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

Author: Ghosh, Atiyo
Title: Calculating hazard rates of introgression with branching processes
Date: 2012-05-22

CHAPTER 1: QUANTIFYING STOCHASTIC INTROGRESSION PROCESSES WITH HAZARD RATES

Reprinted with minor edits from Ghosh, Haccou, 2010. Theor. Popul. Biol., 77, 171-180

Abstract

Introgression is the permanent incorporation of genes from one population into another through hybridization and backcrossing. It can have large environmental consequences, such as the spread of insecticide or herbicide resistant genes, the escape of transgenes from genetically modified crops, and the invasion of exotic genes into new habitats. Introgression usually involves strong random components, such as rare hybridization and backcrossing events, and demographic variation in reproduction and survival. Most introgression studies ignore these random effects, and consequently fail to accurately assess the risk of introgression. This paper presents a methodology for quantifying stochastic introgression processes, based on multitype branching process models. We derive a quantity called the hazard rate, which can be used to investigate how the risk of introgression depends on crop management and life history.

1. Introduction

Introgression is the permanent incorporation of genes from one population into another, through hybridization and backcrossing (Riesberg and Wendel, 1993; Ellstrand et al., 1999; Hails and Morley, 2005). This may result in the spread of insecticide or herbicide resistant genes (Snow et al., 1999; Demon et al., 2007), the escape of transgenes from genetically modified crops (Rieger et al., 2002; Reichman et al., 2006), or the incorporation of genes from exotic species into genomes of local species (Huxel, 1999; Allendorf et al., 2001; Abbott et al., 2003). The potential environmental effects of introgression are severe. For instance, there are serious concerns that transgene escape might produce robust weeds (Maan, 1987; Snow et al., 1999; Thompson et al., 2003; Kelly et al., 2005) that can outcompete other species and reduce biodiversity (Levin et al., 1996; Jenczewski et al., 2003; Ellstrand, 2003).

Mathematical models provide a means to study the likelihood of introgression given certain environmental and species-specific conditions. The great advantage of models is that they allow us to perform experiments that are either too dangerous, impractical, or simply impossible to carry out empirically. Furthermore, models can pinpoint which parameters are crucial for introgression risk, and whose values should therefore be determined empirically.

Introgression processes contain strong stochastic components. Hybridization and backcrossing events usually occur at a very low rate. Hybrids and initial backcrosses are often less viable and fertile than the wild type (Hauser et al., 1998), since their genetic backgrounds are adapted to different conditions. Therefore, initial hybrid populations are usually small, and highly affected by demographic stochasticity. As a consequence, even if foreign genes provide a fitness advantage in later backcrosses, it will usually take several invasions before they are established permanently. The number of individuals carrying foreign alleles will be highly variable, and rise and fall during the initial stages of introgression. The use of deterministic models, that fail to take such key features of introgression into account, can be very misleading.

Nevertheless, stochastic models are seldom used in introgression studies. Exceptions are Davis et al. (1999) and Thompson et al. (2003), who considered the effects of stochastic environmental variation, but ignored demographic stochasticity. Haygood et al. (2003) studied the conditions under which transgenes can become fixed due to genetic drift in small populations with repeated invasions. Haygood et al. (2004) examined the repeated invasions of a transgene with a small fitness advantage.

For practical applications, general quantitative measures that characterize stochastic introgression processes are indispensable. The above-mentioned studies do not provide these, since they are all based on simulations. Our aim is to develop such measures.

We consider a situation where foreign genes invade repeatedly into a resident population, and each invasion has a small probability of leading to a permanent lineage. This is similar to the case studied by Haygood et al. (2004), but we consider a more general model, and our main results are not based on simulations. It is obvious that under these conditions a permanent introgressed lineage will be founded eventually. Before such an introgression event happens, however, there can be an extensive period of failed invasions. Introgression risk is largely determined by the duration of this period. We derive a measure that quantifies the distribution of these lengths in an intuitive way, called the hazard rate.

The hazard rate is the probability per time unit that a random event occurs, given that it has not happened before. It quantifies how the instantaneous risk of introgression events changes in time. For instance, how quickly this risk increases after cultivation of a transgene crop has started, or decreases after such cultivation is stopped, and how this relates to life history characteristics of the crop and its wild relatives. Studies of hazard rates can indicate which periods are especially risky, and thus provide information for optimizing monitoring and management programs.

Hazard rates were first used in medical statistics, to analyze mortality risks (Kalbfleisch and Prentice, 2002). For several decades they have also been applied in behavioral data analysis (Haccou and Meelis, 1994). They have not been used in introgression studies before now.

We derive the hazard rate from a multitype branching process model of hybrid population dynamics (e.g., Haccou et al., 2005, section 2.2). Demon et al. (2007) used such a model to study the effects of whitefly life history parameters on the establishment probability of an insecticide resistance gene, but they only considered a single introduction. Serra and Haccou (2007) were the first to calculate hazard rates from a branching process model. They assumed that individuals that have on average less than one offspring (the so-called subcritical type) can produce mutants with an average of more than one offspring (the supercritical type). Escape from extinction occurs as soon as a supercritical individual that initiates a permanent lineage is produced. In this paper we generalize their method to situations with repeated invasion of individuals of the subcritical type, population structure, and density-dependent competition with an established resident type.

As an example, we consider a model for a monocarpic (i.e., a plant which flowers once then dies) monoecious (i.e., flowers have male as well as female functions) nonselfing species. In numerical examples we use parameter ranges that are deemed to be realistic for Daucus carota (the carrot), but we stress that the model is a caricature, and not meant to give a complete description of introgression in this species. We illustrate the set-up of a branching process model and derivation of the hazard rate. Furthermore we give numerical procedures for calculating this function, and analytic approximations.

While we focus on introgression, our approach may be applied in other fields as well, since repeated invasions with fitness bottlenecks occur in many biological contexts. For example, when a virus colonizes a new host species, its initial reproductive ratio will be small, but after a few mutations this can increase. Another example is the initiation and growth of tumors, where cells usually have to go through several mutations before they produce a successful lineage (see e.g., Michor et al., 2006).

2. THE MODEL

We consider a plant species that flowers only once and then dies. To account for an ageing effect, we distinguish one-year old plants from older ones. We assume that there is a large and stable population of wild plants. By pollen flow from a neighboring crop, stochastic numbers of hybrid seeds are produced each year. Hybrid production can be followed by repeated backcrossing with wild plants. Seeds compete to germinate and survive their first year. Hybrid plants are considered to be less fit than wild individuals, whereas backcrossed individuals have some probability of producing a permanent introgressed lineage. Individuals of the first backcross (BC1) and further backcrosses are assumed to have identical life history parameters, and are therefore not distinguished.

In summary, there are six types of plants in the model: one-year old plants (types 1, 3, 5), and plants that are two or more years old (types 2, 4, 6); and in addition to being characterized by their age, plants can also be either wild (types 1 and 2), hybrids (3, 4), or backcrossed (5, 6).

Fig. 1 illustrates the introgression scheme and the life history.

2.1. Dynamics of the wild population.

Dynamics of the wild population. Since it is assumed that the wild population is large, we use a deterministic model for its dynamics:

$$\begin{pmatrix} z_1(t+1) \\ z_2(t+1) \end{pmatrix} = \begin{pmatrix} p_0(z_1(t), z_2(t))\, r_1 m_1 & p_0(z_1(t), z_2(t))\, r_2 m_2 \\ p_1(1-r_1) & p_2(1-r_2) \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} \tag{1}$$

where $z_i(t)$ represents the population density (number per unit area) of type-i plants in year t, r_i its flowering probability, m_i the average number of seeds it produces and p_i the probability that it survives one year if it does not flower. Furthermore, we assume that there is a density-dependent probability that seeds germinate and survive their first year, $p_0(z_1(t), z_2(t))$.

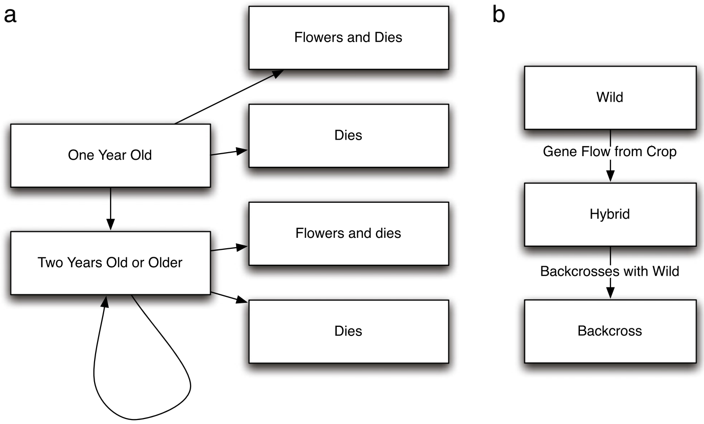

FIGURE 1. (a) Schematic representation of the life history used in the model. (b) A schematic representation of the process by which an introgressed gene moves through a population.

This model has one stable and positive equilibrium. At this equilibrium we have:

$$p_0\left(\hat{z}_1, \hat{z}_2\right) = \frac{1 - p_2\left(1 - r_2\right)}{m_1 r_1 \left(1 - p_2\left(1 - r_2\right)\right) + p_1 m_2 r_2 \left(1 - r_1\right)} \tag{2}$$

We will denote this value by p_0, and refer to it as the germination probability.

Furthermore, we define:

$$\zeta_{WT} = \hat{p}_0^{-1} = r_1 m_1 + \frac{r_2 m_2 \left(1 - r_1\right) p_1}{1 - p_2\left(1 - r_2\right)} \tag{3}$$

which represents the expected number of seeds produced by one wild type individual.

2.2. **Invasion dynamics of hybrids and backcrosses.** Because the population of wild plants is large and the numbers of hybrid and backcrossed individuals 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, implying that the seed germination probability equals $\hat{p}_0$. For convenience, we assume that there are no other factors beside this competition that affect germination probability of hybrids and backcrosses. The model can be easily generalized in this respect. Since hybrid and backcrossed plants do not affect each others reproduction and survival initially, their invasion dynamics can be modeled as a branching process.

In the branching process model, flowering individuals of type $i \in \{3, 4, 5, 6\}$ produce a stochastic number of offspring, denoted by ξ_i, with expectation m_i. The probabilities r_i, and p_i are as defined in the previous section, and are assumed to lie between zero and one. 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 type 0 that produces a stochastic number of hybrid seeds, ξ_0, according to some probability distribution with expectation m_0 in each year. Figure 2 shows a schematic summary of the invasion dynamics.

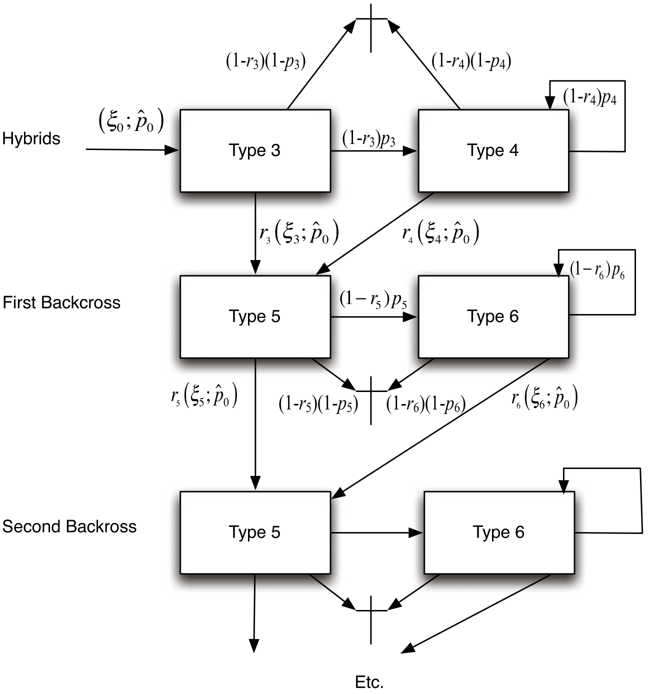

Figure 2. Schematic representation of the branching process model for invasion dynamics. $(\xi;\hat{p}_0)$, $(i = 0, 3, 4, 5, 6)$ represents the production of ξ_i seeds, each of which has a germination probability $\hat{p}_0$. Second and further backcrosses are assumed to have identical life history parameters to BC1.

3. Stochastic versus deterministic invasion dynamics

Most introgression studies are based on deterministic rather than stochastic invasion models. We here demonstrate the difference between the two approaches. Let $Z_i(t)$ denote the number of type-i individuals in year t, then the deterministic equivalent of the branching process model for invasion dynamics is:

$$\begin{pmatrix} Z_3(t+1) \\ Z_4(t+1) \\ Z_5(t+1) \\ Z_6(t+1) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ p_3(1-r_3) & p_4(1-r_4) & 0 & 0 \\ \hat{p}_0 r_3 m_3 & \hat{p}_0 r_4 m_4 & \hat{p}_0 r_5 m_5 & \hat{p}_0 r_6 m_6 \\ 0 & 0 & p_5(1-r_5) & p_6(1-r_6) \end{pmatrix} \begin{pmatrix} Z_3(t) \\ Z_4(t) \\ Z_5(t) \\ Z_6(t) \end{pmatrix} + \begin{pmatrix} m_0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{4}$$

In a population starting with at least one $Z_i(t)$ larger than zero and no immigration (if $m_0 = 0$), the numbers of types 5 and 6 will grow if and only if the dominant eigenvalue of the matrix:

$$\begin{pmatrix} \hat{p}_0 r_5 m_5 & \hat{p}_0 r_6 m_6 \\ p_5 \left(1 - r_5\right) & p_6 \left(1 - r_6\right) \end{pmatrix} \tag{5}$$

is larger than one. This leads to the condition

$$\zeta_{BC} > \zeta_{WT}, \tag{6}$$

where ζ_{BC} is the expected number of seeds produced by a single backcrossed plant:

$$\zeta_{BC} = r_5 m_5 + \frac{r_6 m_6 \left(1 - r_5\right) p_5}{1 - p_6 \left(1 - r_6\right)} \tag{7}$$

With immigration ($m_0 > 0$) and initial condition $Z_3(0) = \ldots = Z_6(0) = 0$, the deterministic model then predicts an exponential growth of $Z_5(t)$ and $Z_6(t)$ from year 2 onwards.

According to the stochastic model, exactly the same condition must hold for a process starting with one individual of type 5 to have a positive establishment chance (see section 4.1). In that case, repeated invasions will eventually lead to exponential growth of the numbers of type-5 and 6 individuals. Before that happens, however, there may be several failed invasions, and even production of backcrossed individuals whose lineage fails. Thus, the numbers of individuals that carry the foreign gene can rise and fall, and periods in which the gene is present or absent from the population may alternate, before initiation of a permanent lineage. This is illustrated in Fig. 3, which shows the trajectory predicted by the deterministic model as well as the results of several simulations of the branching process. As can be seen, the length of the initial period of failed invasions is highly variable and may be quite large. Furthermore, the numbers of individuals with the foreign allele at any given time differ strongly between simulations.

4. DERIVATION OF THE HAZARD RATE

Let T denote the time at which an introgression event occurs, i.e., the time at which the first type-5 individual whose lineage does not die out is produced. In the three examples of Fig. 3, T is respectively about 2, 60, and larger than 100 years. The hazard rate, denoted by $H_n(q)$, equals:

$$H_n\left(q\right) = P\left(T = n \left| T > n - 1\right.\right) = \frac{P\left(T > n - 1\right) - P\left(T > n\right)}{P\left(T > n - 1\right)} \tag{8}$$

From this equation it can be seen that this function fully characterizes the distribution of T.

4.1. **Establishment probability of an introgressed lineage from one type-5 individual.** We first consider the fate of a lineage that starts with one type-5 individual. According to branching process theory, lineages either go extinct or they grow infinitely large. In situations where backcrosses are more fit than the wild type, the extinction probability of a lineage from one type-5 individual, denoted by q_5, is less than one, and the branching process model predicts indeterminate growth with a probability $1 - q_5$. It is unreasonable, however, to assume that populations can grow infinitely large, so we will interpret $1 - q_5$ as the probability

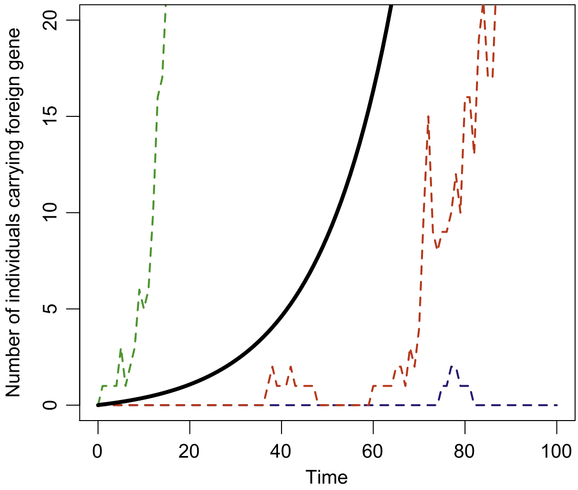

FIGURE 3. Total number of individuals containing the foreign gene, $\sum_{i=3}^{6} Z_i(t)$ according to the deterministic model (solid line), and the three simulation runs of the stochastic model with Poisson-distributed seed productions (dotted lines). Parameter values: $\hat{p}_0 = 0.00079$, $m_0 = 50$, $r_3 = r_4 = r_5 = r_6 = 0.9$, $p_3 = p_4 = p_5 = p_6 = 0.7$, $m_3 = m_4 = 850, m_5 = m_6 = 900$.

that the lineage reaches some size that is large enough to guarantee its permanent presence.

The calculation of extinction probabilities and hazard rate is based on so-called probability generating functions. Let X be a non-negative discrete random variable, then its probability generating function (p.g.f.) is a function from $\mathbb{R}$ to $\mathbb{R}$, which is defined as $E\left[s^X\right]$, where $E[\,]$ denotes expectation and s can take any value on the interval $[0,1]$.

Let $G_i(s)$ be the p.g.f. of the number of seeds produced by such an individual of type i $(i = 0, ..., 6)$, then:

$$\begin{aligned} q_5 =& (1-p_5)(1-r_5) + (1-r_5)p_5 q_6 + r_5 \sum_k P\left(\xi_5 = k\right) \sum_{l \le k} \left(\begin{array}{c} k \\ l \end{array} \right) (1-\hat{p}_0)^{k-l} \hat{p}_0^l q_5^l \\ =& (1-p_5)(1-r_5) + (1-r_5)p_5 q_6 + r_5 \sum_k P(\xi_5 = k) \left(\hat{p}_0 q_5 + (1-\hat{p}_0)\right)^k \qquad (9) \\ =& (1-p_5)(1-r_5) + (1-r_5)p_5 q_6 + r_5 G_5(\hat{p}_0 q_5 + (1-\hat{p}_0)) \end{aligned}$$

The first term represents the probability that a type-5 individual does not flower and does not survive; the second term represents the probability that the individual does not flower, and survives to become a type-6 individual, which initiates a lineage that goes extinct; the last term equals the sum of probabilities that the type-5 individual flowers, produces k seeds of which l germinate and produce type-5 individuals that each produce a lineage which will become extinct. We can write (9) as follows:

$$q_5 = a_5\left(q_5\right) + b_5 q_6 \qquad (10)$$

where we have introduced

$$a_i\left(q\right) = r_i G_i\left(\hat{p}_0 q + \left(1 - \hat{p}_0\right)\right) + \left(1 - r_i\right)\left(1 - p_i\right) \quad \text{and} \quad b_i = \left(1 - r_i\right) p_i \tag{11}$$

for $i = (1, \ldots, 6)$. These quantities will appear frequently in subsequent analyses.

In order to solve (10), we must first derive an expression for q6. Analogous to (10) we can derive:

$$q_6 = a_6\left(q_5\right) + b_6 q_6. \tag{12}$$

The solution of Eqs (10) and (12) depends on the forms of the functions $G_5(s)$ and $G_6(s)$, as well as the parameter values. For ease of notation, we will denote q_5 by q from now on. From (10) and (12) it follows that:

$$q = a_5\left(q\right) + \frac{b_5 a_6\left(q\right)}{1 - b_6} \tag{13}$$

The smallest positive value of q that satisfies (13) represents the extinction probability of a lineage started by one type-5 individual. This value is smaller than one if and only if the derivative of the right-hand side of (13) evaluated at $q = 1$ is larger than one, which leads to inequality (6).

4.2. **Calculation of the hazard rate.** Let $I_i(n)$ ($i \in \{0, 3, 4\}$) be the cumulative number of type-5 individuals produced by type-3 and 4 individuals up to and including time n, given that we start with a population of a single type-i individual. We denote the p.g.f. of $I_i(n)$ by $f_{I_i(n)}(s)$. It is easily seen that:

$$P\left(T > n\right) = E\left[q^{I_0(n)}\right] = f_{I_0(n)}\left(q\right) \tag{14}$$

since $T > n$ implies that all type-5 individuals produced up to and including time n must fail to establish a permanent lineage. In the Appendix it is shown that:

$$\begin{aligned}
f_{I_0(n)}\left(q\right) &= f_{I_0(n-1)}\left(q\right) G_0\left(\hat{p}_0 f_{I_3(n-1)}\left(q\right) + \left(1 - \hat{p}_0\right)\right) \\
f_{I_3(n)}\left(q\right) &= r_3 G_3\left(\hat{p}_0 q + \left(1 - \hat{p}_0\right)\right) + \left(1 - r_3\right) p_3 f_{I_4(n-1)}\left(q\right) + \left(1 - r_3\right)\left(1 - p_3\right) \\
f_{I_4(n)}\left(q\right) &= r_4 G_4\left(\hat{p}_0 q + \left(1 - \hat{p}_0\right)\right) + \left(1 - r_4\right) p_4 f_{I_4(n-1)}\left(q\right) + \left(1 - r_4\right)\left(1 - p_4\right)
\end{aligned} \tag{15}$$

These equations are solved iteratively with initial condition $f_{I_i(0)}(q) = 1$, $i \in \{0, 3, 4\}$ to find $f_{I_0(n)}(q)$. Putting (8), (14) and (15) together gives:

$$H_n\left(q\right) = 1 - \frac{f_{I_0(n)}\left(q\right)}{f_{I_0(n-1)}\left(q\right)} = 1 - G_0\left(\hat{p}_0 f_{I_3(n-1)}\left(q\right) + \left(1 - \hat{p}_0\right)\right) \tag{16}$$

For $n = 0$ or 1, the hazard rate equals 0, since no type-5 individuals can be produced before year 2. For the considered model, it is possible to derive an explicit expression for the hazard rate. For $n \geq 2$, it equals (see Appendix) :

$$H_n\left(q\right) = 1 - G_0\left(\hat{p}_0\left(a_3\left(q\right) + b_3\left(a_4\left(q\right)\frac{1 - {b_4}^{n-2}}{1 - b_4} + {b_4}^{n-2}\right)\right) + \left(1 - \hat{p}_0\right)\right) \tag{17}$$

and its asymptotic value is:

$$\hat{H}\left(q\right) = 1 - G_0\left(\hat{p}_0\left(a_3\left(q\right) + \frac{b_3 a_4\left(q\right)}{1 - b_4}\right) + \left(1 - \hat{p}_0\right)\right) \tag{18}$$

4.3. **Shape of the hazard rate.** The recurrence relations for $f_{I_3(n)}(q)$ and $f_{I_4(n)}(q)$ in (15) can be considered from a different angle. Consider a multi-type branching process with three types, numbered 3, 4 and 5, and let $Q_i(n)$ be the probability that a process starting with one individual of type i will go extinct at or before time n. Furthermore, assume that $Q_5(n)$ is constant and equal to q. Then (15), with $f_{I_i(n)}(q)$ replaced by $Q_i(n)$ $(i = 3, 4)$ specifies recurrence relations for these extinction probabilities, but with different initial conditions: $f_{I_i(0)}(q) = 1$, whereas $Q_i(0) = 0$. This implies that the $f_{I_i(n)}(q)$ decrease with n, whereas the $Q_i(n)$ increase. Of course, it is also evident from their definitions that this should be so. This conclusion is also valid for models describing more complicated life histories.

This equivalence can be used to derive properties of the $f_{I_i(n)}(q)$. From branching process theory (e.g., Athreya and Ney, 1972, chapter 5) it follows that, when q is given, the recurrence relations in (15) have one equilibrium with values in [0,1]. Let $f_{I_i}(q)$ denote the equilibrium values of $f_{I_i(n)}(q)$. If $q = 1$, there is one equilibrium at the point where all $f_{I_i}(q)$ are equal to one. In that case $f_{I_3(n)}(q) = f_{I_4(n)}(q) = 1$ for all n, and the hazard rate equals zero.

If q is less than one, there is one equilibrium point at smaller values of $f_{I_i}(q)$, which is stable. Since hybrid individuals have a positive chance of having no surviving offspring, this equilibrium is larger than zero, even when q equals zero. In this case $f_{I_3(n)}(q)$ decreases monotonically to a constant equilibrium value, and it follows from (16) that the hazard rate increases monotonically to an asymptote between zero and one.

A similar expression to (16) will hold for any process where the numbers of hybrids that are introduced each time unit are independent and identically distributed. Therefore, the same conclusion holds for any introgression process with this immigration structure.

When q is close to one, approximations for the extinction probabilities Q_i in slightly supercritical processes, such as given in Haccou et al. (2005, section 5.6) can be used to approximate the $f_{I_i}(q)$. An example of such an approximation is given in section 5.3.

5. RESULTS

5.1. **The shape of the hazard rate and distribution of T.** In the initial numerical analyses (Fig. 4) we assume that the seed production distributions of types 3 and 4 are either Geometric ($G_i(s) = 1/(1+(1-s)m_i)$) or Poisson ($G_i(s) = e^{-m_i(1-s)}$), and that the numbers of hybrid seeds produced per generation are Poisson-distributed.

As shown in the previous section, the hazard rate is zero in the first year and then increases monotonically in time, to an asymptotic level smaller than one. In this example, the asymptotic value is reached rapidly (see Fig. 4a). If the hazard rate would be constant, the times T would have a Geometric distribution (see e.g., Feller, 1968, section 13.9). The results indicate that we can well approximate the distribution of T with a time-lagged Geometric distribution:

$$P(T = t) \approx P(T_g + 1 = t) = \left(1 - \hat{H}(q)\right)^{t-2} \hat{H}(q), \text{ for } t \geq 2 \qquad (19)$$
$$P(T = t) = 0, \text{ for } 0 \leq t < 2,$$

where T_g represents a Geometrically distributed random variable. Numerical simulations, as shown in Fig. 4b, support the effectiveness of this approximation.

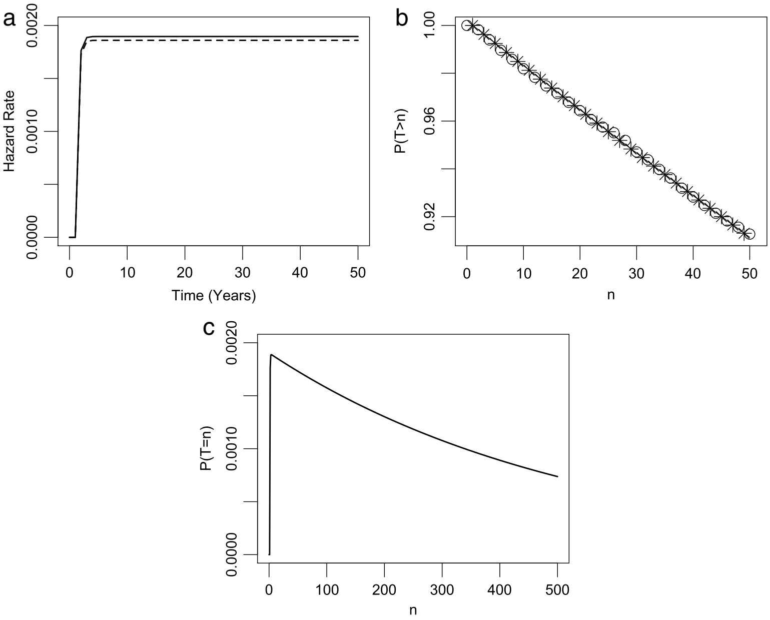

FIGURE 4. (a) The time dependence of the hazard rate (c.f. Eq. (17)) for Poisson (solid line) and Geometrically (dotted line) distributed seed production. Hybrids are formed according to a Poisson distribution. (b) The survival function of T for Poisson-distributed seed numbers (line). Numerical simulations of 10,000 iterations (circles) and the approximation to the survival function of T (stars) with a time-lagged Geometric distribution. (c) The exact distribution of T for Poisson-distributed seed production. In all the three figures the parameter values are: $m_0 = 50$, $m_3 = m_4 = 800$, $\hat{p}_0 = 0.001$, $p_3 = p_4 = 0.7$, $r_3 = r_4 = 0.9$, $q = 0.95$.

As a consequence, a good approximation for the expectation of T is $1 + 1/\hat{H}(q)$. Note, however, that the distribution of T is very skewed (see Fig. 4c). This implies that most introgression events will occur before the expected time. For example, with the parameters used in Fig. 4, the expectation of T equals 529 years, whereas half of the introgression events occur at or before 368 years.

5.2. **Effects of hybrid survival and flowering probabilities, hybrid seed production, and hybrid formation.** It can be shown straightforwardly that the asymptotic hazard rate is monotonically non-decreasing with r_4, p_3, and p_4 (see Appendix). Numerical examples are given in Fig. 5.

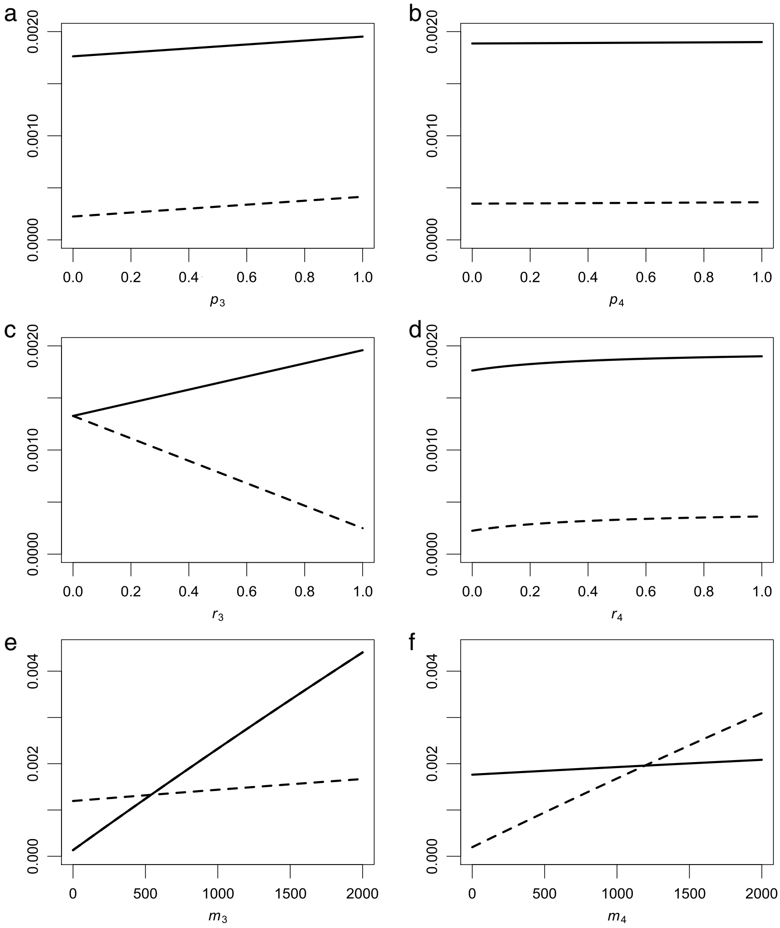

FIGURE 5. Dependence of the asymptotic hazard rate (c.f. Eq. (18)) on parameters p_3 (a), p_4 (b), r_3 (c), r_4 (d), m_3 (e) and m_4 (f). Default parameter values are $m_0 = 50$, $m_3 = 800$, $m_4 = 800$, $\hat{p}_0 = 0.001$, $p_3 = 0.7$, $p_4 = 0.7$, $r_3 = 0.9$, $r_4 = 0.9$, $q = 0.95$, except in dottedlines in (a), (b), (c) and (d) where $m_3 = 100$, and dotted lines in (e) and (f) where $r_3 = 0.1$.

Since $a_i(q)$ decreases in m_i and $G_0(s)$ increases in s, it follows from (18) that the asymptotic hazard rate increases monotonically with m_3 and m_4. The rate of this increase is linked to r_3, as illustrated in Figs. 5e and f. If r_3 is low, a smaller number of type-3 individuals flower, so m_3 has a smaller effect on the asymptotic hazard rate. At the same time, more type-4 individuals are produced, so a decrease in r_3 is associated with an increase in the sensitivity of the asymptotic hazard rate to m_4.

Note that for Poisson and Geometric distributions,

$$\underset{m_i \to \infty}{lim} G_i\left(s\right) = 0 \text{ if } s \neq 1. \tag{20}$$

Thus, when we let m_3 and m_4 tend to infinity, the right hand side of (18) tends to:

$$1 - G_0\left((1-\hat{p}_0) + \hat{p}_0\left((1-r_3)(1-p_3) + \frac{b_3(1-r_4)(1-p_4)}{1-b_4}\right)\right) \tag{21}$$

where the argument of G_0 can be interpreted as the probability that a hybrid seed produces a non-flowering plant, or does not germinate. This result implies that at high hybrid fecundities, the production of hybrids becomes the limiting factor of introgression.

To further examine the effects of the shapes of seed number distributions on the asymptotic hazard rate, we studied its Taylor approximation in the vicinity of $q = 1$ (derivation see Appendix):

$$\hat{H}\left(q\right) = \left(\begin{array}{c} \hat{p}_0^2 m_0 \left(r_3 m_3 + \frac{b_3}{1-b_4} r_4 m_4\right)(1-q) \\ -\frac{1}{2}\beta_0 \hat{p}_0^4 \left(r_3 m_3 + \frac{b_3}{1-b_4} r_4 m_4\right)^2 (1-q)^2 \\ -\frac{1}{2} m_0 \hat{p}_0^3 \left(r_3 \beta_3 + \frac{b_3}{1-b_4} r_4 \beta_4\right)(1-q)^2 + O\left((1-q)^3\right) \end{array}\right) \tag{22}$$

where $\beta_i = G''_i\left(1\right) = E\left[X_i\left(X_i - 1\right)\right] = Var\left[X_i\right] + E\left[X_i\right]\left(E\left[X_i\right] - 1\right)$, X_i represents the number of seeds produced by a type-i individual and $Var[X_i]$ represents its variance. This result indicates that the hazard rate decreases with increasing variance of seed production by hybrids.

The direction of the effect of r_3 depends on the values of other parameters (see Fig. 5c). To get an indication of the parameter ranges where this changes, we studied the first-order term of the Taylor approximation in (22). This term increases in r_3 if the expected number of seeds produced by a flowering one-year old hybrid plant is larger than the expected number of seeds it will produce if it postpones flowering:

$$m_3 > \frac{p_3 r_4 m_4}{1 - \left(1 - r_4\right) p_4}. \tag{23}$$

Numerical work confirmed that this inequality provides a good indicator of the switching boundary.

There will be a steady increase in the asymptotic hazard rate with increasing m_0, unless $q = 1$. This can be seen from (18) and (22). It follows from (20) that for Poisson-distributed hybrid formation, the asymptotic hazard rate approaches one at large m_0. The approximation in (22) indicates that an increased variance in the number of hybrids produced results in a lower asymptotic hazard rate. This can also be seen in Fig. 4a, since a Geometric distribution has a larger variance than a Poisson distribution with the same expectation (see also section 5.4).

5.3. **Effects of backcross fitness.** To examine the effects of backcross fitness relative to the wild type, on introgression success probability, we use the approximation for establishment success of a slightly supercritical branching process, which was derived by Haldane (1927) and later by Eshel (1981) in a more general setting (see also section 4.3). This approximation is based on the second order

Taylor approximation of the right-hand side of (13) in the point $\zeta_{WT} = \zeta_{BC}$, and leads to:

$$(1-q) \approx \frac{2\left(\frac{\zeta_{BC}}{\zeta_{WT}} - 1\right)\zeta_{WT}^2}{\eta_{BC}} \tag{24}$$

where $\eta_{BC} = a''_5(1) + \frac{b_5}{1-b_6}a''_6(1)$. Since η_{BC} increases with increasing variance of backcrossed seed production, such variance decreases establishment success, and therefore decreases the hazard rate. Substituting the approximation in (24) in the Taylor-approximation of the asymptotic hazard rate in (22) gives an approximation of the asymptotic hazard rate for situations where the fitnesses of the backcrosses and wild type are nearly equal. Since the variance in the backcross seed production appears in the first order term of the resulting approximation, it has a larger effect on the hazard rate than the variances of hybrid seed production, or hybrid formation, which only affect the second order terms. This is illustrated in Fig. 6.

5.4. **Effects of variances.** The Taylor approximations of the asymptotic hazard rate presented in sections 5.2 and 5.3 indicate that variance in seed production and hybrid formation reduces the hazard rate. These effects cannot be studied well, however, with the models that we used up to now, because in Poisson and Geometric distributions the variance and the mean are interdependent. Furthermore, in the previous examples we used relatively large values for the mean seed productions, m_i. To study the effects of variance more closely, we use a model with lower mean values, and so-called Linear fractional distributions for seed production and hybrid formation. The probability generating functions for Linear fractional distributions have the form: $G_i(s) = 1 - (b/(1-c)) + (bs/(1-cs))$, with $c \in (0,1)$ and $b \in (0, 1-c)$. The mean and variance, are, respectively: $b/(1-c)^2$ and $\left(b(1-c)^2 - b^2\right)/(1-c)^4$.

As can be seen from Fig. 6, increasing variances can indeed reduce the hazard rate considerably, and changes in variance in backcross seed production have the largest effect. Both results agree with the predictions from the Taylor approximations.

6. DISCUSSION

In this paper we showed how to model and quantify stochastic introgression processes. As illustrated in section 3, predictions of stochastic introgression models differ strongly from those of their deterministic analogues (Fig. 3). Most importantly, deterministic models ignore the initial period before an introgression event occurs. This period, however, strongly determines the risk of introgression, since exponential growth only occurs after initiation of a successful lineage. Furthermore, the high variance between the results of different simulation runs indicates that a representation of population sizes by their expectations, as in deterministic models, is not very useful.

Whereas we focused on the distribution of the lengths of initial periods, previous introgression research, based on deterministic models, considered changes in population sizes of individuals carrying the foreign gene. A complete characterization of the process would involve both aspects. Serra (2006) showed that for a two-type model with one supercritical type, the distribution of the time until the supercritical population reaches a high, fixed level x can be approximated well by

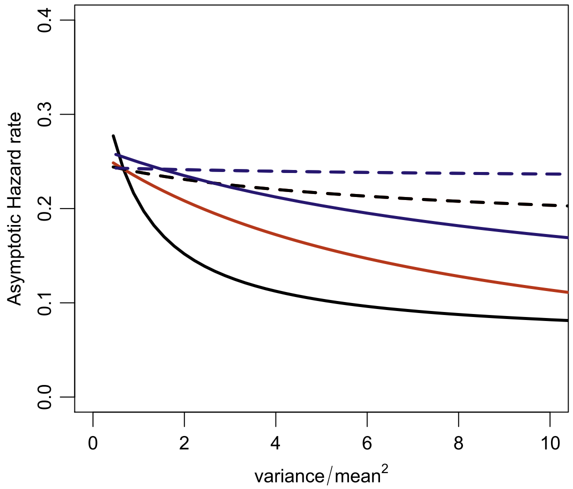

FIGURE 6. The asymptotic hazard rate (Eq. (18)) plotted against the variance of backcrossed seed production (black), hybrid seed production (blue), and hybrid formation (red). Default parameter values: $\hat{p}_0 = 0.8$, $p_3 = p_4 = p_5 = p_6 = 0.7$, $r_3 = r_4 = r_5 = r_6 = 0.9$, $m_0 = m_5 = m_6 = 1.5$, $m_3 = m_4 = 1.0$. All distributions are Linear Fractional, with default variance 1.5. Solid lines represent cases where one-year old plants variances are changing; dotted lines represent the case for changing the variance of older plants. See main text for further details.

the distribution of the time T until an introgression event occurs plus the time it takes a single type supercritical branching process starting with one individual to grow up to level x. The simulation results presented in Fig. 3 indicate that a similar result holds in this case. This is a subject of further research.

We introduced the hazard rate as a measure to quantify the distribution of T. This function specifies the instantaneous risk of an introgression event in the course of time. It is comparable to e.g., the age-dependent mortality risk of humans in demography. We showed that when there is an initial fitness bottleneck and repeated hybrid formation, the hazard rate is initially zero and then increases to an asymptote between zero and one. In the presented example, this increase happens very quickly, and the distribution of times until introgression events is well approximated by a Geometric distribution with a time lag. In situations with, e.g., multi-generation bottlenecks, or the presence of a seed bank with a gradual build-up of introgressed seeds, the hazard rate will increases to its asymptotic level more gradually. We expect, however, that time-lagged Geometric distributions will still provide good approximations. This implies that after the initial lag time, the process can be considered as a lottery where each year there is a constant chance $\hat{H}(q)$ of an introgression event. As a consequence, the distribution of T will be very skewed (as in, e.g., Fig. 5), and the expected value of T will be a misleading measure of introgression risk, since in the majority of cases introgression events will take place before this time.

In the relatively simple example considered here, it is possible to derive an explicit expression for the hazard rate. For more complicated cases this will usually not be possible. The values of q and $f_{I_0(n)}(q)$ then have to be determined numerically. Note that this is a far more efficient procedure than performing simulations to estimate the hazard or survival function of T and, furthermore, it provides exact values (apart from numerical errors, which can be minimized by algorithm optimization) rather than estimates. This is especially important in the tail of the distribution of T (i.e., for large values of n), where the precision of estimators is low. Furthermore, whereas explicit expressions may not always be available, recurrence relations such as (15) can also be used to derive approximations, or perform analytical studies. Procedures for approximating extinction probabilities of multitype processes can be applied for that purpose, by making use of the analogy explained in section 4.3.

As an example, we studied a simple model of a monocarpic plant species. We found that the asymptotic hazard rate increases monotonically in all life history parameters except the first-year flowering probability, r_3 (Fig. 5). Effects of r_3 reflect the trade offs between seed production in the first or later years: the hazard rate decreases with r_3 if postponement of flowering leads to a much higher seed production. Under natural conditions, such life history parameters will evolve to optimal combinations, given the trade offs. So, we would expect a low probability of flowering in the first year when seed production at an older age is relatively high and vice versa, leading to high hazard rates. Cultivar genes might further increase it by lowering the first year flowering probability of hybrids.

As long as adult plant survival chances are positive, their magnitude does not make much difference for the hazard rate, and the same holds for the flowering probability of two-year and older plants, r_4 (Figs. 5a,b,d). The effects of expected seed numbers are much more pronounced (Figs. 5e,f). In addition, the shapes of seed production distributions are also important, for instance, a higher variance decreases the hazard rate (22). As illustrated in Fig. 6, this effect can be significant. We expect this result to hold generally, regardless of the details of the model. Changes in variance of backcross seed production have larger consequences than those in variance of hybrid seed production (e.g., Fig. 6). This result is intuitively clear, since all subsequent generations of offspring from backcrossed individuals are governed by the backcrossed life-history parameters, whereas hybrid life-history parameters only affect one generation.

Based on these results we would advise that in the context of introgression risk, it is better to study effects of hybridization and backcrossing on seed production distributions and first-year flowering probability rather than probabilities of adult survival and flowering at older ages. Since we considered a very simplified model, however, these conclusions are only tentative.

We expect that an important application of the hazard rate will be the study of effects of time-varying environmental conditions, or crop management. In such cases there can be periods where the hazard rate decreases, for instance when crops are rotated. Methods for calculating hazard rates from time-inhomogeneous branching processes (Smith and Wilkinson, 1969) are currently being developed.

The model can be extended easily to incorporate other types of life histories and other modes of density dependent competition, provided that the wild population is large and homogeneous. As long as this holds, interactions between individuals with introgressed genotypes can be ignored initially and branching process models can be used to study their invasion dynamics. In small or spatially structured populations, invaders may affect each other already at low numbers. This necessitates the use of frequency- and density-dependent invasion models. Until now there are not many mathematical results on such generalizations of branching processes (but see Jagers and Klebaner, 2000). Furthermore, foreign genes invading small wild populations may become established by drift, even without a fitness advantage (for $q = 1$), especially when there are repeated invasions (Haygood et al., 2003). Methods for quantifying introgression processes in such situations remain to be developed.

Even in large wild populations, if invasion is successful, the density of invaders will eventually become so large that the possibility that invaders interact directly cannot be neglected. The invasion model that we used can be considered as an approximation of a more complicated model that includes such interactions, valid at low invader densities. In this light, 1? q should be considered as an approximation of the probability that the numbers of type-5 and 6 individuals reach such high levels that the probability that the foreign gene disappears from the population due to demographic stochasticity can be neglected. This type of approach is common in invasion studies (e.g., Garnier and Lecomte, 2006), and generally works well (see Champagnat et al., 2006).

We did not incorporate explicit genetics into the model, but obviously this is an important generalization, which is, for instance, needed to study effects of linkage and hitchhiking. A huge number of types may be needed to represent the different possible introgressed genotypes. Models can be considerably simplified, however, if some genotypes have equal fitnesses. In our model, for instance, we assumed that individuals of BC1 and later backcross generations have the same fitness.

Important other issues for future research are the effects of spatial structure and gene flow between subpopulations (e.g., Hanski, 1999) on the hazard rate.

In conclusion, further development of stochastic models for introgression research is needed, and we expect that this will require much more work. However, it is imperative to use such models, because, as we showed, deterministic models ignore important factors, and give misleading results. To analyze and interpret results of stochastic models, and go beyond simulations, we need measures such as the hazard rate, which adequately quantify essential features of stochastic introgression processes.

Appendix A. Appendix

A.1. **Derivation of (15).** Here, we will use a multi-dimensional extension of the definition of the p.g.f. given in the main text, to deal with multiple types. Thus, the p.g.f. of the offspring of a single type-i individual ($i \in \{0, 3, 4, 5\}$) is defined as:

$$F_i\left(s_0, s_3, s_4, s_5\right) = E\left[s_0^{Z_0(1)} s_3^{Z_3(1)} s_4^{Z_4(1)} s_5^{Z_5(1)} \left| Z_i\left(0\right) = 1, Z_j\left(0\right) = 0 \text{ for } j \neq i \right.\right] \quad \text{(A.1)}$$

where $Z_i(n)$ denotes the number of type-i individuals at time n. In the considered model,

$$F_0\left(s_0, s_3, s_4, s_5\right) = s_0 G_0\left(\hat{p}_0 s_3 + \left(1-\hat{p}_0\right)\right) \tag{A.2}$$

since a type-0 individual produces one of its own type, and a random number of seeds according to a p.g.f. $G_0(s)$, of which a proportion $\hat{p}_0$ of the seeds will germinate.

$$F_3\left(s_0, s_3, s_4, s_5\right) = r_3 G_3\left(\hat{p}_0 s_5 + \left(1-\hat{p}_0\right)\right) + \left(1-r_3\right) p_3 s_4 + \left(1-r_3\right)\left(1-p_3\right) \tag{A.3}$$

since a type-3 individual may flower with a probability r_3 and produce some number of seeds according to a p.g.f. $G_3(s)$, of which a proportion $\hat{p}_0$ will flower to become type-5 individuals.

Following a similar reasoning, we also obtain the following expression:

$$F_4\left(s_0, s_3, s_4, s_5\right) = r_4 G_4\left(\hat{p}_0 s_5 + \left(1-\hat{p}_0\right)\right) + \left(1-r_4\right) p_4 s_4 + \left(1-r_4\right)\left(1-p_4\right) \tag{A.4}$$

Next, we consider the generating function of $I_i(n)$:

$$\begin{aligned} f_{I_i(n)}(s) &= E\left[s^{I_i(n)}\right] \\ &= E\left[E\left[s^{I_i(n)} \mid Z_0(1), Z_3(1), Z_4(1), Z_5(1)\right]\right] \\ &= E\left[E\left[s^{\sum_{k=1}^{Z_0(1)} I_0(n-1)^{(k)} \sum_{k=1}^{Z_3(1)} I_3(n-1)^{(k)} \sum_{k=1}^{Z_4(1)} I_4(n-1)^{(k)} + Z_5(1)} \middle| Z_0(1), \ldots, Z_5(1)\right] \middle| Z_i(0)=1, Z_j(0)=0 \text{ for } j \neq i\right] \end{aligned} \tag{A.5}$$

where $I_j(n-1)(k)$ denotes the total number of type-5 individuals up to and including the next $n-1$ generations produced by type-3 and 4 individuals of the lineage of the kth individual of type j in generation 1. Using the fact that individuals reproduce independently, and that individuals of the same type have identical offspring distributions, we can rewrite (A.5) as follows:

$$\begin{aligned} E\Bigg[& E\left[s^{I_0(n-1)}\right]^{Z_0(1)} E\left[s^{I_3(n-1)}\right]^{Z_3(1)} E\left[s^{I_4(n-1)}\right]^{Z_4(1)} s^{Z_5(1)} \\ & \Bigg| Z_i\left(0\right)=1, Z_j\left(0\right)=0 \text{ for } j \neq i \Bigg] \end{aligned} \tag{A.6}$$

and this equals

$$\begin{aligned} & E\Big[\left(f_{I_0(n-1)}\left(s\right)\right)^{Z_0(1)} \left(f_{I_3(n-1)}\left(s\right)\right)^{Z_3(1)} \left(f_{I_4(n-1)}\left(s\right)\right)^{Z_4(1)} s^{Z_5(1)} \Big| \\ & Z_i\left(0\right)=1, Z_j\left(0\right)=0 \text{ for } j \neq i\Big] \\ & \quad = F_i\left(f_{I_0(n-1)}\left(s\right), f_{I_3(n-1)}\left(s\right), f_{I_4(n-1)}\left(s\right), s\right) \end{aligned} \tag{A.7}$$

where we have used (A.1) in the last line. Using the above relationships, the equations of (15) follow.

A.2. Derivation of Eq. (17). Note that the last equality in Eq. (15) can be written as:

$$f_{I4(n)}(q) = a_4\left(q\right) + b_4 f_{I4(n-1)}(q) \tag{A.8}$$

with $a_4(q)$ and b_4 as defined in Eq. (11). Solving this recursion yields:

$$\begin{aligned} f_{I4(n)}\left(q\right) &= a_4\left(q\right) + b_4 f_{I4(n-1)}\left(q\right) = a_4\left(q\right) + a_4\left(q\right) b_4 + {b_4}^2 f_{I4(n-2)}\left(q\right) \\ &= a_4\left(q\right) + a_4\left(q\right) b_4 + a_4\left(q\right) {b_4}^2 + {b_4}^3 f_{I4(n-3)}\left(q\right) = \ldots = {b_4}^n + \sum_{j=0}^{n-1} a_4\left(q\right) {b_4}^j \\ &= a_4\left(q\right)\left(\frac{1-{b_4}^n}{1-b_4}\right) + {b_4}^n \end{aligned} \tag{A.9}$$

Note that if we use $n=0$ we get $f_{I4(0)}(q) = 1$, so the expression derived in (A.9) is valid for all n. Combining (A.9) with (15) and (16) gives Eq. (17).

A.3. Effects of r_4, p_3 and p_4 on the asymptotic hazard rate. The argument of G_0 in the expression of the asymptotic hazard rate in (18) is:

$$\hat{p}_0 \left(\begin{array}{c} r_3 G_3\left(\hat{p}_0 q + (1-\hat{p}_0)\right) + (1-r_3)(1-p_3) \\ + \frac{(1-r_3)p_3}{1-(1-r_4)p_4}\left(r_r G_4\left(\hat{p}_0 q + (1-\hat{p}_0)\right) + (1-r_4)(1-p_4)\right) \end{array} \right) + (1-\hat{p}_0) \tag{A.10}$$

Furthermore, $G_0(s)$ is a monotonically increasing function of s. Therefore, if the expression in (A.10) increases with a certain parameter, then the asymptotic hazard rate decreases, and vice versa. The derivative of (A.10) with respect to p_4 is positive if

$$\begin{aligned} &-\left(1-\left(1-r_4\right)p_4\right) + \left(r_4 G_4\left(\hat{p}_0 q + (1-\hat{p}_0)\right) + \left(1-r_4\right)\left(1-p_4\right)\right) > 0 \Leftrightarrow \\ &r_4 G_4\left(\hat{p}_0 q + (1-\hat{p}_0)\right) - r_4 > 0 \Leftrightarrow G_4\left(\hat{p}_0 q + (1-\hat{p}_0)\right) > 1 \end{aligned} \tag{A.11}$$

which gives a contradiction. We can conclude that (A.10) is monotonically non-increasing with p_4, and thus the asymptotic hazard rate must be monotonically non-decreasing in p_4. Results for p_3 and r_4 can be derived in a similar way.

A.4. Derivation of (22). Consider $G(s)$ to represent the p.g.f. of some random variable X. It can easily be shown that:

$$G\left(1\right) = 1, G'(1) = \mathrm{E}\left[X\right],\ G''\left(1\right) = \mathrm{E}\left[X\left(X-1\right)\right] \tag{A.12}$$

Looking at (18) and its derivatives at $q=1$, we find the following expressions.

$$\hat{H}\left(1\right) = 0 \tag{A.13}$$

$$\hat{H}'\left(1\right) = -\hat{p}_0^2 m_0 \left(r_3 m_3 + \frac{b_3}{1-b_4} r_4 m_4\right) \tag{A.14}$$

$$\hat{H}''\left(1\right) = -\beta_0 \hat{p}_0^4 \left(r_3 m_3 + \frac{b_3}{1-b_4} r_4 m_4\right)^2 - m_0 \hat{p}_0^3 \left(r_3 \beta_3 + \frac{b_3 r_4}{1-b_4} \beta_4\right) \tag{A.15}$$

where β_i, represents the second derivative of $G_i(s)$ evaluated at 1. Equation (22) follows from substituting these results in the second order Taylor expansion of $\hat{H}(q)$ around the point 1.

References

Abbott, R.J., James, J.K., Milne, R.I., Gillies, A.C.M., 2003. Plant introductions, hybridization and gene ow. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 358, 11231132.

Allendorf, F.W., Leary, R.F., Spruell, P., Wenburg, J.K., 2001. The problems with hybrids: setting conservation guidelines. Trends Ecol. Evol. 16, 613-622.

Athreya, K.B., Ney, P.E., 1972. Branching Processes. Springer, Berlin.

Champagnat, N., Ferriere, R., Mlard, S., 2006. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. Theor. Popul. Biol. 69, 297321.

Davis, S.A., Catchpole, E.A., Pech, R.P., 1999. Models for the introgression of a transgene into a wild population within a stochastic environment, with applications to pest control. Ecol. Model. 119, 267275.

Demon, I., Haccou, P., van den Bosch, F., 2007. Introgression of resistance genes between populations: A model study of insecticide resistance in Bemisia tabaci. Theor. Popul. Biol. 72, 292-304.

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, 539-63.

Ellstrand, N.C., 2003. Dangerous Liaisons? When cultivated plants mate with their wild relatives. The John Hopkins University Press, Baltimore.

Eshel, I., 1981. On the survival probability of a slightly advantageous mutant gene with a general distribution of progeny size: A branching process model. J. Math. Biol. 12, 355-362.

Feller, W., 1968. An Introduction to Probability Theory and its Applications. Third Edition, Volume 1. John Wiley and Sons, New York.

Garnier, A., Lecomte, J., 2006. Using a spatial and stage-structured invasion model to assess the spread of feral populations of transgenic oilseed rape. Ecol. Model. 194, 141149.

Haccou, P., Meelis, E., 1994. Statistical analysis of behavioural data. Oxford University Press, Oxford.

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.

Haldane, J.B.S., 1927. A mathematical theory of natural and artificial selection. V. Selection and mutation. Proc. Cambridge Philos. Soc. 23, 838-844.

Hanski, I., 1999. Metapopulation Ecology. Oxford University Press, Oxford.

Hauser, T.P., Shaw, R.G., Ostergard, H., 1998. Fitness of hybrids between weedy Brassica rapa oilseed rape (B. napus). Heredity, 81, 436-443.

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.

Huxel, G. R., 1999. Rapid displacement of native species by invasive species: effects of hybridization. Biol. Conservat. 89, 143-152.

Jagers, P., Klebaner, F.C., 2000. Population-size-dependent and age-dependent branching processes. Stochastic Process. Appl. 87, 35254.

Jenczewski, E., Ronfort, J., Chvre, A., 2003. Crop-to-wil gene flow, introgression and possible fitness effects of transgenes. Environ. Biosafety Res. 2, 9-24.

Kalbfleisch, J.D., Prentice, R.L., 2002. The statistical analysis of failure time data. 2nd edition. 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.

Levin, D.A., Francisco-Ortega, J., Jansen, R.K., 1996. Hybridizationand the extinction of rare plant species. Conserv. Biol. 10, 1016.

Maan, S.S., 1987. Interspecific and intergeneric hybridisation in wheat. In Heyne, E.G., (Ed.), Wheat and wheat improvement. ASA, CSSA and SSSA, Madison, 453-461.

Michor, F., Nowak, M.A., Iwasa, Y., 2006. Stochastic dynamics of metastasis formation. J. Theor. Biol. 240, 521-530.

Reichman, J.R., Watrud, L.S., Lee, E.H., Burdick, C.A., et al., 2006. Establishment of transgenic herbicide-resistant creeping bentgrass (Agrostis stolonifera L.) in nonagronomic habitats. Mol. Ecol. 15, 13, 4243-4255.

Rieger, M.A., Lamond, M., Preston, C., Powles, S.B., Roush, R.T., 2002. Pollen-mediated movement of herbicide resistance between commercial canola fields. Science, 296, 2386-2388.

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., 2006. On the waiting time to escape. J. Appl. Prob. 43, 296-302

Serra, M.C., Haccou, P., 2007. Dynamics of escape mutants. Theor. Popul. Biol., 72, 167-178.

Smith, W.L., Wilkinson, W.E., 1969. On branching processes in random environments. Ann. Math. Statist. 40, 814827.

Snow, A.A., Andersen, B., Jorgensen, R.B., 1999. Costs of transgenic herbicide resistance introgressed from Brassica napus into weedy B. rapa. Mol. Ecol. 8, 605-15.

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.