical example of the shape of the hazard rate for two different stopping times (10 and 20 years) is given in Fig. 2a. In this example, the hazard rate increases quickly, and, as a consequence, its maximum level does not noticeably differ for the two chosen stopping times. The probability distribution of T can be expressed in terms of the hazard rate as follows (see e.g. Kalbfleisch and Prentice, 2002):

$$P(T = x) = \prod_{i=0}^{x-1} (1 - H_i(q)) H_x(q). \quad (15)$$

For small values of $H_n(q)$, the product term is close to one, and the probability becomes nearly equal to the hazard rate. This is demonstrated in Fig. 2b. As can be seen from the figure, the probabilities of introgression events happening quite early are relatively large, i.e. the probability distributions are very skewed, similar to the situation with constant crop cultivation examined before in Ghosh and Haccou (2010). For the numerical examples in Fig. 2b, the probabilities that no introgression occurs at all are respectively 0.985 ($S = 10$) and 0.970 ($S = 20$).

4.2. Crop rotation. Crop rotation is often used to maintain soil quality and prevent the build up of pathogens. It may also be used as a management strategy to lower introgression risk. In this section we study the situation where periods

with hybridization at a constant rate alternate with periods without hybridization. The duration of hybridization periods is denoted by S , and the durations of the *hybridization pauses* by R . Thus we have:

$$m_0(j) = \begin{cases} m_0 & \text{if } v(R+S) \leq j < v(R+S)+S \\ 0 & \text{if } v(R+S)+S \leq j < (v+1)(R+S) \end{cases} \quad (16)$$

with $v \in \mathbb{N}_0$.

It can be shown (see Appendix A.4) that in the long run the hazard rate tends to a periodic function with period $R+S$, i.e. if we define the time:

$$k = n - v(R+S) - 2 \quad (17)$$

then, for n tends to infinity the hazard rate becomes:

$$\mathcal{H}_k(q) = \begin{cases} 1 - e^{-m_0 p_0 \beta_1(q) \left(1 - b_1^{k+1} \frac{1 - b_1^R}{1 - b_1^{R+S}}\right)} & \text{if } 0 \leq k < S \\ 1 - e^{-m_0 p_0 \beta_1(q) b_1^{k+1-S} \frac{(1 - b_1^S)}{1 - b_1^{(R+S)}}} & \text{if } S \leq k < R+S \end{cases} \quad (18)$$

The time in (17) is the time after the v th crop rotation shifted by two time units. The shift of two units is for mathematical convenience, and corresponds for the first two years where the hazard rate is zero.

This result implies that periods in which instantaneous introgression risk is high alternate with periods in which it is low. Figure 2c illustrates that this asymptotic behavior can be reached very quickly. Figure 2d shows the corresponding probabilities of introgression events happening at time x . As noted previously, the probability distribution is nearly equal to the hazard rate initially, but (inevitably) decreases with x .

There are different ways to quantify the effect of a given crop rotation scheme on the hazard rate. The asymptotic maximum hazard rate can be found by substituting $k = S - 1$ in (18), leading to:

$$1 - e^{-m_0 p_0 \beta_1(q) \frac{1 - b_1^S}{1 - b_1^{(R+S)}}}, \quad (19)$$

and the minimum by substituting $k = R + S - 1$, which gives:

$$1 - e^{-m_0 p_0 \beta_1(q) b_1^R \frac{1 - b_1^S}{1 - b_1^{(R+S)}}}. \quad (20)$$

For the numerical example in Figure 2c the asymptotic maximum hazard rate equals 0.00154, and the minimum is of the order 10^{-6} . As can be seen from the figure, these values are reached quite soon.

An alternative measure is the long-run average hazard rate. This is found by fitting the survivor function of a constant hazard rate to the survivor function of the hazard rate from (18). This approach leads to the following value for the long-run average hazard rate (see Appendix A.5 for details):

$$\lambda \approx 1 - e^{-p_0 m_0 \beta_1(q) \frac{S}{R+S}}. \quad (21)$$

Thus, the long-run average hazard rate is the same as the asymptotic hazard rate with a continuous crop and a constant expected number of newly produced hybrids equal to $S/(R+S)$ times m_0 . In Fig. 2d we have indicated the time-distributions

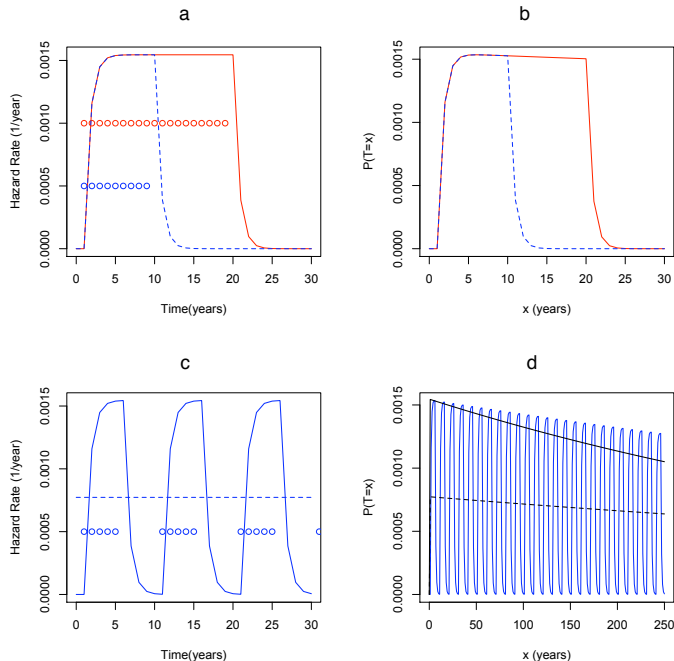


FIGURE 2. (a) Hazard rates when crops are terminated after a period of $S = 10$ (blue), or $S = 20$ (red). Parameter values: $m_0 = 50$, $p_0 = 0.001$, $p_1 = r_1 = 0.5$, $m_1 = 950$, $q = 0.95$, (b) The distributions of time until an introgression event, corresponding to the situations in (a). (c) The hazard rate with crop rotation (see (A.9) and (A.10)) (solid line) for $R = S = 5$ and all other relevant parameters the same as in (a). The average hazard rate (see (21)) (dotted line). (d) Distribution of times until an introgression event for the crop rotation scenario of (c) (blue line), for a constant average hazard rate (dotted black line), and for a constant maximum hazard rate (see (19), solid black line). In (a) and (c), circles indicate periods when hybridization occurs, but not the amount of immigration.

corresponding to a continuous immigration with the maximum hazard rate (c.f. (19)) and the long-run average hazard rate.

5. RANDOMLY VARYING HYBRIDIZATION

Until now we have considered deterministic variation in hybridization rates. In many cases, however, there will also be random variation. For instance, weather conditions will vary over different years, and this may affect pollen dispersal from the crop to local wild populations. Such random variations can be independent, or (positively or negatively) autocorrelated. In this section, we consider the effect of random variation according to different regimes.

Random temporal variation of m_0 can be included in the model by using different type-0 individuals. Thus, we consider γ different types, denoted by type-(0, i) ($i = 1, \dots, \gamma$). A type-(0, i) individual produces a number of type-1 seeds according to a p.g.f. $G_{0,i}(s)$, and with probability $\kappa_{i,j}$ also exactly one individual of type-(0, j) ($j = 1, \dots, \gamma$), so $\sum_{j=1}^{\gamma} \kappa_{i,j} = 1$ for all i .

As an illustration, consider the case where the environment alternates between two states according to a two-type Markov chain. In that case $\gamma = 2$. When the environment is state 1, a Poisson-distributed number of hybrids is formed, i.e. $G_{0,1}(s) = e^{-m_0(1-s)}$ and when the environment is in state 2, no hybrids are produced, i.e. $G_{0,2}(s) = 1$. The transition probability from state 1 to state 2 equals $\kappa_{1,2}$ and that from state 2 to state 1 equals $\kappa_{2,1}$. An independently varying environment corresponds to the situation where $\kappa_{1,2} + \kappa_{2,1} = 1$. In the case of positive autocorrelation, this sum is smaller than one whereas it is larger than one for negatively autocorrelated environments.

As a special case, consider an independently varying environment, with $\kappa_{1,1} = \kappa_{2,1} = S/(R+S)$ and $\kappa_{1,2} = \kappa_{2,2} = R/(R+S)$. Note that the expected proportion of years with positive hybridization numbers is the same as in the crop rotation scenario considered in (16). We assume that the process is stationary. The hazard rate is then given by (see Appendix A.6)

$$H_n(q) = \frac{S}{R+S} \left(1 - e^{-m_0 p_0 (1 - f_{I_1(0, n-1)}(q))} \right). \quad (22)$$

Using the solution of $f_{I_1(0, n)}(q)$ from (6) and taking large n leads to the asymptotic value:

$$H_{\infty}(q) = \frac{S}{R+S} \left[1 - e^{-m_0 p_0 \beta_1(q)} \right]. \quad (23)$$

To examine the effects of autocorrelation, let $\kappa_{1,2} = \kappa_{2,1} = 1 - \kappa_{1,1} = 1 - \kappa_{2,2} = \alpha$. The environment is negatively autocorrelated if $\alpha > 0.5$, positively autocorrelated if $\alpha < 0.5$, and independent if $\alpha = 0.5$. The equations given in Appendix A.6 can be used to calculate the hazard rate for these models numerically. Figure 3a shows the resulting asymptotic hazard rate for different values of α . As can be seen, there is not much difference between negatively autocorrelated or independent environments. The asymptotic hazard rate is much reduced, however, when there is a strong positive autocorrelation. With this choice of parameters, the probability of a year with hybridization is 1/2, and so the situation is comparable to a crop rotation scenario with $S = R$, as in Fig. 2(c). Note that the situation where $\alpha = 1$ corresponds to deterministic alternation between one-year periods with and without a positive hybridization probability. In this scenario, the hazard rate still approaches an asymptotic hazard rate because the process is initiated by the stationary-distribution of type-(0, 1) and type-(0, 2) individuals, as depicted in Fig. 3b. In a specific realisation, the hazard rate then oscillates as previously observed, which is also shown in Fig. 3b, where the process is initiated by a single type-(0, 1) individual.

6. EFFECTS OF BOTTLENECKS

Until now we have considered the situation where all backcrossed generations are more fit than the wild type. However, often there is outbreeding depression, which implies that several backcrosses are needed before a fitness advantage is

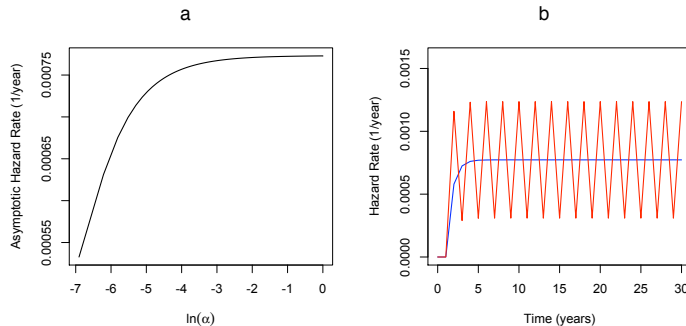


FIGURE 3. (a) The effect of autocorrelation on the asymptotic hazard rate when $k_{1,2} = k_{2,1} = \alpha = 1 - k_{2,2} = 1 - k_{1,1}$, $m_{0,1} = 50$, $m_{0,2} = 0$, and other parameter values as in Fig.2. The environment is positively autocorrelated when $\ln \alpha < \ln 2 (\approx -0.69)$ and negatively autocorrelated when $\ln \alpha > \ln 2$. Periods with and without positive hybridization probabilities alternate deterministically when $\ln \alpha = 0$. (b) The hazard rate at $\alpha = 1$ when the process is started with a stationary distribution of type-(0, 1) and type-(0, 2) individuals (blue), and when the process is started with a single type-(0, 1) individual (red).

observed (e.g. Edmands, 2002). In this section we extend the model to account for such situations, and investigate effects of the length of the bottleneck on the hazard rate.

The generalized model involves $L + 2$ ($L \in \mathbb{N}$) different types: types 0, 1, ..., L , and type- E . Type-0 individuals are defined as before. The flowering probability of type- i ($i \in \{1, 2, \dots, L\}$) is denoted by r_i , the p.g.f. of their seed production by $G_i(s)$ and their seeds will produce type- $(i + 1)$ adults. The survival probability of non-flowering type- i individuals is p_i , and survivors remain of type i . The offspring of type- L individuals will be of type- E . Type- E individuals and q are defined as in previous sections. The scheme is represented in Fig. 4.

The hazard rate in this scenario follows a similar method to the derivation in the previous case, but see Appendix A.7 for full details. Numerical solutions of the supremum of the hazard rate against L are shown in Fig. 5a for the crop-rotation situation described in (16).

To further examine the effect of bottlenecks, we consider a Taylor approximation of the hazard rate around the point $q = 1$, for the case that plants are annual (i.e. $r_i = 1$ for $i = 1, 2, \dots, L$). The resulting Taylor approximation is (see A.8 for details):

$$H_n(q) \approx \left(p_0 m_0 (n - L - 1) \prod_{i=1}^L p_0 m_i \right) (1 - q) \quad (24)$$

where m_i , $i = 1, 2, \dots, L$, represents the average number of seeds produced by a type- i individual.

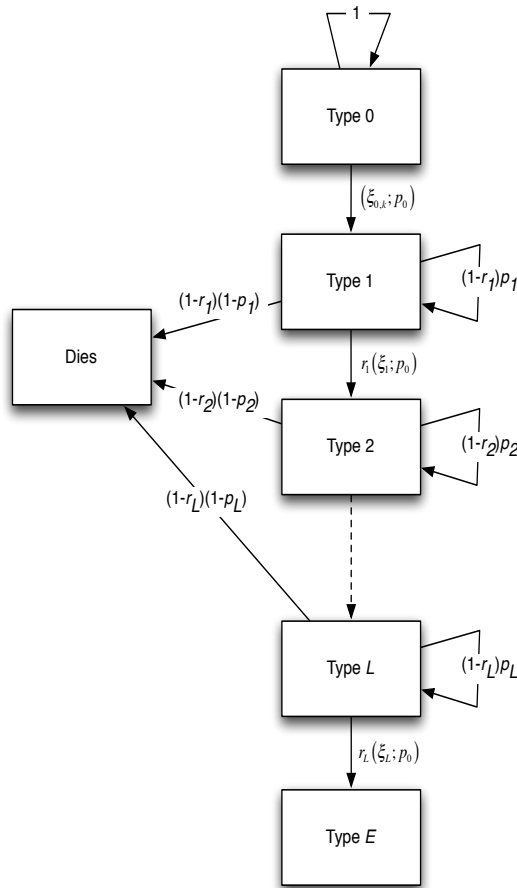


FIGURE 4. Schematic representation of the bottleneck model. $(\xi_i; p_0)$ represents the production of ξ_i seeds $i \in (0, k) \cup \{1, 2, \dots, L\}$, where each seed has a germination probability p_0 . Each type- E individual initiates a lineage which eventually becomes extinct with probability q .

When the values of m_i are similar, this expression decreases geometrically with L , which corresponds to the shape observed in Fig. 5a.

Bottlenecks not only reduce the maximum hazard rate, but also induce a delay in the changes of the hazard rate in reaction to changes in crop cultivation. This is illustrated in Fig. 5b.

7. DISCUSSION

In this paper we generalize our previous results on hazard rates of introgression (Ghosh and Haccou, 2010) to situations with time-varying hybridization. Whereas in our previous paper we considered a model with two age classes and a bottleneck of one generation, the present paper concerns situations without age dependence,

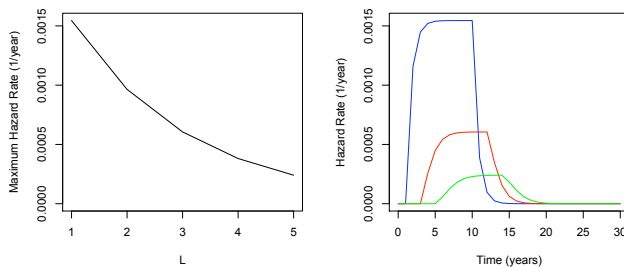


FIGURE 5. (a) The maximum hazard rate as a function of the length of the bottleneck L for a crop rotation scenario with $R = S = 5$, $m_0 = 50$, $p_0 = 0.001$, $p_i = r_i = 0.5$, $m_i = 950$ for $i = 1, 2, \dots, L$ and $q = 0.95$. (b) The hazard rate against time with hybridization as described in (12) with $S = 10$ and all other parameters as in (a). The behaviour for $L = 1$ (blue) $L = 3$ (red) and $L = 5$ (green) is shown.

and effects of extended bottlenecks. The general methodology that we present can be extended straightforwardly to other types of life histories. Furthermore, there are several general conclusions that are valid for a wide range of situations.

First of all, the results shed light on the meaning of the hazard rate as a measure of stochastic introgression rate, and its practical implications. As illustrated in this paper, hazard rates may increase and decrease in time, in relation to changes in the magnitude of hybridization rates. When the hybridization rate is high, the instantaneous risk of introgression events is also high. During such periods, increased vigilance is advisable, to prevent the successful establishment of crop genes in wild populations. When hybridization frequency drops, the hazard rate decreases, and accordingly, vigilance might be decreased. Our results show, however, that managers must take care not to let their guards down too soon, since increased fitness bottlenecks delay the changes in the hazard rate. This implies, for instance, that even after crop cultivation has been terminated for a considerable time, the risk of introgression events may still be quite high (see Fig. 5b), reaffirming a conclusion from Haygood et al. (2003).

The risk that introgression occurs is determined by the interaction between life history and fitness characteristics of hybrids, and crop management. As we illustrated, changes in gene flow induce changes in the level of the hazard rate. The speed at which such changes take place, as well as the magnitude of the hazard rate depends on life-history characteristics. For instance, increases in fitness bottlenecks not only cause a delay in adjustment of the hazard rate, but also decelerate the adjustments, and lower the maximum level. Furthermore, in all scenarios, the maximum level of the hazard rate is affected by the factor $\beta_1(q)$, which is determined by the fitness of the backcrosses (see (7)).

We examined the effect of several possible scenarios. With temporary crops, there is a positive probability that introgression does not occur, that depends on the duration of the crop cultivation. Furthermore, in this situation, the hazard

rate at a given time x is nearly equal to the probability of an introgression event at that time, and thus provides a good approximation for the probability distribution (see e.g. Fig. 2b). This is a general result, that can be derived from the relation between the hazard rate and the time-distribution.

With crop rotation, the hazard rate becomes periodic, and fluctuations also occur in the time-distribution of introgression events (Fig. 2c and d). In such situations, a simpler measure of risk might sometimes be needed. One option is to use the hazard rate that in the long run would lead to the same introgression risk over a given period as the crop rotation scheme. This value is given in (21), and indicated in Fig. 2c. We refer to this value as the long-run average hazard rate. However, please note that it is not the same as the arithmetic time-average of the asymptotic hazard rate. From (21) it can be seen that the average risk level is determined by the proportion of years that crop cultivation occurs. Thus, the average hazard rate remains the same when S and R are multiplied by the same factor. For instance, alternating between one year 'on' and 'off' would in the long run give the same average hazard rate as alternating between, say, ten years 'on' and 'off'. Larger values of S and R would, however, lead to a larger amplitude of the fluctuations in the hazard rate. The magnitude of this effect can be calculated by means of (19) and (20). In situations with large fluctuations the use of the average hazard rate as a risk indicator might be misleading, since the maximum hazard rate is much higher than the average. This is illustrated in Fig. 2c. In such a situation, the time-distribution of introgression events corresponding to the average hazard rate is also radically different from the real one (see Fig. 2d).

Another possible way to quantify the risk is to use the long-run maximum hazard rate, which provides a conservative measure of risk. Figure 2d also shows the time-distribution of introgression events corresponding to the maximum hazard rate, illustrating that in an example with large amplitude of the hazard rate this might be a better risk measure.

We also derived methods to calculate the hazard rate in situations with randomly varying hybridization rates. As a specific example, we considered a situation where the environment alternates between two states, one with and one without hybridization, according to a Markov chain. In the absence of environmental autocorrelation, the hazard rate becomes constant in the long run, and an explicit expression can be derived. This value is given in (23), and corresponds to the arithmetic time-average of the asymptotic hazard rate in a deterministic crop rotation scheme with the same proportion of years of hybridization as the random environment. It can be shown that this value is lower than the long-run average hazard rate given in (21). Therefore, random variation in gene flow appears to reduce the probability that introgression occurs. This also appears to be true in autocorrelated environments, as illustrated in Fig.3. Positive autocorrelation reduces the hazard rate, whereas negative autocorrelation does not seem to have much effect. In any case, the long-run hazard rate is smaller than the long-run average for the deterministically alternating environment. Thus, we expect that hazard rates for deterministic scenarios provide conservative measures for introgression risk. This is a fortunate result, since in many situations there is likely to be random variation in gene flow, which is beyond control of management measures.

We examined several specific gene flow scenarios, to illustrate the methodology and its possibilities. For mathematical tractability, we used a relatively simple life-history and Poisson distributions for the numbers of hybrids. Our methods can readily be adjusted to examine other types of gene flow variation, more complicated life histories, and hybrid number distributions. In such cases, however, no explicit expressions for (asymptotic) hazard rates will be possible. Instead, numerical methods will have to be used, based on the adjusted equations. Such calculations generally do not take much time on a standard computer.

Other generalizations, which are the subject of ongoing research, include the introduction of time-inhomogeneity in backcross fitness, multi-locus genetics, and meta-population dynamics. Another type of generalization concerns small populations. As long as wild receptor populations are assumed to be large enough to exclude direct interactions between initial invaders, the approach that we used up to now, based on branching processes, can be applied. For small populations however, different methods need to be developed, based on density-dependent models (see (e.g. Jagers and Klebaner, 2000)). This is another line of ongoing research.

The use of stochastic models in introgression studies is quite rare, although not completely absent (e.g. Haygood et al., 2004; Thompson et al., 2003). The general methodology for handling such models, and quantifying introgression timing events is, however, still in its infancy. The use of hazard rates is, in our opinion, an important step forward. Serra and Haccou (2007) introduced the concept of the hazard rate for studying branching processes with mutation, and Ghosh and Haccou (2010) were the first to use it in the context of introgression. The work presented here represents the next step of a research program that is aimed at developing a full-fledged toolbox for studying stochastic introgression processes. Such tools are indispensable in introgression risk management, since stochastic elements are inevitably present, and, furthermore, adding stochasticity changes the features of introgression processes considerably.

APPENDIX A. APPENDIX

A.1. **Derivation of (2).** Using (1) and the definition of p.g.f.'s we find:

$$\begin{aligned}
 f_{I_i(k,n)}(s) &= E \left[E \left[s^{I_i(k,n)} | Z_0^{(i)}(k+1), Z_1^{(i)}(k+1), Z_E^{(i)}(k+1) \right] \right] \\
 &= E \left[E[s^{I_0(k+1,n)}]^{Z_0^{(i)}(k+1)} E[s^{I_1(k+1,n)}]^{Z_1^{(i)}(k+1)} E[s^{Z_E^{(i)}(k+1)}] \right] \\
 &= E \left[f_{I_0(k+1,n)}(s)^{Z_0^{(i)}(k+1)} f_{I_1(k+1,n)}(s)^{Z_1^{(i)}(k+1)} s^{Z_E^{(i)}(k+1)} \right] \quad (\text{A.1})
 \end{aligned}$$

We can manipulate (A.1) as above because the individual lineages are independent of each other, and individuals of the same type have identical offspring distributions.

Now we introduce the joint p.g.f of the reproduction distribution of a type- i individual belonging to a year k which, for $i \in \{0, 1\}$ and $k \geq 0$, is defined as

$$F_i(k; (s_0, s_1, s_E)) = E \left[s_0^{Z_0^{(i)}(k+1)} s_1^{Z_1^{(i)}(k+1)} s_E^{Z_E^{(i)}(k+1)} \right] \quad (\text{A.2})$$

for $(s_0, s_1, s_E) \in [0, 1]^3$.

Putting (A.1) and (A.2) together, we find that

$$f_{I_i(k,n)}(s) = F_i(k; (f_{I_0(k+1,n)}(s), f_{I_1(k+1,n)}(s), s)) \quad (\text{A.3})$$

In our specific model, we have the following reproduction laws:

$$F_0(k; (s_0, s_1, s_E)) = s_0 G_0(k; p_0 s_1 + (1 - p_0)) \quad (\text{A.4})$$

$$F_1(k; (s_0, s_1, s_E)) = (1 - r_1)(1 - p_1) + (1 - r_1)p_1 s_1 + r_1 G_1(p_0 s_E + 1 - p_0). \quad (\text{A.5})$$

Substituting (A.5) and (A.4) into (A.3) gives (2).

A.2. **Derivation of (6).** Since the population initiated by a type-1 individual is time-homogeneous, $f_{I_1(k,n)}(s) = f_{I_1(0,n-k)}(s)$. Using this in the second equation of (2) results in:

$$f_{I_1(0,n-k)}(s) = (1 - r_1)(1 - p_1) + (1 - r_1)p_1 f_{I_1(0,n-k-1)}(s) + r_1 G_1(p_0 s + 1 - p_0) \quad (\text{A.6})$$

Introducing $b_1 = (1 - r_1)p_1$ and $a_1(s) = (1 - r_1)(1 - p_1) + r_1 G_1(p_0 s + 1 - p_0)$, allowing $k = 0$, this can be rewritten as follows:

$$\begin{aligned}
 f_{I_1(0,n)}(s) &= a_1(s) + b_1 f_{I_1(0,n-1)}(s) \\
 &= a_1(s) + b_1 (a_1(s) + b_1 f_{I_1(0,n-2)}(s)) \\
 &= \dots \\
 &= b_1^n + a_1(s) \sum_{i=0}^{n-1} b_1^i. \quad (\text{A.7})
 \end{aligned}$$

Computing the geometric sum above, and taking the quantities defined in (6) gives the required result.

A.3. Derivation of (8). Deriving (8) follows from repeating equation (2) in the following way:

$$\begin{aligned}
f_{I_0(0,n)}(s) &= f_{I_0(1,n)}(s)G_0(0; p_0 f_{I_1(1,n)}(s) + 1 - p_0) \\
&= f_{I_0(2,n)}(s)G_0(1; p_0 f_{I_1(2,n)} + 1 - p_0)G_0(0; p_0 f_{I_1(1,n)}(s) + 1 - p_0) \\
&\vdots \\
&= \prod_{j=1}^{n-1} G_0(j-1; p_0 f_{I_1(0,n-j)}(s) + 1 - p_0)
\end{aligned} \tag{A.8}$$

The expression in (8) follows from substituting (A.8) into (5).

A.4. Derivation of (18). Substituting (16) into (10) gives the hazard rate. During the $(v+1)^{th}$ period that hybridization is introduced, i.e. if $v(R+S) + 2 \leq n < v(R+S) + S + 2$, the following holds:

$$H_n(q) = 1 - e^{-m_0 p_0 \beta_1(q) \left(1 - b_1^{n-(1+v(S+R))} + b_1^{n-(S+1)} (1 - b_1^S) \left(\frac{1 - b_1^{v(R+S)}}{b_1^{(v-1)(R+S)} (1 - b_1^{R+S})} \right) \right)} \tag{A.9}$$

and for the $(v+1)^{th}$ period that hybridization is stopped, i.e. if $v(R+S) + S + 2 \leq n < (v+1)(R+S) + 2$,

$$H_n(q) = 1 - e^{-m_0 p_0 \beta_1(q) b_1^{n-(S+1)} (1 - b_1^S) \left(\frac{1 - b_1^{(v+1)(R+S)}}{b_1^{v(R+S)} (1 - b_1^{R+S})} \right)} \tag{A.10}$$

and, as in (10), the hazard rate equals zero for $n \in \{0, 1\}$. Substituting (17) into (A.9) leads to the following for $0 \leq k < S$:

$$H_{v(R+S)+2+k}(q) = 1 - e^{-m_0 p_0 \beta_1(q) \left(1 - b_1^{k+1} + b_1^{v(R+S)+k+1-S} (1 - b_1^S) \left(\frac{1 - b_1^{v(R+S)}}{b_1^{(v-1)(R+S)} (1 - b_1^{R+S})} \right) \right)} \tag{A.11}$$

and substituting (17) into (A.10) leads to, for $S \leq k < S + R$:

$$H_{v(R+S)+2+k}(q) = 1 - e^{-m_0 p_0 \beta_1(q) b_1^{v(R+S)+k+1-S} (1 - b_1^S) \left(\frac{1 - b_1^{(v+1)(R+S)}}{b_1^{v(R+S)} (1 - b_1^{R+S})} \right)}. \tag{A.12}$$

To reach the asymptotic behaviour described in (18), take $v \rightarrow \infty$ in both (A.11) and (A.12).

A.5. Derivation of (21). First, note that the survival function of T and the hazard rate are related as follows. For any $t \in [0, +\infty)$:

$$P[T > t] = \prod_{j \in \mathbb{N}_0 : j \leq t} (1 - H_j(q)). \tag{A.13}$$

Define the sequence $\{c_n, n \in \mathbb{N}_0\}$:

$$c_n = \frac{P[T > n + R + S]}{P[T > n]}. \tag{A.14}$$

The use of (A.8) with (4) and (A.14), gives:

$$\begin{aligned}
c_n &= \frac{f_{I_0(0,n+R+S)}(q)}{f_{I_0(0,n)}(q)} \\
&= \frac{\prod_{i=1}^{n+R+S-1} G_0(i-1; p_0 f_{I_1(0,n+R+S-i)}(q) + 1 - p_0)}{\prod_{i=1}^{n-1} G_0(i-1; p_0 f_{I_1(0,n-i)}(q) + 1 - p_0)} \\
&= \frac{\prod_{i=1}^{R+S} e^{-p_0 m_0(i-1)(1-f_{I_1(0,n+R+S-i)}(q))} \prod_{i=R+S+1}^{n+R+S-1} e^{-p_0 m_0(i-1)(1-f_{I_1(0,n+R+S-i)}(q))}}{\prod_{i=1}^{n-1} e^{-p_0 m_0(i-1)(1-f_{I_1(0,n-i)}(q))}} \\
&= e^{-p_0 m_0 \sum_{i=1}^S (1 - f_{I_1(0,n+R+S-i)}(q))}
\end{aligned} \tag{A.15}$$

Note how the second product in the numerator is identical to the denominator. This is a result of the periodicity of the hybridization rate in (16). Also, note that for $S+1 \leq i \leq R+S$, $m(i) = 0$, which is used to reduce the number of terms in the sum.

When $n \rightarrow \infty$, c_n converges to

$$C = e^{-p_0 m_0 S \beta_1(q)}. \tag{A.16}$$

Thus, in the long run, a process with a constant hazard rate, λ , and such that $\lim_{n \rightarrow \infty} \frac{P[T > n+R+S]}{P[T > n]} = C$, would have the same probability of an introgression event occurring within a period from n to $n+R+S$, with sufficiently large n . Using (A.13) and (A.16) we find that λ must satisfy

$$\lim_{n \rightarrow \infty} \prod_{i=n+1}^{n+R+S} (1 - \lambda) = C, \tag{A.17}$$

and the required result follows by combining (A.16) and (A.17) and solving for λ .

A.6. Derivation of (22). Take the definitions of $f_{I_i(k,n)}(s)$, $I_i(k,n)$ and $Z_m^{(i)}$ as before, but extend it to include $i = (0,1)$ and $(0,2)$. As before, a joint p.g.f. of the offspring distribution of a single type- i ($i = (0,1), (0,2), 1, E$) is defined:

$$F_i(k; (s_{0,1}, s_{0,2}, s_1, s_E)) = E \left[s_{0,1}^{Z_{0,1}^{(i)}(k+1)} s_{0,2}^{Z_{0,2}^{(i)}(k+1)} s_1^{Z_1^{(i)}(k+1)} s_E^{Z_E^{(i)}(k+1)} \right] \tag{A.18}$$

Then, following the same methodology established in A.1, we get:

$$f_{I_i(k,n)}(s) = F_i(k; (f_{I_{0,1}(k+1,n)}(s), f_{I_{0,2}(k+1,n)}(s), f_{I_1(k+1,n)}(s), s)) \tag{A.19}$$

Following further the methodology in A.1, the following recursive relationships hold:

$$f_{I_{0,1}(0,n-k)}(s) = G_{0,1} (p_0 f_{I_1(0,n-k-1)}(s) + 1 - p_0) \times (\kappa_{1,1} f_{I_{0,1}(0,n-k-1)}(s) + \kappa_{1,2} f_{I_{0,2}(0,n-k-1)}(s)) \quad (\text{A.20})$$

$$f_{I_{0,2}(0,n-k)}(s) = G_{0,2} (p_0 f_{I_1(0,n-k-1)}(s) + 1 - p_0) \times (\kappa_{2,1} f_{I_{0,1}(0,n-k-1)}(s) + \kappa_{2,2} f_{I_{0,2}(0,n-k-1)}(s)) \quad (\text{A.21})$$

where the simplifying expression $f_{I_i(k,n)}(s) = f_{I_i(0,n-k)}(s)$ has been applied. Using the forms of $G_{0,1}(s)$ and $G_{0,2}(s)$ as specified in section 5, and setting $k = 0$, gives:

$$f_{I_{0,1}(0,n)}(s) = e^{-m_0 p_0 (1 - f_{I_1(0,n-1)}(s))} \times (\kappa_{1,1} f_{I_{0,1}(0,n-1)}(s) + \kappa_{1,2} f_{I_{0,2}(0,n-1)}(s))$$

$$f_{I_{0,2}(0,n)}(s) = \kappa_{2,1} f_{I_{0,1}(0,n-1)}(s) + \kappa_{2,2} f_{I_{0,2}(0,n-1)}(s) \quad (\text{A.22})$$

Since the environmental process is stationary:

$$P(T > n) = \frac{\kappa_{2,1}}{\kappa_{1,2} + \kappa_{2,1}} f_{I_{0,1}(0,n)}(q) + \frac{\kappa_{1,2}}{\kappa_{1,2} + \kappa_{2,1}} f_{I_{0,2}(0,n)}(q), \quad (\text{A.23})$$

and the hazard rate can be calculated from this.

For the considered analog of the deterministic process without autocorrelation, $f_{I_{0,1}(0,n)}(s) = e^{-m_0 p_0 (1 - f_{I_1(0,n-1)}(s))} f_{I_{0,2}(0,n)}(s)$. Using (5) (A.23) and (A.22) then gives the required result.

A.7. Derivation of the hazard rate in the bottleneck scenario. We start by defining the random variable $I_i(k, n)$ as before, except with $i \in \{0, 1, \dots, L\}$. Also, we define p.g.f.'s, $f_{I_i(k,n)}(s)$, of these random variables in the same way as previously done.

Since an individual belonging to a generation greater than n can produce no type- E individuals before n , write the following for any $i \in \{0, 1, \dots, L\}$,

$$I_i(k, n) = 0, \quad \text{if } k \geq n. \quad (\text{A.24})$$

Let us now turn to the case $k < n$. For a fixed $i \in \{0, \dots, L\}$, and a general scenario, where individuals can have offspring of any type, the following decomposition holds

$$I_i(k, n) = Z_E^{(i)}(k+1) + \sum_{m=0}^L \sum_{j=1}^{Z_m^{(i)}(k+1)} I_m^{(j)}(k+1, n), \quad (\text{A.25})$$

where the random variables

$$Z_0^{(i)}(k+1), Z_1^{(i)}(k+1), \dots, Z_L^{(i)}(k+1), Z_E^{(i)}(k+1)$$

represent the number of offspring of types $0, 1, \dots, L, E$, respectively, that the initial type i produced. Also, as the notation suggests, the random variables

$$I_0^{(j)}(k+1, n), \quad j = 1, \dots, Z_0^{(i)}(k+1),$$

represent the number of type- E individuals with non-type- E parents, appearing up to and including year n , in the line of descent of the j^{th} type-0 offspring of the

initial type- i individual. Notice that, since the initial type- i individual belongs to year k , its offspring belongs to year $k + 1$. The random variables

$$\begin{aligned} I_1^{(j)}(k+1, n), & \quad j = 1, \dots, Z_1^{(i)}(k+1), \\ I_2^{(j)}(k+1, n), & \quad j = 1, \dots, Z_2^{(i)}(k+1), \\ & \quad \vdots \\ I_L^{(j)}(k+1, n), & \quad j = 1, \dots, Z_L^{(i)}(k+1), \end{aligned}$$

are defined in an analogous way, but now for the type-1, type-2, ..., type- L , respectively, offspring of the initial type- i individual.

First manipulate the generating functions of (A.25) as follows:

$$\begin{aligned} f_{I_i(k,n)}(s) &= E \left[E \left[s^{I_i(k,n)} \mid Z_0^{(i)}(k+1), Z_1^{(i)}(k+1), \dots, Z_L^{(i)}(k+1), Z_E^{(i)}(k+1) \right] \right] \\ &= E \left[f_{I_0(k+1,n)}(s)^{Z_0^{(i)}(k+1)} f_{I_1(k+1,n)}(s)^{Z_1^{(i)}(k+1)} \dots f_{I_L(k+1,n)}(s)^{Z_L^{(i)}(k+1)} s^{Z_E^{(i)}(k+1)} \right] \end{aligned} \quad (\text{A.26})$$

We can manipulate (A.26) as above because the individual lineages are independent of each other, and individuals of the same type have identical offspring distributions.

Introduce the joint p.g.f of the reproduction distribution of a type- i individual belonging to a year k which, for $i \in \{0, 1, \dots, L\}$ and $k \geq 0$, is defined as

$$F_i(k; (s_0, s_1, \dots, s_L, s_E)) = E[s_0^{Z_0^{(i)}(k+1)} s_1^{Z_1^{(i)}(k+1)} \dots s_L^{Z_L^{(i)}(k+1)} s_E^{Z_E^{(i)}(k+1)}] \quad (\text{A.27})$$

for $(s_0, s_1, \dots, s_L, s_E) \in [0, 1]^{L+2}$.

Putting (A.26) and (A.27) together, we find that

$$f_{I_i(k,n)}(s) = F_i(k; (f_{I_0(k+1,n)}(s), f_{I_1(k+1,n)}(s), \dots, f_{I_L(k+1,n)}(s), s)) \quad (\text{A.28})$$

In our specific model, we have the following assumptions regarding the reproduction:

- the reproduction law of a type 0 individual depends on the year number and the corresponding p.g.f. is given by

$$F_0(k; (s_0, s_1, \dots, s_L, s_E)) = s_0 G_0(k; p_0 s_1 + (1 - p_0)) \quad (\text{A.29})$$

- for a type i individual, with $i \in \{1, \dots, L\}$, the reproduction law does not depend on the year number and the corresponding p.g.f. is given by

$$\begin{aligned} F_i(k; (s_0, s_1, \dots, s_i, s_{i+1}, \dots, s_L, s_E)) &= (1 - r_i)(1 - p_i) + (1 - r_i)p_i s_i \\ &\quad + r_i G_i(p_0 s_{i+1} + 1 - p_0) \end{aligned} \quad (\text{A.30})$$

with $s_{L+1} \equiv s_E$. The fact that the reproduction law of these individuals is independent of time implies that

$$f_{I_i(k,n)}(s) = f_{I_i(0,n-k)}(s).$$

This relation will be used more or less explicitly in the following calculations.

The use of (A.30) and (A.28) with $i = L$, gives

$$f_{I_L(0,n)}(s) = (1 - r_L)(1 - p_L) + (1 - r_L)p_L f_{I_L(0,n-1)}(s) + r_L G_L(p_0 s + 1 - p_0).$$

The use of initial condition $f_{I_L(0,0)}(s) = 1$ results in the following for any $n \geq 0$, which is :

$$f_{I_L(0,n)}(s) = 1 - \beta_L(s) + \beta_L(s) b_L^n, \quad (\text{A.31})$$

with

$$b_L = (1 - r_L) p_L \quad \text{and} \quad \beta_L(s) = \frac{r_L (1 - G_L(p_0 s + 1 - p_0))}{1 - b_L}. \quad (\text{A.32})$$

The calculation of (A.31) above follows the same reasoning shown in Appendix A.2.

Now that we can calculate the p.g.f.'s of $I_L(0, n)$, we proceed by finding expressions for the p.g.f.'s of $I_i(0, n)$ for $i = 0, 1, \dots, L - 1$.

Note that, in the line of descent of a single type- i individual belonging to year 0, new type- E individuals can only appear after $L - i + 1$ years (this is intuitively clear from Fig. 4). Hence, for $i \in \{1, \dots, L - 1\}$,

$$f_{I_i(0,1)}(s) = f_{I_i(0,2)}(s) = \dots = f_{I_i(0,L-i)}(s) = 1.$$

Now, for $n > L - i$, the use of (A.30) and (A.28), gives

$$f_{I_i(0,n)}(s) = (1 - r_i)(1 - p_i) + r_i G_i(p_0 f_{I_{i+1}(0,n-1)}(s) + 1 - p_0) + (1 - r_i) p_i f_{I_i(0,n-1)}(s).$$

Repeating the procedure gives

$$\begin{aligned} f_{I_i(0,n)}(s) &= [(1 - r_i) p_i]^{n-(L-i)} + (1 - p_i) \sum_{j=1}^{n-(L-i)} (1 - r_i)^j p_i^{j-1} \\ &\quad + \sum_{j=1}^{n-(L-i)} r_i [(1 - r_i) p_i]^{j-1} G_i(p_0 f_{I_{i+1}(0,n-j)}(s) + 1 - p_0). \end{aligned}$$

Computing the sums above gives us the following p.g.f.'s:

$$f_{I_i(0,n)}(s) = 1 - \alpha_i + \alpha_i b_i^{n-(L-i)} + r_i \sum_{k=L-i}^{n-1} b_i^{n-k-1} G_i(p_0 f_{I_{i+1}(0,k)}(s) + 1 - p_0), \quad (\text{A.33})$$

where

$$b_i = (1 - r_i) p_i \quad \text{and} \quad \alpha_i = \frac{r_i}{1 - b_i}. \quad (\text{A.34})$$

We have $f_{I_0(0,n)}(s) = 1$ for $n \leq L$, since a type-0 individual requires at least L generations to produce a type-E individual. For $n > L$ we combine (A.29) and (A.28) to give:

$$f_{I_0(0,n)}(s) = \prod_{j=1}^{n-L} G_0(j - 1; p_0 f_{I_1(0,n-j)}(s) + 1 - p_0) \quad (\text{A.35})$$

which can be calculated using (A.33) and (A.31).

The use of (A.35) and noting that, as before, $P(T > n) = f_{I_0(0,n)}(q)$ yields the hazard rate:

$$H_n(q) = \begin{cases} 0 & \text{if } 0 \leq n \leq L \\ 1 - \frac{\prod_{j=1}^{n-L} G_0(j-1; p_0 f_{I_1(0,n-j)}(q) + 1 - p_0)}{n-1-L} & \text{if } n \geq L+1. \\ \prod_{j=1}^{n-L} G_0(j-1; p_0 f_{I_1(0,n-1-j)}(q) + 1 - p_0) & \end{cases} \quad (\text{A.36})$$

A.8. Derivation of (24). Taking $r_1 = 1$ in (A.31) to (A.34) gives:

$$f_{I_L(0,n)}(s) = 1 - \beta_L(s) \quad (\text{A.37})$$

$$f_{I_i(0,n)}(s) = G_i(p_0 f_{I_{i+1}(0,n-1)}(s) + 1 - p_0) \quad (\text{A.38})$$

where $i = 1, 2, \dots, L-1$. Differentiating these expressions with respect to s and evaluating the results at the point $s = 1$ gives:

$$\begin{aligned} f'_{I_L(0,n)}(1) &= p_0 m_L \\ f'_{I_i(0,n)}(1) &= p_0 m_i f'_{I_{i+1}(0,n-1)}(1) \end{aligned} \quad (\text{A.39})$$

where we have used the fact that the derivative of a p.g.f. evaluated at one is the mean of the random variable.

Taking logarithms in (A.35) and differentiating at $s = 1$ yields the following expression:

$$\begin{aligned} f'_{I_0(0,n)}(1) &= \sum_{j=1}^{n-L} p_0 m_0 (j-1) f'_{I_1(0,n-j)}(1) \\ &= \sum_{j=1}^{n-L} p_0 m_0 (j-1) p_0^L \prod_{i=1}^L m_i \end{aligned} \quad (\text{A.40})$$

where the last equality uses the expressions in (A.39).

Consider the representation of the hazard rate in (5). It is apparent that the constant-term in the Taylor approximation will be zero, due to the fact that p.g.f.'s evaluated at one are one. Taking the derivative of (5) around one yields:

$$H'_n(1) = f'_{I_0(0,n-1)}(1) - f'_{I_0(0,n)}(1). \quad (\text{A.41})$$

Using the above with (A.40) gives the required result.

REFERENCES

- Bos, P.M.J., Jan, P-J, van Raaij, M.T.M., 2004. Risk assessment of peak exposure to genotoxic carcinogens: a pragmatic approach. *Toxicol. Lett.* 151, 43-50.
- Demon, I., Haccou, P., van den Bosch, F., 2007. Introgression of resistance genes between population: A model study of insecticide resistance in *Bemisia tabaci*. *Theor. Popul. Biol.* 72, 292-304.
- Edmands, S., 2002. Does parental divergence predict reproductive compatibility? *Trends Ecol. Evol.* 17, 520-527.

- Ellstrand, N.C., Prentice, H.C., Hancock, J.F., 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annu. Rev. Ecol. Systemat.* 30, 539-563.
- Ghosh, A., Haccou, P., 2010. Quantifying stochastic introgression processes with hazard rates. *Theor. Popul. Biol.* 77, 171-180.
- Haccou, P., Jagers, P., Vatutin, V.A., 2005. *Branching Processes: Variation Growth and Extinction of Populations*. Cambridge University Press, Cambridge.
- Hails, R.S., Morley, K., 2005. Genes invading new populations: A risk assessment perspective. *Trends Ecol. Evol.* 20, 245-252.
- Haygood, R., Ives, A.R., Andow, D.A., 2003. Consequences of recurrent gene flow from crops to wild relatives. *Proc. Biol. Sci.* 270, 1879-1896.
- Haygood, R., Ives, A.R., Andow, D.A., 2004. Population genetics of transgene containment. *Ecol. Lett.* 7, 213-220.
- Jagers, P., Klebaner, F.C., 2000. Population-size-dependent and age-dependent branching processes. *Stoch. Proc. Appl.* 87, 235-254.
- Kalbfleisch, J.D., Prentice, R.L., 2002. *The Statistical Analysis of Failure Time Data*, 2nd ed., John Wiley & Sons, New York.
- Kelly, C.K., Bowler, M.J., Breden, F., Fenner, M., Poppy, G.M., 2005. An analytical model assessing the potential threat to natural habitats from insect resistance transgenes. *Proc. Biol. Sci.* 272, 1759-1767.
- Maan, S.S., 1987. Interspecific and intergeneric hybridisation in wheat. In: Heyne, E.G. (Ed.), *Wheat and Wheat Improvement*. ASA, CSSA and SSSA, Madison, pp. 453-461.
- Michor, F., Nowak, M.A., Iwasa, Y., 2006. Stochastic dynamics of metastasis formation. *J. Theoret. Biol.* 240, 521-530.
- Reluga, T., Meza, R., Walton, D.B., Galvani, A.P., 2007. Reservoir interactions and disease emergence. *Theor. Popul. Biol.* 72, 400-408.
- Riesberg, L.H., Wendel, J.F., 1993. Introgression and its consequences in plants. In: Harrison, R.G. (Ed.), *Hybrid Zones and the Evolutionary Process*. Oxford University Press, Oxford, pp. 70-109.
- Serra, M.C., Haccou, P., 2007. Dynamics of escape mutants. *Theor. Popul. Biol.* 72, 167-178.
- Snow, A.A., Andersen, B., Jorgensen, R.B., 1999. Costs of transgenic herbicide resistance introgressed from *Brassica napus* into weed *B. Rapa*. *Mol. Ecol.* 8, 605-615.
- Thompson, C.J., Thompson, B.J.P., Ades, P.K., Cousens, R., Carinier-Gere, P., Landman, K., Newbigin, E., Burgman, M.A., 2003. Model-based analysis of the likelihood of gene introgression from genetically modified crops into wild relatives. *Ecol. Model.* 162, 199-209.
- Welliver, R., 2009. The relationship of meteorological conditions to the epidemic activity of respiratory syncytial virus. *Paediatr. Respir. Rev.*, 10 (2009) 6-8.

