Heredity (2010) 104, 28–39

www.nature.com/hdy

ORIGINAL ARTICLE

Genetic diversity in a crop metapopulation

J van Heerwaarden[1], FA van Eeuwijk[1] and J Ross-Ibarra[2]
[1]*Biometris, Department of Plant Sciences, Wageningen University, Wageningen, The Netherlands and* [2]*Department of Plant Sciences, University of California, Davis, CA, USA*

The need to protect crop genetic resources has sparked a growing interest in the genetic diversity maintained in traditional farming systems worldwide. Although traditional seed management has been proposed as an important determinant of genetic diversity and structure in crops, no models exist that can adequately describe the genetic effects of seed management. We present a metapopulation model that accounts for several features unique to managed crop populations. Using traditional maize agriculture as an example, we develop a coalescence-based model of a crop metapopulation undergoing pollen and seed flow as well as seed replacement. In contrast to metapopulation work on natural systems, we model seed migration as episodic and originating from a single source per population rather than as a constant immigration from the entire metapopulation. We find that the correlated origin of migrants leads to surprising results, including a loss of invariance of within-deme diversity and a parabolic relationship between F_{ST} and migration quantity. In contrast, the effects of migration frequency on diversity and structure are more similar to classical predictions, suggesting that seed migration in managed crop populations cannot be described by a single parameter. In addition to migration, we investigate the effects of deme size and extinction rates on genetic structure, and show that high levels of pollen migration may mask the effects of seed management on structure. Our results highlight the importance of analytically evaluating the effects of deviations from classical metapopulation models, especially in systems for which data are available to estimate specific model parameters.
Heredity (2010) **104,** 28–39; doi:10.1038/hdy.2009.110; published online 9 September 2009

Keywords: metapopulation; genetic diversity; maize; crop diversity

Introduction

The need to protect the genetic resources of the world's most important cultivated plants has sparked a growing interest in crop diversity (Brush, 2000, 2004). Considerable effort has been expended in cataloging agricultural genetic diversity (Brush, 2004; Pressoir and Berthaud, 2004; Perales *et al*., 2005; Jarvis *et al*., 2008), and debating the practical, political and scientific bases for maintaining diversity of our major food crops (Brush, 2000; Fowler and Hodgkin, 2004; Esquinas-Alcazar, 2005). It has become clear that a large amount of genetic diversity is contained within the small-scale agricultural systems that are typical of the developing world (Jarvis *et al*., 2008). In contrast to commercial farmers, small-scale farmers generally obtain their crop varieties through a traditional system of seed management that is based on saving and exchange of local germplasm (Almekinders *et al*., 1994). The dependence on this traditional seed system means that genetic diversity is affected by seed management dynamics (Louette *et al*., 1997). For this reason, traditional seed management has become an important research topic for crop conservationists (Badstue *et al*., 2007).

Borrowing from population biology, researchers have noted parallels between traditional crops and subdivided populations (Brush, 1999; Alvarez *et al*., 2005; Dyer and Taylor, 2008). Surprisingly little effort, however, has been spent on developing models of population subdivision that are suitable for traditionally managed crops. To date, discussion of agroecosystems in a metapopulation context has been predominantly metaphorical (Louette, 2000; Pressoir and Berthaud, 2004; Alvarez *et al*., 2005), and the few attempts to treat population dynamics mathematically (Heisey and Brennan, 1991; Dyer and Taylor, 2008) have ignored population genetics completely.

Population genetic models of subdivided species have been instrumental to our understanding of neutral genetic diversity and structure in nature. General results from models such as Wright's island model (Wright, 1951) and more recent metapopulation models (Slatkin, 1977; Maruyama and Kimura, 1980; Lande, 1992; Whitlock and Barton, 1997; Wakeley and Aliacar, 2001) have served to predict the genetic effects of population size, migration rates and extinction/colonization in natural populations. And in spite of vast differences in natural history, most species of animals and plants present patterns of demography and migration that can be interpreted and modeled in a metapopulation framework (Harrison and Taylor, 1997).

In contrast to natural populations, demography and seed migration in cultivated plants are subject to conscious intervention by farmers. Traditional agricultural practices are rather well documented, particularly for grain crops such as maize, rice, sorghum and millet (Longley and Richards, 1993; Louette *et al*., 1997; Brocke *et al*., 2003; Barnaud *et al*., 2008). The available literature suggests that farmer-managed crops differ from a classic metapopulation in several respects. First, many crops are characterized by having a large number of seeds per

Correspondence: Dr J van Heerwaarden, Department of Plant Sciences, University of California, One Shields Ave, Davis, CA 95616, USA.
E-mail: jvheerwaarden@gmail.com
Received 27 October 2008; revised 26 July 2009; accepted 27 July 2009; published online 9 September 2009

inflorescence (Harlan *et al*., 1973). This has made the inflorescence the focus of seed management (Li and Wu, 1996; Louette and Smale, 2000; Brocke *et al*., 2003; Perales *et al*., 2003; Barnaud *et al*., 2007), and because the seed is derived from a limited number of maternal plants, effective deme size is expected to be much smaller than census size (Louette *et al*., 1997). Second, seed migration into demes is not random. Whereas existing models assume that migrants are drawn from the entire metapopulation, farmers generally obtain seed from a very limited number of familiar sources (Almekinders *et al*., 1994; Zeven *et al*., 1999; Badstue *et al*., 2007; Barnaud *et al*., 2008). In cases where detailed data are available, farmers are reported to receive seed from a single source each time (Rice *et al*., 1998). Third, seed is often recycled for several years without any influx of foreign germplasm (Perales *et al*., 2003), so seed migration into individual demes is episodic rather than continuous. Finally, the process of extinction and recolonization generally occurs without passing through the bottleneck that is assumed in most metapopulation models, because farmers will generally obtain enough seed to plant the desired acreage of land instead of reducing the planted area. There is thus good reason to believe that farmer-managed seed systems may deviate substantially from classical metapopulation models, yet the population-genetic implications of farmer–seed management and the validity of models of subdivided populations have not been effectively explored.

In this paper, we explicitly address the population genetic dynamics of managed crops in a metapopulation framework. We show that incorporating characteristics that distinguish crop metapopulations from most natural species leads to predictions that are different from those emanating from classical models. We begin by generalizing a common approach to modeling neutral genetic diversity in metapopulations and adapt it to include several important features unique to farmer-managed crops. We present results on the effects of seed migration quantity, migration frequency and extinction on patterns of genetic diversity in crop metapopulations. Notably, we find that two predictions of classical models—namely geographical invariance of within-deme diversity and the reduction of genetic structure through migration—do not necessarily hold in farmer-managed systems. Finally, we end with analysis of other factors of potential importance in crop metapopulations, including deme size and pollen migration. We frame our work in the context of maize cultivation to take advantage of the large body of knowledge of seed management practices and diversity at the farm level, but we expect our results to be representative for other sexually propagated crops.

Materials and Methods

Defining the model

To illustrate the features that are unique to our crop metapopulation, we define what we will refer to as a classic metapopulation model. Our definition is based on Slatkin's (1977) model II. This model describes a number of discrete sub-populations, or demes, consisting of N sexually reproducing diploid organisms. Demes are linked by a constant flow of migrants sampled from the entire metapopulation. In case of extinction of demes, there is instant colonization by a limited number of colonists. Colonists are either drawn at random from the metapopulation (migrant pool model), or each deme receives colonists from one randomly chosen source deme (propagule pool model).

We start by presenting a generalization of the recurrence methods initially developed by (Latter, 1973; Slatkin, 1977; Maruyama and Kimura, 1980), and reframed in terms of average coalescence times by Pannell and Charlesworth (1999). A subdivided population is described in terms of the mean time to coalescence for two alleles sampled at generation t from either a single deme (T_0), or two different demes (T_1). Mean coalescence time can be defined as the time that has elapsed since two sampled alleles were derived from the same ancestral allele, and is directly proportional to genetic diversity under the infinite sites model without recombination (Hudson, 1990). T_0 and T_1 thus represent the equilibrium values of genetic diversity for alleles sampled within and between demes, respectively. Average diversity for the entire metapopulation may be expressed as $T = (T_0/n) + T_1(1 - (1/n))$, where n is the total number of demes (Pannell and Charlesworth, 1999). Genetic structure, defined as the relative reduction in within-deme diversity, is estimated by $F_{ST} = (T - T_0)/T$ (Slatkin, 1991).

Coalescence of a pair of alleles can only occur when they are present in the same deme. We will refer to this condition as co-location. If we define T'_0 and T'_1 as the mean coalescence times for alleles in the previous generation, then for an allele pair sampled at generation t, three possible coalescence times exist: one generation for those that co-located and coalesced in the previous generation, $1 + T'_0$ generations for alleles that co-located in the previous generation but did not coalesce, and $1 + T'_1$ generations for two alleles from different demes. Mean values of T_0 and T_1 may then be calculated by the following recursion equations:

$$T_0 = \sum_i a_i P_i + \sum_i a_i (1 - P_i)(1 + T'_0) + \left(1 - \sum_i a_i\right)(1 + T'_1) \quad (1)$$

$$T_1 = \sum_i b_i P_i + \sum_i b_i (1 - P_i)(1 + T'_0) + \left(1 - \sum_i b_i\right)(1 + T'_1) \quad (2)$$

where P_i is the probability of coalescence for two parental co-locating alleles. The subscript reflects the fact that coalescence probabilities may be different for different combinations of alleles. The terms a_i and b_i are compound terms expressing the proportion of all possible allele pairs that co-locate and have a coalescence probability of P_i. The sums $\sum_i a_i$ and $\sum_i b_i$ thus represent the mean co-location probabilities for allele pairs sampled within and between demes.

At equilibrium $T_0 = T'_0$ and $T_1 = T'_1$, we may therefore substitute T'_0 and T'_1 with T_0 and T_1 in equations (1) and

(2) such that

$$T_0 = \frac{1}{\bar{P}}\left(\frac{1-\sum_i a_i}{\sum_i b_i}+1\right) \quad (3)$$

and

$$T_1 = T_0(1-\bar{P}b)+\frac{1}{\sum_i b_i} \quad (4)$$

where

$$\bar{P} = \left(1-\sum_i a_i\right)\bar{P}b + \left(\sum_i a_i\right)\bar{P}a \quad (5)$$

with $\bar{P}b = \sum_i b_i P_i / \sum_i b_i$ being the mean coalescence probability for co-locating allele pairs from different demes, and $\bar{P}a = \sum_i a_i P_i / \sum_i a_i$ representing the mean coalescence probability for co-locating allele pairs from the same deme. Equation (5) may be interpreted as the mean coalescence probability for pairs of alleles that are present in the same deme. The equilibrium expression for T_0 may therefore be understood as follows. At any point in time a fraction $1-\sum_i a_i$ of alleles sampled from a deme will have left their deme and will return to a single deme with a probability of $\sum_i b_i$, or on average each $1/\sum_i b_i$ generations. A fraction $\sum_i a_i$ will not leave the deme and will be in the same deme in the previous generation with probability $\sum_i a_i$. This is equivalent to a fraction $\sum_i a_i$ sharing the same deme every $1/\sum_i a_i$ generations. On average, alleles will thus share a deme every $(1-\sum_i a_i)(1/\sum_i b_i)+(\sum_i a_i)(1/\sum_i a_i) = ((1-\sum_i a_i)/\sum_i b_i)+1$ generations. Each generation that two alleles spend in the same deme they have an average coalescence probability $\bar{P}$, so the mean time to coalescence is given by equation (3). Within-deme coalescence is thus essentially a function of the time that allele pairs spend within a single deme.

The average coalescence time for allele pairs sampled from two different populations is given in equation (4) by the average time $1/\sum_i b_i$ it takes for two non-colocating alleles to reach the same deme and the mean time needed for two alleles entering the same deme to coalesce. A fraction $\bar{P}b$ of allele pairs coalesces upon entering the same deme and a fraction $1-\bar{P}b$ coalesces in T_0 generations.

Metapopulation model for farmer-managed maize

We will proceed by presenting the parameters of our crop metapopulation model that will allow the estimation of T_0 and T_1 as described above. We consider a diploid, monecious plant species with random mating within demes. There are n demes, each of which consists of seed from N_f ears, yielding N mature plants, with $N_f \ll N$ and a fixed number of N/N_f seeds per ear. Generations are discrete, and the life cycle of each deme consists of two consecutive phases: a reproductive phase and a seed phase. During the reproductive phase random pollination, pollen migration and zygote formation occur. Each new seed that is formed contains a maternal allele inherited from one of N_f ears and a paternal allele derived from one of N pollen parents. A proportion of $1-m_g$ of all paternal alleles will result from random pollination by pollen from the same deme whereas a proportion m_g result from migrant pollen from other demes. Pollen migration follows an island model with migrants originating from any of the other $n-1$ demes.

The seed phase begins after flowering and lasts until the onset of the next reproductive phase. It is in this phase that extinction, recolonization and seed migration take place. Extinction occurs with probability e. Each generation, ne demes go extinct and $n(1-e)$ demes remain. An extinct deme is replaced by introducing N_f ears from the non-migrant fraction of any of $n-1$ extant demes (propagule pool model), with no subsequent migration during the seed phase. Seed migration into individual demes is episodic, occurring with probability p_m. Consequently, an expected fraction p_m of all $n(1-e)$ extant demes receive seed migrants from any of $n(1-e)-1$ potential source demes. There is a single seed source per generation for each deme. For demes in this fraction, N_{fm} migrant ears are planted in addition to $N_f - N_{fm}$ ears taken from the resident deme. The fraction of migrant seeds is thus $m = N_{fm}/N_f$ in demes undergoing migration and $\bar{m} = p_m m$ in all extant demes. For mathematical simplicity, we will assume that $n(1-e)$ is large so that $n(1-e) \approx n(1-e)-1$ and we will use $n(1-e)-1$ as the number of seed sources for both migrants and colonists.

At the end of the seed phase the metapopulation consists of a set of $2Nn$ gene copies that can be divided into non-overlapping subsets of paternal and maternal alleles that did or did not undergo seed extinction, seed migration or pollen flow (Table 1). The proportions represented by these subsets are assumed to remain constant over time. Genetic diversity within this system may now be described as the average time to coalescence for pairs of lineages sampled from the total collection of allele subsets. As outlined in the general model, different combinations of alleles may have different coalescence probabilities when co-locating. Table 2 presents these different probabilities and the corresponding expected fractions a_i and b_i of co-locating allele pairs. Derivation of these terms is given in the Appendix, and R code to calculate T_0, T_1 and F_{ST} under our model is available upon request.

Simulation study

We compared our theoretical results to expectations from simulated data. We made use of a stochastic, biallelic simulation algorithm developed for maize and described in Piñeyro-Nelson *et al*. (2009), modifying it to exactly match the assumptions of our metapopulation model (C++ code is available upon request). Reproduction and seed management were explicitly modeled, with N, N_f, N_{fm} and m_g included as deterministic parameters and p_m and e as binomial probabilities. F_{ST} in each run was

Table 1 Representation of maternal (seed) and paternal (pollen) allele fractions in a metapopulation

	e		$1-e$					
	1		$1-p_m$		p_m			
Seed ($\frac{1}{2}$)	1		1		$1-m$		m	
Pollen ($\frac{1}{2}$)	$1-m_g$	m_g	$1-m_g$	m_g	$1-m_g$	m_g	$1-m_g$	m_g

Table 2 Coalescence probabilities for allele-pairs sampled within and between demes and corresponding co-locating fractions

Sample	*Type*	P_i	*Co-locating fraction for* P_i
Within demes	Seed × seed non-migrants	$P_1 = \frac{1}{2(N_f - N_{fm})}$	$a_1 = \frac{1}{4}p_m(1-e)(1-m)^2$
	Seed × seed migrants	$P_2 = \frac{1}{2N_{fm}}$	$a_2 = \frac{1}{4}(1-e)p_m m^2$
	Seed × seed no migration	$P_3 = \frac{1}{2N_f}$	$a_3 = \frac{1}{4}(1-(1-e)p_m)$
	Pollen × pollen pollen × seed	$P_4 = \frac{1}{2N}$	$a_4 = \left(\frac{3}{4} - m_g\left(1-\frac{m_g}{4}\right)\right)(1-2p_m m(1-e)(1-m)) + \frac{m_g(2-m_g)(1-e)p_m m(1-m)+\frac{1}{4}m_g^2}{(n-1)}$
Between demes	Pollen × pollen pollen × seed	$P_4 = \frac{1}{2N}$	$b_4 = \frac{\left(\frac{3}{4}-m_g\left(1-\frac{1}{4}m_g\right)\right)\left(1-(1-e)^2(1-p_m m)^2\right)}{n(1-e)-1} + \frac{m_g\left(1-\frac{1}{4}m_g\right)}{n-1}$
	Seed × seed	$P_5 = 0$	$b_5 = \frac{\frac{1}{4}\left(1-(1-e)^2(1-p_m m)^2\right)}{n(1-e)-1}$

calculated directly from the simulated allele frequencies as $F_{ST} = \sigma_{\bar{p}}^2/(\bar{p}(1-\bar{p}))$.

Results

We can now use the metapopulation model outlined above to investigate the effects of farmer-mediated demographic processes on genetic diversity and structure in crop metapopulations. In the following sections, we discuss the behavior of the model by deriving analytical approximations, and present graphical results from the full model using the general equilibrium solutions presented in the Appendix.

Effective size of individual demes and coalescence time

The common practice of selecting a limited number of ears per deme as the source of the next generation's seed reduces the effective size of individual demes with respect to the census size N. The inbreeding effective size of a panmictic population is related to the mean probability of coalescence P in the previous generation by $N_e = 1/2P$ (Kimura and Crow, 1963). In our model, we may hence calculate the effective size of a single deme without migration by setting pollen and seed migration to zero and substitute the terms a_i and P_i from Table 2 in equation (5). This yields:

$$N_e = \frac{4N_f N}{3N_f + N} \tag{6}$$

which is identical to Crossa and Vencovsky's (1994) variance effective size with female gametic control. We will use the term N_e to describe the effective size of a single deme in the absence of migration throughout the paper, rather than as effective size of the metapopulation. In the classical metapopulation model without extinction, there is only a single coalescence probability for any pair of parental co-locating alleles. Therefore, $\bar{P} = P$ and we may write:

$$T_0 = \frac{1 - \sum_i a_i}{\sum_i b_i} 2N_e + 2N_e \tag{7}$$

In case of different coalescence probabilities, $\bar{P}$ does not need to be equal to P. It can be shown numerically, however, that $\bar{P}$ in our model closely approximates $1/2N_e$ under a wide range of parameter values. We may thus use equation (7) as an approximation to T_0. Moreover, assuming N_e is large we will use

$$T_1 \approx T_0 + \frac{1}{\sum_i b_i} \tag{8}$$

for between-deme coalescence time. We will make further use of expressions (7) and (8) as they greatly simplify comparison to previous results.

Extinction, seed migration and within-deme coalescence time

Under most models of subdivided populations, the weighted mean within-deme coalescence time can be shown to be unaffected by the rate of migration (reviewed in Nagylaki, 2000). For the classical metapopulation model, with conservative migration and equal deme sizes, this means that T_0 has an expected value of $2N_e n$ (Nagylaki, 1998). Pannell and Charlesworth (1999) showed that including extinction leads to a breakdown of this invariance result. Under extinction, T_0 increases with migration rate because genetic diversity that is lost in the process of extinction and recolonization is partially restored by diversity contained in the migrant pool. When extinction is assumed absent, invariance follows directly from the equilibrium solution for T_0 in the classical metapopulation model. Substituting $a = (1-m)^2 + m^2/(n-1)$ and $b = m^2/(n-1) + 2m(1-m)/(n-1)$ from Pannell and Charlesworth (1999) into equation (7), we may thus write:

$$T_0 = \left(\frac{1-(1-m)^2 - \frac{m^2}{n-1}}{1-(1-m)^2}\right)2N_e(n-1) + 2N_e \tag{9}$$

The term m^2 represents the fraction of allele pairs sampled from two migrant alleles. As migrants are assumed to be a random sample from the metapopulation, they have a co-location probability of $1/(n-1)$. When n is large, $m^2/(n-1)$ can be ignored and equation (9) reduces to

$2N_e n$. Invariance to migration rate may thus be understood as the balance between the fraction $1 - \sum_i a_i$ of allele pairs that do not co-locate and the fraction $\sum_i b_i$ that relocates from different demes.

Seed migration in our model differs in two key aspects from migration in a classical metapopulation. First, migrants are sampled from single source demes rather than from the entire metapopulation. Second, migration is defined by both a frequency (p_m) and a quantity (m) instead of by a single parameter. The response of T_0 to changes in the quantity of exchanged seed m under different rates of extinction is shown in Figure 1a. Clearly, the invariance result does not hold with respect to m, even in the absence of extinction.

We will explain this result mathematically by substituting the terms a_i and b_i from Table 2 into (3), and then setting p_m to unity and m_g to zero and defining $\varepsilon = (1/(1-e)^2) - 1$:

$$T_0 \approx \left(\frac{1-(1-m)^2-m^2}{1-(1-m)^2+\varepsilon}\right)(n-1)2N_e + 2N_e \qquad (10)$$

The term m^2 in the numerator in (10) is now divided by unity instead of by $n-1$ as was the case for the classical metapopulation model. This difference arises because under single source migration such as assumed in our model, two alleles that are sampled from seed migrants in the same deme always co-locate. The term m^2 can thus not be ignored when m is high. At $e=0$, increasing m leads to a decrease in T_0 from approximately $2N_e n$ when m is close to zero to $2N_e$ when m is one. When $e>0$, the term ε in the denominator lowers T_0. This is partially reverted as $1-(1-m)^2$ in the numerator becomes larger with larger m. The numerator in equation (10) equals zero at both $m=0$ and $m=1$ and has a maximum at $m=0.5$. Therefore, T_0 increases monotonically with m until reaching a maximum, which is dependent on e, and then decreases for higher m. The relation between T_0 and migration quantity thus deviates strongly from what would be expected under the classical metapopulation model.

If we no longer assume $p_m=1$ as above, we can see the effect of seed migration frequency on within-deme coalescence time as well (Figure 1b):

$$T_0 \approx \left(\frac{2p_m m - 2p_m m^2}{2p_m m - p_m^2 m^2 + \varepsilon}\right)2N_e(n-1) + 2N_e \qquad (11)$$

We note that when m is small so that we may ignore terms containing m^2, T_0 is invariant with respect to p_m when $e=0$ and increases with p_m when $e>0$ as predicted. At higher m, m^2 may no longer be ignored. As $2p_m m^2 \geqslant p_m^2 m^2$, migration will always lead to a value of T_0 that is below $2N_e n$. The term $-p_m^2$ in the denominator of (11) decreases with p_m more rapidly than the term $-2p_m$ in the numerator, causing T_0 to rise in response to migration frequency. Again, the invariance result breaks down, and we may conclude that single source migration causes dependence of within-deme coalescence time on both seed migration quantity and frequency.

The above results follow directly from the interpretation of T_0 as the ratio between co-location and relocation from different demes. After extinction, the co-location probability for alleles in the same deme is not affected because all colonists derive from the same deme. But recolonization increases the probability that alleles in different demes co-locate, which decreases the time alleles spend in different demes and reduces within-deme coalescence time. This effect is exacerbated by the lower number of extant source demes, which further increases the probability of co-location for alleles in different demes. A similar explanation underlies the effects of m and p_m. Under single source migration, the origin of immigrants within a deme is completely correlated and migration quantity, m, is therefore not equal to the rate at which lineages separate into different demes. Increasing m will decrease the proportion of co-locating alleles within a deme but at a decreasing rate until half of the alleles in a deme are migrants. Increasing m beyond this point will result in a higher proportion of co-locating alleles until all alleles co-locate at $m=1$. At the same time, the rate of relocation of alleles from different demes increases monotonically over the entire range of m, leading to a loss of invariance with respect to m. The response to p_m is different because seed sources for each deme are independent. As more demes receive migrants, there is a proportionally higher

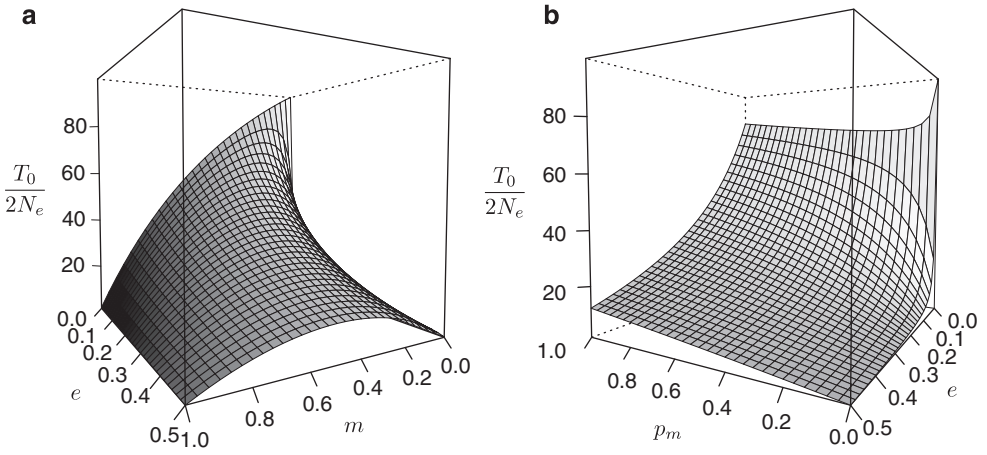

Figure 1 Within-deme diversity ($\frac{T_0}{2N_e}$) as a function of extinction rate e and (**a**) quantity of migrating seed m (with $p_m=1$) or (**b**) seed migration frequency p_m (with $m=0.25$). Assumes $N=1000$, $N_f=200$, $m_g=0.01$, $n=100$.

probability that two demes receive migrants from the same source and thus co-locate. The time that alleles spend in different demes thus remains approximately unchanged.

Expectations for F_{ST} in the classical model without extinction

Many empirical studies of subdivided populations use Wright's (1951) fixation index F_{ST} or similar measures as an estimator of the amount of gene flow between demes. Under the island model with infinite demes and low migration rates, the expectation for F_{ST} is given by $1/(4Nm+1)$. Although recognized as too simplistic (Whitlock and McCauley, 1999), this formula serves as the basis for two general predictions with respect to genetic structure. First, an increase in the number of migrants, Nm, always reduces genetic structure. Second, F_{ST} will be approximately independent of deme size provided that Nm remains constant.

As expected, both expectations hold in the classical metapopulation model without extinction. Substituting $T = T_0/n + T_1(1-(1/n))$ into the equation $F_{ST} = (T - T_0)/T$, we obtain:

$$F_{ST} \approx \frac{1}{4N_e m(1-\frac{1}{2}m)+1} \tag{12}$$

which is identical to the result obtained by (Wright, 1951), and to his reduced equation when m is small.

Seed migration and F_{ST}

Figure 2 shows the response of F_{ST} to the quantity and frequency of seed migration in our model. The response of F_{ST} to m differs strongly from what is predicted by the classical model. Instead of the usual hyperbolic relation, the response of F_{ST} to migration is parabolic with a minimum at $m=0.5$. We can derive that this result is because of the assumption of single source migration by analyzing the equilibrium solution for F_{ST} without extinction or pollen flow.

Ignoring extinction and pollen flow, and assuming $n \to \infty$, the relation between F_{ST} and migration in our model is given by:

$$F_{ST} \approx \frac{1}{4N_e p_m m(1-m)+1}. \tag{13}$$

Because migrating seed derives from a single source in each generation, $m=0.5$ represents the point where the proportion of alleles that come from different demes is maximal and inbreeding is lowest. Any further increase in m increases the proportion of co-locating alleles within demes and will therefore cause an increase in the genetic structure. In contrast, migration frequency determines the amount of migrant seed that comes from different demes. For small m, the effect of p_m can be explained under the classical model, as $p_m m(1-m) \approx \bar{m}$ and thus $F_{ST} \approx 1/(4N_e\bar{m}+1)$. A combination of high m and low p_m, however, may result in a higher value of F_{ST} than expected on the basis of the number of migrants N_m. Nonetheless, the negative relation between p_m and F_{ST} will hold regardless of the magnitude of m.

Deme size and F_{ST}

In our model, the quantity (m) and frequency (p_m) of seed flow may vary independently of one another. If we redefine m in equation (13) as N_m/N_e, where N_m is the effective number of seed migrants, we see that, in practice, the average effective number of seed migrants, $p_m N_m$, may be low whereas the effective number of seed migrants entering a receiving deme, N_m, is high. An important consequence of this model property is that F_{ST} becomes dependent on deme size, illustrated in equation 14:

$$F_{ST} \approx \frac{1}{4p_m N_m\left(1-\frac{N_m}{N_e}\right)+1}. \tag{14}$$

When N_m is relatively large with respect to N_e, greater deme size thus causes a reduction in F_{ST} similar to that caused by migration. Figure 3 illustrates this by showing F_{ST} as a function of p_m and N_f, given a fixed number of migrants N_{fm}. This effect is of potential importance in agricultural systems because quantities of migrant seed can be high.

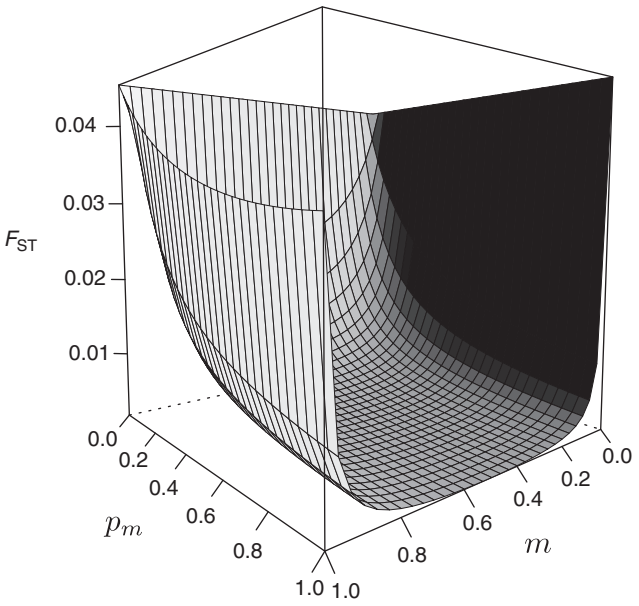

Figure 2 F_{ST} as a function of seed migration frequency (p_m) and quantity of migrating seed (m). Assumes $n=100$, $N=5000$, $N_f=200$, $e=0.2$, and $m_g=0.01$.

Extinction and F_{ST}

In metapopulations with extinction, Wright's (1951) classic formula for F_{ST} no longer provides an adequate description of the relation between seed flow and genetic structure. Under Slatkin's (1977) model II with propagule pool recolonization, extinction increases differentiation among demes (Wade and McCauley, 1988; Whitlock and McCauley, 1990; Pannell and Charlesworth, 1999). This result is mostly due to the strong drift which occurs during recolonization bottlenecks. The present model does not share Slatkin's assumption of a bottleneck after extinction. Consequently, our results on the effect of extinction on F_{ST} are rather different. Figure 4 shows the full model results for F_{ST} as a function of the extinction rate at different frequencies of seed migration and for different numbers of demes.

As seed migration becomes more frequent, F_{ST} is indeed increased by extinction until total diversity becomes so low that any further increase in extinction will lead to an effective decrease in F_{ST}. At low migration frequencies, however, F_{ST} is decreased by extinction. The reason for this can be seen in expression (15):

$$F_{ST} \approx \frac{1}{4N_e p_m m(1-m)(1-e) + \frac{2N_e}{n}\left(\frac{1}{1-e} - (1-e)(1-p_m m)^2\right) + 1} \quad (15)$$

The denominator consists of the sum of two terms that respond inversely to changes in e. When p_m is small the first term becomes negligible compared with the second and F_{ST} will decrease with increasing e. On the other hand, when n becomes very large, the second term tends to zero and F_{ST} will respond positively to extinction. Equations (16) and (17) present the cases for $p_m = 0$ and $n \to \infty$, respectively.

$$F_{ST} \approx \frac{1}{\frac{2N_e}{n}\left(\frac{1}{1-e} - (1-e)\right) + 1} \quad (16)$$

$$F_{ST} \approx \frac{1}{4N_e p_m m(1-m)(1-e) + 1} \quad (17)$$

This result shows that the effects of extinction on differentiation depend both on n and migration frequency; the conclusions drawn by (Wade and McCauley, 1988) thus hold for large n but become dependent on migration frequency when n is low.

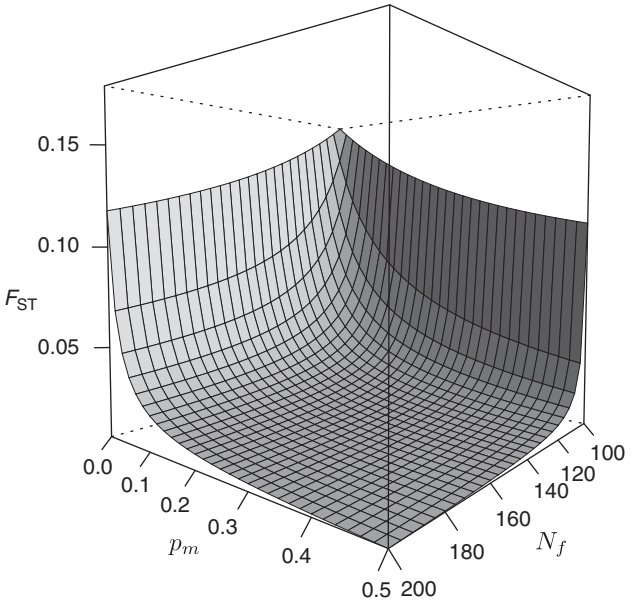

Figure 3 F_{ST} as a function of seed migration frequency p_m and number of planted ears N_f. Assumes $N = 1000$, $n = 100$, $N_{fm} = 100$, $e = 0.3$, and $m_g = 0$.

Seed management in the presence of pollen flow

In the results presented so far, pollen migration was assumed low to explore the effects of human-mediated gene flow on genetic diversity and structure. In reality, pollen migration may be extensive, and our ability to detect the effects of seed-related factors will depend on their interaction with pollen flow. It thus becomes relevant to know the sensitivity of genetic structure to seed management under different levels of pollen flow. Figure 5 shows results for our full model on the response of F_{ST} to extinction, migration frequency, migration quantity and number of ears planted at different levels of pollen flow ($m_g = 0.001$, 0.005, 0.01, 0.04). For the effect of deme size, pollen flow was defined by a fixed number of pollen migrants for each level. At the lowest level of pollen flow the response to the seed-related parameters is quite strong. At the highest level, however, the presence of pollen flow is dominant and overrides most effects of seed management on genetic structure.

Simulation study

Our model predicts expected coalescence times based on fractional coalescence probabilities. In doing so it treats probabilities, such as p_m and e, as fractions of possible allele pairs. This makes the model mathemati-

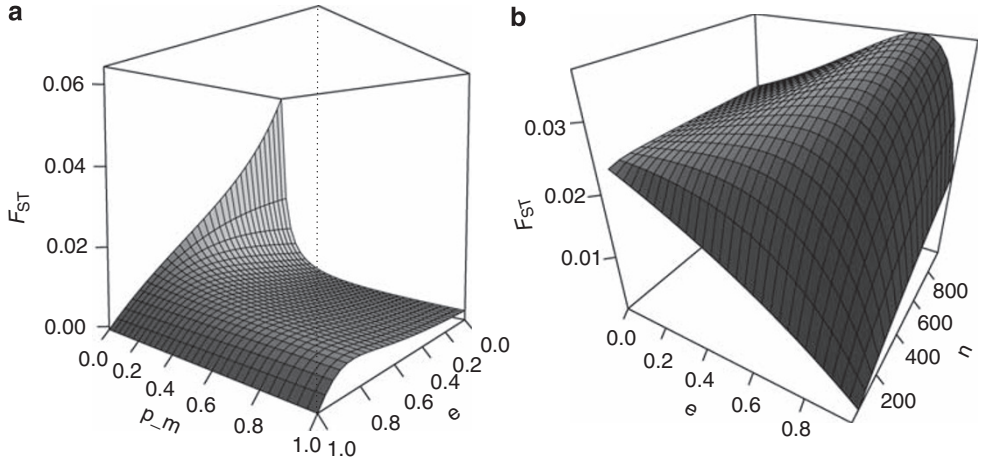

Figure 4 F_{ST} as a function of extinction rate e and (**a**) seed migration frequency p_m (with $m = 0.25$ and $n = 100$) or (**b**) number of demes n (with $m = 0.25$ and $p_m = 0.05$). Assumes $N = 5000$, $N_f = 200$, and $m_g = 0.01$.

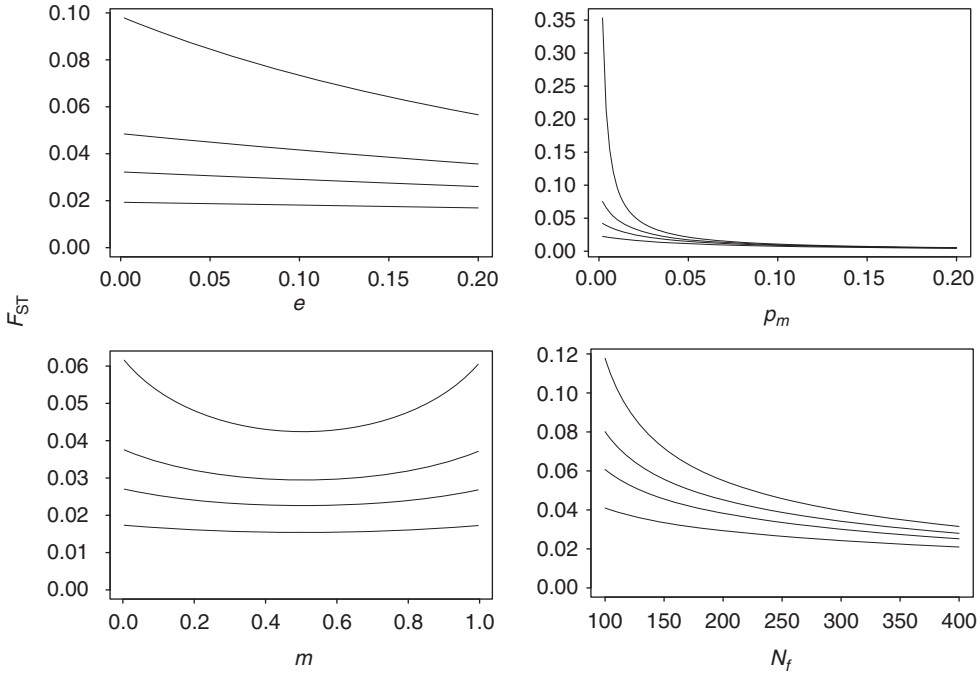

Figure 5 Clockwise from upper left: F_{ST} as a function of extinction rate e, migration frequency p_m, number of planted ears N_f, and quantity of migrating seed m at different levels of pollen flow ($m_g = 0.001$, 0.005, 0.01, 0.04). In each pane, pollen flow increases from higher to lower curves.

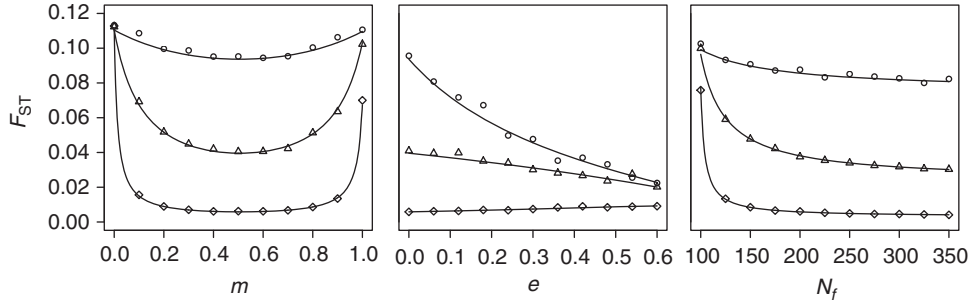

Figure 6 Comparison of simulated data (open symbols) to analytical predictions (solid lines). From left to right: F_{ST} as a function of quantity of migrating seed m, extinction rate e, and number of planted ears N_f (with constant number of pollen migrants) at different levels of migration frequency ($p_m = 0.2$, diamonds; 0.02, triangles; 0.002, circles). All graphs use $m_g = 0.005$, $e = 0$, $N_f = 200$, $N_{fm} = 100$ and $n = 100$ when these parameters are fixed.

cally tractable but raises the question whether the deterministic predictions hold when stochasticity is introduced. Furthermore, we define F_{ST} as a ratio of expected coalescence times rather than as the relative reduction of heterozygosity that forms the basis for most empirical F_{ST} estimates. Although the two measures are theoretically equivalent (Slatkin, 1991), it would be desirable to confirm that F_{ST} estimated from allele frequencies indeed concurs with our calescence-based predictions. We therefore performed a simulation study to evaluate the accuracy of our model. Using a modified version of the simulation algorithm of Piñeyro-Nelson *et al*. (2009), we performed stochastic simulations of a single biallelic locus under our metapopulation model. We calculated F_{ST} across the metapopulation for a range of values of m, e and N_f for each of the three values of p_m. Results were obtained by averaging over 100 simulations of 2000 generations each. These simulated data match our theoretical predictions almost perfectly (Figure 6), providing strong corroboration of our analytical results under the specified metapopulation model and assumptions.

Discussion

The determinants of neutral genetic diversity and structure are of substantial interest to evolutionary and conservation biology. But while molecular markers can be used to describe the observed distribution of genetic diversity within and among populations, the interpretation of such data relies on models of subdivided populations that adequately represent the population

genetics of the system under study. Following the introduction of Wright's infinite island model, a large body of theory has accumulated showing how deviations from basic assumptions can affect model behavior. Examples include the introduction of stepping stone migration (Kimura and Weiss, 1964; Slatkin, 1991), extinction-recolonization dynamics (Maruyama and Kimura, 1980), seed and pollen migration (Wang, 1997) and stochastic and kin-structured migration (Levin, 1988; Whitlock and McCauley, 1990). The results from these model refinements suggest that when a system is well defined, incorporating system-specific model features can provide a better understanding of population genetic processes. Although there has been a growing interest in understanding the genetic diversity of agricultural plant species under traditional management and the population dynamics of many crops is well documented, explicit models describing the population genetics of subdivided crop populations are currently unavailable. To our knowledge, the adapted metapopulation model presented here represents the first attempt to incorporate aspects unique to farmer-managed metapopulations into an explicit population genetic framework, and as such presents a significant step forward in our understanding of the effects of management practices on patterns of genetic diversity.

The main property that sets our present model apart from existing metapopulation models is that seed dynamics is mediated by conscious human intervention. Specifically, this translates into the assumptions of single source migration and the absence of a population bottleneck following extinction. Both assumptions are supported by empirical data (Rice *et al*., 1998; Badstue *et al*., 2007) and follow naturally from the basic need to obtain enough seed to ensure a successful harvest at minimal cost. A farmer's response to personal seed shortage is usually to look for a sufficient seed from a reliable source, often a friend or family member (Almekinders *et al*., 1994; Zeven *et al*., 1999; Badstue *et al*., 2007; Barnaud *et al*., 2008). Occasional departures from these assumptions may of course occur (Brocke *et al*., 2003), but providing they are infrequent, such deviations are unlikely to qualitatively change our results. Although, we use data on traditional maize agriculture to define the parameters of our model, studies on other crops have reported similar dynamics (Longley and Richards, 1993; Almekinders *et al*., 1994; Brocke *et al*., 2003; Barnaud *et al*., 2008), and we expect that our basic model predictions should apply to many traditionally managed species. A minor difference between our model and more general models is that migration is essentially kin structured because of the movement of ears rather than individual seeds. Other models have explored the effects of kin-structured migration in detail (Whitlock and McCauley, 1990). In our case, kin structured migration is of little theoretical interest because genetic sampling in the resident proportion of demes is similarly kin structured, so that migration remains a simple proportion of effective deme size.

By evaluating approximations for equilibrium coalescence times, our model provides insight into the mechanisms shaping neutral genetic diversity in crop metapopulations. Our predictions deviate significantly from those emanating from classical models of subdivided populations in several respects. First, the effects of single source migration on within-deme diversity and F_{ST} suggest that it is impossible to characterize gene flow by a single migration parameter, because the magnitude and frequency of seed migration have different and sometimes opposing consequences (Figures 1 and 2). The correlated origin of migrants causes a relative decrease in the time that alleles spend in different demes, leading to a loss of invariance of within-deme coalescent time with respect to migration as well as deviation from the monotonic relationship between F_{ST} and migration quantity. Second, the independence of migration frequency and quantity means that deme size may affect genetic structure (Figure 3), especially when migration is rare but involves large numbers of seeds. Dependence of differentiation on deme size has been reported in some theoretical studies on specific systems, for example by Ingvarsson (1997), who reported lower differentiation in small demes in a model of delayed population growth. Deme size is often ignored as a determinant of genetic structure however, based on the classical prediction that for low migration rates only the number of migrants affects F_{ST} (Wright, 1951). When faced with a shortage of planting material, however, farmers are likely to incorporate large quantities of migrant seed, suggesting that deme size is a factor that should be accounted for in order to understand the genetic structure in agroecosystems. Third, because a farmer can be expected to obtain sufficient seed in case of seed loss, the effects of extinction take a different form than in classical metapopulation models. The absence of a bottleneck after recolonization means that F_{ST} does not always increase with extinction as predicted by other models (Wade and McCauley, 1988; Pannell and Charlesworth, 1999), but instead depends on migration and the number of demes (Figure 4). Although finite deme number has been considered as a factor influencing F_{ST} Wade and McCauley, 1988; McCauley, 1991, previous work has not investigated the interaction of deme number and extinction.

The primary purpose of our model is to describe the genetic consequences of seed management. Our results suggest that researchers interested in linking empirical observations of genetic structure to data on farming practice should distinguish between replacement, migration quantity and migration frequency when collecting data, and that estimates of the number of planted ears and pollen migration are also required. Given our results showing the dampening effect of pollen migration, consideration of pollen flow should be of particular importance in any empirical study. Unfortunately, few estimates of pollen flow in traditional agroecosystems exist (for example Louette *et al*., 1997), and the dynamics of pollen migration are likely to be location specific (Messeguer *et al*., 2006).

We are not aware of any single study giving precise estimates of the above parameters; to show the use of field data to explain genetic differentiation, however, we compile seed management data on six maize farming communities in the Central Valley of Oaxaca from published articles and unpublished interview data obtained by the International Maize and Wheat Improvement Center (CIMMYT). We compare the predicted genetic structure from these data to empirical estimates of structure from the same region (Pressoir and

Berthaud, 2004). Records of average planting area (2.5 Ha) (Smale *et al*., 1999), seed quantity planted (16 kg Ha^{-1}) (Badstue *et al*., 2007) and grain weight of a single ear (70, 0.38 g per kernel) (Soleri and Smith, 2002) yield an estimated $N_f \approx 560$ and $N \approx 100\,000$. Futhermore, CIMMYT interview data suggest that $p_m \approx 0.02$ (M Bellon, personal communication), estimates of seed lot replacement provide $e \approx 0.1$ (Smale *et al*., 1999), and the mean quantity of exchanged seed by farmers (12.5 kg) leads to $m \approx 0.30$ (Badstue *et al*., 2007). In the absence of pollen flow data for the region we use the average estimate $m_g \approx 0.018$ from adjacent fields reported by Messeguer *et al*. (2006). From these data, we calculate an equilibrium $F_{ST} = 0.008$, quite close to the reported value of $F_{ST} = 0.011$ (Pressoir and Berthaud, 2004).

The above example shows that our model produces reasonable values of population structure based on farming system data. Although the parameter values used are relatively rough estimates, the value of the model is precisely that it provides a means of assessing the effects of parameter variation. In this particular case, relatively high pollen flow causes close agreement with approximations from classical models (island model: $F_{ST} \approx 0.007$, Slatkin's model II: $F_{ST} \approx 0.009$), but we now know that specifics of seed management may cause deviations from these models when pollen migration is limited. Our model allows, for the first time, clear identification of the specific data required to explain observed population structure in traditional agricultural systems. Knowing whether farmers have ever mixed seed, for example, is insufficient; predicting genetic structure requires quantitative estimates of amounts and frequencies of seed migration. Our model thus serves as a guide to the kinds of data that breeders or conservationists interested in genetic structure in crop systems must collect.

It is important to point out that the current model is framed in terms of fixed parameter values and equilibrium conditions. Exact prediction of genetic structure under specific field conditions, however, may be better served by explicit computer simulations (Piñeyro-Nelson *et al*., 2009). Nonetheless, the close correspondence between our model results and computer simulations shows that our main predictions are robust to stochasticity. Rather than serve as a detailed predictive method, however, we feel the value of the present work lies in providing a better understanding of the general behavior of genetic diversity in crop metapopulations. It is our hope that this work will be the first step toward a more quantitative approach to the study of crop metapopulations, paving the way for explicit—rather than metaphorical—analysis of the role that farmers have in shaping genetic diversity.

Finally, our study provides an example of the benefits of incorporating information from well-defined systems to create more refined population genetic models. Although we have focused on the genetic structure within traditionally managed crops, we suggest that similar analytical evaluation of well-defined natural systems may also lead to interesting and potentially novel results.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank RA Cartwright, BS Gaut, M Slatkin, and three anonymous reviewers for helpful comments and revisions to the article. JR-I acknowledges funding from National Science Foundation Grant DBI0321467. JVH wishes to thank the Rockefeller foundation for financial support and the International Maize and Wheat Improvement Center for use of data and facilities.

References

Almekinders C, Louwaars N, Debruijn G (1994). Local seed systems and their importance for an improved seed supply in developing countries. *Euphytica* **78**: 207–216.

Alvarez N, Garine E, Khasah C, Dounias E, Hossaert-McKey M, McKey D (2005). Farmers' practices, metapopulation dynamics, and conservation of agricultural biodiversity on-farm: a case study of sorghum among the Duupa in sub-sahelian Cameroon. *Biol Conserv* **121**: 533–543.

Badstue L, Bellon M, Berthaud J, Ramirez A, Flores D, Juarez X (2007). The dynamics of farmers' maize seed supply practices in the Central Valleys of Oaxaca, Mexico. *World Dev* **35**: 1579–1593.

Barnaud A, Deu M, Garine E, McKey D, Joly HI (2007). Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. *Theor Appl Genet* **114**: 237–248.

Barnaud A, Joly HI, McKey D, Deu M, Khasah C, Monne S *et al*. (2008). Management of sorghum (Sorghum bicolor) genetic resources among Duupa farmers (northern Cameroon). *Cah Agric* **17**: 178–182.

Brocke K, Christinck A, Weltzien R, Presterl T, Geiger H (2003). Farmers' seed systems and management practices determine pearl millet genetic diversity patterns in semiarid regions of India. *Crop Sci* **43**: 1680–1689.

Brush S (1999). Genetic erosion of crop populations in centers of diversity: A revision. In: Serwinski J, Faberova I (eds). *Genetic erosion of crop populations in centers of diversity: A revision*. Vyzkumny Ustav Rostlinne Vyroby: Prague.

Brush S (2000). *Genes in the Field: On-Farm Conservation of Crop Diversity*. IPGRI: Rome.

Brush S (2004). *Farmers' Bounty: Locating Crop Diversity in the Contemporary World*. Yale University Press: New Haven.

Crossa J, Vencovsky R (1994). Implications of the variance effective population size on the genetic conservation of monoecious species. *Theor Appl Genet* **89**: 936–942.

Dyer G, Taylor J (2008). A crop population perspective on maize seed systems on Mexico. *Proc Natl Acad Sci USA* **105**: 470–475.

Esquinas-Alcazar J (2005). Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nat Rev Genet* **6**: 946–953.

Fowler C, Hodgkin T (2004). Plant genetic resources for food and agriculture: Assessing global availability. *Annu Rev Env Resour* **29**: 143–179.

Harlan J, Wet J, Price E (1973). Comparative evolution of cereals. *Evolution* **27**: 311–325.

Harrison S, Taylor A (1997). Empirical evidence for metapopulation dynamics. In: Hanski I, Gilpin M (eds). *Metapopulation Biology: Ecology, Genetics, and Evolution*. Academic Press: San Diego, pp 27–42.

Heisey P, Brennan J (1991). An analytical model of farmers' demand for replacement seed. *Am J Agr Econ* **73**: 1044–1052.

Hudson RR (1990). Gene genealogies and the coalescent process. *Oxford Surv Evol Biol* **7**: 1–44.

Ingvarsson P (1997). The effect of delayed population growth on the genetic differentiation of local populations subject to frequent extinctions and recolonizations. *Evolution* **51**: 29–35.

Jarvis DI, Brown AHD, Cuong PH, Collado-Panduro L, Latournerie-Moreno L, Gyawali S *et al*. (2008). A global perspective of the richness and evenness of traditional

crop-variety diversity maintained by farming communities (vol 105, pg 5326, 2008). *Proc Natl Acad Sci USA* **105**: 8160.
Kimura M, Crow J (1963). The measurement of effective population number. *Evolution* **17**: 279–288.
Kimura M, Weiss G (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**: 561-&.
Lande R (1992). Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution* **46**: 381–389.
Latter B (1973). The island model of population differentiation: a general solution. *Genetics* **73**: 147–157.
Levin D (1988). Consequences of stochastic elements in plant migration. *Am Nat* **132**: 643–651.
Li Y, Wu S (1996). Traditional maintenance and multiplication of foxtail millet (Setaria italica (L) P. Beanv) landraces in China. *Euphytica* **87**: 33–38.
Longley C, Richards R. (1993). Selection strategies of rice farmers in Sierra Leone. In: de Boef W, Amanor A, Wellard K, Bebbington A (eds). *Cultivating knowledge: genetic diversity, farmer experimentation and crop research*. Intermediate Technology Publication: London, pp 51–57.
Louette D (2000). Traditional management of seed and genetic diversity: what is a landrace?. In: Brush S (ed). *Genes in the Field: On-farm Conservation of Crop Diversity*. IPGRI: Rome.
Louette D, Charrier A, Berthaud J (1997). *In situ* conservation of maize in Mexico: Genetic diversity and maize seed management in a traditional community. *Econ Bot* **51**: 20–38.
Louette D, Smale M (2000). Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. *Euphytica* **113**: 25–41.
Maruyama T, Kimura M (1980). Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proc Natl Acad Sci USA* **77**: 6710–6714.
McCauley D (1991). Genetic consequences of local population extinction and recolonization. *Trends Ecol Evol* **6**: 5–8.
Messeguer J, Penas G, Ballester J, Bas M, Serra J, Salvia J *et al*. (2006). Pollen-mediated gene ow in maize in real situations of coexistence. *Plant Biotechnol J* **4**: 633–645.
Nagylaki T (1998). The expected number of heterozygous sites in a subdivided population. *Genetics* **149**: 1599–1604.
Nagylaki T (2000). Geographical invariance and the strong-migration limit in subdivided populations. *J Math Biol* **41**: 123–142.
Pannell J, Charlesworth B (1999). Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* **53**: 664–676.
Perales H, Benz B, Brush S (2005). Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. *Proc Natl Acad Sci USA* **102**: 949–954.
Perales H, Brush S, Qualset C (2003). Dynamic management of maize landraces in Central Mexico. *Econ Bot* **57**: 21–34.
Piñeyro-Nelson A, Van Heerwaarden J, Perales H, Serratos-Hernandez J, Rangel A, Hufford M *et al*. (2009). Transgenes in Mexican maize: molecular evidence and methodological considerations for GMO detection in landrace populations. *Mol Ecol* **18**: 750–761.
Pressoir G, Berthaud J (2004). Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* **92**: 95–101.
Rice E, Smale M, Blanco J (1998). Farmers' use of improved seed selection practices in Mexican maize: Evidence and issues from the Sierra de Santa Marta. *World Dev* **26**: 1625–1640.
Slatkin M (1977). Gene ow and genetic drift in a species subject to frequent local extinctions. *Theor Popul Biol* **12**: 253–262.
Slatkin M (1991). Inbreeding coefficients and coalescence times. *Genet Res* **58**: 167–175.
Smale M, Aguirre A, Bellon M (1999). Farmer management of maize diversity in the Central Valleys of Oaxaca. CIMMYT/INIFAP 1998 baseline. CIMMYT Economics Working Paper, Mexico 99?09, DF: CIMMYT.
Soleri D, Smith S (2002). Rapid estimation of broad sense heritability of farmer-managed maize populations in the Central Valleys of Oaxaca, Mexico, and implications for improvement. *Euphytica* **128**: 105–119.
Wade M, McCauley D (1988). Extinction and recolonization—their effects on the genetic differentiation of local-populations. *Evolution* **42**: 995–1005.
Wakeley J, Aliacar N (2001). Gene genealogies in a metapopulation. *Genetics* **159**: 893–905.
Wang J (1997). Effective size and F-statistics of subdivided populations.1. Monoecious species with partial selfing. *Genetics* **146**: 1453–1463.
Whitlock M, Barton N (1997). The effective size of a subdivided population. *Genetics* **146**: 427–441.
Whitlock M, McCauley D (1990). Some population genetic consequences of colony formation and extinction—genetic correlations within founding groups. *Evolution* **44**: 1717–1724.
Whitlock M, McCauley D (1999). Indirect measures of gene ow and migration: Fst not equal 1/(4 Nm+1). *Heredity* **82**: 117–125.
Wright S (1951). The genetical structure of populations. *Ann Eugen* **15**: 323–354.
Zeven A, Waninge J, van Hintum T, Singh S (1999). Phenotypic variation in a core collection of common bean (Phaseolus vulgaris L. in the Netherlands). *Euphytica* **109**: 93–106.

Appendix

Different combinations of non-migrant maternal, migrant maternal, and paternal gene copies have different coalescent and co-location probabilities. We calculate these probabilities and the compound terms a_i and b_i for allele pairs from the same deme and from different demes, then use these probabilities to calculate the general equilibrium solutions for T_0, T_1 and F_{ST}.

Alleles sampled from the same deme

For two non-migrant maternal alleles sampled from an extant deme that has undergone seed migration there are $N_f - N_{fm}$ possible maternal plants so P_i becomes $P_1 = 1/(2(N_f - N_{fm}))$. The co-locating probability in this case is one. Hence,

$$a_1 = \frac{1}{4} p_m (1-e)(1-m)^2 \tag{18}$$

Two maternal alleles that are both sampled from migrant seed can originate from N_{fm} maternal plants, yielding $P_2 = 1/2N_{fm}$. As there is only a single source of migrant seed for each deme in each generation the co-location probability again equals one, giving

$$a_2 = \frac{1}{4}(1-e) p_m m^2 \tag{19}$$

Two maternal alleles sampled from a deme that has not received seed migrants may have originated from any of N_f ears, so that $P_3 = 1/2N_f$ with

$$a_3 = \frac{1}{4}(1-(1-e)p_m) \tag{20}$$

When a pair of sampled alleles includes a paternal allele, the coalescent probability is determined by the total number of plants in a deme and becomes $P_4 = 1/2N$. Both seed and pollen migration now have to be taken into account. If the sample does not contain a pollen migrant, the two alleles will co-locate unless one of them is sampled from resident seed and the other from migrant seed; this probability can be written as $\frac{1}{4}(2(1-m_g) + (1-m_g)^2)(1-(1-e)p_m 2m(1-m))$. After

rearranging terms, we can write:

$$\left(\frac{3}{4}-m_g\left(1-\frac{1}{4}m_g\right)\right)(1-2mp_m(1-e)(1-m)) \quad (21)$$

The term $(1-m_g)^2$ represents combinations of two paternal alleles that are both pollen residents, and $2(1-m_g)$ represents samples that contain a paternal and a maternal allele of which the paternal allele is a pollen resident.

Allele pairs containing a pollen migrant have a co-location probability of $1/(n-1)$ when both alleles are pollen migrants or if one of the alleles is sampled from migrant seed, giving

$$\frac{1}{4}\left(m_g^2+(2m_g(1-m_g)+2m_g)2mp_m(1-e)(1-m)\right) \quad (22)$$

Combining (21) and (22) we get

$$\begin{aligned}a_4=&\left(\frac{3}{4}-m_g\left(1-\frac{m_g}{4}\right)\right)(1-2p_mm(1-e)(1-m))\\&+\frac{m_g(2-m_g)(1-e)p_mm(1-m)+\frac{1}{4}m_g^2}{(n-1)}\end{aligned} \quad (23)$$

Alleles sampled from two different demes

When the sample from two demes contains at least one paternal allele the coalescence probability is again given by P_4. When such a sample contains a pollen migrant the co-location probability equals $1/(n-1)$. The proportion of allele pairs that contain one or two pollen migrants is:

$$\frac{1}{4}\left(1-(1-m_g)^2\right)+\frac{m_g}{2}=m_g\left(1-\frac{1}{4}m_g\right) \quad (24)$$

The remaining fraction are allele combinations that do not contain a pollen migrant, whose frequency is given by

$$\frac{1}{4}\left(2(1-m_g)+(1-m_g)^2\right)=\frac{3}{4}-m_g\left(1-\frac{1}{4}m_g\right) \quad (25)$$

Samples from this fraction co-locate with probability $1/(n(1-e)-1)$ unless both alleles are sampled from resident seed. Combining (24) and (25) we then have:

$$\begin{aligned}b_4=&\frac{\left(\frac{3}{4}-m_g\left(1-\frac{1}{4}m_g\right)\right)\left(1-(1-e)^2(1-p_mm)^2\right)}{n(1-e)-1}\\&+\frac{m_g\left(1-\frac{1}{4}m_g\right)}{n-1}\end{aligned} \quad (26)$$

For two maternal alleles sampled from different demes, the probability that both originated from the same ear is equal to zero given that ears are assumed to be the units of seed migration. Thus $P_5=0$. Co-location probability is $1/(n(1-e)-1)$ unless both alleles are sampled from resident seed, and b_5 can thus be written:

$$b_5=\frac{1}{4}\left(\frac{\left(1-(1-e)^2(1-p_mm)^2\right)}{n(1-e)-1}\right) \quad (27)$$

General equilibrium solutions

The general equilibrium solutions for T_0 and T_1 are:

$$T_0=\frac{1-\sum_i a_i}{\bar{P}\sum_i b_i}+\frac{1}{\bar{P}} \quad (28)$$

and

$$T_1=T_0(1-\bar{P}_i)+\frac{1}{\sum_i b_i}. \quad (29)$$

For our model these become:

$$T_0=\frac{\frac{1-\sum_i a_i}{b_4+b_5}+1}{P_4Q(1-\sum_i a_i)+\sum_i a_iP_i} \quad (30)$$

and

$$T_1=T_0(1-QP_4)+\frac{1}{b_4+b_5} \quad (31)$$

where $Q=b_5/(b_4+b_5)$. Furthermore, we assume $1/2N_e \approx P_4Q(1-\sum_i a_i)+\sum_i a_iP_i$.

Derivations of F_{ST}

We begin by defining $\alpha=\sum_i a_i$, $\beta=\sum_i b_i$, $T=T_0/n+T_1(1-(1/n))$, and assume $T_1\approx T_0+(1/\beta)$ such that

$$F_{ST}=\frac{T_0\frac{1}{n}+(T_0+\frac{1}{\beta})(1-\frac{1}{n})-T_0}{\frac{T_0}{n}+(T_0+\frac{1}{\beta})(1-\frac{1}{n})}. \quad (32)$$

Equation (32) can be simplified to

$$F_{ST}=\frac{\frac{1}{\beta}(1-\frac{1}{n})}{\frac{1}{\beta}(1-\frac{1}{n})+T_0} \quad (33)$$

and when n is large

$$F_{ST}=\frac{1}{T_0\beta+1}. \quad (34)$$

Substituting in the value of T_0, we then get:

$$F_{ST}\approx\frac{1}{((1-\alpha)\frac{1}{\beta}2N+2N)\beta+1}. \quad (35)$$

If we assume $m_g=0$, then

$$\alpha=1-(1-e)p_m2m(1-m) \quad (36)$$

and

$$\begin{aligned}\beta&=\frac{1-(1-e)^2(1-p_mm)^2}{n(1-e)-1}\\&\approx\frac{1-(1-e)^2(1-p_mm)^2}{n(1-e)}.\end{aligned} \quad (37)$$

With some rearrangement, F_{ST} can then be written as

$$F_{ST}\approx\frac{1}{4N_ep_mm(1-m)(1-e)+1+\frac{2N_e}{n}(\frac{1}{1-e}-(1-e)(1-p_mm)^2)}. \quad (38)$$

When $n\to\infty$ so that $2N_e\beta\to0$, we can write the approximation

$$F_{ST}\approx\frac{1}{(1-\alpha)2N_e+1}. \quad (39)$$

For the island model with infinite n, $1-\alpha=1-(1-m)^2=2m+m^2$, and ignoring m^2 we get $F_{ST}\approx1/(4N_em+1)$. For our model, if we set both m_g and e to zero, $1-\alpha\approx1-(1-p_m2m(1-m))=2p_mm(1-m)$ and then

$$F_{ST}\approx\frac{1}{4N_ep_mm(1-m)+1}. \quad (40)$$