Population genetics of traditionally managed maize

Farming practice as a determinant of genetic structure and identity of maize landraces in Mexico

Joost van Heerwaarden

Proefschrift

Ter verkrijging van de graad van doctor

Op gezag van de rector magnificus

van Wageningen Universiteit,

Prof. Dr. M.J. Kropff,

in het openbaar te verdedigen

op dinsdag 11 december 2007

des morgens te elf uur in de Aula

Population genetics of traditionally managed maize

Farming practice as a determinant of genetic structure and identity of maize landraces in Mexico

Joost van Heerwaarden

PhD thesis Wageningen University, the Netherlands

ISBN: 978-90-8504-862-6

Propositions:

1. When migration occurs strictly between individual demes, increasing gene flow may augment genetic differentiation. (This thesis)

2. Replacement of traditional landraces with improved varieties need not constitute an overall loss of genetic diversity. (This thesis)

3. The biological reasons behind possible suboptimal adaptation of landraces that have been evolving under domestication for thousands of generations should be of great interest to both breeders and evolutionists.

4. It is virtually impossible to do a better job than an organism is doing in its own environment (Lewontin 1967).

5. A quantitative model should be employed as an extension of our explicit hypotheses and common sense, not as a magician's hat to replace knowledge or data.

6. Rigorous cross-disciplinary research may be the best recipe for avoiding excessive scientific inbreeding.

7. After years of wondering about the causes of poverty one starts to realize that it is wealth that requires an explanation.

Propositions belonging to the thesis: "Population genetics of traditionally managed maize: Farming practice as a determinant of genetic structure and identity of maize landraces in Mexico" by Joost van Heerwaarden, Wageningen, December 11, 2007.

Through and through the world is infested with quantity. To talk sense is to talk quantities. It is no use saying the nation is large- how large? It is no use saying that radium is scarce- how scarce? You can not evade quantity. You may fly to poetry and music and quantity and number will face you in your rhythms and your octaves. (Alfred North Whitehead)

Contents:

About the cover:

Detail taken from a replica of the stair mural found within the Red Temple at the Cacaxtla archeological site, located in the state of Tlaxcala, Mexico. The mural was painted between 800 and 900 A.D. It serves as a beautiful reminder of the importance of maize as the nutritional foundation of Mesoamerican civilization.

Chapter I

General introduction

The importance of maize genetic resources

The importance of agricultural genetic diversity can hardly be overstated. Ever since crop domestication began some 10,000 years ago, our subsistence has mostly relied on the cultivation, adaptation and improvement of a small number of plant species. Continuing breeding success in these species depends on access to a sufficient amount of natural variation on which selection can operate. For this reason, there is increasing concern that agricultural modernization will lead to diversity loss from centers of crop origin (Harlan 1975; Brush 1999). Conservation of genetic resources has hence become an important topic to both research and policy.

Mexico is an important centre of origin and diversification for several cultivated plant species (Vavilov 1951). The main crop to originate in this region is maize (*Zea mays* spp. *mays*); the world's third most important food plant after rice and wheat (faostat.fao.org). Maize was domesticated from an annual species of teosinte (*Zea mays* spp. *parviglumis*) some 9,000 years B.P. (Beadle 1939; Matsuoka et al. 2002) (Figure 1), and was spread throughout the Americas soon thereafter.

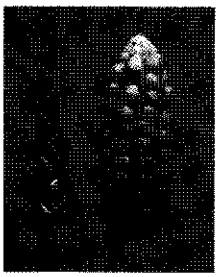


Figure 1. Ears of Teosinte (left) and maize (right).

Mexico still harbors a large amount of maize diversity, as traditional landraces continue to be grown by about 2 million smallholder farmers throughout the country (Aquino et al. 2001). Efforts to describe and collect these varieties started in the late nineteen forties. Groundbreaking work on maize racial classification was done by Wellhausen and collaborators (Wellhausen et al. 1952), resulting in some thirty described races. Although racial types have been of great value to the study and conservation of maize genetic resources, Wellhausen himself admitted that much heterogeneity exists within each race (Wellhausen et al. 1952). Later studies have confirmed this (Sanchez et al. 2000) and the last decades have seen a growing interest into patterns of genetic diversity beyond and within the traditional races (Louette et al. 1997; Sanou et al. 1997; Pressoir et al. 2004; Perales et al. 2005).

Farmers as determinants of maize genetic diversity

Describing the patterns of diversity in maize is of great interest to conservation and use of genetic resources. The underlying mechanisms that have shaped these patterns may be considered equally important however. Understanding the determinants of genetic structure is key to judging the importance of observed differences as well as to our ability to predict the evolutionary fate of existing and novel genetic diversity in the field. The latter has become particularly relevant now that the introduction of genetically modified crop varieties has created the potential for novel genes to become incorporated into the genetic background of maize landraces (Quist and Chapella 2001).

It has become evident that farmers are important actors in crop evolution (Bellon 1996). Contrary to commercial maize producers in most of the industrialized world, Mexican smallholders generally select seed from their own harvest or from that of other farmers (Louette et al. 1997). This millennial process of seed recycling and exchange is likely to affect patterns of genetic diversity in maize. A number of pioneering studies have described the seed management practices of Mexican smallholder farmers (Louette et al. 1997; Rice et al. 1998; Perales et al. 2003). Many of these practices have probably remained unchanged since the beginning of agriculture and may be considered representative of the genetic processes that maize has undergone for thousands of years. We present a synthesis of those processes that we consider essential to a proper understanding of maize population genetics.

Variety choice

Farmers usually have access to a set of local varieties of different agronomical characteristics such as color, growing cycle, or grain type. The type or types that a farmer grows may hence depend on individual choices related to consumption or production characteristics (Bellon 1996; Smale et al. 2001). Recent work has established that the set of varieties in any particular region is not static but may change in the face of changing preferences or availability of new types of seed (Louette 1999; Bellon et al. 2001). Sometimes the choice of a certain variety may reflect local climatic or soil conditions (Bellon et al. 1993).

Seed selection

A farmer generally manages a maize variety as a single entity that has been called a seedlot (Louette et al. 1997). A seedlot comprises the seed of a certain type that is selected and planted. More than one seedlot may be planted in the same field (Figure 2). The maize harvest is usually stored in the form of unshelled ears, either with or without husk leaves (Figure 3). Ears of different seedlots are usually, but not always, stored separately. Seed is selected from stored ears, and farmers generally choose big, clean ears that have an appearance that corresponds to its variety (Louette et al. 2000). Most seed of every ear is used for planting, although some grain from the tip and the base of the ear is usually discarded (Perales et al. 2003). Since maize ears contain a large number of seeds, only about ten ears are needed for a kilogram of seed. A hectare of land is planted with 12 to 20 kg. of seed. It is important to stress that seed from a single ear is derived from a single maternal plant so that the selection of a small number of ears represents a limited genetic sample. The significance of this will be discussed in chapter II and V.

Figure 2. Don Evarista Matias Galvez, holding a tradtitional landrace in his left hand and a commercial hybrid in his right.

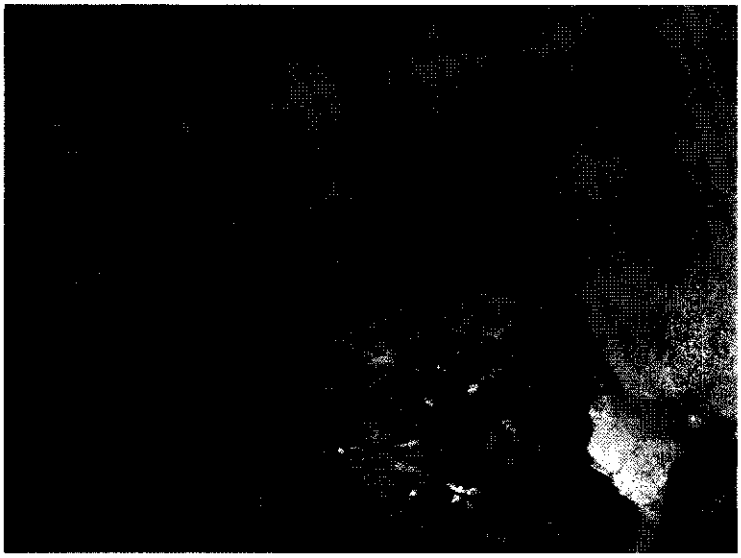


Figure 3. Example of traditional storage of unshelled maize ears

Seed replacement

A farmer will generally recycle a seedlot for several years. Eventually seedlots will be lost however. Storage losses or the need to consume or sell the last remaining seed may cause a farmer to loose a particular seedlot. Fortunately, farmers can usually obtain new seed of the same variety from another farmer from within or outside his community (Louette et al. 1997; Rice et al. 1998; Perales et al. 2003). This way of obtaining seed is common and is referred to as the informal seed system (Almekinders et al. 1994). The most common suppliers are neighbors or family but seed may also be obtained from more distant sources (Perales et al. 2003). Recent data (M. Bellon, unpublished) show that seed replacement is usually a matter of necessity and that seedlots are only rarely abandoned by voluntary choice.

Seed migration

Apart from complete seed replacement as described above, a seedlot may also undergo partial replacement when a farmer includes seed from a different seedlot in his planting material (Aguirre Gomez 1997). We will refer to this process as seed migration. The incentives for seed migration may be the same as for replacing a seedlot however. Although some farmers introduce new seed out of curiosity, the most prominent reason is seed shortage (M. Bellon, unpublished data). Most movement of germplasm between farmers is hence motivated by the need of having sufficient seed rather than being the result of conscious decisions.

Pollen flow

Seed movement is not the only process leading to genetic exchange between seedlots. Maize is an open-pollinating crop and pollen flow occurs readily among different maize populations. Although pollen flow is not strictly under a farmer's control, planting location and time will affect the amount of pollen that can migrate between seedlots. Differences in land use and individual planting decisions may hence affect the genetics of maize seedlots by inhibiting or promoting pollen-mediated gene flow.

The processes described above are all expected to affect genetic differences between seedlots. The way in which these different factors shape genetic structure in maize landraces will be the main topic of the present work. Existing studies on genetic differentiation in maize have often assigned a prominent role to farmer seed management in the explanation of observed patterns of diversity (Sanou et al. 1997; Pressoir et al. 2004; Perales et al. 2005). Unfortunately, not many attempts have been made to link observations on farming practice to genetic data. This has partly been due to the lack of models that can adequately describe the population genetics of maize landraces as a function of seed management. One of the goals of this study is to develop the necessary theory to allow the translation of information on seed management into expected patterns of genetic diversity.

A word on diversity

This study deals with genetic diversity in traditional maize farming systems and the processes that determine this diversity. Genetic diversity is a rather complex quantity that may be measured at different scales and on different characteristics. It is hence necessary to categorize diversity in a way that allows sensible analysis.

First, it is important to point out that genetic diversity is a relative concept. It is only meaningful to speak about diversity if a proper frame of reference is defined. In the present work we will mainly look at diversity above the seedlot level. That is to say that we are interested in genetic differences between different seedlots or groups of seedlots rather than in any measure of variability within populations. This seems justified, as it is the seedlot that is subject to human management.

Second, we will distinguish between molecular differentiation and phenotypic differentiation. Molecular differentiation is measured by genetic markers that are usually presumed to be selectively neutral. Both allozyme and microsatellite (SSR) markers are often used for this purpose. Molecular differentiation is affected by random drift and gene flow. When there is little flow, populations of limited size will diverge genetically through drift and become different, genetic diversity will thus decrease within populations and increase between populations. As gene flow increases, diversity is distributed more evenly and populations will be more similar while containing more diversity within each population.

The classical measure of molecular genetic differentiation is Wright's F_{st} (Wright 1951), which essentially reflects the amount of variation that is maintained between populations relative to overall genetic diversity within the set of studied populations. Alternatively, it may be defined as the correlation of allele frequencies within populations (Weir et al. 1984). F_{st} theoretically ranges from 0 in the absence of any genetic structure to 1 when all diversity is maintained between populations. We will use this measure and related quantities throughout this work to quantify neutral genetic differences between populations.

As we are interested in the determinants of genetic differentiation, we need ways of linking genetic processes to expected values of F_{st}. This may be achieved by using computer models that simulate the processes of drift and gene flow in a set of populations. Although we will follow this approach to generate predictions based on complex information, it is not the most insightful way of describing genetic processes. For analytical purposes therefore, we have chosen to use coalescent theory (Kingman 1982) as a means of obtaining expectations for genetic structure.

The coalescent describes the genealogical process of a sample of alleles backwards in time. Under this theoretical framework it is possible to predict the expected time to the nearest common ancestor, or coalescence, for alleles sampled at random from a population. Since alleles are identical at the moment of coalescence, the time elapsed since they separated provides an estimate of how much mutational change has occurred. Under assumptions of zero recombination this provides a direct relation between coalescence time and the amount of genetic differences that separate alleles in a sample (Hudson 1990). Alleles can only coalesce if they are present in the same population. Consequently, gene flow between populations is expected to affect the mean time to coalescence for alleles sampled from different populations. In case of zero gene flow, alleles from different populations will never coalesce. On the contrary, very high levels of gene flow will make alleles sampled from different populations coalesce equally fast as those sampled from the same population. The relationship between coalescence time and genetic diversity may thus be used to translate coalescence times to expectations of levels of between- and within population diversity and hence F_{st}.

Phenotypic differentiation may be defined as genetic differentiation measured on traits that, apart from being subject to the aforementioned random genetic processes, are potentially under selection. Differences between populations as measured by these traits can be strongly increased when selection acts differentially or may be reduced when selection is homogenous among populations.

A widely used measure of the distribution of genetic differentiation for phenotypic traits is Q_{st}. This measure is defined analogously to F_{st} and is expected to have the same value as the latter measure when the considered trait is not under selection (Spitze 1993). Under divergent selection the expectation is that $Q_{st} > F_{st}$ whereas $Q_{st} < F_{st}$ in the case of homogenizing selection (Merila et al. 2001).

The comparison between Q_{st} as estimated for agronomical traits and F_{st} based on neutral molecular markers may thus yield information on the role of selection versus neutral genetic processes in the structuring of genetic diversity. Genetic structure caused by selection is obviously more relevant from an agronomic point of view and will bear relation to diversity of useful traits that is available to maize producers. Although some attempts at quantifying selection on maize traits have been made (Louette et al. 2000), it is hard to reliably estimate all selective forces that may act on agronomical traits. Estimating the quantitative relation between observed values of Q_{st} and specific processes is therefore difficult to achieve and will not be attempted in this study. Where possible more qualitative explanations will be sought for observed patterns of phenotypic differentiation.

Objectives and outline of the thesis:

The present work aims to improve our understanding of the main determinants of genetic structure in maize landraces. In contrast to previous studies, we specifically focus on explaining the relation between observed patterns of diversity and farmer practice. Where possible, we study this relationship quantitatively. To this effect, we develop and apply new models that enable the combined effects of all relevant processes on neutral genetic structure to be evaluated. We will present results on patterns of molecular and phenotypic structure in both subsistence and commercialized smallholder agriculture in order to infer the extent of human impact on diversity in both traditional and modernized seed systems. Finally, we will use the insights and information obtained in these studies to make predictions about the inadvertent spread of transgenes in traditionally managed maize populations.

In chapter II we will define a mathematical model describing the effects of farmer seed management on neutral genetic diversity and structure. We use a straightforward description of average coalescence times in a metapopulation to give analytical expectations of genetic diversity and structure. We will show that a specific model is needed to adequately describe maize seed systems in terms of genetic processes.

Chapter III describes a study on determinants of molecular and quantitative genetic structure in a collection of seedlots, sampled from both highland and lowland environments. We provide a description of patterns of genetic differentiation at different hierarchical levels. A newly developed computer model is used to evaluate if observed genetic structure in the two environments can be explained by seed management and pollen flow. Comparisons of Q_{st} against the baseline provided by F_{st} will be employed to find evidence of divergent selection at the seedlot and village levels.

Chapter IV deals with the topic of genetic erosion due to the replacement of traditional landraces by modern varieties in the state of Chiapas, Mexico. We will outline and execute a methodology for estimating changes in genetic diversity in a system where modern and traditional maize coexist. Again we will characterize seedlots for both molecular markers and agronomic traits. We will compare levels of differentiation within and between different types of modern varieties, traditional landraces and local varieties that were derived from improved germplasm. We will discuss the consequences of the increased presence of modern varieties for local levels of genetic diversity.

Chapter V presents a theoretical study aimed at applying our current state of knowledge on maize reproductive biology and population genetics to the issue of transgene detection in field samples. We address potential factors that may lead to overestimation when using current methods for calculating detection probabilities. We evaluate the separate effects of unequal parental contribution, pollination restriction and transgene frequency distribution. We employ population genetic simulations to predict the type of frequency distributions that can be expected under realistic scenarios of transgene introduction.

By combining models, farmer interviews, molecular- and phenotypic data, we provide an integrated assessment of the role that smallholder farmers play in the population genetic processes that define the structure of maize in Mexico.

Chapter II

Neutral genetic diversity in a metapopulation of farmer-managed germplasm

Abstract

The population genetics of traditionally managed crop landraces is of interest to in-situ conservation of genetic resources. Although it is widely recognized that seed management is an important determinant of genetic diversity and structure in crops, no models exist that can adequately describe the effects of management on genetic structure among crop populations. We present a metapopulation model that accounts for several features that are unique to managed crop populations. We use maize as an example to develop a coalescence-based model of a metapopulation undergoing pollen and seed flow as well as extinction in the form of seed replacement. Within- and between deme diversity are described by mean coalescence times that can be used to predict genetic structure. Contrary to previous models, seed migration is modeled as episodic, partial replacement with seed from single seedlots rather than as constant immigration from the entire metapopulation. This particular form of migration led to novel results. Within-deme coalescence time was not invariant to the amount of migrating seed as predicted by classical models. Genetic structure showed a parabolic relationship to the amount of migrating seed instead of presenting the expected exponential decrease. In contrast, the effects of seed migration frequency on diversity and structure were in line with classical predictions. These results imply that seed migration in managed maize populations cannot be described by a single parameter. We showed that genetic structure depends on deme size when the amount of migrant seed is large with respect to the size of the population. Extinction could decrease or increase genetic structure depending on the level of migration and number of demes. By studying the effect of seed related parameters on genetic structure in the presence of different levels of pollen flow, we demonstrated that higher levels of pollen migration can mask the effects of seed management on structure.

Introduction

The need to protect the genetic resources of the world's most important cultivated plants has sparked a growing interest in the patterns of crop genetic diversity and in the cultural practices that affect these patterns (Louette 1997; Sanou et al. 1997; Dje et al. 1999; Brocke et al. 2003; Pressoir et al. 2004; Perales et al. 2005). A lack of appropriate models has meant an inability to link knowledge on farmer practice to genetic data however. Metapopulation models have been proposed to describe the population dynamics of managed crop populations, since apart from pollen and seed migration there is frequent extinction and recolonization in the form of seed loss and replacement (Brush 1999; Louette 1999; Pressoir et al. 2004; Alvarez et al. 2005). To date, the metapopulation concept in crop population genetics has been used mainly metaphorically and its aptness for describing patterns of genetic diversity has not been evaluated (Louette 1999).

Population genetic models of subdivided species have been instrumental to our understanding of neutral genetic diversity and structure. General results from classical models such as Wright's island model and the more recent metapopulation models have served to predict the genetic effects of population size, migration rates and extinction/colonization in natural populations (Slatkin 1977; Maruyama et al. 1980; Lande 1992; Whitlock et al. 1997; Wakeley et al. 2001). In spite of vast differences in natural history, most species of animals and plants present patterns of demography and migration that often approximate assumptions underlying population genetic models. Crop species are different in this respect. Demography and seed migration in cultivated plants are subject to conscious intervention by farmers and hence deviate substantially in quantity and pattern from what may be expected in most natural populations. The consequences of seed management for the validity of models of subdivided populations have yet to be explored.

We will begin by generalizing a common approach to modeling neutral genetic diversity in metapopulations and extend it to include several important features that are unique to farmer-managed crops. We use maize as an example since there is a good body of knowledge both on farm level diversity and seed management practices. We expect our results to be representative for other sexually propagated crop species however.

To illustrate the features that are unique to our maize metapopulation, we define what we will refer to as a classic metapopulation model. Our definition is based on Slatkin's model II (Slatkin 1977). The latter 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 a randomly chosen source deme (propagule pool model).

Farmer-managed maize differs from a classic metapopulation in several respects. First, like other grain crops, maize was selected for having many, non-detaching seeds per panicle (Harlan et al. 1973). This has made the ear the focus of seed management (Louette et al. 2000; Perales et al. 2003). Seed is generally planted from a limited number of ears so effective population size is expected to be much smaller than census size (Louette 1997). Second, seed migration into a deme generally involves a batch of seed taken from a single source rather than a mix from different sources (Rice et al. 1998). This aspect of migration is equivalent to the propagule pool model of recolonization as introduced by Slatkin (Slatkin 1977) but in this case it applies to seed migration as well. In addition farmers tend to recycle their seed for several years without any inflow 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 population bottleneck that is assumed in most metapopulation models. In case of total loss of seed, farmers will generally obtain enough seed to plant the desired acreage of land instead of reducing the planted area.

In this paper we will show that the characteristics that distinguish crop metapopulations from most natural species lead to predictions that are different from those emanating from the classical metapopulation model. We present results on the effect of different measures of seed migration on two classical predictions about genetic diversity and structure in subdivided populations, namely the invariance principle and the reduction of genetic structure through migration. In addition we will explore the effect of deme size and extinction rate on genetic structure. We will conclude by evaluating to what extent seed related parameters in our model are expected to leave measurable imprints on genetic diversity when pollen flow is incorporated.

Calculating mean diversity and structure in a metapopulation

We start by presenting a generalization of the recurrence equations developed by Maruyama , Latter, and Slatkin (Slatkin 1977) and reframed in terms of average coalescence times by Pannell and Charlesworth (Pannell et al. 1999). A subdivided population is described in terms of the mean time to coalescence for two alleles sampled in at time 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. It provided a direct measure of 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 $T = \frac{T_0}{n} + T_1\left(1 - \frac{1}{n}\right)$, where n is the total number of demes (Pannell et al. 1999). Genetic structure, when defined as the relative reduction in within population diversity is estimated by $F_{st} = \frac{T - T_0}{T}$ (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. For a an allele pair sampled at generation t, three possible coalescence times for pairs of alleles thus exist. A coalescence time of 1 generation for those that co-located and coalesced at time t'. A coalescence time of $1 + T_0'$ for alleles that co-located in the previous generation but did not coalesce. Finally, two alleles that have come from two different demes have a coalescence time of $1 + T_1'$. Mean values of T_0 and T_1 may then be calculated by the following recursion equations:

$$T_0 = \sum a_i P_i + \sum a_i (1 - P_i)(1 + T_0') + \left(1 - \sum a_i\right)(1 + T_1') \qquad (1)$$

$$T_1 = \sum b_j P_j + \sum b_j (1 - P_j)(1 + T_0') + \left(1 - \sum b_j\right)(1 + T_1') \qquad (2)$$

Where P_i and P_j are probabilities of coalescence for two co-locating alleles sampled within the same deme and from two different demes respectively. The subscript reflects the fact that coalescence probabilities may be different for different combinations of alleles.

The terms a_i and b_j are compound terms expressing the proportion of all possible allele pairs that co-locate and have a coalescence probability of P_i and P_j respectively. The sums $\sum a_i$ and $\sum b_j$ 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_1' and T_0' with T_1 and T_0 in equations (1) and (2) such that:

$$T_0 = \frac{\left(1-\sum a_i\right)}{\sum b_j}\overline{P}^{-1} + \overline{P}^{-1} \qquad (3)$$

and

$$T_1 = T_0\left(1-\overline{P}_j\right) + \left(\sum b_j\right)^{-1} \qquad (4)$$

where

$$\overline{P} = \left(1-\sum a_i\right)\overline{P}_j + \left(\sum a_i\right)\overline{P}_i \qquad (5)$$

with

$$\overline{P}_j = \frac{\sum b_j P_j}{\sum b_j}$$

being the mean coalescence probability for co-locating allele pairs from different populations, and:

$$\overline{P}_i = \frac{\sum a_i P_i}{\sum a_i}$$

Representing the mean coalescence probability for co-locating allele pairs from the same population.

These expressions can be interpreted as follows. Looking back in time, a fraction $\sum a_i$ of allele pairs sampled from the same deme in generation t contains alleles that co-located in t-1, and a fraction $1-\sum a_i$ that contains alleles that did not co-locate. The coalescence probability for these fractions is given by the probability of co-location in t-2, times the mean probability of coalescence for alleles sampled from the same deme. This combined probability is $\sum a_i\overline{P}$ for fraction $\sum a_i$, and $\sum b_j\overline{P}$ for fraction $1-\sum a_i$.

Since the expected time to coalescence is given by the inverse of the coalescence probability per generation we may write: $T_0 = \left(1-\sum a_i\right)\left(\sum b_j \overline{P}\right)^{-1} + \sum a_i \left(\sum a_i \overline{P}\right)^{-1} = \frac{1-\sum a_i}{\sum b_j}\overline{P}^{-1} + \overline{P}^{-1}$

At a given value of $\overline{P}$, average within-deme coalescence is thus as a function of the ratio between the probability that two alleles move to different demes in t-1 and the probability that they trace back to the same deme in t-2. This ratio is essentially the time that allele pairs spend outside of a single deme relative to the time spent within a single deme.

Average coalescence time for allele pairs sampled from two different population is given by the sum of the average time $\frac{1}{\sum b_j}$ 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 $\overline{P}_j$ of allele pairs coalesces upon entering the same deme and a fraction of $1-\overline{P}_j$ coalesces in T_0 generations.

These general equilibrium expressions will be used to generate a specific model that incorporates features specific to maize metapopulations.

Metapopulation model for farmer-managed maize

We will proceed by describing the describing the parameters of our maize metapopulation model that will allow the estimation of T_0 and T_1 as described above. We describe a diploid, monoecious plant species with random selfing. There are n demes each of which consists of seed from N_f ears, yielding N mature plants with $N_f << N$ and a fixed number of $\frac{N}{N_f}$ seeds per ear. Generations are discrete. The life cycle of each deme consists of two consecutive phases: a reproductive phase and a seed phase. During the reproductive phase zygote formation, random pollination and pollen migration occur. Each new seed that is formed contains a maternal allele inherited from one of N_f plants and a paternal allele derived from one of N pollen fathers. A proportion of $1-m_g$ of all paternal alleles will result from random pollination by pollen from the same population while a proportion m_g will represent migrant pollen from other populations. Pollen migration follows an island model with migrants originating from any of the $n-1$ populations. 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 populations go extinct and $n(1-e)$ populations remain. An extinct population is replaced by introducing N_f ears from the non-migrant fraction of any of $n(1-e)$ extant populations. There will be no subsequent migration into this deme. 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 population. The fraction of migrant seeds thus equals $m = \frac{N_{fm}}{N_f}$ in populations undergoing migration and $\overline{m} = p_m m$ in all extant populations. 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_j of co-locating allele pairs. The derivation of these terms is given in appendix I.

Table 1. Representation of maternal (F) and paternal (M) allele fractions in a metapopulation.

	e		1-e					
			1-p_m		p_m			
F (1/2)					1-m		m	
M (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 of allele pair	P_i, P_j	Co-locating fraction for P_i, P_j	
Within Demes	maternal × maternal non-migrants	$P_1 = \frac{1}{2(N_f - N_m)}$	a_1	$\frac{1}{4}(1-e)p_m(1-m)^2$
	maternal × maternal migrants	$P_2 = \frac{1}{2N_m}$	a_2	$\frac{1}{4}(1-e)p_m m^2$
	maternal × maternal no migration	$P_3 = \frac{1}{2N_f}$	a_3	$\frac{1}{4}(1-(1-e)p_m)$
	paternal × paternal paternal x maternal	$P_4 = \frac{1}{2N}$	a_4	$\left(\frac{3}{4} - m_s\left(1-\frac{1}{4}m_s\right)\right)\left(1-(1-e)2p_m m(1-m)\right) + \frac{\frac{1}{4}\left(2m_g(2-m_s)(1-e)2p_m m(1-m) + m_g^2\right)}{n-1}$
Between Demes	maternal × maternal	$P_5 = 0$	b_1	$\frac{\frac{1}{4}\left(1-(1-e)^2(1-\overline{m})^2\right)}{n(1-e)-1}$
	paternal × paternal paternal × maternal	$P_4 = \frac{1}{2N}$	b_2	$\frac{\left(\frac{3}{4} - m_g\left(1-\frac{1}{4}m_s\right)\right)\left(1-(1-e)^2(1-\overline{m})^2\right)}{n(1-e)-1} + \frac{m_g\left(1-\frac{1}{4}m_s\right)}{n-1}$

Results

Effective deme size and coalescence time

As was mentioned in the introduction, the practice of selecting a limited number of ears per deme reduces effective population size with respect to the census size N. Inbreeding effective size is related to the mean probability of coancestry P in the previous generation as follows: $N_e = 1/2P$ (Kimura et al. 1963). For a single deme this probability is given by equation (5). We may hence calculate N_e by setting pollen- and seed migration to 0 and substitute the terms a_i and P_i from Table 2 in equation (5). This yields:

$$N_e = \frac{4N_f N}{3N_f + N} \qquad (6)$$

This result is identical to Venkovsky and Crossa's variance effective size with female gametic control (Crossa et al. 1994).

In the classical metapopulation model without extinction, there is only a single coalescence probability for any pair of co-locating alleles. Therefore $\overline{P} = P$ and we may write:

$$T_0 = \left(1 - \sum a_i\right)\left(\sum b_i\right)2N_e + 2N_e \qquad (7)$$

In case of different coalescence probabilities $\overline{P}$ does not need to be equal to P. It can be shown numerically however, that $\overline{P}$ in our model closely approximates $2N_e^{-1}$ 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:

$$T_1 \approx T_0 + \sum b^{-1} \qquad (8)$$

for between deme coalescence time. Expressions (7) and (8) greatly simplify comparison to previous results and will be used subsequently in this paper.

Extinction, seed migration and within deme coalescence time

Under the classical model of subdivided populations, within-deme coalescence time has an expected value of $2Nn$ or two times the total number of breeding individuals, irrespective of the rate of migration (Nagylaki 1982; Slatkin 1987; Strobeck 1987; Hey 1991; Nagylaki 1998; Wilkinson 1998; Nagylaki 2000). Pannell and Charlesworth (Pannell et al. 1999), showed that extinction leads to a breakdown of this so-called invariance principle. Under population replacement, 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:

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

The term m^2 represents the fraction of allele pairs sampled from two migrant alleles. Since migrants are assumed to be a random sample from the metapopulation they have a co-location probability of $(n-1)^{-1}$. When n is large, $\dfrac{m^2}{n-1}$ 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 $1-(1-m)^2$ of allele pairs that do not co-locate and the fraction $\left(1-(1-m)^2\right)(n-1)^{-1}$ that re-locates 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 1. Clearly, the invariance principle does not hold with respect to m, even in the absence of extinction. At $e=0$, increasing m leads to a linear decrease in T_0 from approximately $2N_e n$ when m is close to zero to $2N_e$ when m is one. At higher rates of extinction, T_0 is first increased until reaching a maximum and then decreases until reaching $2N_e$ at $m=1$.

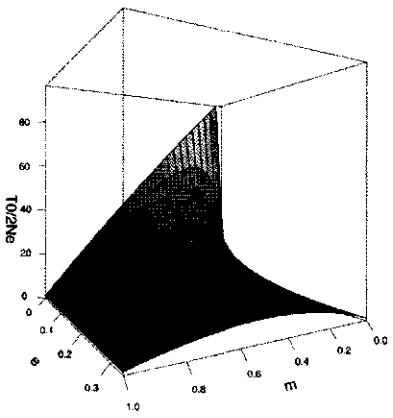


Figure 1. Within-deme diversity ($T_0/2N_e$) as a function of extinction rate (e) and quantity of migrating seed (m), ($n = 100$, $m_g = 0$).

We will explain this result mathematically by setting p_m to unity and m_g to zero in (3), which gives:

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

With: $\varepsilon = \frac{1}{(1-e)^2} - 1$.

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. Increased quantity of migrating seed will thus decrease T_0 in the absence of extinction. 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, as m increases further, T_0 starts to decrease. The maximum value of T_0 converges to $m = 0.5$ for large ε.

The relation between T_0 and migration quantity thus deviates strongly from what would be expected based on the classical metapopulation model.

The effect of seed migration frequency on within-deme coalescence time can be appreciated in the following equation:

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

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. Since $2p_m m^2 > p_m{}^2 m^2$ for $m > 0$, 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 stronger with p_m than the term $-2p_m$ numerator, causing T_0 to rise in response to migration frequency. Single source migration may therefore be said to cause 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 re-location from different demes. Extinction augments the probability that alleles in different demes co-locate. At the same time, co-location probability for alleles within the same deme is not affected because all colonists derive from the same population. Therefore, extinction is expected to decrease the time that alleles spent outside a single deme, causing a reduction in within-deme coalescence time. This effect is exacerbated by the lower number of extant source demes, which increases the probability for co-location of alleles in different demes even further. A similar explanation underlies the effects of m and p_m. Under single source migration, immigrants within a deme share the same population of origin. Increasing the quantity of migrating seed will therefore decrease the proportion of co-locating alleles but at a decreasing rate until half of the alleles in a deme consists of migrants. Increasing m beyond this point will result in a higher proportion of co-locating alleles until all alleles co-locate at $m = 1$. In contrast, the probability of drawing co-locating alleles from two different demes keeps increasing with m. As a result, higher values of m will cause alleles to spend less time in different demes, causing a relative reduction in T_0. The response to p_m is different because seed sources for each deme are independent.

As more demes receive migrants and have a lower proportion of co-locating alleles, there is a proportionally higher 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 classical model without extinction

Many empirical studies of subdivided populations use Wright's fixation index F_{st} (Wright 1951), 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 $\frac{1}{4Nm+1}$. Although recognized as overly simplified (Whitlock et al. 1999), this formula serves as the basis of 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 population size provided that Nm remains constant. As expected, both expectations hold in the classical metapopulation model without extinction. Assuming an infinite number of demes we obtain:

$$F_{st} \approx \frac{1}{4N_e\left(m-\frac{1}{2}m^2\right)+1} \qquad (12)$$

Which is identical to the result obtained by Wright and to his reduced equation when m is small.

Seed migration and F_{st}

Figure 2 shows the response of F_{st} to both seed migration quantity and frequency 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 migrating seed quantity is parabolic with a minimum at $m = 0.5$. We can derive that this result is due to the assumption of single source migration by analyzing the equilibriums solution for F_{st} without extinction or pollen flow. When the number of demes is very large F_{st} is determined by $1-\sum a_i$ (appendix II).

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

$$F_{st} \approx \frac{1}{4N_e p_m m(1-m)+1} \qquad (13)$$

Migrating seed derives from a single source in each generation. Hence, $m = 0.5$ 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 genetic structure. In contrast, migration frequency determines the amount of migrant seed that comes from different demes. For small m therefore, the effect of p_m is expected under the classical model, since $p_m m(1-m) \approx \overline{m}$ and $F_{st} \approx \frac{1}{4N_e\overline{m}+1}$. A combination of high m and low p_m may result in a higher value of F_{st} than expected on the basis of the number of migrants $N_e\overline{m}$. The negative relation between p_m and F_{st} will hold regardless of the magnitude of m.

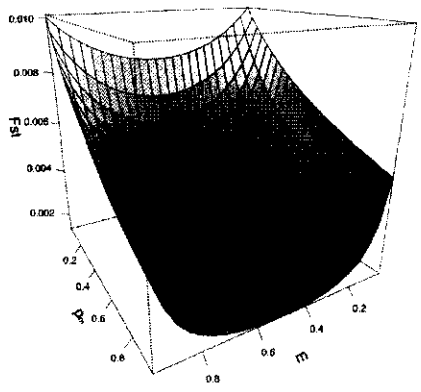


Figure 2. F_{st} as a function of seed migration frequency (p_m) and quantity of migrating seed (m), ($n = 100$, $e = 0.3$, $m_g = 0.04$).

Deme size and F_{st}

The parameters m and p_m may vary independently. Consequently, the average number of migrants per deme, $p_m N_m$ may be low while the number of ears entering a receiving deme, N_m is high. An important consequence of this model property is that F_{st} becomes dependent on deme size. To illustrate this we write m in equation (13) as $N_m N_e^{-1}$ so that we may write:

$$\frac{1}{4p_m N_m\left(1-N_m N_e^{-1}\right)+1} \qquad (14)$$

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

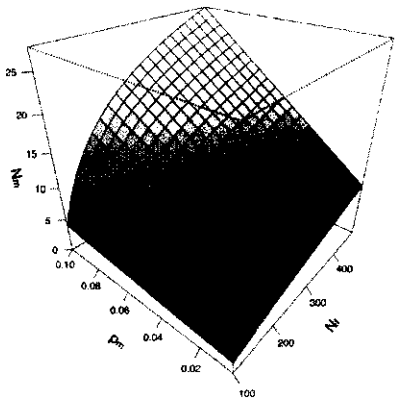


Figure 3. Estimated number of migrants as a function of seed migration frequency (p_m) and number of planted ears (N_f), ($n = 100$, $N_{fm} = 90$, $e = 0.3$, $m_g = 0$).

The effect of extinction on F_{st}

In metapopulations with extinction, Wright's classic formula for F_{st} no longer provides an adequate description of the relation between seed flow and genetic structure. In an analysis of Slatkin's model II, Wade & McCauley showed that with propagule pool recolonization, extinction increases population differentiation (Wade et al. 1988; Whitlock et al. 1990; Pannell et al. 1999). This result was due to the strong drift occurring during recolonization. The present model does not share Slatkin's assumption of a population bottleneck after extinction. Consequently, our results on the effect of extinction on F_{st} are rather different. Figure 4. shows our model's results for F_{st} as a function of extinction rate at different frequencies of seed migration in a metapopulation of a hundred demes.

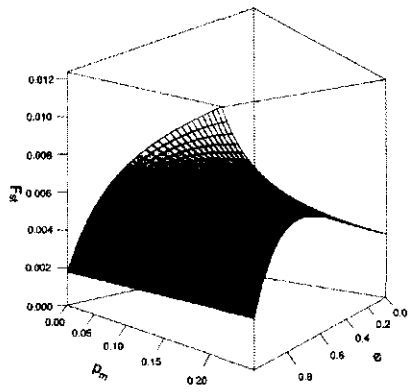


Figure 4. F_{st} as a function of seed migration frequency (p_m) extinction rate (e), ($n = 100$, $N_{fm} = 75, m_g = 0.04$).

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} (Pannell et al. 1999).

At low migration frequencies however, F_{st} is decreased by extinction. The reason for this can be seen in expression (15)(Appendix II).

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

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 to the second and F_{st} will decrease with increasing e. When on the other hand 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 $n = \infty$ respectively.

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

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

This result shows that in our model the conclusion drawn by Wade & McCauley on the effect of extinction on F_{st} holds for large n, but that at lower n extinction may either increase or decrease population differentiation depending on migration rates.

The effects of seed management in the presence of pollen flow.

In the results presented so far pollen migration was assumed absent in order to explore the effects of human mediated gene flow on genetic diversity and structure. In reality, both seed- and pollen migration will occur simultaneously 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 = 0.005, 0.01, 0.02, and 0.04).

For the effect of deme size, the 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 swamps any effect of seed management on genetic structure. It is important to note that in many agricultural systems, the potential for pollen flow between neighboring fields is high. Considering the published estimates of around 1% for each neighboring field (Messeguer et al. 2006), values of up to 4% may be realistic for situations were fields are planted contiguously.

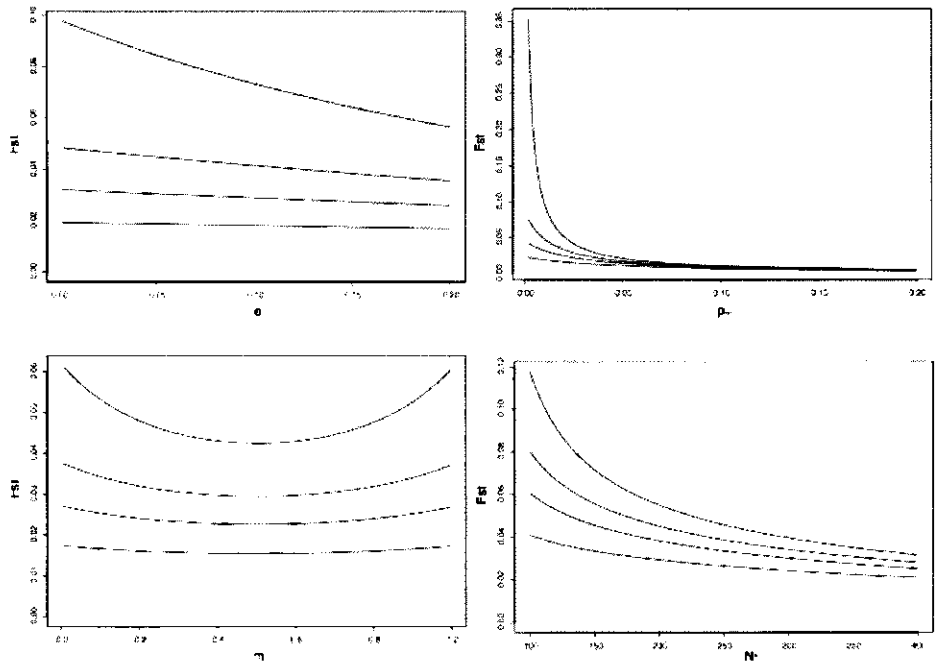


Figure 5. Clockwise from upper left pane to lower right pane: F_{st} as a of function extinction rate (e), migration frequency (p_m), quantity of migrating seed (m) and number of planted ears (N_f) at different levels of pollen flow (m_g = 0.005, 0.01, 0.02, 0.04). In each pane, pollen flow increases from higher to lower curves.

Discussion

The determinants of neutral genetic diversity and structure are of interest to evolutionary- and conservation biology. Molecular marker data can describe the distribution of genetic diversity but their correct interpretation relies on models that adequately represent the population genetics of the species under study. Although there has been a growing interest in the population genetics of agricultural plant species, no specific models describing the dynamics of subdivided crop populations have been available. The adapted metapopulation model presented in this study represents the first attempt to accommodate aspects specific to managed germplasm. It provides a means to exploring the effects of observed cultural practices on patterns of genetic diversity. The approximations for equilibrium coalescence times have provided insight into the mechanisms shaping neutral genetic diversity in maize metapopulations.

Our model confirms that given the specified assumptions, managed metapopulations deviate significantly from classical models of subdivided populations. The effects of single source migration on within-deme diversity and F_{st} show that it is impossible to characterize gene flow by a single migration parameter. The amount of migrating seed and seed migration frequency have different and sometimes opposing consequences. The response of coalescence time and F_{st} to changes in migrant seed quantity defies classical predictions because the correlated origin of migrants reduces the time spent in different demes. Furthermore, the combination of low migration frequency and high migration quantity causes deme size to affect genetic structure. Deme size has been traditionally ignored as a determinant of genetic structure, based on the classical prediction that only the number of migrants affects F_{st} (Wright 1951). Farmers are likely to incorporate rather large quantities of migrant seed whenever they are faced with a shortage of planting material. This suggests that deme size is a factor that should be accounted for in order to understand genetic structure in crop populations.

Because a farmer can be expected to obtain sufficient seed in case of seed loss, extinction takes 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 was predicted previously (Wade et al. 1988). Genetic structure may either increase or decrease in response to seed loss, depending on migration rate and number of demes.

The characteristics of seed migration and extinction set apart our metapopulation model from models based on natural populations. Maize is an out-breeding crop however and pollen flow needs to be taken into account. Our results show that parameters related to seed management, especially extinction rate and migration quantity, may be swamped at high but realistic rates of pollen migration. Studies relating field data with measures of genetic structure should therefore first establish the potential for pollen flow. In case pollen flow is low enough to expect significant effects of seed dynamics, it is important to quantify number of planted ears seed replacement, migration frequency and quantity.

As an example of using data on farmer practice to explain genetic differentiation, we compare the value of within village F_{st} reported by Pressoir and Berthaud (2004) to seed management data from their sample area (Bellon unpublished). The following mean parameter values were obtained: $N_f = 260$, $N = 6500$, $p_m = 0.03$, $e = 0.22$, $m = 0.3$, $n = 500$. Based on published pollen flow estimates (Messeguer et al. 2006) and the observation of field contiguity in the study area we assume $m_p = 0.04$. The resulting predicted value of 0.010 is indeed very close to the published value of 0.011.

Admittedly, the latter correspondence is conditional on the level of pollen flow assumed. This conclusion in itself confirms the value of our model however. It proves that we are now able to the model the interaction of a set of parameters and assumptions that are unique to maize metapopulations. Obtaining a result that is in close agreement with reported values and the fact that we may explain why this is so implies a significant step forward in our ability to understand observed patterns of crop genetic diversity.

Appendix I

Alleles sampled from the same population

For two non-migrant maternal alleles sampled from an extant population that has undergone seed migration there are $N_f - N_{fm}$ possible maternal plants so P_i becomes:

$$P_1 = \frac{1}{N_f - N_{fm}} \times \frac{1}{2} = \frac{1}{2(N_f - N_{fm})} \qquad (18)$$

The co-locating probability in this case is one.

Hence:

$$a_1 = (1-e)p_m \frac{1}{2}(1-m) \times \frac{1}{2}(1-m) \times 1 = \frac{1}{4}(1-e)p_m(1-m)^2 \qquad (19)$$

The factor $\frac{1}{2}$ represents the subdivision of the metapopulation into maternal and paternal alleles.

Two maternal alleles that are both sampled from migrant seed can originate from N_{fm} maternal plants, yielding:

$$P_2 = \frac{1}{2N_{fm}}$$

Since there is only a single source of migrant seed for each deme in each generation the co-location probability again equals one, giving:

$$a_2 = \frac{1}{4}(1-e)p_m m^2 \qquad (20)$$

Two maternal alleles sampled from a population that has not received seed migrants may have originated from any of N_f ears, so that:

$$P_3 = \frac{1}{2N_f}$$

with $a_3 = \frac{1}{4}(1-(1-e)p_m)$ (21)

When a pair of sampled alleles includes a paternal allele the coalescent probability is determined by the total number of plants in a population and becomes:

$$P_4 = \frac{1}{2N}$$

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.

Hence:

$$\frac{1}{4}\left(2(1-m_g)+(1-m_g)^2\right)\left(1-(1-e)p_m 2m(1-m)\right)\times 1$$

which after rearanging terms gives,

$$\left(\frac{3}{4}-m_g\left(1-\frac{1}{4}m_g\right)\right)\left(1-(1-e)p_m 2m(1-m)\right) \quad (22)$$

The term $(1-m_g)^2$ represents combinations of two paternal alleles that are both pollen residents. The term $2(1-m_g)$ is the fraction of 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 $(n-1)^{-1}$ when both alleles are pollen migrants or if one of the alleles is sampled from migrant seed, giving:

$$\frac{1}{4}\left(m_g{}^2+\left(2m_g(1-m_g)+2m_g\right)(1-e)p_m 2m(1-m)\right) \quad (23)$$

Combining (22) and (23) we get:

$$a_4 = \left(\frac{3}{4}-m_p\left(1-\frac{1}{4}m_p\right)\right)\left(1-(1-e)p_m 2m(1-m)\right)+\frac{2m_p(2-m_p)(1-e)p_m 2m(1-m)+m_p{}^2}{4(n-1)} \quad (24)$$

Alleles sampled from two different populations

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. Co-location probability is $n(1-e)^{-1}-1$ unless both alleles are sampled from resident seed. So we may write:

$P_5 = 0$

with

$$b_1 = \frac{1}{4}\left(\frac{1-(1-e)^2(1-\overline{m})^2}{n(1-e)-1}\right) \quad (25)$$

When the sample from two populations 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 $(n-1)^{-1}$. The proportion of allele pairs that contain one or two pollen migrants is:

$$\frac{1}{4}\left(2m_g(1-m_g)+2m_g+m_g^2\right) = m_g\left(1-\frac{1}{4}m_g\right) \quad (26)$$

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

$$\frac{1}{4}\left(2(1-m_g)+(1-m_g)^2\right) = \frac{3}{4}-m_g\left(1-\frac{1}{4}m_g\right) \quad (27)$$

As in the case of b_1, samples from this fraction co-locate with probability $n(1-e)^{-1}-1$ unless both alleles are sampled from resident seed. Hence:

$$b_2 = \frac{m_g\left(1-\frac{1}{4}m_g\right)}{n-1} + \frac{\left(\frac{3}{4}-m_g\left(1-\frac{1}{4}m_g\right)\right)\left(1-(1-e)^2(1-\overline{m})^2\right)}{n(1-e)-1} \quad (28)$$

Appendix II

The general equilibrium solutions for T_0 and T_1 are:

$$T_0 = \frac{\left(1-\sum a_i\right)}{\sum b_j}\overline{P}^{-1} + \overline{P}^{-1}$$

and

$$T_1 = T_0\left(1-\overline{P}_j\right) + \left(\sum b_j\right)^{-1}$$

Which for our model becomes:

$$T_0 = \frac{\dfrac{1-\sum_{i=1}^{i=4} a_i}{b_1+b_2} + 1}{P_4 Q\left(1-\sum_{i=1}^{i=4} a_i\right) + \sum_{i=1}^{i=4} a_i P_i}$$

$$T_1 = T_0\left(1-QP_4\right) + \frac{1}{b_1+b_2}$$

With:

$$Q = \frac{b_2}{b_1+b_2}$$

We assume:

$$P_4 Q\left(1-\sum_{i=1}^{i=4} a_i\right) + \sum_{i=1}^{i=4} a_i P_i \simeq \frac{1}{2N_e}$$

Derivations of F_{st}:

We define:

$$\alpha = \sum a_i$$

$$\beta = \sum b_i$$

$$T = T_0 n^{-1} + T_1\left(1 - n^{-1}\right)$$

$$T_1 \simeq T_0 + \beta^{-1}$$

$$F_{st} = \frac{T - T_0}{T}$$

So that:

$$F_{st} = \frac{T_0 n^{-1} + \left(T_0 + \beta^{-1}\right)\left(1 - n^{-1}\right) - T_0}{T_0 n^{-1} + \left(T_0 + \beta^{-1}\right)\left(1 - n^{-1}\right)}$$

$$F_{st} = \frac{\beta^{-1}\left(1 - n^{-1}\right)}{\beta^{-1}\left(1 - n^{-1}\right) + T_0}$$

when n is large:

$$F_{st} \approx \frac{1}{T_0\beta + 1}$$

Which equals:

$$F_{st} \approx \frac{1}{\left((1-\alpha)\beta^{-1}2N + 2N\right)\beta + 1}$$

For $m_g = 0$, this leads to:

$$\alpha = 1 - (1-e)p_m 2m(1-m)$$

$$\beta = \frac{1-(1-e)^2(1-\overline{m})^2}{n(1-e)-1} \approx \frac{1-(1-e)^2(1-\overline{m})^2}{n(1-e)}$$

$$F_{st} = \frac{1}{4N_e\overline{m}(1-m)(1-e) + 1 + 2N_e\beta}$$

$$F_{st} = \frac{1}{4N_e p_m m(1-m)(1-e) + 1 + \dfrac{2N_e\left(1-(1-e)^2(1-\overline{m})^2\right)}{n(1-e)}}$$

Which gives:

$$F_{st} = \frac{1}{4N_e p_m m(1-m)(1-e) + 1 + \dfrac{2N_e}{n}\left(\dfrac{1}{1-e} - (1-e)(1-p_m m)^2\right)}$$

When $n \to \infty$ so that $2N_e\beta \to 0$, we have:

$$F_{st} = \frac{1}{(1-\alpha)2N_e + 1}$$

For the island model with infinite n:

$$1-\alpha = 1-(1-m)^2 = 2m + m^2$$

Which, ignoring m^2, gives:

$$F_{st} \approx \frac{1}{4N_e m + 1}$$

For our model, setting m_g and e to 0, we have:

$$1-\alpha = 1-\left(1 - p_m 2m(1-m)\right) = 2p_m m(1-m)$$

Which gives:

$$F_{st} \approx \frac{1}{4N_e p_m m(1-m) + 1}$$

Chapter III

Determinants of regional genetic structure in Mexican maize landraces

Abstract

In this study we aim to determine the different environmental and human factors that affect genetic structure in maize landraces in central Mexico. Although the importance of both environment and humans in shaping crop genetic diversity is well established, few studies have looked at both types of determinants simultaneously. We describe 60 seedlots sampled from 20 villages in highland and lowland environments for both agronomical traits and molecular markers. Within-and between village F_{st} and Q_{st} values are used as measures of neutral and agronomic genetic structure respectively. We apply a newly developed computer model in combination with data on local seed management practice and planting patterns to predict F_{st} in the two environments. Genetic differences were strong between highland and lowland maize, for both markers and traits. Three highland villages planted maize varieties showing evidence of admixture in molecular markers and phenological traits. This provided evidence for the occurrence of gene flow from lowland to highland environments. Genetic structure was low for molecular markers but was notably higher in the lowlands. This difference was predicted by our model and was explained by lower pollen flow and smaller seedlot sizes in the lowlands. Genetic structure was higher for agronomical traits, especially those related to flowering time. This suggests that selection on flowering time is an important determinant of genetic structure. Field data suggested a relation between phenology and planting dates. Phenological differentiation was highest in the transect containing the admixed seedlots, proving that genetic structure may result from the introgression of traits that diverged in a foreign environment.

Introduction

Maize was domesticated around 6,000 to 9,000 years ago in Mexico (Piperno et al. 2001; Pohl et al. 2007), a country that today represents the crop's main center of diversity (Sanchez et al. 2000). Morphological and molecular diversity is high compared to other crop species (Doebley et al. 1985; Buckler et al. 2001). Part of this large diversity is represented by racial types adapted to different environments (Wellhausen et al. 1952). In Mexico alone, cultivation takes place from the tropical lowlands to altitudes up to 3000 meters. Several studies attest to the role of growing environment and most importantly altitude in shaping the genetic differences between maize types (Doebley et al. 1985; Perales et al. 2003). The factors that cause genetic differentiation within different environments are less well known however.

Understanding the determinants of genetic diversity is relevant to germplasm conservation and management. Over the last decade, the use of in-situ conservation strategies (Brush 2001) as a complement to ex-situ germplasm collections have gained importance (Hammer 2003). This has generated an interest in local and regional patterns of genetic structure as well as appreciation for the role of farmers in shaping these patterns. Maize grown in traditional farming systems is subject to seed recycling, selection and exchange which may affect genetic differentiation between seedlots (Louette et al. 2000). Work performed in two environmentally homogeneous regions in Mexico has shown that differences in agronomical traits are maintained by diversifying selection acting at the farm- and village level (Pressoir et al. 2004; Perales et al. 2005). This suggests that even in the absence of clear environmental contrasts selection creates genetic differences between seed-lots from different farmers and villages. Human selection has been suggested as the main explanation for observed differentiation (Pressoir et al. 2004; Perales et al. 2005), but direct links between farmer practice and genetic data have not been made.

An issue that has been ignored thus far, is the way that human practice and environment may act together to shape genetic structure. Farmers are known to adopt foreign germplasm (Louette et al. 2000) so seed exchange between contrasting growing environments may occur. Such exchange could lead to the introduction of material that is distinct from local varieties and hence cause an increase in genetic structure. Moreover, it is not known if farmer practice may vary across environments, and to what extent this may lead to different patterns of genetic diversity.

The present study aims at elucidating the determinants of between- and within village genetic structure across two different altitudinal environments in Mexico. We focus on the central highlands above 2000 meters and the adjacent lowlands below 500 meters. Climatic differences between the two environments are large. Highlands are characterized by a short growing season, low spring temperatures and lower rainfall. Lowlands are tropical, with high temperatures and rainfall, allowing year-round cultivation. Maize landraces in the two environments belong to different germplasm groups (Jiang et al. 1999). Highland maize has a shorter growing cycle and plant type as well as conical ears with a high kernel row number. Lowland maize is generally late flowering and tall, with cylindrical ears with fewer kernel rows. In spite of these differences, both types of maize can grow in each of the two environments albeit at the cost of considerable yield reduction (Jiang et al. 1999). No absolute barriers to gene flow thus appear to exist.

We addressed the occurrence of gene flow between highland and lowland environments and its potential role in shaping regional genetic structure. Seedlots sampled from both environments were characterized for molecular marker frequencies and quantitative traits to detect evidence of admixture. Comparison of genetic structure observed for markers and traits can serve to measure the occurrence of diversifying selection (Spitze 1993; Pressoir et al. 2004). By measuring between- and within village differentiation for both markers and traits we were able to evaluate the importance of both neutral and selective forces in shaping regional and local genetic diversity. We developed a population genetic simulation model that incorporates data on agricultural practice and planting patterns to investigate whether model predictions for local farming systems are in line with observed levels of molecular marker structure. Available data on farming practice was used to propose explanations for patterns of differentiation in agronomical traits.

Materials and methods

Seed sampling

Our study area was chosen to include both the central highlands and the eastern lowland areas of Mexico, between 19.3-20.9°N and 96.7-99.1° W. Transition from highlands to lowlands in this region is abrupt, with little physical distance separating the two environments. Four sampling transects running roughly parallel to the transition zone were defined. Transects were designated as follows: highlands (HH), highlands close to lowlands (HL), lowlands close to highlands (LH) and lowlands (LL) (Table 1, Figure 1). The sampling layout was designed to enable distinction between the effects of distance and environment while maximizing climatic homogeneity within transects and environments. In March 2004, five villages were visited in each transect and three farmers per village were asked to provide seed. Twenty five to fifty ears were sampled from each of the 60 farmers.

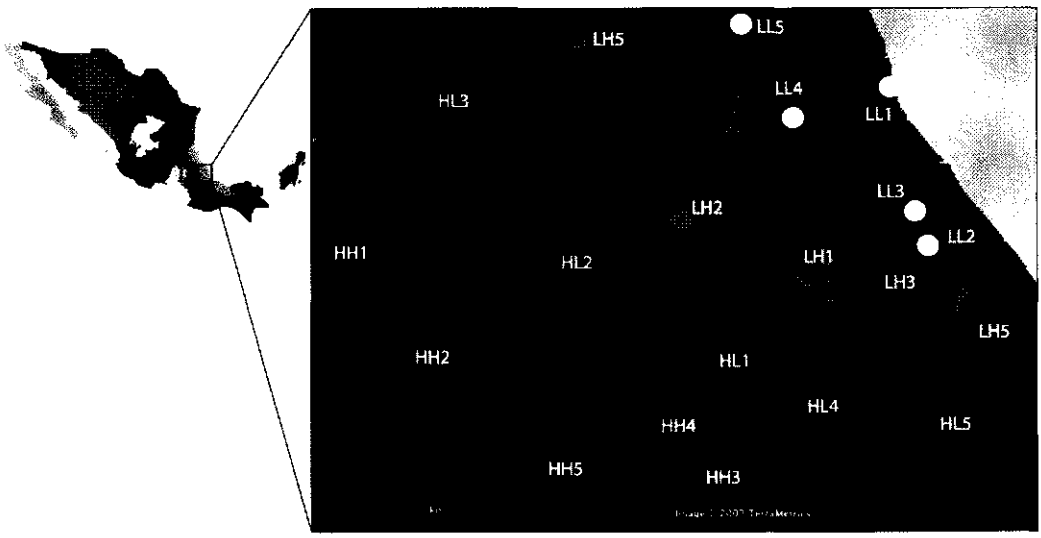


Figure 1. Seed sampling locations. The four transects (HH,HL,LH and LL) are shown. Three seedlots were sampled from each locality.

Table 1. Seed sampling locations with their respective transect codes (ID), altitude and population.

ID	State	Municipality	Locality	Altitude	Category	Population
HH1	Mexico	Hueypoxtla	Tianguistongo	2500.0	HH	1965
HH2	Mexico	Otumba	Ahuatepec	2350.0	HH	1372
HH3	Puebla	Tepeyahualco	Tetipanapa	2340.0	HH	788
HH4	Puebla	Cuyoaco	La gloria	2440.0	HH	425
HH5	Tlaxcala	Panotla	Texantla	2340.0	HH	673
HL1	Puebla	Zacapoaxtla	Cuacuilco	2280.0	HL	588
HL2	Puebla	Huauchinango	Teopancingo	2360	HL	966
HL3	Veracruz	Huayacocotla	Viborillas	2280.0	HL	322
HL4	Veracruz	Altotonga	Adolfo moreno	2180.0	HL	1277
HL5	Veracruz	Landero y coss	Landero y coss	1980.0	HL	1155
LH1	Puebla	Cuetzalan	Tacuapan	300.0	LH	575
LH2	Puebla	Zihuateutla	Tecpatlan	520.0	LH	580
LH3	Veracruz	Tlapacoyan	Piedra pinta	230.0	LH	1798
LH4	Veracruz	Yecuatla	La defensa	260.0	LH	374
LH5	Veracruz	Zontecomatlan	Otlatzintla	380.0	LH	451
LL1	Veracruz	Cazones	Buenavista	10.0	LL	636
LL2	Veracruz	Martinez de la torre	Arroyo blanco	60.0	LL	833
LL3	Veracruz	Tecolutla	Cañada rica	80.0	LL	1031
LL4	Veracruz	Tihuatlan	Huizotate	80.0	LL	614
LL5	Veracruz	Temapache	Emiliano zapata	80.0	LL	1579

Microsatellite genotyping

DNA was extracted from 24 individuals for each of the 60 seedlots, using CIMMYT's standard protocol. Extracted samples were genotyped for 11 microsatellite loci that are listed in Table 2. Fluerescently labeled primers were obtained for these loci (Applied Biosystems, Sigma-Aldrich). Sequences may be retreived from the Maize Genetics and Genomics Database (http://www.agron.missouri.edu/ssr.html). Markers were selected based on chromosomal location, scorability and lack of size overlap. One of the loci, bnlg1784 was chosen because prior information suggested high differentiation between lowland and highland germplasm for this locus (Matsuoka et al. 2002, supplementary material). PCR reactions were performed in a 10µl reaction volume, containing 1-2µl of 2µM primer, 1.2µl of 10mM dNTP, 0.4µl of 50mM $MgCl_2$ and 1µl of 10X PCR buffer.

We used two amplification programs. Q: 94°C for 2 min; followed by 30 cycles of 94°C for 30 s, X°Cfor 1 min, and 72°C for 1 min; followed by extension at 72°C for 5 min. And SSR: 93°C for 2 min; followed by 30 cycles of 93°C for 1 min, X°C for2 min, and 72°C for 2 min; followed by extension at 72°C for 5 min. Where X°C indicates the specific annealing temperature (Table 2).
PCR products were pooled for each individual. 1.5µl of pooled PCR product was denatured in 9 µl of HiDi (applied biosystems) formamide containing 1µl of ROX500 (Applied Biosystems) size standard. Samples were analyzed on an ABI 3100 capillary sequencer (Applied Biosystems). Fragment sizes were scored using Genotyper 2.1 (Perkin Elmer/ Applied Biosystems) software.

Table 2. SSR markers used in this study.Chr.: Chromosome on which markers are located.

Locus	Chr.	Size range	Repeat	Program
phi227562	1	309-325	ACC	SSR-54
phi96100	2	269-305	ACCT	SSR-56
bnlg1784	4	237-264	AG	SSR-56
phi029	3	148-162	AG/AGCG	SSR-56
phi093	4	278-314	AGCT	SSR-56
phi024	5	357-375	CCT	Q-60
umc1061	10	89-113	TCG	Q-60
phi034	7	113-149	CCT	SSR-52
phi014	8	417-435	GGC	Q-56
phi061	9	81-97	TTCT-GTAT	Q-62
bnlg2047	3	132-150	AG	Q-60

Molecular data Analysis

Allelic frequency data was analyzed with the program MSA (Dieringer et al. 2003). Pairwise genetic differences between populations i and j were calculated as $d_{ij} = -\ln(1-\theta)$ (Reynolds et al. 1983), where θ is the coancestry coefficient as defined by Weir and Cockerham (Weir et al. 1984). The measure θ is equivalent to Wright's F_{st} and is defined as the ratio of between population to total genetic variance calculated over all alleles and loci. For neutral alleles, θ provides a measure of differentiation due to drift. Hierarchical θ-statistics were calculated to describe genetic structure within and between villages using the program Arlequin (Excoffier et al. 2005). We write differentiation between villages, seedlots within villages and between all seedlots as θ_v, θ_f and θ respectively.

Pairwise distances d_{ij} were visualized by plotting the two first coordinates obtained from a principal coordinate analysis (procedure pco as implemented in the ecodist package, R statistical software).

Evidence for admixture between highlands and lowlands was evaluated using Bayesian-clustering as performed by the program Structure (Pritchard et al. 2000). This analysis assigns individuals to a predefined number of groups based on posterior group membership probabilities given observed genotypes. We assumed the presence of two groups (K=2) that given the existence of strong genetic differences between environments should correspond to lowlands and highlands. Prior information on genotypes belonging to specific seedlots was included. For seedlots their membership to the two final groups, supposed to roughly coincide with lowlands and highlands, was determined. The group memberships were expressed in relation to the inferred highland cluster (PMH).

Field experiment

A field experiment was planted at CIMMYT's Tlaltizapan field station in November 2004. Fields were managed, irrigated and fertilized following CIMMYT standard procedures. Seed was planted in a split plot like design in two replicates. The 60 collected seedlots were represented by 18 half-sib families consisting of 6 seeds sampled from a single ear. Each replicate field was divided into four blocks of 5 × 21 plots, where each plot consist of a 5-meter row. Each block was randomly assigned to one of the four transects (HH, HL, LH or LH). Within each block, five villages from a single transect were randomly assigned to the 5 sets of 21 rows. The 21 plots per village were divided into three groups of 7 plots. Six randomly selected plots in each group were planted to a single seedlot with three half-sib families per plot. Each plot hence contained 18 plants belonging to three families. Three border plants were planted per plot. The seventh remaining plot within each seed-lot was planted with 18 plants of a spatial control (CML264 x CML311).

A set of phenological and ear / kernel traits were measured for individual plants (Table 3). Low grain yields were obtained for highland populations in this experiment due to lack of adaptation. A second trial including only the highland material was planted at CIMMYT's El Batan station on the central highlands in May 2005. This trial used a commercial hybrid (Promesa, ColPos), adapted to highland conditions, as a spatial control. Means were calculated for all sixty populations. Standardized principal components (PCA) of population means were calculated for phenological and ear traits separately. The first two Principal components were used to plot the phenotypic data.

Table 3. Measured phenotypic traits with their respective units of measurement.

Trait type	Trait name	Units	Trait
Phenological	DA	days	Days from planting to anthesis
	DS	days	Days from planting to anthesis
	PH	5cm	Plant height
	EH	5cm	Ear height
	TLN	#	Total leaf number
	LN	#	Leaves above the ear
	LD	cm	Width of ear leaf
	LL	cm	Length of ear leaf
	SD	mm	Stem diameter (above the ear)
Ear/kernel	EW	g.	Ear weight
	EL	mm	Ear length
	ED	mm	Ear diameter
	KT	mm	Ten kernel thickness
	KN	#	Kernel row number
	GW	g.	Total grain weight
	KW	g.	Hundred kernel weight
	KD	mm.	Ten kernel width
	KL	mm.	Ten kernel length
	CW	g.	Cob weight
	CD	mm.	Cob diameter

Quantitative genetic analysis

For each sampling transect, variance components for half-sib families within populations, populations within villages, and villages within transect were calculated by fitting the following mixed model using the lmer procedure as implemented by the lme4 package in the R statistical software (R_DevelopmentCoreTeam 2005):

Response = replicate + row + column + village + seedlot within village + HS family within farmer's population + error

All terms were random, except replicate. Narrow-sense heritability was estimated as: $h^2 = 4\frac{\sigma^2_f}{\sigma^2_f + \sigma^2_e}$. Each variance component was corrected for effects caused by field heterogeneity by subtracting the village and seedlot components as estimated for the control genotype.

Genetic differentiation for quantitative genetic traits within and between villages was estimated by calculating the ratio of within population genetic variance to total genetic variance or Q_{st} (Lande 1992; Spitze 1993). Trait differentiation between populations is given by $Q_{st} = \frac{\sigma^2_b}{\sigma^2_b + 2\sigma^2_w}$ (Lande 1992; Spitze 1993), where σ^2_b is the genetic variance between populations and σ^2_w the genetic variance within populations. Under the assumptions of neutrality and additivity, Q_{st} is expected to equal θ as estimated using neutral molecular markers (Lande 1992; Spitze 1993; Podolsky et al. 1995). Values of Q_{st} exceeding θ are therefore considered evidence for diversifying selection whereas values lower than θ indicate that stabilizing selection has operated on the trait in question (Merila et al. 2001).

Following (Pressoir et al. 2004), we may define the following hierarchical measures of quantitative trait differentiation:

$$Q_{st,v} = \frac{\sigma^2_v}{\sigma^2_v + 2\left(\sigma^2_p + \sigma^2_g\right)}$$

For between village Q_{st}. Where σ^2_v is the variance component due to village, σ^2_p is the population within village variance and $\sigma^2_g = 4\sigma^2_f$ is the within population genetic variance.

$$Q_{st,f} = \frac{\sigma^2_p}{\sigma^2_p + 2\sigma^2_g}$$

For within village Q_{st}.

$$Q_{st,f} = \frac{\sigma^2_v + \sigma^2_p}{\sigma^2_v + \sigma^2_p + 2\sigma^2_g}$$

For total between-population Q_{st}.

For quantitative traits that are not under selection, $Q_{st,v}$, $Q_{st,f}$ and Q_{st} are expected to equal θ_v, θ_f and θ respectively (Pressoir et al. 2004). By calculating Q_{st} within and between villages it is possible to infer to what extent diversifying selection occurs within and between villages.

Farmer surveys

Farmer surveys were conducted in 2004 as part of a larger project with the aim of quantifying differences in farming practice between highland and lowland communities. Eight highland and nine lowland communities were selected at random. Selected locations were different than those chosen for seed sampling. Twenty farmers per locality were asked questions on seed management and seed history. A shorter version of the same questionnaire was applied to farmers during seed sampling. In each sampled village, two farmers that had provided seed were interviewed. Data on planting dates were reported per week. A conversion to individual planting dates was made by assuming an equal planting probability for each day of the week and assigning a random day of the week to each farmer. This was done to allow the approximation of the number of simultaneously flowering fields.

Remote sensing and meteorological data

Remote sensing data was used to estimate potential pollen flow between fields. From our seed collection sites, four highland and four lowland villages were selected based on the availability of land use data. Ortho-photos, taken at 2-meter resolution between 1994 and 1999, were acquired for these sites. Data on the boundaries of agricultural lands for each location were obtained from Mexican National Statistics and Geography Institute (http://www.inegi.gob.mx/). Ortho-photos were overlain with publicly available multi-spectral Landsat images (glcfapp.umiacs.umd.edu) of the same area. Dates for the multi-spectral images were chosen to fall at the time when maize is not being grown in the studied areas (winter for highlands, spring for lowlands). As a result, empty fields are expected to be visible as grey to purple areas on the image whereas green vegetation is shown as green. We randomly chose a 2.25 $km.^2$ area within the agricultural boundaries of each location for visual inspection of field distributions. For these selected areas, a binary image of inferred fields and vegetation was created using the software MultiSpec© (version 3.0, www.ece.purdue.edu/~biehl/MultiSpec/). Meteorological data was obtained for CIMMYT's highland El Batan and lowland Poza Rica Station for the years 1997-1999.

Simulations

We developed a computer model to simulate the population genetic dynamics of single bi-allelic locus within a metapopulation consisting of n demes (seedlots) under farmer management. Seedlots *i...n* consist of $N_{f(i)}$ ears with N_s seeds per ear, yielding a total of $N_{(i)}$ diploid individuals represented by a single locus genotype. Each seedlot is assigned a position within a square grid of fields and will have up to four neighbor fields (Figure 2).

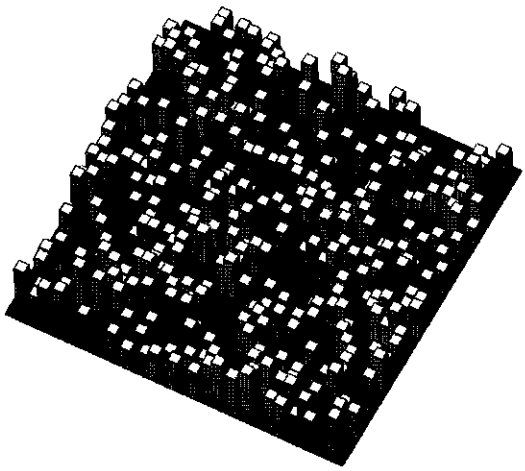


Figure 2. Visualization of population genetic simulation simulation. Bars indicate gene frequencies in individual fields (seedlots). Dark gray squares indicate empty space separating fields.

Reproduction from one generation to the next for each of the n populations is simulated as follows. $N_{f(i)}$ maternal genotypes are sampled without replacement from the population of $N_{(i)}$ genotypes. A new set of $N_{(i)}$ haploid maternal genotypes is generated by sampling N_s times with replacement from each of the $N_{f(i)}$ selected diploid genotypes. A new gene pool of $N_{(i)}$ alleles is generated by combining the $N_{(i)}$ maternal haplotypes with $N_{(i)}$ paternal alleles that are drawn at random with replacement from the population. Pollen flow is incorporated by replacing part of the $N_{(i)}$ resident paternal alleles with a sample of fixed size taken from the neighboring populations. The proportion of migrant pollen was assumed independent of population size.

Seed replacement- and mixing are simulated by replacing all or a proportion of the $N_{f(i)}$ selected ears with migrant ears. Seed for mixing or replacement is sampled from a seedlot selected randomly from any of the seedlots that do not undergo seed replacement in the same generation. We tested the model's accuracy by comparing the effects of drift and gene flow on F_{st} to theoretical predictions.

Simulations were run using parameter values derived from the survey data on seed management and remote sensing data for highlands and lowlands. Simulated villages consisted of 324 farmers. Values for $N_{f(i)}$ were defined by drawing values randomly from a vector of reported values. Frequency of seed replacement and mixing were estimated by the proportion of seedlots that were reported as being replaced or mixed in the previous year. The quantity of mixed seed was taken as the mean reported value (Table 6). The number of neighboring fields was inferred from our remote sensing images. The binary image created from the images was superimposed with a square grid of cells, where the surface area of each cell equaled the average field size in the simulated environment. A cell was assigned as being a field if more than 50% of pixels inside belonged to a field in the binary image (Figure 5).

For both highland and lowland parameters, hundred generations were simulated for seven villages. θ_f was calculated for a thousand random samples of five villages with three seedlots per village sampled from the simulation results.

Results

Evidence for gene flow across environments

Highland and lowland seedlots were clearly differentiated for molecular marker frequencies, phenological traits and ear and kernel traits (Figure 3). The principal coordinate analysis of pairwise genetic distances revealed two distinct groups for highlands and lowlands. Differentiation between environments measured by θ was relatively low however, equaling 0.12 over all markers and 0.07 with the bnlg1784 locus excluded. The separation between environments was thus strongest with locus bnlg1784 included (Figure 3A). As was mentioned, prior information showed bnlg1784 to be informative of altitude. In our sample, highland populations were almost fixed for a single 227 bp allele, whereas lowland populations were polymorphic for this locus but did not present the highland allele. It thus seems that this locus is under strong directional selection related to adaptation to highland conditions. All other loci showed much lower levels of differentiation between environments and were assumed to be neutral. The bnlg1784 locus was excluded from further analysis.

Excluding bnlg1784 did not affect the overall distance between the two environments but samples from HL1, 4 and 5 now showed high similarity to the lowland group, suggesting gene flow from the lowlands (Figure 3B). In line with the molecular data, seedlots from HL1, 4 and 5 proved to be phenologically similar to lowland maize (Figure 3C). Without exception, plants from these localities were tall, late flowering with a high final leaf number. Ear leaf length was the only measured plant trait for which these populations resembled the other highland samples. By contrast, ear and kernel traits showed no evidence of admixture between environments (Figure 3D). All HL populations grouped together and somewhat separate from the HH populations due to smaller kernel width and lower grain weight. In our highland experiment, agronomic performance as judged by total grain weight was similar between HL and HH populations.

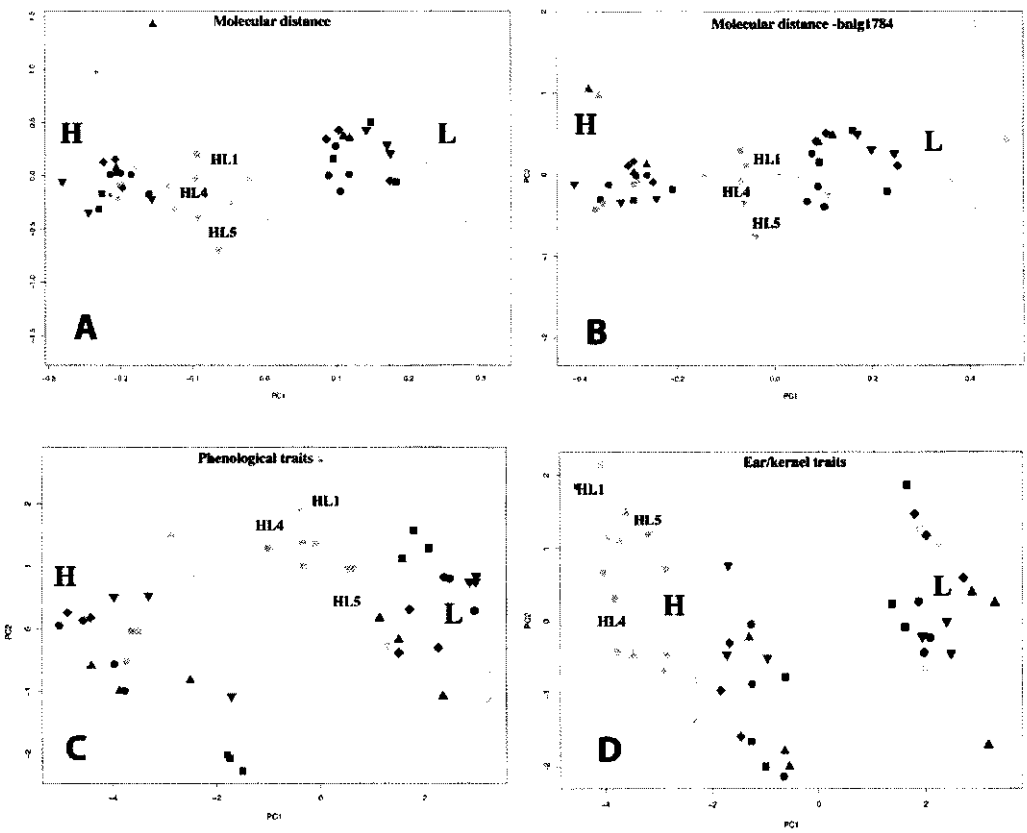


Figure 3. Upper panes show the two first Principal Coordinates (PCo) calculated from matrix of pairwise molecular distances. A: including bnlg1784, B: excluding bnlg1784. Lower panes show the first two standardized principal components (PCA) for C: phenological and D: ear/kernel traits. HH: dark gray, HL: light gray, LH: lightest gray, LL: black. Like symbols of the same grayscale belong to the same locality.

The above results were corroborated by Structure (Pritchard et al. 2000) analysis. Based on the proportion of membership to the inferred highland cluster (PMH), four genetic groups were identified (Figure 4). Group A, with an average PMH of 0.79-0.88, included all HH populations as well as the two northernmost HL sites. Group B, showing values of 0.08-0.16, contained all LL populations and LH5. Group C, comprising the remaining LH sites, had somewhat higher PMH compared to B. LH1, 4 and 5 formed a cluster of medium PMH (0.53-0.59). PMH values for all groups except D were thus in agreement with the environmental origin of the included samples. The intermediate values observed for group D provide additional support for the admixed origin of this group. The relatively elevated PMH in group C suggests that some gene flow from the highlands has occurred.

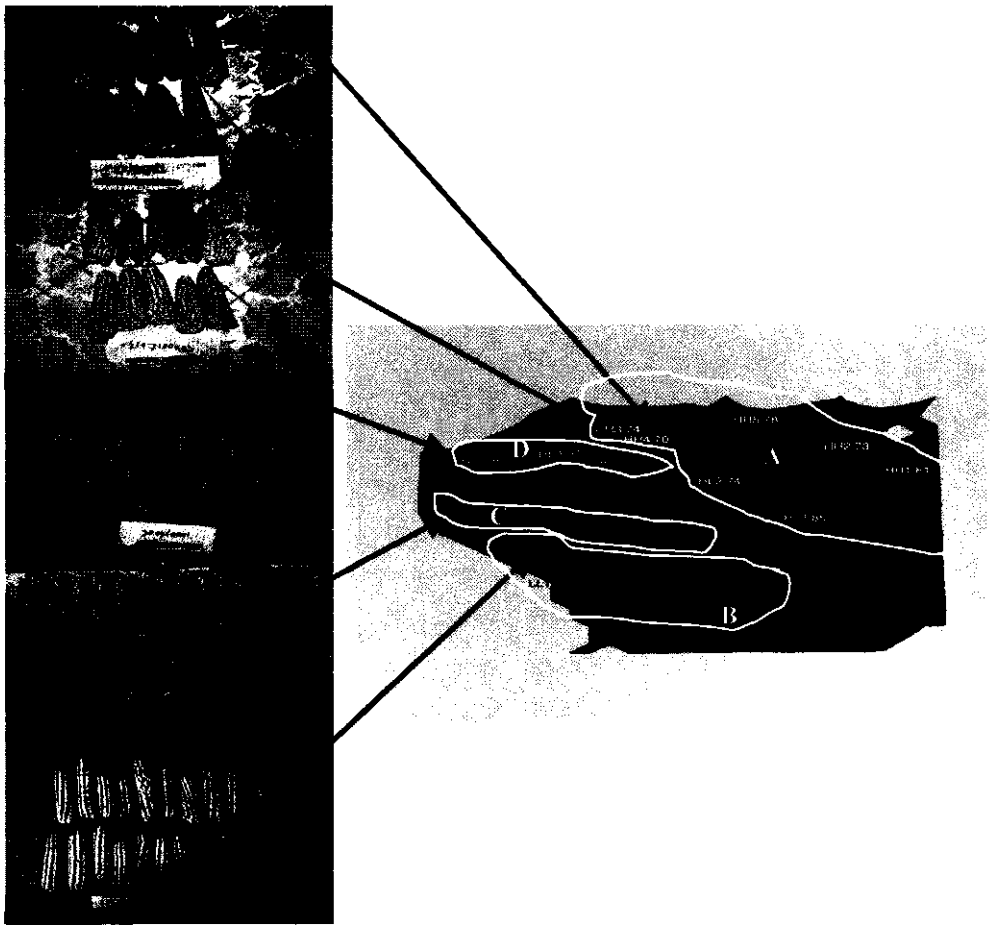


Figure 4. Geographic projection of mean days to anthesis for all 20 localities. White lines delineates groups of localities with different proportions of membership to the inferred highland cluster (PMH).A: 0.79-0.88, B: 0.08-0.16, C:0.2-0.24, D: 0.53-0.59. Groups of ears as observed in the field are shown with black arrows indicating their origin.

Genetic structure within environments

Between- and within village differentiation for molecular markers was low (Table 4), which is in line with previous reports (Pressoir et al. 2004; Perales et al. 2005). Over the entire sample, only 3-9% of genetic diversity (as expressed by θ) was contained between seedlots in a single environment. Differentiation within villages proved to by substantially higher in the lowland environment. As a result, pairwise genetic distances between seedlots from the lowlands are high compared to the average distance for highland populations (Figure 3A,B).

Table 4. Between-village, within-village and total θ observed for Highland (H) and lowland (L) environments. θ_v: beween village differntiation.: θ_f within village differentiation. θ: total differentiation.

Environment	θ_v	θ_f	θ
H	0.026	0.008	0.034
L	0.027	0.064	0.088

Genetic differentiation for quantitative traits was higher than what was observed for molecular markers (Table 5). Q_{st} was larger than θ for all considered traits in the four sampling transects. Differences were mainly found between villages. Within-village values were mostly in the range of observed θ_f. The HL transect showed high $Q_{st,v}$ for phenological traits, reflecting the presence of the admixed populations in HL1, 4 and 5. Phenological $Q_{st,v}$ was lower in the HH and LL transects but generally exceeded θ_v. The only exception was found for plant- and ear height in the LL transect which did not show any differentiation. The latter was related to a lack of correlation between plant height and flowering time as was observed in the highlands (data not shown). This suggests that flowering time rather than plant height that is under diversifying selection.

Kernel and ear traits showed lower levels of differentiation. $Q_{st,v}$ was elevated in HL and LH, possibly due to correlation with phenological traits (e.g. correlation coefficients of 0.7 and 0.56 for population means of **da** and **kd** in HL and LH respectively). High levels of within village differentiation for kernel row number and kernel width were observed in the lowlands, both correlated to differentiation in cob width (data not shown). Interestingly, lowland farmers indeed made mention of varieties with narrow and wide cobs.

Table 5. Between-village ($Q_{st,v}$), within-village ($Q_{st,f}$) and total (Q_{st}) Q_{st} for the four sampling transects. Values are shown for a selection of traits. Ph: phenological traits. Kern.: kernel and ear traits. Values for narrow sense heritability h^2 are given. Values higher than unity are found for days to anthesis as was reported by others (Pressoir et al. 2004). Assortative mating for flowering time was proposed as a possible explanation (Pressoir et al. 2004). Measurements on kernel and ear traits for HH and HL populations were measured in a separate experiment and are marked by an asterisk (*). Highest values are indicated in boldface.

Env.		Trait	h^2	$Q_{st,v}$	$Q_{st,f}$	Q_{st}
HH	Ph.	da	1.06	**0.21**	0.14	0.34
		ph	0.54	**0.23**	0.14	0.35
		eh	0.83	**0.21**	0.15	0.34
	Kern.*	kn	0.38	**0.11**	0.04	0.16
		kd	0.59	**0.03**	**0.03**	0.06
		kw	0.34	**0.09**	0.03	0.12
HL	Ph.	da	1.29	**0.61**	0.08	0.65
		ph	0.49	**0.52**	0.02	0.54
		eh	0.57	**0.61**	0.05	0.64
	Kern*.	kn	0.38	**0.26**	0.05	0.31
		kd	0.62	**0.30**	0.01	0.31
		kw	0.56	**0.40**	0.00	0.40
LH	Ph.	da	1.03	**0.51**	0.04	0.54
		ph	0.81	**0.10**	0.04	0.14
		eh	0.88	**0.15**	0.05	0.21
	Kern.	kn	0.35	0.11	**0.40**	0.48
		kd	0.45	**0.22**	**0.19**	0.39
		kw	0.45	**0.25**	0.01	0.26
LL	Ph.	da	0.86	**0.24**	0.03	0.26
		ph	0.60	0.00	**0.12**	0.12
		eh	0.65	0.00	**0.09**	0.08
	Kern.	kn	0.27	0.00	**0.48**	0.47
		kd	0.39	0.13	**0.19**	0.31
		kw	0.37	**0.04**	0.03	0.07

Potential determinants of genetic structure within environments

The difference in within-village differentiation for molecular markers between highlands and lowlands was unexpected. Assuming neutrality for our markers, the observed discrepancy is caused by differences in the balance between drift and gene flow. We analyzed the available data on seed management for the two environments in order to identify structural differences in agronomical practice that might explain the observed levels of differentiation.

In maize, gene flow occurs by means of pollen migration, seed mixing and seed replacement (Louette et al. 1997). The effect of gene flow on the reduction of drift depends in part on the size of the individual seedlots (see chapter II). Our interview data showed differences in all relevant parameters (Table 6). Most notably, the size of the average seedlot was more than twice as large in the highlands than in the lowland environment. Running our simulation model with similar field distributions revealed that only the difference in population size accounted for a significant part of the difference in differentiation (data not shown). Our model predicted lower structure between highland seedlots due to lower drift and a higher absolute number of pollen migrants. The predicted difference was not as high as observed in our data however, suggesting a possible role for pollen flow.

Table 6. Characteristics of seed management in the two environments.

	Highlands	**Lowlands**
Seedlot size in kg. (trimmed mean)	29.1	12.6
Proportion of seedlots replaced	0.28	0.22
Proportion of seedlots mixed	0.01	0.02
Mix Proportion (trimmed mean)	0.55	0.37

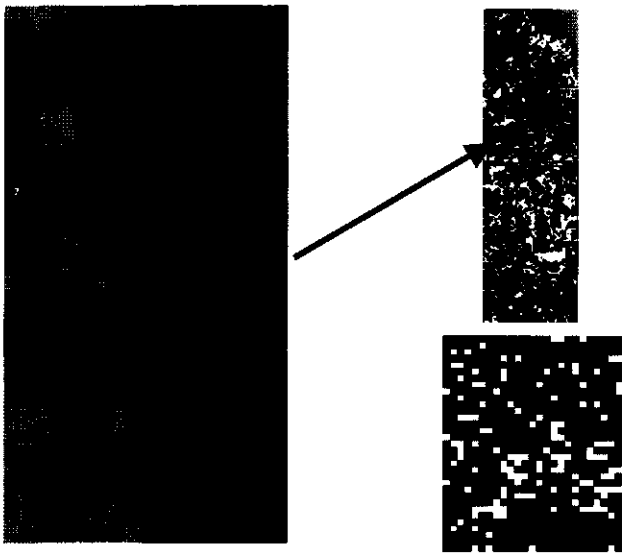


Figure 5. Inference of field layout from remote sensing images. Left part of the figure presents the mosaic of eight orthophotos overlain with multispectral image. H: Highlands, L: Lowlands. Arrow indicates the conversion of the lowland image to a binary image. Lower right part shows the resulting field layout used to model pollen flow in simulations.

Analysis of remote sensing images (Figure 5) revealed that highland planting areas form a densely cultivated area of contiguous fields. Lowland sites in contrast, showed a pattern of dispersed fields separated by large tracts of citrus orchards and pasture. We calculated the mean number of neighboring fields as 4 in the highlands and 1 in the lowlands. We corrected these figures for an estimated probability of 0.22 of flowering overlap between neighboring fields as estimated from the interview data. Including inferred field layout predicted θ_f values that were in the range as those observed for the two environments (Figure 6).

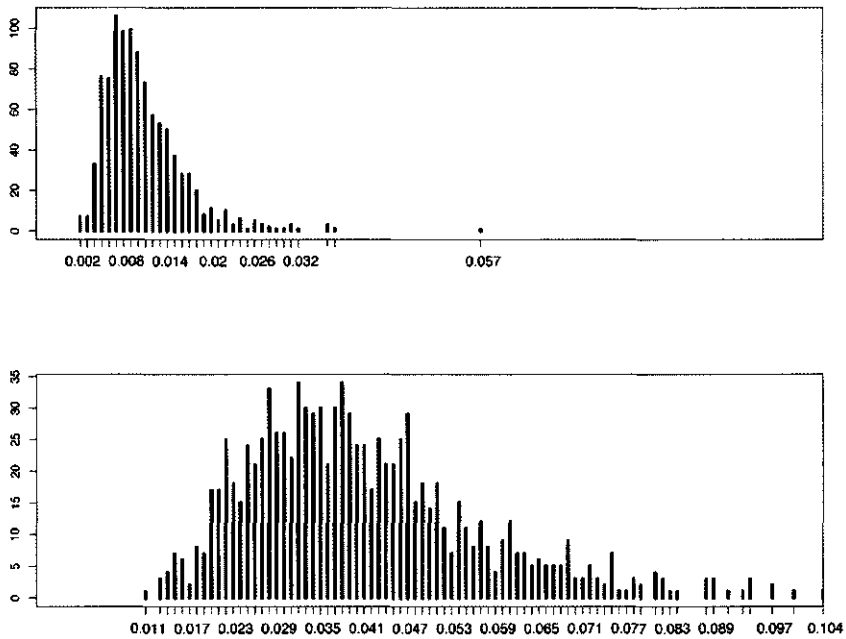


Figure 6. Histogram of model results for highland and lowland molecular within-village structure (θ_f). Horizontal axes shows values of θ_f, while vertical axes show counts of each value in 1000 simulations. Highlands, upper pane. Lowlands lower pane.

Mean simulated θ_f for highlands was 0.011, as compared to the observed value of 0.008. The range of the 95% most common values was 0.0044-0.0197. The average predicted value for the lowlands was 0.041 with a 95% range of 0.02-0.072, compared to a measured value of 0.064. It thus seems that the difference in genetic structure between the two environments can be adequately explained by the levels of pollen flow and drift inferred from local farming practice.

The second salient feature of the observed genetic structure is the high level of between-village differentiation for flowering time and related phenological traits. As was mentioned, the difference in days to anthesis can be up to 30 days between villages within the same environment. Farmer interviews suggested some possible determinants of these differences. As can be seen from Figure 7A, a significant negative relationship exists between mean reported planting dates and mean reported flowering times calculated per village. Villages that plant late in the season reported earlier flowering times than those that plant early. This relation was confirmed by our experimental data. In the highlands, where strong flowering differences were observed, earlier flowering times were found for seedlots that were reportedly planted late (Figure 7B).

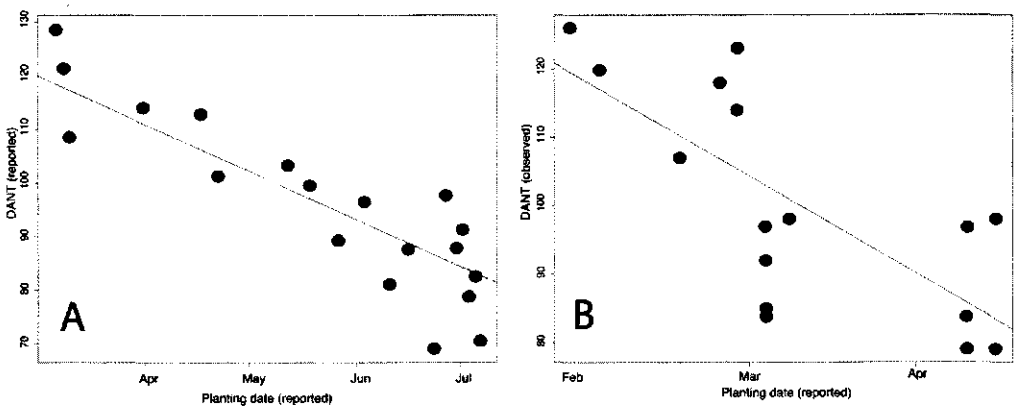


Figure 7. Relation between reported planting date (horizontal axis) and days to anthesis (vertical axis). A: reported flowering dates, B: flowering times observed in field experiment.

A possible explanation for the relation between planting date and flowering time was found by analyzing expected flowering dates in relation to rainfall. We plotted the distribution of flowering dates as estimated from the interviews against average precipitation in the two environments (Figure 8). The highest density of inferred flowering dates coincided with the weeks of maximum rainfall in both environments. The correspondence probably reflects the fact that maize is very sensitive to drought during flowering (Bolanos et al. 1992). Maize that is planted too late in the season will therefore run the risk of suffering severe yield reduction. Different planting dates that may exist for either cultural or environmental reasons could hence impose directional selection on flowering time by favoring plants that flower at the time when the probability of drought is lowest.

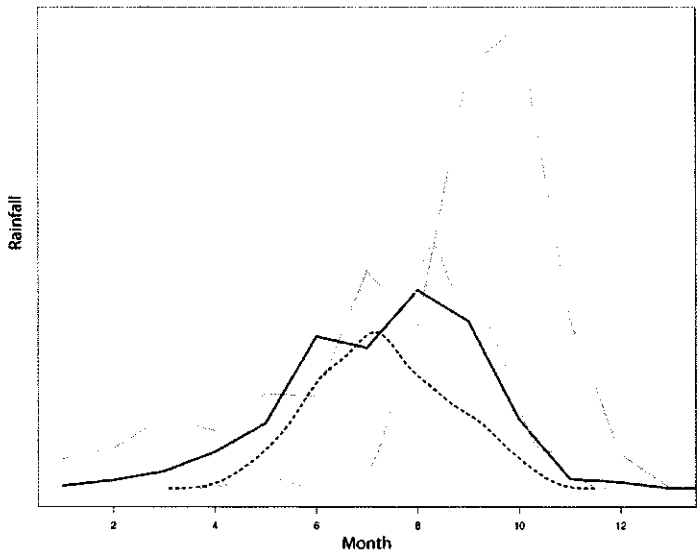


Figure 8. Distribution of rainfall and estimated flowering dates in highland and lowland environments. Solid lines represent rainfall, dashed lines estimated flowering dates. Black = highland, Gray = lowland

Discussion

Genetic structure across environments

We confirmed the importance of altitude in determining genetic structure in maize landraces at a regional scale. Molecular marker data showed that gene flow is restricted between the two environments in spite of their geographic proximity. One locus, bnlg1784, displayed an extreme level of differentiation between the two environments suggesting it is involved in adaptation to altitude. Large morphological differences were observed between the two germplasm types but samples from three highland villages resulted to be phenologically and genetically intermediate. The presence of admixed seedlots in the highland area proves the existence of historical gene flow between the two environments. The direction of gene flow seems to be predominantly from lowlands to highlands. This directionality is not unexpected, given the relative ease with which lowland germplasm can be grown in highland environments compared to the reverse (Jiang et al. 1999).

Our structure analysis revealed higher proportions of membership to the highland cluster in four of the LH localities as compared to the LL sites, suggesting gene flow from highlands to lowlands does take place. The sudden transition from admixed to non-admixed localities within the HL transect suggests that the observed pattern reflects seed migration rather than pollen flow. This hypothesis was supported by the observation of a farmer in HL4 who planted lowland maize that he had received as a gift. Human mediated seed migration from the lowlands is hence the most likely explanation for the observed admixture.

Genetic structure between and within villages

Differentiation for molecular markers was low compared to quantitative trait divergence, as was found by other authors (Sanchez et al. 2000; Pressoir et al. 2004; Perales et al. 2005). Maize is an open pollinated species with high rates of gene flow (Louette et al. 1997) and a large historical population size (EyreWalker et al. 1998). Low levels of neutral genetic differentiation between types of germplasm are therefore expected. Similarly, our comparison between highland and lowland maize revealed relatively modest values of θ. On the other hand, we have shown that under conditions of limited gene flow and small population size, within village differentiation can be relatively high. In a recent study on maize races based on single accessions a mean θ of 0.21 was reported (Reif et al. 2006). At this level of differentiation, structure analysis showed that 86% of individuals were correctly assigned to their respective accessions, which was taken as confirmation of racial identity. For our lowland data we found an average θ of 0.09 in the lowlands, while values close to 0.2 were observed for several pairwise comparisons. This suggests that part of the structure observed by these authors may have represented differentiation at the individual seedlot level rather than racial differences. Additional inbreeding during seed regeneration (Parzies et al. 2000) may add to this effect. We would therefore recommend that genetic studies on different germplasm types include a number of accessions per type, as was the case in several older studies (Doebley et al. 1985; Bretting et al. 1990; Sanchez et al. 2000).

High values of quantitative genetic structure compared to that measured by molecular markers suggests that diversifying selection on agronomical traits is the dominant force in generating genetic differentiation in our study area.

This coincides with previous results for maize in Mexico (Pressoir et al. 2004; Perales et al. 2005) and is in line with the general observation that genetic structure for quantitative traits exceeds differentiation in neutral markers (Merila et al. 2001). In part, high trait differentiation was caused by the presence of admixed populations. The highest values of $Q_{st,v}$ were observed within the HL transect where three of the five villages planted admixed seedlots of intermediate phenology. Traditionally, Q_{st} values that exceed θ_f are interpreted as evidence for diversifying selection (Spitze 1993; Podolsky et al. 1995; Lynch et al. 1999; Merila et al. 2001). Our results showed that high differentiation might result from the inheritance of traits that diverged in response to adaptation to a foreign environment. Apparently, historical gene flow between highlands and lowlands leads to relatively low levels of θ, while strong directional selection in each environment creates high levels of quantitative genetic differentiation. Localized introduction of lowland germplasm into the highlands may hence increase $Q_{st,v}$ much more than θ_v.

This is not to say that the observed phenology in the admixed highland populations is not adaptive. Our data showed that late flowering populations are planted earlier. The success of early planting depends to a large extent on the availability of enough moisture for germination and early plant growth. Wellhausen describes the planting of later, higher yielding varieties in areas with sufficient soil moisture or irrigation in the highlands (Wellhausen et al. 1952). Assuming that a longer growing season constitutes a yield advantage, it may be beneficial to farmers to plant late flowering varieties whenever early planting is possible. As can be seen from Figure 1, April rainfall is highest in the region where the admixed seedlots were sampled. Conditions in the highland change abruptly from moist to dry as one moves further inland, so the appropriate type of germplasm in transition zone may depend on geographical conditions that differ strongly from location to location.

The HH and LL transects showed values of $Q_{st,v}$ that although moderate, were considerably higher than θ_v. As these localities showed no evidence of admixture, local adaptation seems to provide the best explanation for our data. Flowering time is known to respond strongly to directional selection (Paterniani 1969). Selection against late flowering genotypes imposed by yearly autumn drought may therefore lead to shifts in flowering time between populations when different planting dates are applied. Our observation of high $Q_{st,v}$ for phenological traits compared to ear traits differs from results reported by Pressoir and Berthaud for maize in Oaxaca.

Genetic divergence in their study was highest for ear and kernel traits and $Q_{st,v}$ exceeded $Q_{st,f}$. In our case, only lowland maize presented high differentiation in ear/kernel for these traits and differences were observed within rather than between villages. Differentiation in ear morphology was recognized by farmers and it may be farmer preference that is responsible for maintaining this diversity as was suggested by other authors (Louette et al. 2000; Pressoir et al. 2004).

Final remarks

The effect of environment on genetic structure in crop-landraces and their wild ancestors has been well established (Nevo et al. 1979; Doebley et al. 1985; Verhoeven et al. 2004). Only recently has there been recognition of the role of farmers in shaping patterns of genetic diversity. Although previous work has compared genetic structure across regions (Brocke et al. 2003), our study seems the first to address the combined effects of environment and agricultural practice on measures of genetic differentiation. By linking field and genetic data we were able to shed light on some of the probable determinants of genetic differentiation. We have shown that both environmental and human factors need to be considered. Our study demonstrates that through detailed knowledge on local farming practice we may achieve a better understanding of observed patterns of genetic diversity.

Chapter IV

Measuring genetic erosion in modernized smallholder agriculture

A case study on maize in Mexico

Abstract

There has long been concern that traditional landraces of our most important food crops may disappear due to the large-scale adoption of modern varieties. The idea that such replacement will cause a loss of valuable genes and genotypes is known as genetic erosion. Actual proof of genetic erosion for any particular area or crop has rarely been found, in part due to the complex nature of the processes involved. Recent years have seen evidence that instead of disappearing, local germplasm often coexists with improved varieties. Moreover, the composition of the set of local varieties is subject to constant change. In particular, the adoption of modern varieties into the traditional seed supply system may blur the distinction between modern and traditional varieties. The inability to classify germplasm into discrete types makes it hard to measure diversity. We address these problems by means of a case study on modernized smallholder maize agriculture in southern Mexico. Thirty seedlots obtained from both farmers and commercial seed vendors were characterized for agronomical traits and molecular markers. Farmer interviews were used as a tool to distinguish between traditional landraces and recycled modern varieties. Based on this classification we calculated genetic diversity, defined as the mean differentiation between individual seedlots, for different types of germplasm. We showed that modern germplasm is clearly distinct from traditional landraces. Although recycled modern varieties had probably evolved since their adoption, they retained close resemblance to their ancestral stocks. Defining the group containing the highest level of diversity resulted to be complicated because levels of relative diversity were different for different traits. The group of recycled modern varieties presented the lowest diversity for all measured traits. Complete replacement of landraces by these varieties would thus reduce diversity in the traditional seed system. Under current patterns of coexistence however, the distinctness of modern and traditional varieties limit the reduction of genetic diversity.

Introduction

Since the advent of modern plant breeding, there has been concern that the substitution of improved germplasm for traditional crop varieties reduces genetic diversity (Harlan et al. 1936; Harlan 1975). This process is commonly referred to as genetic erosion (Frankel et al. 1970) and has been defined as: "the loss of genetic diversity, in a particular location and over a particular period of time, including the loss of individual genes, and the loss of particular combinations of genes such as those manifested in landraces or varieties" (FAO/IPGRI 2002). This broad description hides a complex phenomenon that is hard to measure in practice (Brush 1999).

First, most studies on genetic erosion part from the traditional assumption (Frankel et al. 1970) that the introduction of improved seed invariably causes the disappearance of local varieties (Hawkes 1983; Brush 1999). It has become clear however, that farmers continue to plant their own seed in many areas where improved germplasm has been introduced (Bellon 1996). The modern, formal seed system thus often coexists with the traditional, informal seed system based on seed recycling and exchange (Almekinders et al. 1994). Persistence of the informal seed system is no guarantee for the conservation of traditional varieties however. Improved varieties are known to be adopted into the informal system, a process that is known as creolization (Almekinders et al. 1994; Bellon et al. 2001). Traditional landraces may thus be replaced indirectly by creolized varieties, which to complicate matters further, are often managed under local names. Evaluation of genetic erosion in areas where the formal and informal seed system coexists thus requires proper identification of creolized and traditional seed.

Second, there is the challenge of measuring diversity. The following quantities are often suggested: (1) numbers of different types or richness (2) evenness of distribution of these types and (3) the extent of the difference between types (FAO/IPGRI 2002). All three of these measures require the definition of separate types. Modern varieties can be classified into distinct types, since the formal system supplies certified seed of known identity. Seed from the informal system however, is often of unknown origin and characteristics and is not easily grouped into types (Cromwell 1990; Almekinders et al. 1994; Louette et al. 1997). Richness, evenness and difference are therefore hard to measure.

Third, while it is usually assumed that the loss of traditional landraces implies a reduction in genetic diversity, this is not necessarily the case. Local diversity will only decrease if improved varieties are less diverse than the traditional varieties that are being replaced. Although recent studies have reported lower levels of genetic diversity in modern materials than in landraces (Reif et al. 2005; Reif et al. 2005; Huang et al. 2007), these results cannot be generalized to any specific region or crop. Coexistence of modern and traditional varieties could even increase diversity if new germplasm offers a set of traits that are not present in the traditional landraces (Wood et al. 1997; Louette et al. 2000). Also, modern varieties adopted into the informal seed system will undergo differentiation from their parental stock by local gene flow and local selection (Pressoir et al. 2004; Perales et al. 2005) which would lead to the generation of new diversity.

This paper presents a case study on genetic erosion in maize agriculture in Mexico. Production in most parts of the country is dominated by smallholder agriculture that relies mainly on traditional landraces. Our study was performed in La Frailesca in Southern Chiapas. Although largely dominated by smallholder agriculture, this region has seen a strong increase in the use of formal seed (Bellon et al. 1994). The informal seed sector does persist in this region but replacement of landraces by creolized varieties has occurred (Bellon et al. 2001). This particular situation provides an excellent opportunity to test the hypotheses of local genetic erosion as a result of agricultural modernization.

We will use information obtained by local farmers and seed companies to make an a-priory distinction between creolized and traditional varieties within the informal seed system. We address the problem of defining discrete types in the informal sector by treating each seedlot as a different type. Seedlots may be defined as the basic entity of farmer seed management (Louette et al. 2000). We measure diversity as the molecular or phenotypical distance between different seedlots. Average values of between seedlot differentiation may thus be calculated for different classes of seedlots such as formal vs. informal seed.

Although work on genetic diversity in landraces and modern germplasm exists (Reif et al. 2005; Huang et al. 2007), this study is unique in that it compares diversity between traditional and modern varieties at the local level, including local varieties that have been derived from commercial germplasm. We test if replacement of traditional maize varieties by modern germplasm has a negative impact on biological diversity.

We present a general approach that combines biological data with information on seed history and local abundance to arrive at a description of diversity within different classes of seed. Based on careful classification of local seed types, we compare seedlots of locally available landraces, creolized and modern varieties for differentiation in agronomical and morphological traits as well as in allelic frequencies for molecular markers. We address the hypothesis that commercial varieties are distinct but less diverse compared to traditional landraces currently present in the area under study. In addition, we investigate if creolized varieties have maintained their original characteristics or have become altered over time. Based on our results, we evaluate the potential consequences of increased adoption of improved maize varieties, both directly and through creolization, on different measures of biological diversity.

Materials and methods

Seedlot sampling

In March 2006 we conducted a two-week field visit to the La Frailesca region in southern Chiapas, Mexico. The area lies at an average altitude of 600m and comprises several municipalities that lie south of the state capital Tuxtla Guttierrez. A total of 30 seedlots were collected from farmers, local resellers and CIMMYT's gene-bank (Table 1). Seed from the informal system was obtained from 16 farmers in six communities (Figure 1). Forty ears of each seed lot were sampled. Information on local nomenclature, seed history and planted area was attained by semi-structured interviews with the farmers that provided seed. A 2005 survey on local planting materials was combined with data on 2003 sales volumes of maize seed volunteered by local seed companies to estimate the regional frequency of each collected seed type (Table 1).

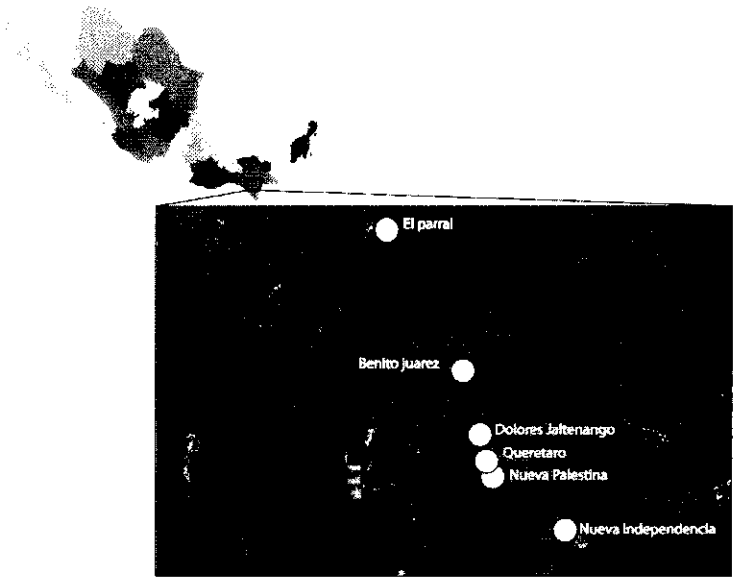


Figure 1. Sampling area in La Frailesca Chiapas. Names of the sampled villages are marked in the magnified area.

Table 1. Sampled seedlots. Local names are given as well as the classification made by farmers in terms of their seedlot being an improved variety or a landrace. Freq 1., Freq 2., Freq 3., correspond to weights used to calculate mean weighted diversity for grouping I, II and III respectively.

	Seed-lot	Class	Local name	Farmer class.	Race	Origin	Freq. 1	Freq. 2	Freq. 3
formal	OV1	OPV	V424	improved	Tuxpeño	Buena Vista A.C.	0.14	0.033	0.033
	OV2	OPV	V424	improved	Tuxpeño	CIMMYT	-	-	-
	OV3	OPV	V524	improved	Tuxpeño	CIMMYT	-	-	-
	OV4	OPV	V526	improved	Tuxpeño	PROASE	0.40	0.095	0.095
	OV5	OPV	V534	improved	Tuxpeño	PROASE	0.46	0.110	0.110
	HB1	Hybrid	Nutria	improved	Tuxpeño	ASGROW	0.267	0.203	0.203
	HB2	Hybrid	S-3G	improved	Tuxpeño	Cristiani Burkard	0.004	0.002	0.002
	HB3	Hybrid	S-5G	improved	Tuxpeño	Cristiani Burkard	0.069	0.053	0.053
	HB4	Hybrid	Z-30	improved	Tuxpeño	Hartz	0.083	0.063	0.063
	HB5	Hybrid	Z-31	improved	Tuxpeño	Hartz	0.021	0.016	0.016
	HB6	Hybrid	3086	improved	Tuxpeño	Pioneer	0.140	0.106	0.106
	HB7	Hybrid	30F94	improved	Tuxpeño	Pioneer	0.420	0.318	0.318
informal	CC1	Landrace	*Conejito*	Landrace	Zapalote chico	El Parral	-	0.11	0.03
	CO1	Landrace	*Olotillo*	Landrace	Olotillo	Dolores Jaltenango	0.200	0.11	0.03
	CO2	Landrace	*Olotillo*	Landrace	Olotillo	El Parral	0.200	0.11	0.03
	CO3	Landrace	*Olotillo*	Landrace	Olotillo	Guadalupe Victoria	0.200	0.11	0.03
	CO4	Landrace	*Olotillo*	Landrace	Olotillo	Nueva Palestina	0.200	0.11	0.03
	CO5	Landrace	*Olotillo*	Landrace	Olotillo	Nueva Palestina	0.200	0.11	0.03
	CT1	Landrace	*Jarocho*	Landrace	Tuxpeño	El Parral	0.330	0.11	0.03
	CT2	Landrace	*Jarocho*	Landrace	Tuxpeño	N. Independencia	0.330	0.11	0.03
	CT3	Landrace	*Jarocho*	Landrace	Tuxpeño	Nueva Palestina	0.330	0.11	0.03
	CT4	Landrace	*Jarocho*	Landrace	Tuxpeño	Queretaro	-	0.11	0.03
	RV1	Creolized	V424	Landrace	Tuxpeño	Benito Juarez	0.125	0.125	0.09
	RV2	Creolized	*Precoz*	Landrace	Tuxpeño	Dolores Jaltenango	0.125	0.125	0.09
	RV3	Creolized	*Tuxpeño precoz*	Landrace	Tuxpeño	Dolores Jaltenango	0.125	0.125	0.09
	RV4	Creolized	San Gregoreo	Landrace	Tuxpeño	Dolores Jaltenango	0.125	0.125	0.09
	RV5	Creolized	*Pronase*	Landrace	Tuxpeño	N. Independencia	0.125	0.125	0.09
	RV6	Creolized	*Pronase*	Landrace	Tuxpeño	Queretaro	0.125	0.125	0.09
	RV7	Creolized	*Tuxpeño*	Landrace	Tuxpeño	Queretaro	0.125	0.125	0.09
	RV8	Creolized	*Sardina*	Landrace	Tuxpeño	Queretaro	0.125	0.125	0.09

Field experiment

A field experiment was planted at CIMMYT's Tlaltizapan field station in May 2006. The experiment consisted of five adjacent replicate blocks containing all thirty seedlots. In order to minimize effects of competition between different seed types, seedlots were grouped into the different germplasm classes. Four groups were distinguished: landrace, creolized, hybrid and OPV. Groups were randomly assigned within blocks and populations were randomized within groups. Each plot consisted of 50 plants of a single population planted at 20 cm intervals in two 5-meter rows. The first two plants in each row were discarded. The field was irrigated and fertilized throughout the experiment according to CIMMYT standard protocols. Vegetative traits were measured after flowering. Tassels were harvested and stored in a cold room before measurement. Measured traits are given in Table 2.

Analysis of phenotypic data

Data from the field experiment was aggregated at the seedlot level by fitting a mixed model and calculating Best Linear Unbiased Estimates (BLUE). The information for each phenotypic trait was thus summarized in a vector of seedlot means. The mixed model for individual phenotypic traits included a fixed seedlot effect besides random terms for block, row and column. The row and column random effects were included to correct for local fertility trends in the field.

For further statistical analysis, the seedlot BLUEs for each trait i, x_{ij}, were standardized by subtracting the mean for the trait, after which a division followed by the standard error of the seedlot mean. $x_{ij}' = \frac{x_{ij} - \bar{x}_i}{SE_i}$. This scaling of the farmer population BLUEs serves to assess the discriminatory power of individual traits from a biplot. Longer representations of variables in the biplot indicate stronger discrimination between farmer populations. Tests for differentiation of specific populations for each trait were performed by fitting the above mixed model comparing to reduced datasets containing the populations of interest. Significance was tested by comparing log-likelihood values between the full mixed model and a reduced model without the population term.

Table 2. Traits measured in the field experiment and their units of measurement. Traits marked with an asterisk are derived measures.

	Trait	Unit	Code
Plant/Tassel	Days to anthesis	Days	DA
	Days to silking	Days	DS
	*Anthesis silking interval	DS-DA	ASI
	Plant height	5cm	PH
	Ear height	5cm	EH
	Leaf number above the ear	#	LN
	Width of ear leaf	mm	LW
	Length of ear leaf	mm	LL
	Stem Diameter	mm	SD
	Tassel branch number	#	TN
	Tassel length	#	TL
	Primary tassel branch length	cm	PL
	Secondary tassel branch length	cm	SL
	Tassel fresh weight	0.1g	TW
Ear	Ear diameter	Mm	ED
	Kernel row number	#	KN
	Kernel thickness	mm	KT
	Ear weight	g	EW
	Kernel weight	0.01g	KW
	Cob length	mm	CL
	Cob diameter	mm	CD
	Cob weight	0.01g	CW
	*Total grain weight	EW-CW	TGW
	*Kernel length	ED-CD	EKL
	*Relative ear diameter	ED/CL	RED
	*Estimated kernel width	π(CD+ED)/2KN	EKI

SSR genotyping

For all sampled populations, seed from 30-35 individuals was germinated under greenhouse conditions. Hybrid seed was considered to be of a single genotype and only a single seed was planted for each seedlot. DNA was extracted using CIMMYT's standard protocol from ground, lyophilized Lyophilized leaf tissue. Ten easily scorable SSR loci were selected from previous studies (Matsuoka et al. 2002; Warburton et al. 2002) based on BIN number and product size, in order to achieve the highest possible coverage while allowing for multiplexing of individual PCR products. Fluorescently labeled primers (Applied Biosystems, Sigma-Aldrich) were ordered for the following markers: phi034, phi093, phi061, phi014, umc1061, phi227562, phi96100, bnlg1784, phi029 and bnlg2047 (Maize GDB, http://www.agron.missouri.edu/ssr.html). Reaction conditions were described in chapter III. After PCR, 1.5µl of pooled product was denatured in 9 µl of HiDi formamide containing 1µl of ROX500 (Applied Biosystems) size standard. Samples were analyzed on an ABI 3100 capillary sequencer (Applied Biosystems). Fragment sizes were scored using Genotyper 2.1 (Perkin Elmer/Applied Biosystems) software.

Phenotypic distance

In order to express between-population differences in ear and plant traits we used the Gower distance (Gower 1971) as a measure of pairwise phenotypic distance:

$$d_{ij} = n^{-1} \sum_{k=1}^{k=n} \frac{\left| x_{ik} - x_{jk} \right|}{R_k}$$

Where x_{ik} and x_{jk} are the values for trait k in seedlot i and j respectively, n is the number of measured traits and R_k is the range of trait k among all seedlots.

Molecular distance

Pairwise molecular distance between seedlots i and j was defined as $d_{ij} = -\ln(1-\theta)$, where θ is Weir and Cockerham's measure of coancestry (Weir et al. 1984). θ is estimated as the between population to total genetic variance (Weir et al. 1984).

This is an efficient measure of genetic divergence as long as drift is the sole evolutionary factor involved (Reynolds et al. 1983). The latter assumption seems appropriate given the relatively short time of genetic isolation expected for out-breeding populations from the same region. The use of θ instead of the standard Reynolds distance is preferable in this case, as deviations from HW have been reported in maize (Pressoir et al. 2004; Reif et al. 2006). All genetic analyses were done using the program MSA (Dieringer et al. 2003).

Dendrograms

Phenotypic and genetic relationships between the thirty seedlots were studied by cluster analysis. Neighbor-joining trees (Saitou et al. 1987) were constructed based on pairwise distances for ear, plant and molecular data. This method was chosen because it is expected to perform better than other methods when rates of evolutionary change differ between populations (Saitou et al. 1987).

Analysis of within-class diversity

Mean phenotypic and molecular diversity was evaluated using an approach similar to the one presented by Cox in a 1986 study on wheat diversity in the United States (Cox et al. 1986). The idea is to give a weighted mean of between seedlot distance that is representative of a set of seedlots sampled at random from farmer's fields. Values were calculated for each of a set of seedlot classes that will be specified in the results section. Mean diversity D_g within n seedlots is given by:

$$D_g = \sum_{i}^{n} \sum_{j}^{n} p_i p_j d_{ij}$$

Where d_{ij} is the difference between seedlots i and j, and p_i and p_j are their within-class proportions expected in a random sample (i.e. Freqs. 1, 2 and 3 in Table 1). The sampling distribution of D_g was estimated using a bootstrap approach. Data was resampled 5,000 times. For phenotypic data, 5 experimental replicates and 48 plants for each seedlot within a replicate were sampled with replacement. For molecular data, ten loci were sampled. D_g was calculated for each iteration. We applied a bias correction to account for the fact that distances between repeated samples of identical seedlots are expected to be higher than zero.

The minimum difference for phenotypic and molecular data is inflated by experimental and sampling error. We therefore adjusted D_g by setting $d_{ij}, i = j$ equivalent to the mean difference between bootstrap iterations for that population.

Results

Description of biological material

All thirty collected seedlots, together with their estimated frequency are listed in Table 1. The composition of the seed supply in La Frailesca (Figure 2) can be summarized as follows: forty-eight percent of all planted seedlots were commercial hybrids; fifteen percent were open pollinated varieties; twenty-seven percent were creolized seed; and ten percent were traditional landraces.

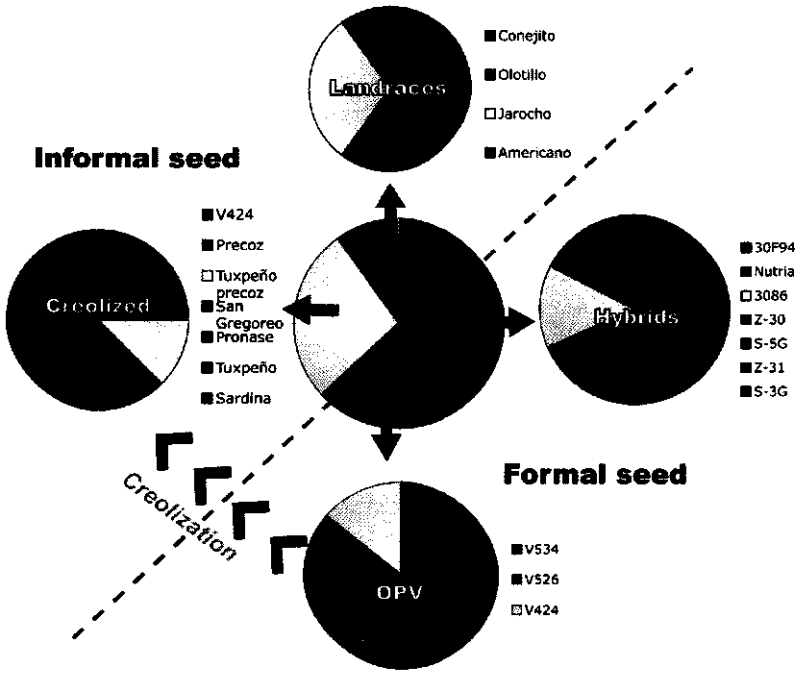


Figure 2. Schematic representation of the composition of the seed supply in la frailesca in terms of variety names. The circle in the center shows the proportions represented by the different seed types in the total seed supply.

Formal system

Certified seed falls into hybrids and open-pollinated varieties (OPVs). Hybrid maize is relatively new in La Frailesca, the first types dating backs to the late nineties. Four seed companies dominate hybrid seed sales in the region (Table 1.). We obtained one kilogram of seed for seven of the most popular hybrid types. Open pollinated varieties have a much longer history in the region. They mainly originate from public breeding programs and have been released in central Chiapas since the late 1970s by the National Agricultural Research Service (INIFAP). Four of the most common varieties were obtained. V-524 is a variety created by CIMMYT and released by INIFAP in 1975. It was very popular until it was removed from the market in 2001. Previous studies suggest that many creolized populations are derived from this variety (Bellon et al. 2001). V-424 is a variety selected for earliness by CIMMYT from the same population and released in 1981 by INIFAP. Of this variety, both a sample of commercially sold seed as well as seed from CIMMYT's gene bank were included in this study. The varieties V-526 and V-534 were released by INIFAP in 1982 and 1989 respectively.

Informal system

As mentioned, informal seed can be classified into traditional landraces and creolized varieties. A seedlot was considered a traditional landrace if none of the interviewed farmers remembered it as being either foreign or derived from certified seed. Seed was assumed to be creolized if a farmer reported having obtained it as certified seed. In case of doubt seed was not included in the sample.

Three named landrace varieties were collected (Figure 3). The two main varieties *Olotillo* and *Jarocho* belong to two distinct races, Olotillo and Tuxpeño respectively (Wellhausen et al. 1952). The name *Jarocho* suggests that it may be an introduced variety since the same word is commonly used to indicate the inhabitants of the neighboring state of Veracruz. It has long history in the area however and there was no evidence of it being a creolized variety. The name *Olotillo*, meaning thin cob, refers to the most obvious trait that distinguishes this race from most other races that have been described (Wellhausen et al. 1952). A single seedlot called *Conejo* was collected: based on its ear and plant traits, it probably belongs to the bolita race.

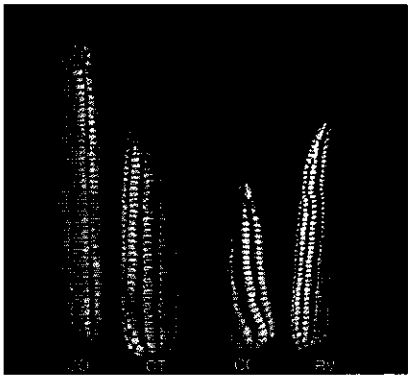


Figure 3. Maize types in informal seed system (CO: *Olotillo* landrace, CT: *Jarocho* landrace, CC: *Conejito* landrace; and RV: Creolized variety

Creolized varieties had the most diverse nomenclature, reflecting their introduction history or population of origin. The varieties named V424, *Tuxpeño precoz* and *Precoz*, were most probably derived from V-424 which was sold under the popular name *Tuxpeño precoz*. Populations called *Tuxpeño* are likely to originate from V-524 and to a lesser extent from V-526 both of which went under this name. *Pronase* refers to the former state-owned seed company *PRONASE* and probably reflects the name mentioned on the distributed seed bags. This company sold different types of OPVs, and, hence, the name *Pronase* sheds little light on the seed's identity. Similarly, the name San Gregoreo could be traced to the label on the seed bags that were distributed by the government in 1989. The name probably refers to an irrigation district that produced a number of different OPVs. The variety *Sardina*, allegedly owes its name to the man who introduced it in the early eighties and had praised it because its ears produced grain like "sardines in a can". Although this variety was rumored to derive from V-424 the first owner, who we were able to locate, claimed that he never actually knew the seed's identity when he bought it from a local store. Table 3 shows data on mean seedlot size and antiquity of the different types of seed. Most varieties are planted in quantities of around 30 kg (equivalent to 1.5 hectares) regardless if it is from the formal or informal sector. *Olotillo* and *Conejo* are planted in very small volumes however and are apparently used for special purposes.

Table 3. Areas planted per farmer and years without seed replacement for different seed types

Maize type	n	Mean kg. planted	SD	Years	SD
CC	1	1.0	-	-	-
CO	5	6.4	7.9	24	28.5
CT	4	26.7	11.5	39	32.1
RV	8	28.4	15.0	10.5	4.6
OV	5	30.3	16.0	1	-
HB	6	29.2	23.9	1	-

Phenotypic description

Figure 4 presents the Principal component analysis of all measured traits for the different seedlots. The trait values having the highest correlation with the first and second principal components are shown. The different populations separated into five groups consisting of the three landraces, a group of improved varieties including all hybrids, creolized varieties and most OPVs and an outlier pair formed by the two populations of V-424. Both *Conejo* and V-424 are early maturing varieties with a short plant height and small, sturdy ears. *Conejo* is clearly different from V424 however, with a taller plant, narrower leaf and stem, shorter ears with fewer kernel rows and less grain. The two main landraces, *Olotillo* and *Jarocho*, were separated from the other types by tallness and lateness, longer tassel branches as well as a lower number of kernel rows with slightly wider kernels and longer slimmer ears. Especially compared to hybrid maize these two types had relatively narrow ear-leaves and lower ear and grain weight. They differed mainly in cob diameter and weight, relative ear diameter and lesser extent by ear/grain weight. Total grain weight differed quite substantially between the different seed classes under experimental conditions (Figure 5). Hybrids have the highest grain weight per ear as expected. With the exception of the early maturing V-424 (OV1/2) and the creolized variety RV6, all improved and creolized material had a higher individual grain yield than the traditional landrace populations. Their total grain weight was generally equivalent to that of the classic V-524 variety (OV3, indicated with an asterisk).

There was significant differentiation within the three main groups as well as within the set of creolized varieties for most traits (data not shown). The creolized varieties RV3 and RV6 fell in between OV1-2 and the taller later open pollinated varieties (Figure 4). When tested against the most probable OPV ancestors, OV3 and OV1, significant differences were found for several traits (data not shown). This indicates that RV3 and RV6 have become differentiated from their original source populations. Surprisingly, one population of the *Jarocho* landrace grouped together with the improved and creolized varieties instead of within the other *Jarocho* populations. It is relatively early flowering and short and has 14 kernel rows instead of the typical 10-12. This points to this population being a creolized variety in spite of its name.

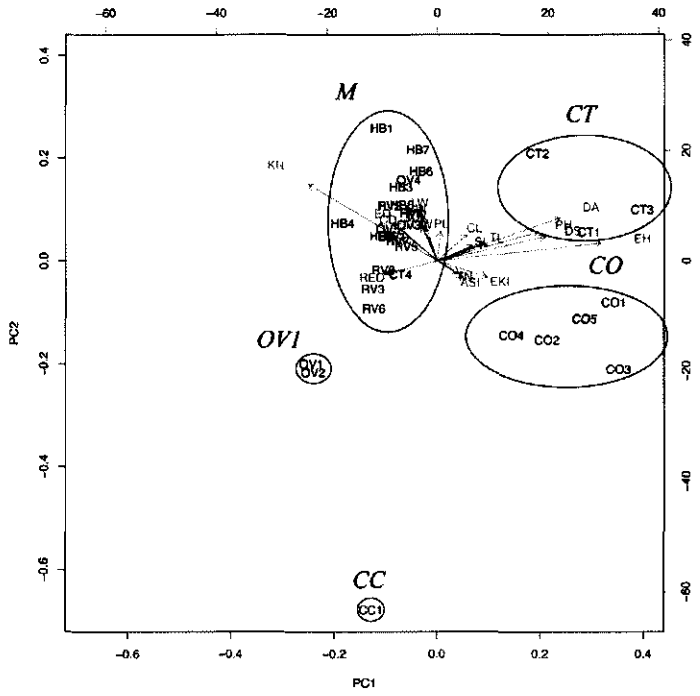


Figure 4. Biplot showing the scores of the two first principal components. Length of arrows are proportional to the discriminative value of the different traits. Circles indicate different types of germplasm. M: modern and creolized varieties, OV1: V424, CT: *Jarocho* landrace, CO: *Olotillo* landrace, CC: *Conejo* landrace.

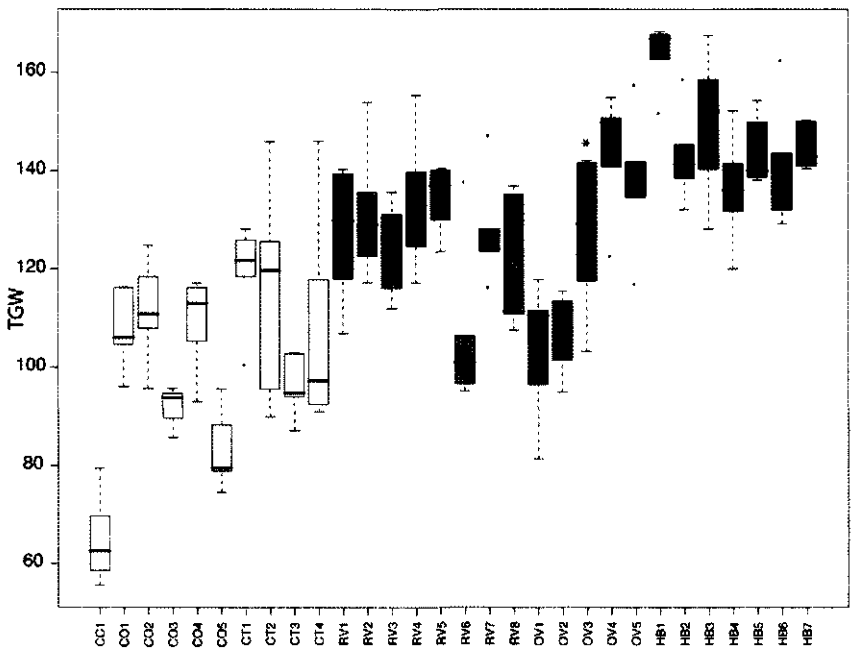


Figure 5. Total grain weight (g.) for all seedlots. Landraces are given in white, Creolized varieties in light grey, OPVs in dark grey and hybrids in black. The classic OPV, V-524 (OV3) is indicated by an asterisk.

Phenotypic relationships

The dendrogram based on pairwise distances for ear traits (Figure 6) confirmed the clear separation between landraces and improved and creolized varieties. Most hybrids, OPVs, and creolized varieties formed a tight cluster. A sub-cluster could be distinguished that comprised the seedlots with the highest grain yield, namely HB1,2,3,5 and 7 as well as OV4. The two populations of V-424 (OV1 and OV2) formed a separate cluster due to shorter cobs and lower grain weight. The *Jarocho* CT4 grouped closely with the latter cluster, just as RV3, RV6 and RV8. This again suggests a close relation of CT4 to improved maize. Among the landraces, all *Olotillos* grouped together, thereby confirming their racial identity. Distances within this cluster were quite large however, reflecting considerable variability in ear traits. Similar heterogeneity was present among the *Jarocho* populations.

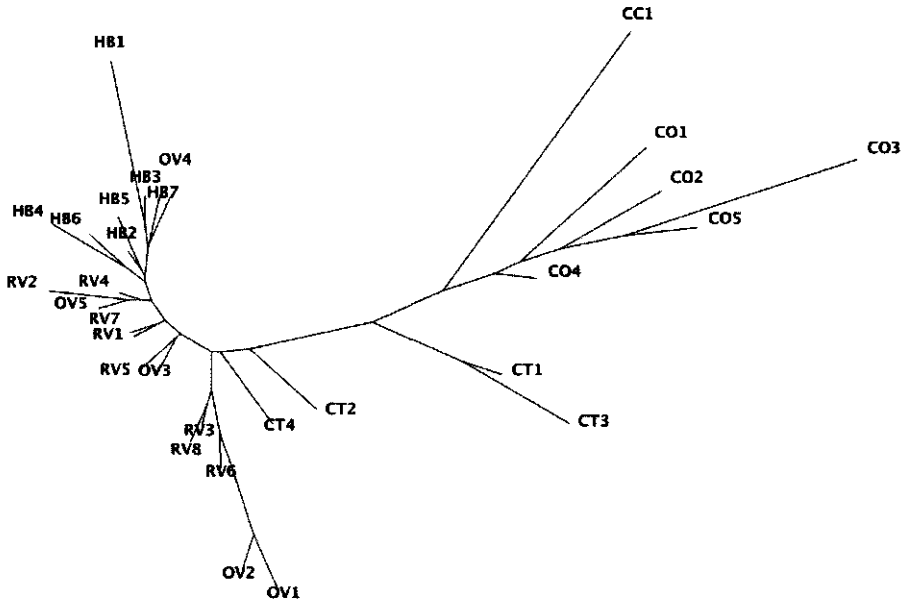


Figure 6. Dendrogram based on ear traits (Gower distance).

Distances calculated from plant traits also revealed a distinction between improved and traditional varieties, with the exception of *Conejo* (Figure 7). Contrary to the results based on ear traits, there was no clear separation between *Olotillo* and *Jarocho* populations. Both types appeared in the same cluster. Moreover, the distances between the seedlots of the two main landraces turned out to be quite low. There seemed to be relatively little differentiation for vegetative and tassel traits within the two landraces. In contrast, hybrids harbored a large amount of diversity, with long branches separating the different types. Variability was notably less between OPVs and between creolized populations. Most of these populations formed a poorly differentiated group that again included CT4. Only V-424 was clearly different and clustered together with *Conejo* due to its earliness and low plant height.

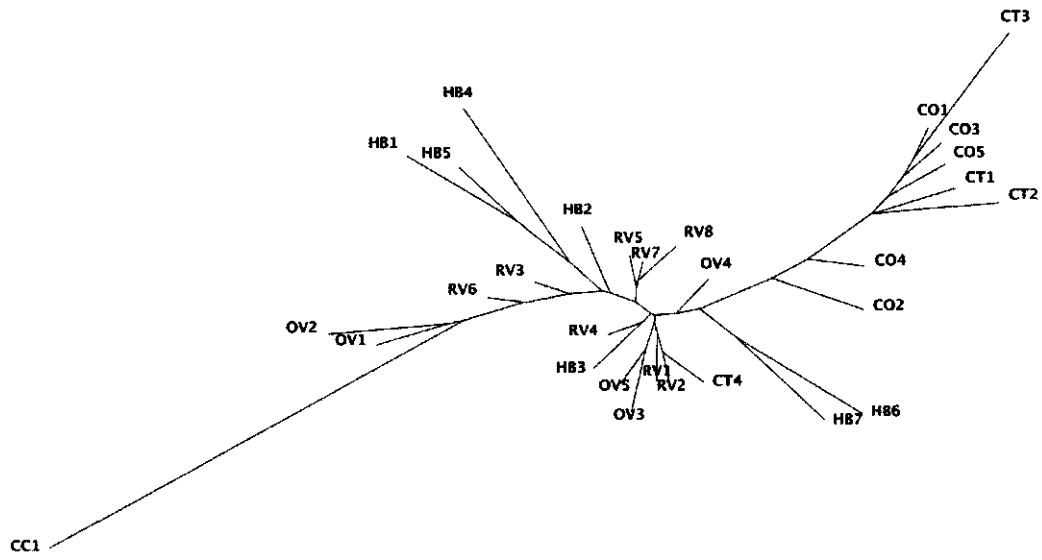


Figure 7. Dendrogram based on vegetative and flowering traits (Gower distance).

Molecular relationships

Genetic distance between the different hybrid types proved to be superior to the distance between different landrace, creolized and open pollinated varieties (Figure 8). This was not unexpected given the fact that hybrids are produced by crossing two inbred lines. The relationships between the remaining populations was thus best appreciated by excluding hybrids from the analysis (Figure 9). OPVs and creolized varieties fell in a separate cluster from the landrace populations. Also here, the identity of CT4 as a creolized variety was confirmed by it falling within the cluster of creolized populations and OPVs.

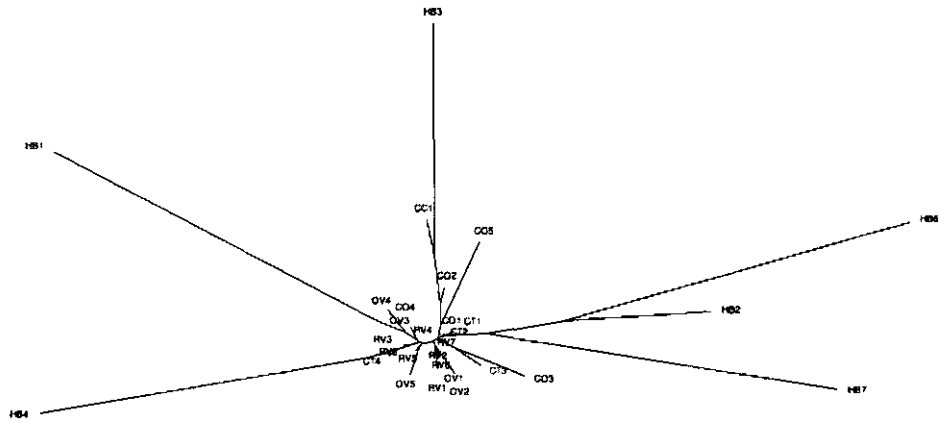


Figure 8 Dendrogram based on molecular distance, hybrids included.

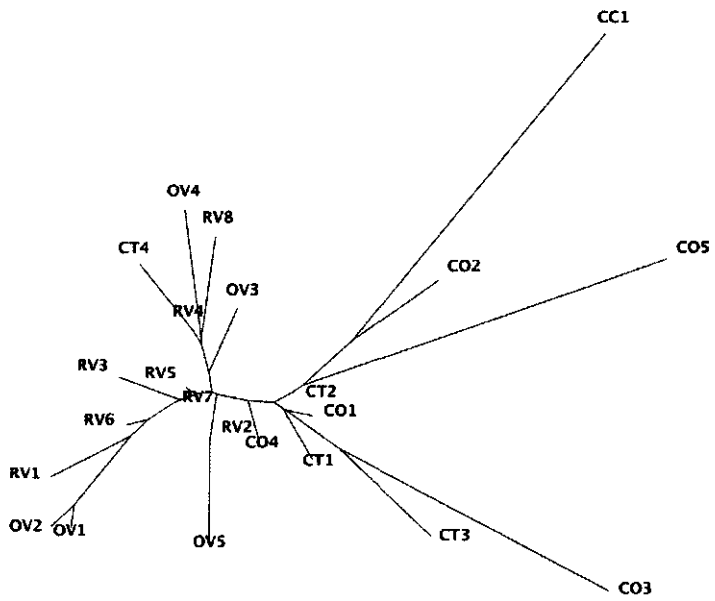


Figure 9. Dendrogram based on molecular distance, hybrids excluded.

It should be noted that the actual genetic difference separating the two clusters was very small, especially compared to the branch lengths found within each cluster. Pairwise θ between the OPVs and the *Olotillo* group was only 0.027 for example. Genetic differentiation between *Olotillo* and *Jarocho* was also surprisingly low. The two races did not cluster separately in the dendrogram. Pairwise θ was in fact only 0.015, which is much lower than the differentiation found between populations within seed groups. The lowest differentiation was found within the OPVs, creolized varieties and *Jarocho* landrace, which showed an average θ of 0.05. Differentiation within the *Olotillo* group was higher, with a θ of 0.092. The single population of *Conejo* had a pairwise θ of 0.09 with respect to the other landrace groups. The relatively high level of differentiation of these landrace populations was probably due to relatively strong drift caused by small population sizes (Table 3).

Diversity within seed types

Seedlots were assigned to different classes based on three types of groupings that reflect different subdivisions of the formal and informal seed systems. Seedlots were subsequently assigned weights according to their expected proportion in farmers' fields in La Frailesca. Within-class proportions for each seedlot (Table 1) were estimated by using information obtained from previous farmer interviews (Bellon et al. 2005) and sales information provided by seed companies (Flores et al. 2004). Seedlots from the informal system were treated as if they represented different seed types with a proportion equal to their share in the seed sample.

Weighted within class diversity was calculated for the following groupings:

(I.) *Jarocho* landrace (CT), *Olotillo* landrace (CO), creolized populations (RV), open pollinated varieties (OV) and hybrids (HB).
(II.) All landrace varieties (C) , creolized varieties (RV), and all seed from the formal system (F), containing all OPVs and Hybrids.
(III) Formal seed (F), versus informal seed (I), containing all landraces and creolized varieties.

Diversity was determined separately for ear traits, vegetative / flowering traits and molecular markers. Results on the average within class diversity are given in Figure 10.

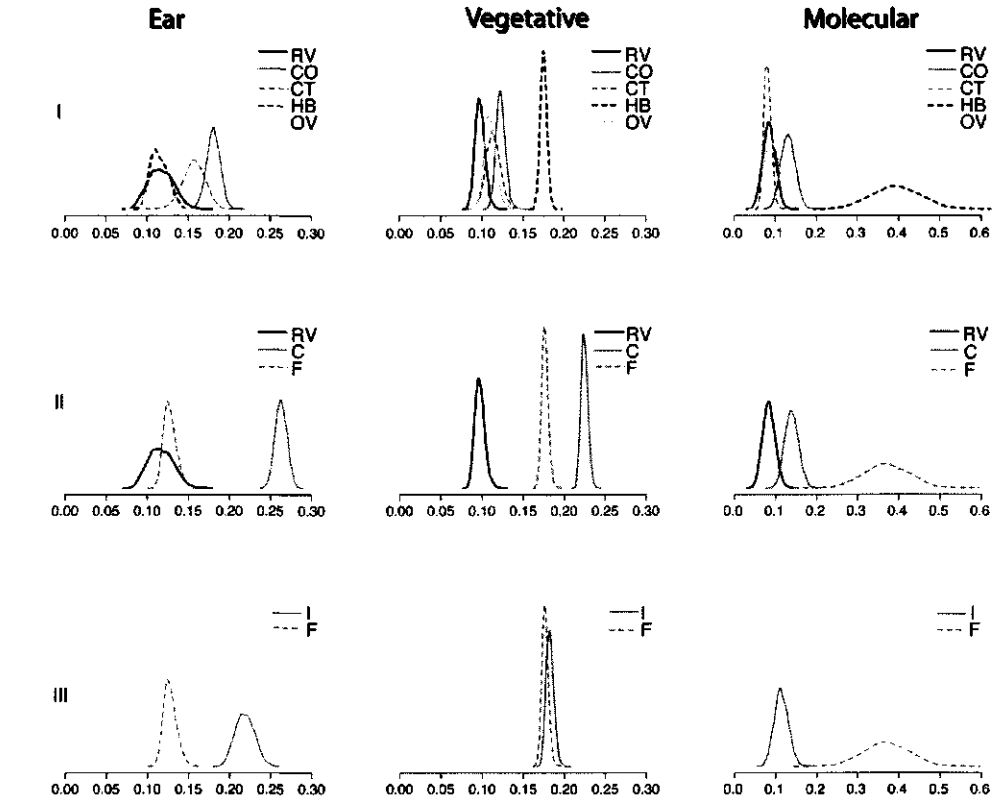


Figure 10. Diversity distribution within different categories of germplasm for three different groupings (I-III). Results are shwown for ear traits, vegetative/flowering traits and molecular markers. Diversity is shown on the horizontal axis. CO: *Olotillo*, CT: *Jarocho*, RV: Creolized, OV: OPV, HB: Hybrid, C: Landrace, F: Formal seed, I: Informal seed

In grouping I, the *Olotillo* class is more diverse for ear traits than any of the other classes. results were different for plant and molecular traits however, the. As was suggested by Figure 7 and Figure 8, hybrids were the most diverse class in terms of vegetative / flowering characteristics and molecular distance. Levels of diversity within the RV, OV and CT classes were similar for all characteristics. The *Olotillo* class showed higher molecular differentiation than all classes except HB. As was mentioned, this probably reflects the effect of drift due to smaller population sizes.

In grouping II, the traditional landraces were much more diverse for ear characteristics than formal and creolized seed. This was also true for plant traits but the difference between landraces and formal seed was less pronounced. Creolized seed was the least diverse class for plant / flowering traits. In terms of molecular differentiation formal seed was the most diverse due to the high differentiation of Hybrids.

Comparison of the grouping II and III serves to evaluate if the adoption of modern varieties into the informal seed sector has caused a decrease in diversity within this class. Differentiation between seedlots in the informal sector was lower than in the landrace class for ear traits, plant and vegetative traits and molecular markers. The reduction, although significant for ear and plant traits, was minor relative to the diversity reduction that would result from a complete replacement of landraces with creolized varieties.

Discussion

The aim of this study was to estimate local diversity impacts of modern maize germplasm adoption in La Frailesca. Replacement of traditional varieties by improved germplasm can occur in two ways. Farmers may replace their local material with commercial seed obtained directly from the formal seed system, thereby loosing both their varieties and the practices of seed management and exchange that have traditionally shaped genetic diversity. In this case the impact on diversity will depend on the variability offered by the commercial varieties in the market (Louette et al. 1997). Alternatively, farmers may continue to rely on the informal seed system but adopt modern varieties as part of their traditional seed system. Diversity consequences of this process of creolization will depend on the heterogeneity between seedlots adopted by different farmers as well as on the extent to which seedlots become subsequently diverged through local evolution.

The informal seed system in La Frailesca was dominated by creolized seed. Incorporation of modern germplasm into the informal seed supply has thus been substantial in the last decades. Creolized varieties actually present the highest diversity in nomenclature (Figure 2). Farmers generally consider creolized seed as being local and researchers should be aware of the risk of misclassifying these varieties as local landraces. This point is made evident by our observation of the seedlot CT4 that revealed striking similarity in both morphology and marker frequencies to modern varieties, in spite of it being classified as a traditional landrace by farmers. Creolized maize populations have undergone years of local seed management with the potential of becoming differentiated from their parental seed type. The observation of creolized populations that were distinct from their most probable parents confirmed that local evolution may indeed have occurred.

However, creolized seedlots formed a rather homogeneous group and were poorly differentiated phenotypically and genetically from the open pollinated varieties from which they derived. This result contradicts earlier suggestions that creolized varieties combine traits from both modern and traditional varieties (Wood et al. 1997; Bellon et al. 2001). The group of modern varieties formed by OPVs hybrids and creolized seedlots differed phenotypically from the two main traditional landraces by plant height, leaf width, grain yield and row number. Total grain weight was consistently higher in improved and creolized varieties as compared to the traditional landraces. This may explain why the latter are becoming rarer in the commercialized agriculture that is currently being practiced.

Among the traditional landraces, *Olotillo* and *Jarocho* were distinguishable only by ear characteristics. Although they are considered to belong to two different races (Wellhausen et al. 1952), they did not cluster separately for plant traits and molecular markers. This suggests that the two types represent the outcome of differential selection on ear shape by farmers (Louette et al. 2000) rather than forming two separate genetic entities. In this light it is relevant to point out the low genetic differentiation between the two races and between modern and landrace germplasm, compared to the differentiation between individual seedlots. Genetic difference between seedlots, as measured by molecular markers, is likely to be strongly affected by local drift due to limited population size. The relatively high differentiation shown by *Olotillo* and *Conejo* seedlots seem to confirm this notion. This may explain why a recent marker study, based on single seedlots, clearly identified Olotillo and Tuxpeño as two separate races (Reif et al. 2006). Our result raises questions about the meaning of race as a unit of germplasm conservation.

Genetic diversity within different classes of germplasm was measured as the weighted mean distance between seedlots. Different classifications were used to describe diversity contained in the formal and informal seed system. Although this measure does not provide an estimate of richness, it represents the mean distinctness of seedlots within a class. This is appropriate since it reflects the difference in a set of traits that farmers or breeders may obtain by accepting a different seedlot from a randomly chosen source of seed. Our results show clearly that different classes of germplasm may be more or less diverse depending on the type of traits studied. The question if replacement of one class by the other affects agronomical and genetic diversity in La Frailesca can thus not be answered unequivocally. Hybrids for example offer large diversity of plant and flowering types that is not paralleled by any of the landraces taken separately or even the two main varieties taken together (data not shown).

From a conservation perspective however, the evolutionary potential of local material should be taken into account. In this sense, hybrids do not constitute a source of new traits and genes since they are seldom incorporated into the informal seed system because of inbreeding depression upon replanting. Genetic diversity present in the informal seed system is therefore of greater relevance to conservation. Given the lack of differentiation between creolized varieties, the complete replacement of traditional landraces by creolized seed would hence constitute a loss of diversity. However, these creolized varieties offer traits that are distinct from those found in local landraces. Hence the coexistence of the two classes of seed at their actual frequencies has caused only limited reduction of diversity in the informal seed system.

In conclusion, this study shows that testing the hypothesis of genetic erosion in smallholder agriculture is indeed complex. On the one hand one should be careful in assuming genetic erosion based on the sole observation that the formal system has become the primary source of seed. As we have seen, commercial seed may be diverse for certain traits. On the other hand one should not claim that a regions' genetic resources have been conserved because the informal seed system persists, since the adoption of creolized seed can still reduce local diversity. The present work has allowed for actual levels of diversity loss to be assessed for different traits. Admittedly, our results are based on a relatively arbitrary set of traits and markers. In order to address the issue of genetic erosion in a way that is meaningful to farmers and breeders diversity needs to be described for traits and trait combinations that are considered valuable (Bellon 1996; Bellon et al. 2003). We hope however, that the methodology presented here will contribute to a more quantitative approach to the problem of genetic erosion in dynamic seed systems that are typical of smallholder agriculture.

Chapter V

Limitations of GMO detection in traditionally managed maize populations

Abstract

Mexico is the centre of origin and diversification of maize. A much debated report published in 2001 (Quist et al. 2001) suggested that a *de facto* moratorium on the introduction of genetically modified maize imposed in 1998 had failed to prevent the inadvertent spread of transgenic elements to locally collected traditional maize varieties in the state of Oaxaca. The only peer reviewed paper to contribute new data on this issue found no evidence of transgenes in maize sampled in subsequent years. Although some criticism was voiced regarding the interpretation of the data, the authors' conclusion of a strong reduction or disappearance of transgenes from the region found resonance in the media. Recent unpublished results from another multi-year study performed in the same area both confirm the presence of transgenes and suggest their continued presence until 2004. Shortcomings of molecular detection essays were identified as a possible cause of disagreement between the two studies. The present work intends to contribute to the correct interpretation of contrasting results between GMO detection studies. We discuss three main aspects related to sampling that may affect the detection probability. We present theoretical and simulation results that show that maize reproductive biology can lead to a reduction in sample size. We show that the strongest potential limitation on detection lies in the expected aggregated frequency distribution that is a consequence of farmer-mediated introduction. Analysis of recent sampling efforts reveals that detection probabilities may be much lower than previously assumed.

Introduction

Mexico is the centre of origin and diversification of maize. Maize remains the main staple in Mexico, with a pivotal place in the country's economic, cultural and agricultural spheres. In contrast to the United States and Europe, commercial seed accounts for only one-fourth of the maize area planted in Mexico (Aquino et al. 2001). As in other developing countries, maize is mostly grown by smallholders who rely on their own harvest or on that of other farmers for their planting material. This practice creates an open seed system, subject to evolutionary processes of drift, gene flow, and selection, in which the fate of newly introduced genes is hard to predict (Bellon et al. 2004). For this reason, a *de facto* moratorium on field-testing and commercial planting of genetically engineered (GE) maize was established in 1998 in order to avoid unintended gene flow into local landraces.

Despite the restrictions imposed on the introduction of transgenic maize varieties into Mexico, a study published in 2001 reported the presence of the 35S cauliflower mosaic virus (CaMV) promoter and Nopaline Synthase terminator (NOS) recombinant sequences, in local landraces sampled from the Sierra de Juárez region in the state of Oaxaca (Quist et al. 2001). The paper contained several methodological shortcomings for which it was criticized (Metz et al. 2002). Initial reports issued by the Mexican Government seemed to confirm transgene presence in Mexican native seed stocks however (Ezcurra et al. 2002). In 2003 and 2004 a large scale sampling effort conducted in the same region sampled by Quist and Chapela yielded no evidence of transgenes (Ortiz-Garcia et al. 2005). The authors used arguments based on simple calculations of detection probability to suggest that transgenes were absent or present at very low frequencies. A strong reduction in transgene frequency was invoked to explain the contrast with earlier reports, a suggestion that found some resonance in the scientific media (Marris 2005; raven 2005). Recent unpublished results by our collaborators (Piñeyro et al. submitted) have established the presence of GMOs in Oaxaca in 2001. A subsequent analysis performed in 2002 came out negative, thereby confirming the results by Ortiz-García et al. However, in 2004 several positive fields were found in a set of samples from the same villages where positives were detected in 2001. This lack of agreement between studies is puzzling and presents a potential source of continuing polemic surrounding this politically sensitive theme. Comparison of PCR-based 35S detection between two laboratories has revealed that differences in laboratory procedure might explain the observed inconsistencies (Piñeyro et al. submitted).

It is important to establish if standardizing laboratory procedures will be sufficient to avoid controversy. Low probabilities of detection, leading to large sampling errors, also present a potential source of inconsistencies. Adequate estimation of detection probability is therefore key to interpreting results from transgene assays in a particular sample.

Detection probabilities for rare alleles have been calculated by different authors (Gregorius 1980; Crossa 1989; Crossa et al. 1993; Wang et al. 2004). The detection probability, P_d, for samples of fixed size taken from different populations, is given by (Lockwood et al. 2007):

$$P_d = 1 - \prod_{i=1}^{i=m} (1 - p_i)^S \qquad (1)$$

Where m is the number of sampled seedlots or fields, p_i the frequency of individuals containing the rare allele and S is the sample size defined as the number of diploid individuals collected per field. Assuming the allele occurs with a uniform frequency p in all populations, this equation reduces to:

$$P_d = 1 - (1 - p)^{mS} \qquad (2)$$

This is the recommended formula for calculating GMO detection thresholds in bulked seed samples (USDA 2001) used by Ortiz et al. (2005). Detection probability is thus assumed to be a simple function of the number of sampled seeds and the mean transgene frequency.

We will show that the latter assumption is likely to be violated for maize samples taken from land-race populations. First, we argue that sample size should be corrected for unequal paternal and maternal parentage. Second, we use simulation of maize reproduction to estimate the effect of restricted pollination on sample size. Finally, we present analytical and simulation results that suggest that the use of expression (2) will lead to overestimation of detection probability when allele frequencies are not equal across populations, but follow a skewed distribution.

Methods

Simulation of pollination process

The distribution of paternity of seeds sampled from a single ear was simulated as a spatially explicit, competitive sampling process determined by flowering synchronicity between male and female inflorescences and distance between plants. A field of N = 60,000 plants was modeled assuming three plants per hill and 0.75 meters between hills. Each plant was randomly assigned an anthesis and silking date based on actual field data (Van Heerwaarden, unpublished). Data on day-to-day silk emergence and pollen production were derived from the study by Uribelarrea et al. (Uribelarrea et al. 2002). A total of 505 silks were assumed to emerge in discrete groups over 7 days. Silks emerging on a single day were assigned pollen parents by drawing with replacement from a probability vector representing the entire set of plants. Probability of paternity p_i for each plant was defined as follows $p_i = \frac{G_i d_i}{\sum_{i=1}^{i=n} G_i d_i}$,

where G_i is the amount of pollen produced by plant i on that day and $d_i = e^{-0.4098x}$ representing the reduction of pollen concentration with distance x (Ma et al. 2004). The mean number of unique paternal alleles n_u in a sample of n_s seeds was determined by drawing samples of size n_s from 100 simulated vectors of sires.

Definition of frequency distributions

Aggregation of allele frequency was described by means of a gamma distribution with the following parameters: the shape parameter k, which determines the skewness, and the scale parameter θ which was defined as $\theta = \frac{p}{k}$, so that the distribution mean θk equaled the mean allele frequency p at any value of k. The value of k was set to range from 0.5 to 0.005 to achieve a range of increasing levels of aggregation.

Simulation of transgene introduction and diffusion

The process of transgene introduction was modeled by simulating the population genetic dynamics of single bi-allelic locus in a set of 1000 independent villages through time. Each village was modeled as a square grid of 81 fields. Pollen flow was assumed to occur only between neighboring fields and was set at 1.5% per synchronously flowering neighbor field (Louette 1997; Messeguer et al. 2006). Seed migration was simulated as complete or partial replacement with individual, randomly selected farmers as a source. Interviews conducted in two of the sampled localities (H. Perales, unpublished data) served to estimate the following model parameters: average population size (40 selected ears, 300 seeds per ear), average number of neighbors with synchronized flowering (1) and replacement frequency (0.07). Partial replacement was not observed in the survey but was assumed to occur with a frequency of 0.01 and involve 20 migrant ears. We modeled seven years of random introduction. The probability of planting a transgene in a single year was set at 0.005 per farmer. Farmers planting GMO maize in a single year were excluded as a source for seed migration and were set to abandon the seed in the next season.

Sampling from simulated distribution

Detection probabilities for samples taken from the simulated frequency distribution were based on 10,000 random samples from 1000 villages with $n_{f(i)}$ fields per village, $n_{c(i,j)}$ ears per field and $n_{s(i,j)}$ seeds per ear. The number of represented paternal alleles in a sample $n_{u(i,j)}$ was set to $n_{s(i,j)}$ in case of unrestricted mating. Restricted mating was introduced by setting $n_{u(i,j)}$ to a value lower than $n_{s(i,j)}$ based on simulation of the pollination process as described above. The detection probability for each sampled field was thus given by: $P_{d(i,j)} = 1 - \left(1 - p_{(i,j)}\right)^{2n_{c(i,j)} \cdot n_{u(i,j)}}$. Complete selfing was defined by a sample size of $n_{c(i,j)}$ alleles. Sampling was assumed to take place from stored planting material so primary introductions were assumed to be absent from the population given they were rejected after harvest. Sample sizes (number of villages, fields per villages, ears per field and seeds per ear) were set according to reported values (Ortiz et al. 2005, Piñeyro et al. submitted.)

Results

Unequal Paternal vs. Maternal Contribution and sample Size:

Cleveland and collaborators (2005) recently argued that the number of individuals is not an appropriate measure of sample size. When seed is harvested from a limited number of ears, the unequal contribution of paternal and maternal parents has to be taken into consideration (Cleveland et al. 2005). The authors proposed replacing S in (2) with the variance effective population size N_e. This measure is related to the variance of an allele in a sample of diploid individuals such that $\sigma = p(1-p)/2N_e$. It can be written as a function of the number of maternal and paternal parents as: $N_e = \frac{1}{\frac{1}{4n_c} + \frac{1}{n}}$ (Vencovsky et al. 1999). Where n_c is the number of maize ears and n the total number of seeds sampled. When $n_c << n$, N_e has a maximum value at $4n_c$, so S is at most four times the number of sampled ears.

We note here that although unequal parental contribution affects detection probability, effective population size is not a correct measure of sample size. To explain this, it is convenient to redefine the transgene frequency as the allele frequency rather than the frequency of positive individuals. Variance in allele frequency between samples, and hence N_e, is dominated by n_c due to the genetic correlation of materal alleles sampled from the same ear. Samples containing ears derived from transgenic maternal plants will have very high frequencies of transgenic alleles, thereby increasing the variance. As transgenes are expected to be rare however, most samples will not derive from transgenic maternal plants. Hence, the probability of finding a transgenic allele is largely determined by the chance of occurrence in n independent pollination events. Figure 1 presents this graphically by showing the distribution of allele frequency in 10,000 simulated ear samples from a single population with a transgene frequency of 1%. The sample frequency has a variance of 0.00026, which is as expected given the effective population size of 19.7 diploid individuals or 39.4 alleles ($n_c = 5$, $n = 1500$). The distribution has a long right tail due to the sampling of positive ears. Three separate distributions can be observed. These correspond to seed derived from negative homozygous maternal plants, positive heterozygotes, and positive homozygotes respectively.

Negative homozygotes are the most frequent maternal plants, so transgene frequency in most samples falls within the narrow distribution on the left. Using effective population size yields: $P_d = 1-(1-0.01)^{39.4} = 0.33$. In contrast, all 10,000 samples contained at least one transgenic allele.

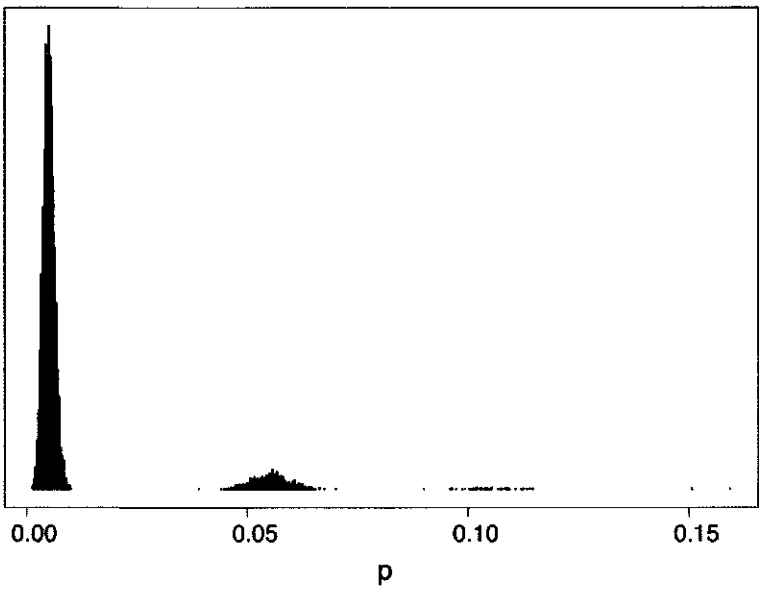


Figure 1. Histogram showing the distribution of transgene frequency in 10,000 simulated samples consisting of 5 ears and 1500 seeds. Mean frequency in the population 0.01.

This result shows that for the purpose of detection, variance effective size is not a proper substitute for sample size. As we are interested in detecting a single transgenic allele that was inherited from a population of parental plants, the number of represented parental alleles provides a more appropriate measure. As was pointed out by Crossa (Crossa 1989), determining the number of parents in a sample is an occupancy problem. In our case, we define a single population of N diploid parental plants, containing $2N$ alleles. A sample of n seeds taken from these plants will contain S_a alleles that are represented at least once in the sample. These alleles may be divided into paternally and maternally inherited alleles such that:

$$S_a = S_m + S_f$$

Where S_m is the number of paternal alleles and S_f the number of maternal alleles.

We assume an infinite amount of pollen and random mating. The probability of including any paternal allele in a sample of n seeds is:

$$1-\left(1-\frac{1}{2N}\right)^{n}$$

The expectation of the number of alleles represented in the sample thus becomes:

$$E(S_m) = 2N - 2N\left(1-\frac{1}{2N}\right)^{n} \text{ (Crossa 1989)}$$

When $n << 2N$ this equation approaches:

$$S_m \approx 2N - 2N\left(1-\frac{n}{2N}\right) = n$$

A sampled ear yields n_s seeds such that $n = n_s n_c$. For maternal alleles we need to account for the fact that n_s seed from the same ear will contain only one or two maternal alleles. The probability of having only one allele represented is given by:

$$\pi_1 = \left(\frac{1}{2}\right)^{n_s-1}$$

The expectation for the number of alleles in a sample of n_c ears becomes:

$$E(S_f) = n_c\left(\pi_1 \times 1 + (1-\pi_1)\times 2\right) = n_c\left(2-\left(\frac{1}{2}\right)^{n_s-1}\right)$$

Under the assumptions of $n << 2N$ and large n_s may thus simplify S_a to:

$$S_a = 2n_c + n \qquad (3)$$

Assuming $S_a << 2N$, we may consider S_a a sample with replacement from the total set of $2N$ alleles.

Redefining p as the allelic frequency of the transgene, we may substitute S_a for S in equation (1) and (2). When only one seed per ear is sampled, S_a takes a maximum value of $2n$ alleles. This is equivalent to the conventional sample size of n diploid individuals. A very small number of ears will yield a sample of approximately half this size. Unequal allelic contribution thus reduces sample size, but not by as much as calculated by effective population size.

The effect of restricted Pollination

The preceding estimation of S_a assumes that any paternal plant has an equal probability of siring a kernel. Although maize is a highly allogamous species, some studies suggest that pollination is restricted to some extent (Bijlsma et al. 1986). Strong effects of distance and flowering synchronicity can cause a reduction in the number of paternally derived alleles in a single ear. Ortiz et al (2005) accounted for this factor by including a conservative sample size estimate based on the assumption of complete selfing. The authors suggested that this measure was too conservative, as they expected "that many seeds from the same cob were sired by different paternal plants" (Ortiz-Garcia et al. 2005), a contention that is shared by other authors (Paterniani et al. 1974). We are unfortunately ignorant of the distribution of paternity in maize. We can therefore not evaluate the effect of restricted pollination on the reduction of detection probability. To have an indication of the type of distribution of sires under realistic assumptions of restricted mating, we performed a simulation of the pollen process (See methods). Both distance and flowering synchronicity were imposed as limiting factors on free pollination. As can be seen from Figure 2, simulated pollination is indeed highly restricted in space. Contributing sires come from a relatively small part of the field surrounding the sampling location of the ear. Several paternal plants in close proximity to the sampled ear contribute up to 30 offspring. Flowering synchronicity also plays a role, as can be seen from the fact that a number of plants at the same distance do not contribute any offspring. On average our simulation predicted that a sample of around about 400 seeds from the same ear contains 162 unique paternal alleles. The number of expected alleles in samples of increasing size can be seen in Figure 3. Although sampling efficiency decreases with the number of sampled seeds, the proportion of sampled seeds contributing unique paternal alleles remains higher than 80% until n_s >40 and is still 46% at 300 seeds per ear. This suggests that restricted pollination may not be a strong limitation to the overall detection probability. Corrected estimates of detection probability may be generated by substituting simulated numbers of paternal alleles for n in equation (3).

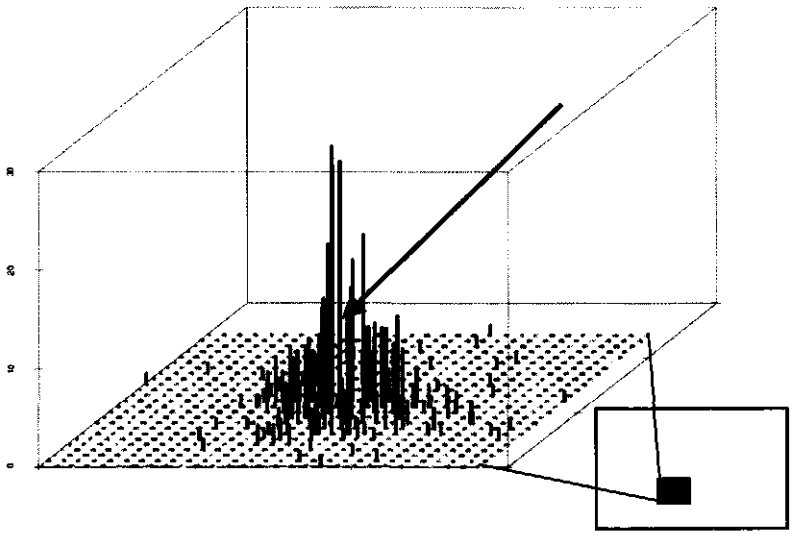


Figure 2. Histogram of simulated paternal contribution to a single ear. An enlarged area from a field of 60,000 plants is shown. Height of the bars indicates the number of times the same parent was represented in the seed. The arrow shows the location from which the ear was sampled.

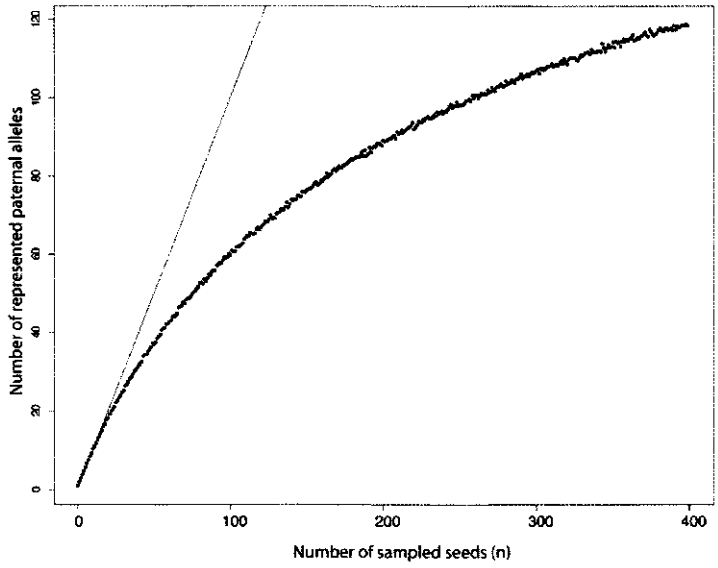


Figure 3. Estimated number of represented paternal alleles as a function of the number of sampled seeds. The solid line indicates the expected number of paternal alleles under unrestricted pollination.

Shape of the Transgene Frequency Distribution Among Populations

Equation (1) assumes that we know the transgene frequency in each sampled seedlot. In practice, seeds are collected from fields with unknown transgene frequencies. We can expect that these frequencies differ strongly among fields; Transgenic seed has been available for only ten years (www.agbios.com). This period is probably too short for pollen and seed flow to have homogenized transgene frequency among populations.

The fact that aggregated distributions may lower detection success for rare species is well known in ecology (Green et al. 1993). However, transgene detection probability in a single field depends on p_i (see equation 1). The effect of absence in some fields may therefore be offset by increased detection probability in fields containing high frequencies. P_d is thus expected to be relatively insensitive to differences in transgene frequency among fields.

For transgene sampling, the effect of aggregation on detection probability can be expressed as follows. The expected detection probability for a sample of size S taken from a single randomly selected field with transgene frequency p_i is given by:

$$E(P_d) = 1 - E\left((1 - p_i)^S\right)$$

Which for low values of p_i, may be written as:

$$E(P_d) = 1 - E\left(e^{-Sp_i}\right)$$

or:

$$E(P_d) = 1 - E\left(e^{-S(p+d_i)}\right)$$

Where:

$$d_i = p_i - p$$

So that we have :

$$E(P_d) = 1 - e^{-Sp}\Psi = 1 - (1 - \bar{p})^S\Psi$$

With:

$$\Psi = E\left(e^{-Sd_i}\right)$$

For m independently sampled fields we have:

$$P_d = 1 - \left((1-\bar{p})^S \Psi\right)^m \qquad (4)$$

When the expected difference in p_i among fields is small, Ψ is close to unity and expression (4) reduces to equation (2). For large S, strong differences in p_i will increase Ψ and hence lower detection probability. Transgenes with a skewed frequency distribution due to high levels of aggregation are thus harder to detect than expected from its mean frequency.

The effect of aggregation on P_d as estimated by (4) can be observed in Figure 4a. A gamma distribution with several values of the shape parameter k was used to estimate Ψ for different levels of aggregation. Detection probability was accurately estimated by mean transgene frequency up to values of k of about 0.1. At higher levels of aggregation however, P_d was significantly reduced. Increasing the number of sampled fields and sampling less seed per field increased detection probability.

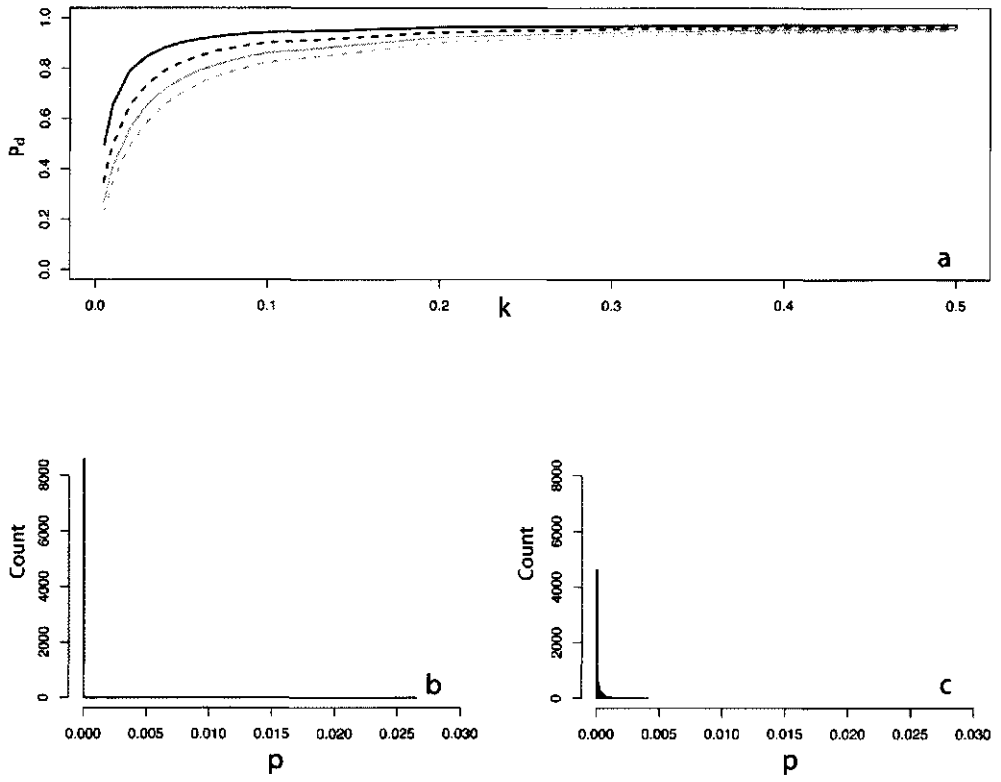


Figure 4. Effect of aggregated transgene distribution on the expected detection probability (a). Pd is shown over a range of values of the shape parameter of the gamma distribution (k= 0.005:0.5) at four different values of *m* (48, 24,16,12). Mean allele frequency p, and total sample size were set at 0.0002) and (24000) respectively. Lower panels show histograms of 10,000 random values of p at k=0.03 (b), and k=0.35 (c).

We have no empirical data on the type of transgene frequency distribution. However, we do have considerable knowledge on smallholder farming practice (Rice et al. 1998; Louette et al. 2000; Perales et al. 2003), which we may employ to generate estimations of transgene distribution as a result of inadvertent introduction. We simulated the unintentional introduction and spread of transgenes over time, using data on pollen (Louette 1997; Messeguer et al. 2006) and seed flow (H. Perales unpublished data). Figure 5 shows the observed distribution of transgenes after seven years. The simulated frequency distribution was highly aggregated, with most fields having frequencies close to zero and with a few having values of over three percent.

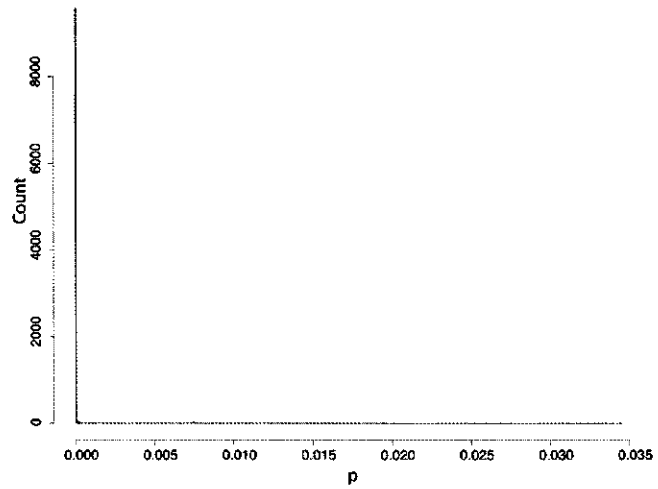


Figure 5. Simulated distribution of transgene frequencies in farmers field after seven years of introduction.

A comparison of detection probabilities for samples taken from this distribution can be seen in Table 2. Results are presented for our 2001 and 2002 collections as well as for Ortiz-García et al. (2005) 2003 and 2004 samples. Values are shown for unrestricted pollination, restricted pollination and complete selfing. Detection probabilities under the assumption of uniform frequency across fields are provided for comparison. Except in the case of complete selfing, aggregated transgene frequencies clearly reduced detection probability for the different studies. These results show that given the actual sample sizes, a frequency of introduction of 0.5% may go undetected as long as contamination by pollen and seed flow remains localized.

Table 1. Estimated detection probabilities (P_d) for three independent samples from studies conducted in Sierra Juárez, Oaxaca.

Study	Mean allelelic freq.	P_d uniform frequency			P_d simulated aggregation		
		Unrestricted pollination	Restricted pollination	Complete selfing	Unrestricted pollination	Restricted pollination	Complete selfing
2001	0.0002	0.33	0.28	0.01	0.19	0.19	0.01
2002	0.0002	0.99	0.97	0.13	0.79	0.77	0.13
2003*	0.0002	1.00	0.99	0.03	0.67	0.52	0.03
2004*	0.0002	1.00	1.00	0.14	0.84	0.82	0.13

* sample from Ortiz et al. 2005.

Discussion

The present controversy surrounding possible transgene introgression into Mexican maize landraces has highlighted the need for methodological consensus on detection methodology. Correct estimates of sample size and detection probability are hence important. The conclusion of absence of transgenes expressed by Ortiz et al (2005) relied entirely on predicted probabilities. They argued that given their sample size, they could detect an allelic frequency of 0.00005, with a probability of 0.99335 and 0.99997 in 2003 and 2004 respectively. Our work suggests that even at frequencies four times higher than the above, their actual detection probabilities could have been much lower at 0.52 and 0.82 respectively. The present paper has analyzed several potential factors that contribute to this discrepancy and that should be taken into account when analyzing results from transgene monitoring surveys.

Currently used calculations for detection probability are designed for detection of transgenes in large seed bulks and are not always appropriate when sampling maize seed in-situ. We have shown that the usual measures of sample size do not correctly account for the unequal contribution of maternal and paternal plants to the offspring of sampled ears. On the one hand Ortiz et al. (2005) ignored this issue altogether while Cleveland et al. (2005) erroneously proposed the use of effective population size. We have argued that effective population size, although it correctly predicts variance in transgene frequency among independent samples, underestimates the actual sample size. We propose a simple approximate calculation of allelic sample size that should be used instead of either number of seeds or effective size.

Restricted pollination has been proposed as a factor that could reduce detection probabilities considerably (Ortiz-Garcia et al. 2005). Our simulations of the pollination process confirm empirical reports that paternity within single maize ears may be restricted. However, the reduction of the number of sampled paternal alleles is expected to be limited unless very large numbers of seeds are sampled per ear. This effect was indeed visible in our simulation of the Ortiz et al.'s 2003 sample. A lower detection probability was observed under the assumption of restricted mating due to the fact that 300 seeds were sampled per ear (vs. 70 in 2004).

The most important factor affecting detection probability was shown to be the shape of the transgene frequency distribution. Farmer-mediated introduction of transgenes may be expected to be highly localized, since farmers will make individual decisions with respect to planting any type of new seed. One of the more plausible scenarios for the introductions of transgenes in Oaxaca is the planting of imported transgenic feed grain obtained from local stores. Existing data on seed sources indicates that seed from stores (Perales et al. 2003), and especially grain (G. Dyer, pers. Comm.) is rarely used by farmers. Introduction of transgenes thus probably occurs through a small number of individual farmers. Our simulations of pollen and seed flow over a limited period of time suggest that gene flow was not sufficient to homogenize frequencies among fields. Seven years after introduction transgenes were mostly concentrated in seedlots that had been planted next to transgenic fields. Frequencies in these fields were in the range of 1 to 3 percent while most other fields did not contain any transgenes. We showed that such an aggregated frequency distribution is expected to lead to lower than expected detection probabilities. We confirmed this by taking samples of the same size as reported for different recent detection efforts from our simulated transgene distribution. In all cases, detection probability was significantly reduced with respect to previous estimates. At present, we may thus expect a much lower probability of detection than expected on the basis of total sample size and mean transgene frequency. Sampling more fields and less seeds per field is therefore recommended.

These results show that population genetic processes are important for generating expectations on transgene frequencies and detection probabilities. By explicitly taking farmer mediated introduction and subsequent pollen and seed flow into account we will improve our ability to evaluate the present and future status of transgene introgression in local landraces.

Chapter VI

General discussion

A quantitative take on traditional seed management

The present study has addressed the role of smallholder farmers as determinants of genetic diversity and structure of maize landraces. The population genetics of managed maize populations is complex, with human and natural selection, population size, pollen flow, seed mixing and replacement all potentially affecting the distribution of genetic variation. An important first step to understanding the relation between farmer practice and genetic diversity would be the ability to understand the fate of neutral genes in managed crop populations. Although quantitative data on both seed management and the genetic structure of neutral molecular markers are available (Louette et al. 1997; Sanou et al. 1997; Rice et al. 1998; Smale et al. 1999; Louette et al. 2000; Perales et al. 2003; Pressoir et al. 2004; Perales et al. 2005), the lack of a proper theoretical framework has prevented successful integration of these two types of information.

Crop species may be thought of as systems of connected demes, subject to pollen and seed migration as well as extinction and recolonization. For this reason, the meta-population concept (Levins 1969) has been proposed as a basis for models of crop diversity (Louette 1999; Alvarez et al. 2005). The application of this concept to population genetics of cultivated species has thus far not transcended the metaphorical or the semi-quantitative however (Brush 1999). We have presented population genetic metapopulation models that are able to translate knowledge on seed management into expectations of neutral genetic structure and diversity. We chose to construct a computational model as well as an analytical model based on coalescent theory. Our computational model is spatially explicit and highly flexible. Complex information obtained from the field can be easily incorporated and few simplifying assumptions need to be made. These features make it suitable for generating testable predictions about patterns of diversity and for studying the spread of introduced genes in traditional farming systems. Like any computational model however, it does not lend it self easily to mathematical analysis.

Our coalescence-based model on the other hand, is well suited for gaining a more formal understanding of the population genetics of maize metapopulations. The latter model has yielded new insights into the determinants of genetic structure in maize. We have shown that the special features of farmer seed systems lead to predictions that differ fundamentally from those emanating from existing meta-population models. In particular, the assumption that seed migration between demes occurs from a single source per deme per generation led to interesting results. The model showed that it is important to distinguish between seed migration frequency and quantity of migrating seed and to treat seed replacement as a separate form of seed exchange. Contrary to classical predictions, higher amounts of migrating seed may lower neutral genetic diversity within seedlots. The so called invariance principle (Nagylaki 1982), i.e. the independence of diversity to migration rate, is hence violated by single source seed migration. Genetic structure will decrease with the quantity of mixed seed as predicted in existing models but only as long as the relative quantity is below 50%, after which structure will increase. The same is true for within-seedlot diversity in the presence of extinction (seed replacement); mixing up to 50% of a seedlot with seed from another farmer will restore genetic diversity while mixing more seed will lead to diversity loss. Seed migration frequency on the other hand, obeys the classical predictions of invariance of within seedlot diversity and decreased structure with increasing migration. These novel results arise from the genetic correlation between alleles originating from a single population which is inherent to the process of single source migration (See chapter II).

Seed replacement decreases genetic diversity as expected. The effect of replacement on genetic differentiation may be either positive or negative depending on migration and the number of seedlots. Another important observation was the impact of population size on genetic structure, whenever the quantity of migrating seed is large with respect to seedlot size. This suggests that seedlot size is an important parameter that has previously been ignored based on the classical prediction of independence of population size and genetic differentiation (Wright 1951).

A final lesson arising from our model is that in the presence of high levels of pollen flow, the effect of seed related parameters on genetic structure becomes negligible. This is especially the case for extinction frequency and quantity of mixed seed. This result is especially relevant to future studies that intend to explain observed molecular differentiation as a function of seed management, and puts a note of caution on published interpretations of low structure as reflecting high levels of seed flow (Brocke et al. 2003; Pressoir et al. 2004).

We would like to stress that results such as these were not obtainable before and that our analytical model provides an attractive tool for exploring the possible consequences of different aspects of seed management for genetic diversity and structure in managed crop populations.

Farmer practice and patterns of diversity

In Mexico, extreme differences in environmental conditions are found over short geographical distances. Such contrasting growing conditions are known to be important determinants of genetic differentiation of maize landraces (Wellhausen et al. 1952; Doebley et al. 1985). The ways that human mediated seed migration and local management can affect diversity patterns created by the environment are not well studied. We analyzed the joint effects of human and environmental forces on genetic structure within maize landraces by looking at molecular and phenotypic data across two contrasting but adjacent environments. We collected seedlots from villages in highland and lowland environments of central Mexico. In addition to biological data, information on farmer practice and field conditions was obtained for the two environments. We were able to describe seed management in both environments by farmer interviews and obtained data on approximate field distributions from aerial imagery. We developed a spatially explicit computer model to generate predictions of molecular differentiation based on this agronomical data.

Maize production systems in highland and lowland environments were found to be different in several respects. Differences were observed in seed- replacement and migration but our computer model showed that these were of little consequence to predicted genetic structure. Relevant differences were found in seedlot size and land use patterns. Field sizes are larger in the highlands and maize is generally planted in large arrays of connected fields. Pollen flow in the lowlands was inferred to be much lower because maize is planted in small plots that are interspersed with citrus plantations and pasture. Our model hence predicted higher neutral genetic structure in the lowlands due to restricted pollen flow and smaller seedlots. Observed levels of molecular differentiation were in agreement with these model predictions. Contrasting levels of between-seedlot divergence among different regions have been reported for pearl millet in India (Brocke et al. 2003). The authors proposed differences in seed flow as a cause but did not consider the possibility of pollen flow nor did they quantify the amount of seed flow occurring in their two study areas. Our study has shown that in allogamous crops, pollen flow may be an important determinant of within-village genetic structure.

Our modeling results, based on quantitative information on management, discarded seed flow as the most likely explanation for observed differences. Our results provide an example of how appropriate quantitative models can aid in the interpretation of genetic data.

Genetic structure measured by molecular markers is important as a measure of divergence caused by neutral processes such as drift and gene flow. It provides the perfect baseline information against which differentiation of selected traits can be compared. Marker differentiation between environments, villages and farmers was weak, as expected based on earlier studies (Pressoir et al. 2004). High differentiation in phenotypic traits was found however. The most pronounced differences were observed between the different climatic zones, confirming the importance of environment in structuring genetic diversity. However, we found three villages in the highlands planting an intermediate phenological phenotype. Our molecular data confirmed the hybrid origin of these populations. Data on ear and kernel traits revealed close similarity to all other highland germplasm, indicating that these admixed populations are adapted to highland conditions. The restricted distribution of these populations suggested that the observed pattern is a result of seed migration from lowlands to highlands.

Phenological traits were highly structured between villages, possibly related to differences in average planting dates. Considerable differentiation was also found for ear and kernel traits but for these traits differences were less pronounced. These results contradict earlier findings by Pressoir and Berthaud (2004) in Oaxaca, who reported low between village differentiation for phenology and high differentiation for ear traits. We did observe strong ear trait differentiation between villages in the HL transect. This sampling transect included the phenologically divergent admixed populations however, suggesting that the observed structure was due to the presence of admixed populations. Our results do support the notion that villages provide an appropriate focus for in-situ conservation (Pressoir et al. 2004).

As was shown in previous studies (Pressoir et al. 2004; Perales et al. 2005) divergence in phenotypic traits was much stronger than neutral differentiation measured by molecular markers. This suggests that divergent selection imposed by local growing conditions is the strongest determinant of genetic structure in our study area. In addition, our results on the admixed highland populations revealed that in some cases farmers may introduce traits that evolved in response to selection in a contrasting environment, thereby creating new patterns of genetic differentiation.

Seed dynamics and genetic resource conservation in the face of agricultural change

The subject of genetic erosion has been the prime motivation for studying the relationship between farmers and their crops. The recent shift from landraces to improved planting materials in many areas in the world is often assumed to represent a loss of genetic diversity, either because improved material is inherently less diverse or because a set of unique and locally adapted varieties are lost (Louette et al. 1997). Reality has proven not to be as clear cut however (Brush 1999). Previous research on smallholder maize agriculture in Mexico has shown that commercial and locally produced seed often coexist instead of the former replacing the other (Bellon et al. 2001). Moreover, farmers do not plant a stable set of traditional varieties. New materials are continuously introduced while others disapear (Louette et al. 1997). Determining what needs to be conserved thus represents a major challenge. Particularly, improved seed may become incorporated into the local repertoire of varieties (Bellon et al. 2001) thereby blurring the distinction between traditional and modern varieties, especially if these new introductions undergo subsequent evolution. Finally, newly introduced germplasm need not be less diverse than traditional varieties and may even add to local diversity if it offers distinct traits. The actual impact of modern germplasm thus represents a complex issue that will depend on both the diversity represented by the different kinds of germplasm as on the quantitative changes in the composition of the seed supply (Louette et al. 1997).

We have presented a case study on genetic erosion in La Frailesca in the state of Chiapas. Our study represents the first attempt to quantify the local diversity impacts of improved maize germplasm within the context of modernized smallholder agriculture in which the formal and informal seed system coexist. We were able to separate creolized from traditional varieties within the informal seed sector and characterize genetic differences at the individual seedlot level for agronomic traits and molecular markers. It has been argued that creolized varieties have changed genetically by selection and gene flow to become better suited to local conditions (Wood et al. 1997; Bellon et al. 2001). Although we found evidence of evolution for some seedlots, creolized maize seemed to have retained its similarity to modern varieties in yield, ear characteristics and phenology. The current popularity of creolized varieties is therefore likely a result of the fact that they represent a cheap source of improved seed, rather than being a consequence of their specific adaptation to local conditions. This seems to be confirmed by the observation that creolized varieties are planted preferentially on better soils than traditional landraces (Bellon et al. 1993). Modern and creolized varieties thus form a group that is clearly different from the traditional landraces in the area.

Our diversity analysis showed that it is not useful to speak about diversity in a general sense, as different traits show varying levels of differentiation. Moreover, not all measures of diversity may have equal relevance to farmers. Molecular differentiation between seedlots for example, seemed to reflect inbreeding related to population size or breeding process rather than meaningful biological differences. This was exemplified by the low molecular differentiation between the two main landrace types.

We have shown that genetic differentiation between seedlots of traditional landraces is higher than that found between improved varieties, except for molecular markers. Commercial varieties, and specifically hybrids, offered a rather high diversity for phenological traits. This diversity was however not reflected in the creolized varieties, which in fact formed the group of least phenological diversity. This might indicate that only a small number of varieties has been adopted into the informal seed system. Alternatively it may be the result of recent evolutionary change that has caused convergent changes in the different creolized seedlots. The low diversity represented by creolized varieties implies that replacing traditional landraces with creolized maize will lead to a loss of genetic diversity. At present however, the coexistence of both traditional and creolized seed in the informal seed sector leads to a level of diversity that is not much lower than that represented by traditional landraces.

New genes in dynamic farming systems

Today, it is hard to imagine a more urgent need for knowledge on maize population genetics than that created by the issue of genetically modified maize. In the United States and Europe, transgenic contamination is a matter of pollen flow during a single generation, since maize producers only use commercial seed. In Mexico, farmer's use of traditional seed management changes the potential consequences of contamination. Ever since the first reports of possible contamination of local Mexican varieties (Quist et al. 2001), farmer seed reproduction and exchange have taken on special relevance. The controversial nature of the initial findings (Metz et al. 2002), combined with the subsequent inability to produce new positive detection results in the same region (Ortiz-Garcia et al. 2005), have generated polemics on this politically sensitive theme (Cleveland et al. 2005). Under these circumstances, a balanced approach to both the interpretation of sampling results and the prediction of future consequences of GMO release is called for. We aimed to contribute to the debate by using current knowledge on maize reproduction and population genetics to evaluate possible limitations of current measures of detection probability.

We studied the effect of unequal parental contribution, restricted pollination and transgene frequency distribution on the performance of traditional estimators of detection probability. We have shown that equating sample size to the total number of seeds leads to overestimation of detection probability. Kernels are sampled from a limited number of maternal plants and the unequal contribution of paternal and maternal parents needs to be accounted for. This had been noted by other authors (Cleveland et al. 2005) but we have shown that the ensuing recommendation of using the variance effective population size is incorrect. We have argued that sample size is approximately equal to the total number of parental alleles represented in the sample. In case of seed sampled from very few ears, sample size approaches the number of paternal alleles. Simulations revealed that limited pollen dispersal and assortative mating within fields reduces the number of sampled paternal alleles as expected. Pollination restriction is expected to cause a relatively minor reduction in detection probability however.

A more serious limitation was posed by aggregated transgene frequencies. Detection probability can be strongly reduced when transgenes are concentrated at high frequencies in a small proportion of fields. Introduction of transgenic seed by individual farmers leads to high local frequencies. On the other hand, pollen and seed flow may homogenize frequencies among fields, thereby increasing the detection probability. Modeling the processes of introduction and gene flow through time we showed that transgene distribution are highly aggregated shortly after introduction. This is not surprising given the fact that pollen flow is only in the order of one percent and seed mixing was observed to be rare. At simulated levels of aggregation, detection probability was shown to be strongly overestimated using conventional calculations. The obvious way to increase detection probability is by sampling more fields at less seed per field. This recommendation has been made before (Cleveland et al. 2005), but under the argument of increased representativeness and not in relation to detection probability. Our results demonstrate the value of using current knowledge on maize reproductive biology and population genetic processes to adjust our expectations with respect to detection probabilities. Taking explicit account of farmer seed management will allow us to generate quantitative predictions on the fate of new genes introduced into the dynamic informal seed system.

Final remarks

Seed management as practiced by smallholder farmers is more than just an academic topic. The conservation of genetic resources and the controversy over the possible spread of transgenes in maize's center of origin have created a growing interest in the subject. Unfortunately, the heightened attention for these two legitimate concerns has generated quite some politically motivated rhetoric that is often based on anecdotal evidence rather than on rigorous investigation. The effect of farmers on the genetics of their crops is real however. Quantitative studies into the effects of seed management on genetic diversity and structure are important in order to avoid the role of farmer practice being reduced to a mere metaphor. We hope that the present work has added to our ability to study the evolutionary forces imposed by humans on their crops. Increased understanding of the dynamic processes that define the population genetics of maize landraces may be of great value to conservationists, breeders and policy makers. These processes have shaped landrace evolution for millennia and will inevitably affect the fate of maize genetic diversity in the future.

References

Aguirre Gomez, J. A. (1997). Analisis regional de la diversidad del maiz en el sureste de guanajuato. PhD thesis. Universidad Nacional Autonoma de Mexico , Facultad de ciencias, Mexio D.F.

Almekinders, C.-J.-M., N.-P. Louwaars and G.-H.-d. Bruijn (1994). Local seed systems and their importance for an improved seed supply in developing countries. Euphytica 78(3): 207-216.

Alvarez, N., E. Garine, C. Khasah, E. Dounias, M. Hossaert McKey and D. McKey (2005). Farmers' practices, metapopulation dynamics, and conservation of agricultural biodiversity on-farm: a case study of sorghum among the Duupa in sub-sahelian Cameroon. Biological Conservation 121(4): 533-543.

Aquino, P., F. Carrión, R. Calvo and D. Flores (2001). CIMMYT 1999-2000 World Maize Facts and Trends, Meeting World Maize Needs: Technological Opportunities and Priorities for the Public Sector. P.-L. Pingali, CIMMYT: 45-57.

Beadle, G. W. (1939). Teosinte and the origin of maize. Journal of Heredity 30: 245–247

Bellon, M. R. (1996). The dynamics of crop infraspecific diversity: a conceptual framework at the farmer level. Economic Botany 50(1): 26-39.

Bellon, M. R. and J. Berthaud (2004). Transgenic maize and the evolution of landrace diversity in Mexico. The importance of farmers' behavior. Plant Physiology 134(3): 883-888.

Bellon, M. R., J. Berthaud, M. Smale, J. A. Aguirre, S. Taba, F. Aragon, J. Diaz and H. Castro (2003). Participatory landrace selection for on-farm conservation: an example from the Central Valleys of Oaxaca, Mexico. Genetic Resources and Crop Evolution 50(4): 401-416.

Bellon, M. R. and S. B. Brush (1994). Keepers of maize in Chiapas, Mexico. Economic Botany. 1994; 48(2): 196 209.

Bellon, M. R. and J. Risopoulos (2001). Small-scale farmers expand the benefits of improved maize germplasm: A case study from Chiapas, Mexico. World Development 29(5): 799-811.

Bellon, M. R. and J. E. Taylor (1993). 'Folk' soil taxonomy and the partial adoption of new seed varieties. Economic Development and Cultural Change 41(4): 763-786.

Bijlsma, R., R. W. Allard and A. L. Kahler (1986). Nonrandom mating in an open-pollinated maize population. Genetics 112: 669-680.

Bolanos, J. and G. O. Edmeades (1992). Eight cycles of selection for drought tolerance in lowland tropical maize: II. Responses in reproductive behavior. Field Crops Research 31(3-4): 253-268.

Bretting, P. K., M. M. Goodman and C. W. Stuber (1990). Isozymatic variation in Guatemalan races of maize. American Journal of Botany 77(2): 211-225.

Brocke, K. v., A. Christinck, E. Weltzien, T. Presterl and H. H. Geiger (2003). Farmers' Seed Systems and Management Practices Determine Pearl Millet Genetic Diversity Patterns in Semiarid Regions of India. Crop Science 43: 1680-1689.

Brush, S.-B. (2001). A farmer-based approachto conserving crop germplasm. Ecocomic Botany 45: 153-661.

Brush, S. B. (1999). Genetic Erosion of Crop Populations in Centers of Diversity: A revision. ed. FAO-WIEWS. Prague.

Buckler, E. S., J. M. Thornsberry and S. Kresovich (2001). Molecular diversity, structure and domestication of grasses. Genetical Research 77(3): 213-218.

Cleveland, D. A., D. Soleri, F. Aragon Cuevas, J. Crossa and P. Gepts (2005). Detecting (trans)gene flow to landraces in centers of crop origin: lessons from the case of maize in Mexico. Environmental Biosafety Research 4(4): 197-208.

Cox, T.-S., J.-P. Murphy and D.-M. Rodgers (1986). Changes in genetic diversity in the red winter wheat regions of the United States. Proceedings of the National Academy of Sciences of the United States of America 83 (15): 5583-5586.

Cromwell, -. E. e. (1990). Seed diffusion mechanisms in small farmer communities. Lessons from Asia, Africa and Latin America. Network Paper Agricultural Administration Research and Extension Network(21): i + 57pp.

Crossa, J. (1989). Methodologies for estimating the sample size required for genetic conservation of outbreeding crops. Theoretical and Applied Genetics 77(2): 153-161.

Crossa, J., C. M. Hernandez, P. Bretting, S. A. Eberhart and S. Taba (1993). Statistical genetic considerations for maintaining germ plasm collections. Theoretical and Applied Genetics 86(6): 673-678.

Crossa, J. and R. Vencovsky (1994). Implications of the variance effective population size on the genetic conservation of monoecious species. Theoretical and Applied Genetics 89(7/8): 936-942.

Dieringer, D. and C. Schlotterer (2003). MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3(1): 167-169.

Dje, Y., D. Forcioli, M. Ater, C. Lefebvre and X. Vekemans (1999). Assessing population genetic structure of sorghum landraces from North-western Morocco using allozyme and microsatellite markers. Theoretical and Applied Genetics 99(1-2): 157-163.

Doebley, J. F., M. M. Goodman and C. W. Stuber (1985). Isozyme variation in the races of maize from Mexico. American Journal of Botany 72(5): 629-639.

Excoffier, E., G. Laval and S. Schneider (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online(1): 47-50.

EyreWalker, A., R. L. Gaut, H. Hilton, D. L. Feldman and B. S. Gaut (1998). Investigation of the bottleneck leading to the domestication of maize. Proceedings of the National Academy of Sciences of the United States of America 95(8): 4441-4446.

Ezcurra, E., S. Ortíz and J. Soberón (2002). In: LMOs and the Environment: Proceedings of an International Conference. OECD: 289-295.

FAO/IPGRI (2002). Review and development of indicators for genetic diversity, genetic erosion and genetic vulnerability (GDEV): Summary report of a joint FAO/IPGRI workshop (Rome, 11-14 September, 2002).

Frankel, O. H. and E. Bennett (1970). Genetic resources in plants-their exploration and conservation, Blackwell Scientific Publications.

Gower, J.-C. (1971). A general coefficient of similarity and some of itsproperties. Biometrics 27: 857-874.

Green, R. H. and R. C. Young (1993). Sampling to Detect Rare Species. Ecological Applications 3(2): 351-356.

Gregorius, H. R. (1980). The probability of losing an allele when diploid genotypes are sampled. Biometrics 36(4): 643-652.

Hammer, K. (2003). A paradigm shift in the discipline of plant genetic resources. Genetic Resources and Crop Evolution 50(1): 3-10.

Harlan, H.-R. and M.-L. Martini (1936). Problems and results of barley breeding. USDA Yearbook of Agriculture. Washington DC: US Government Printing Office.: 303-346.

Harlan, J. R. (1975). Our Vanishing Genetic Resources. Science 188(4188): 618-621.

Harlan, J. R., J. M. J. De Wet and E. G. Price (1973). Comparative Evolution of Cereals. Evolution 27(2): 311-325.

Hawkes, J.-G. (1983). The diversity of crop plants. Cambridge, MA, U.S.A. Harvard University Press.

Hey, J. (1991). A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theoretical Population Biology 39(1): 30-48.

Huang, X., M. Wolf, M. Ganal, S. Orford, R. Koebner and M. Roder (2007). Did modern plant breeding lead to genetic erosion in European winter wheat varieties? Crop Science 47(1): 343-349.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology 7: 1-44.

Jiang, C., G. O. Edmeades, I. Armstead, H. R. Lafitte, M. D. Hayward and D. Hoisington (1999). Genetic analysis of adaptation differences between highland and lowland tropical maize using molecular markers. Theoretical and Applied Genetics 99(7-8): 1106-1119.

Kimura, M. and J. F. Crow (1963). The measurement of effective population number. Evolution 17: 279-288.

Kingman, J. F. C. (1982). On the genealogy of large populations. Journal of applied probabilty 19A: 27-43.

Lande, R. (1992). Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. Evolution 46(2): 381-389.

Levins, R. A. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America 15: 237–240.

Lockwood, D. R., C. M. Richards and G. M. Volk (2007). Probabilistic models for collecting genetic diversity: comparisons, caveats, and limitations. Crop Science 47(2): 861-868.

Louette, D. (1997). Seed Exchange Among Farmers and Gene Flow Among Maize Varieties in Traditional Agricultural Systems. In: Gene Flow Among Maize Landraces, Improved Maize Varieties, and Teosinte: Implications for Transgenic Maize, pp.4043, Serratos, J.A., Wilcox, M.C., and Castillo-Gonzalez, F. (Eds.), Mexico, D.F., CIMMYT.

Louette, D. (1999). Traditional management of seed and genetic diversity: what is a landrace? Genes in the field: on farm conservation of crop diversity. S.-B. Brush, IPGRI, IDRC, Lewis.

Louette, D., A. Charrier and J. Berthaud (1997). In situ conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. Economic Botany 51(1): 20-38.

Louette, D. and M. Smale (2000). Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. Euphytica 113(1): 25-41.

Lynch, M., M. Perender, K. Spitze, N. Lehman, J. Hicks, D. Allen, L. Latta, M. Ottene, F. Bouge and J. Colbourne (1999). The quantitative and molecular genetic architecture of a subdivided species. Evolution 53(1): 100-110.

Ma, B. L., K. D. Subedi and L. M. Reid (2004). Extent of cross-fertilization in maize by pollen from neighboring transgenic hybrids. Crop Science 44(4): 1273-1282.

Marris, E. (2005). Four years on, no transgenes found in Mexican maize Nature 436: 760.

Maruyama, T. and M. Kimura (1980). Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. Proceedings of the National Academy of Sciences of the United States of America 77(11): 6710-6714.

Matsuoka, Y., Y. Vigouroux, M. M. Goodman, G. J. Sanchez, E. Buckler and J. Doebley (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences of the United States of America 99(9): 6080-6084.

Merila, J. and P. Crnokrak (2001). Comparison of genetic differentiation at marker loci and quantitative traits. Journal of Evolutionary Biology 14(6): 892-903.

Messeguer, J., G. Penas, J. Ballester, M. Bas, J. Serra, J. Salvia, M. Palaudelmas and E. Mele (2006). Pollen-mediated gene flow in maize in real situations of coexistence. Plant Biotechnology Journal 4(6): 633-645.

Metz, M. and J. Futterer (2002). Suspect evidence of transgenic contamination. Nature 416(6881): 600-601.

Nagylaki, T. (1982). Geographical invariance in population genetics. Journal of Theoretical Biology 99(1): 159-172.

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. Journal of Mathematical Biology 41(2): 123-142.

Nevo, E., D. Zohary, A. H. D. Brown and M. Haber (1979). Genetic diversity and environmental associations of wild barley, Hordeum spontaneum, in Israel. Evolution 33(3): 815-833.

Ortiz-Garcia, S., E. Ezcurra, B. Schoel, F. Acevedo, J. Soberon and A.-A. Snow (2005). Absence of detectable transgenes in local landraces of maize in Oaxaca, Mexico (2003-2004). Proceedings of the National Academy of Sciences of the United States of America 102(35): 12338-12343.

Pannell, J. R. and B. Charlesworth (1999). Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. Evolution 53(3): 664-676.

Parzies, H.-K., W. Spoor and R.-A. Ennos (2000). Genetic diversity of barley landrace accessions (Hordeum vulgare ssp. vulgare) conserved for different lengths of time in ex situ gene banks. Heredity 84(4): 476-486.

Paterniani, E. (1969). Selection for reproductive isolation between two populations of maize, Zea Mays L. Evolution 23(534-547).

Paterniani, E. and A. C. Stort (1974). Effective Maize Pollen Dispersal in the Field. Euphytica 23(1): 129-134.

Perales, H., S. B. Brush and C. O. Qualset (2003). Dynamic management of maize landraces in Central Mexico. Economic Botany. 2003 57(1): 21-34.

Perales, H., S. B. Brush and C. O. Qualset (2003). Landraces of maize in Central Mexico: An altitudinal transect. Economic Botany 57(1): 7-20.

Perales, H. R., B. F. Benz and S. B. Brush (2005). Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. Proceedings of the National Academy of Sciences of the United States of America 102(3): 949-954.

Piñeyro-Nelson, A., J. van Heerwaarden, H.R. Perales, J. A. Serratos-Hernández, A. Rangel, M. B. Hufford, P. Gepts, A. Garay-Arroyo, R. Rivera-Bustamante and E.R. Álvarez-Buylla (submitted). Presence of transgenes in Mexican maize landraces of Oaxaca: a reopened case and a precautionary note on analytical and sampling methods for biomonitoring.

Piperno, D. R. and K. V. Flannery (2001). The earliest archaeological maize (Zea mays L.) from highland Mexico: New accelerator mass spectrometry dates and their implications. Proceedings Of The National Academy Of Sciences Of The United States Of America. Feb 98(4): 2101-2103.

Podolsky, R.-H. and T.-P. Holtsford (1995). Population structure of morphological traits in Clarkia dudleyana: I. Comparison of F-ST between allozymes and morphological traits. Genetics 140(2): 733–744.

Pohl, M. E. D., D. R. Piperno, K. O. Pope and J. G. Jones (2007). Microfossil evidence for pre-Columbian maize dispersals in the neotropics from San Andres, Tabasco, Mexico. Proceedings Of The National Academy Of Sciences Of The United States Of America. 104(16): 6870-6875.

Pressoir, G. and J. Berthaud (2004). Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity 92(2): 88-94.

Pressoir, G. and J. Berthaud (2004). Population structure and strong divergent selection shape phenotypic diversification in maize landraces. Heredity 92(2): 95-101.

Pritchard, J. K., M. Stephens and P. Donnelly (2000). Inference of population structure using multilocus genotype data. Genetics 155(2): 945-959.

Quist, D. and I. H. Chapela (2001). Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico. Nature 414(6863): 541-543.

R_DevelopmentCoreTeam (2005). R: A language and environment for statistical computing. h. w. R.-p. o. R Foundation for Statistical Computing.

raven, P. H. (2005). Transgenes in Mexican maize: Desirability or inevitability? Proceedings of the National Academy of Sciences of the United States of America 102(37): 13003–13004.

Reif, J., S. Hamrit, M. Heckenberger, W. Schipprack, H. Maurer, M. Bohn and A. Melchinger (2005). Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years. Theoretical and Applied Genetics 111(5): 838-845.

Reif, J., M. Warburton, X. Xia, D. Hoisington, J. Crossa, S. Taba, J. Muminovic, M. Bohn, M. Frisch and A. Melchinger (2006). Grouping of accessions of Mexican races of maize revisited with SSR markers. Theoretical and Applied Genetics 113(2): 177-185.

Reif, J., P. Zhang, S. Dreisigacker, M. Warburton, G. M. van, D. Hoisington, M. Bohn and A. Melchinger (2005). Wheat genetic diversity trends during domestication and breeding. Theoretical and Applied Genetics 110(5): 859-864.

Reynolds, J., B. S. Weir and C. C. Cockerham (1983). Estimation of the Co Ancestry Coefficient Basis for a Short-Term Genetic Distance. Genetics . 1983; 105 (3): 767 779.

Rice, E., M. Smale and J. Blanco (1998). Farmers' use of improved seed selection practices in Mexican maize: Evidence and issues from the Sierra de Santa Marta. World Development 26(9): 1625-1640.

Saitou, N. and M. Nei (1987). The Neighbor-Joining Method a New Method for Reconstructing Phylogenetic Trees. Molecular Biology and Evolution 4(4): 406-425.

Sanchez, J. J., M. M. Goodman and C. W. Stuber (2000). Isozymatic and morphological diversity in the races of maize of Mexico. Economic Botany 54(1): 43-59.

Sanou, J., B. Gouesnard and A. Charrier (1997). Isozyme variability in West African maize cultivars (Zea mays L). Maydica 42 (1): 1-11.

Slatkin, M. (1977). Gene flow and genetic drift in a species subject to frequent local extinctions. Theoretical Population Biology 12(3): 253-262.

Slatkin, M. (1987). The average number of sites separating DNA sequences drawn from a subdivided population. Theoretical Population Biology 32(1): 42–49.

Slatkin, M. (1991). Inbreeding coefficients and coalescence times. Genetical Research 58(2): 167-175.

Smale, M., A. Aguirre, M. Bellon, J. Mendoza and I. M. Rosas (1999). Farmer management of maize diversity in the Central Valleys of Oaxaca, Mexico: CIMMYT/INIFAP 1998 Baseline Socioeconomic Survey. Working Paper CIMMYT Economics Program.

Smale, M., M. R. Bellon and J. A. Aguirre Gomez (2001). Maize diversity, variety attributes, and farmers' choices in Southeastern Guanajuato, Mexico. Economic Development and Cultural Change 50(1): 201-225.

Spitze, K. (1993). Population structure in Daphnia obtusa: Quantitative genetic and allozymic variation. Genetics . 1993; 135 (2) 367 374.

Strobeck, C. (1987). Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics 117(1): 149-153.

Uribelarrea, M., J. Carcova, M. E. Otegui and M. E. Westgate (2002). Pollen production, pollination dynamics, and kernel set in maize. Crop Science 42(6): 1910-1918.

Vavilov, N. I. (1951). The origin, variation, immunity, and breeding of cultivated plants. Chronica Botanica 13: 1-366.

Vencovsky, R. and J. Crossa (1999). Variance effective population size under mixed self and random mating with applications to genetic conservation of species. Crop Science 39(5): 1282-1294.

Verhoeven, K.-J. F., T.-K. Vanhala, A. Biere, E. Nevo and J.-M. M. Van-Damme (2004). The genetic basis of adaptive population differentiation: A quantitative trait locus analysis of fitness traits in two wild barley populations from contrasting habitats. Evolution . 2004; 58(2): 270 283.

Wade, M. J. and D. E. McCauley (1988). Extinction and Recolonization Their Effects on the Genetic Differentiation of Local Populations. Evolution 42(5): 995-1005.

Wakeley, J. and N. Aliacar (2001). Gene genealogies in a metapopulation. Genetics 159(2): 893-905.

Wang, J., J. Crossa, M. v. Ginkel and S. Taba (2004). Statistical genetics and simulation models in genetic resource conservation and regeneration. Crop Science 44(6): 2246-2253.

Warburton, M. L., X. C. Xia, J. Crossa, J. Franco, A. E. Melchinger, M. Frisch, M. Bohn and D. Hoisington (2002). Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. Crop Science 42(6): 1832-1840.

Weir, B. S. and C. C. Cockerham (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution 38(6): 1358-1370.

Wellhausen, E.J., L.M. Roberts, X. Hernandez, E and P.-C. Mangelsdorf (1952). Races of maize in Mexico. Their origin, characteristics and distribution. Bussey Institution, Harvard University.

Whitlock, M.C. and D.E. McCauley (1999). Indirect measures of gene flow and migration: FST ? 1/(4Nm + 1). Heredity 82(2): 117-125.

Whitlock, M. C. and N. H. Barton (1997). The effective size of a subdivided population. Genetics 146(1): 427-441.

Whitlock, M. C. and D. E. McCauley (1990). Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. Evolution 44(7): 1717-1724.

Wilkinson, H. H. M. (1998). Genealogy and subpopulation differentiation under various models of population structure. Journal of Mathematical Biology 37(6): 535-585.

Wood, D. and J. M. Lenne (1997). The conservation of agrobiodiversity on-farm: Questioning the emerging paradigm. Biodiversity and Conservation 6(1): 109-129.

Wright, S. (1951). The genetical sructure of populations. Ann. Eugen. 15: 323-353.

Summary

A large amount of crop genetic diversity is being maintained in farmers' fields worldwide. The population genetics of traditionally managed landraces is therefore of interest to the conservation of genetic resources. The growing trend towards agricultural modernization and the prospect of introducing genetically modified varieties into centers of origin have increased the need to understand the determinants of genetic structure in landraces of our basic food crops. Patterns of genetic diversity are known to be affected by environmental and geographic factors, but there has been an increasing interest in the role of farmers. Recent years have seen work on both genetic differentiation between seedlots, as well as on the agricultural practices that are expected to influence this differentiation. Unfortunately, few studies have been able to link observed patterns of differentiation to farming practice. The lack of a proper analytical framework has probably contributed to this omission. The population genetics of landraces is complex, with many human and environmental factors affecting the distribution of genetic variation. In this thesis, we aim at achieving a better understanding of the processes that underlie the genetic structure maize landraces in their centre of origin, Mexico. We combine a wide range of theoretical and empirical methods in order to provide explanations for observed patterns of genetic structure. In addition, we use these tools to predict some present and future consequences of seed management by farmers on the genetic identity of landrace populations.

In chapter II, we present a metapopulation model that accounts for several features that are unique to managed maize populations. We developed a coalescence-based model of a metapopulation undergoing pollen and seed flow as well as extinction in the form of seed replacement. Unlike previous models, our model treats seed migration as episodic-, partial replacement from a single source rather than as constant immigration from the entire metapopulation. We showed that this particular form of migration leads to novel results. Contrary to classical predictions, within-deme coalescence time was not invariant to the amount of migrating seed. Genetic structure had a parabolic relationship to the amount of migrating seed instead of showing the expected exponential decrease. In contrast, the effects of seed migration frequency on diversity and structure were in line with classical predictions. We concluded that is impossible to describe seed migration by a single parameter. Genetic structure was shown to depend on deme size when the amount of migrant seed is large. Extinction decreased or increased genetic structure depending on the level of migration and number of demes.

Finally, we demonstrated that higher levels of pollen migration can mask the effects of seed management. This model provides an important first step in our ability to understand the effects of farming practice on the population genetics of maize landraces.

In chapter III, we study the joint role of the environment and humans as determinants of genetic differentiation. We present results on the hierarchical genetic structure in a sample of seedlots in highland and lowland environments in central Mexico. Within-and between village F_{st} and Q_{st} values were used as measures of neutral and agronomic genetic differentiation respectively. We developed and used a new computer model to predict F_{st} in the two environments on the basis of data on local seed management practice and planting patterns. Strong genetic differences were found between highland and lowland maize, for both markers and traits. Three highland villages planted maize of admixed origin, as evidenced by both molecular markers and phenological traits. This suggested that human mediated gene flow from lowland to highland environments has taken place. Molecular differentiation was low for molecular markers but was notably higher in the lowlands. Our model correctly predicted this difference based on lower pollen flow and smaller seedlot sizes in the lowlands. Agronomical traits showed higher differentiation between villages and were probably subject to diversifying selection. Phenological traits showed the strongest differentiation. Field data suggested that different planting dates may explain the observed differences. Phenological differentiation was highest in the transect containing the admixed seedlots, proving that genetic structure may result from the introgression of traits that diverged in a foreign environment.

In chapter IV, we address the issue of genetic erosion in modernized subsistence agriculture. Genetic erosion is thought to occur when modern varieties replace traditional landraces. Actual proof of genetic erosion for any particular area or crop has been rarely found however. A complicating factor in the study of diversity loss in traditional agriculture is the often-noted coexistence between traditional and improved varieties. Moreover, adoption of modern varieties into the traditional seed supply system may blur the distinction between modern and traditional varieties. The inability to classify germplasm into discrete types makes it hard to measure diversity. We addressed these problems by means of a case study on modernized smallholder maize agriculture in southern Mexico. We characterized seedlots obtained from both farmers and commercial seed vendors, for agronomical traits and molecular markers. Farmer interviews were used to distinguish between traditional landraces and recycled modern varieties.

We calculated genetic diversity, defined as the mean differentiation between individual seedlots, for different types of germplasm. Modern germplasm was clearly distinct from traditional landraces. Close resemblance between modern- and recycled modern varieties proved that despite years of independent evolution, recycled varieties have not diverged much from their ancestral stocks. We showed that different traits reveal different levels of relative diversity, demonstrating the inherent difficulty of assessing diversity loss. The group of recycled modern varieties presented the lowest diversity for all measured traits. We could therefore predict that complete replacement of landraces by these varieties will reduce diversity in the traditional seed system. Under current patterns of coexistence however, the distinctness of modern and traditional varieties caused only a limited reduction of genetic diversity.

Chapter V deals with the effects of reproductive and population genetic processes on the probability of detecting inadvertently introduced transgenes in maize landraces. This subject has become relevant since initial findings suggesting contamination of Mexican landraces with transgenes were followed by contradictory results in subsequent years. Theoretical and simulation results showed that certain aspects of maize reproductive biology negatively affect the detection probability. We demonstrated that the strongest potential limitation on detection was caused by the aggregated frequency distribution that is a consequence of farmer-mediated introduction of transgenes. Analysis of recent sampling efforts reveals that detection probabilities may be much lower than previously assumed, partly explaining the recent inconsistent results

Samenvatting

Een grote mate van de genetische diversiteit van landbouwgewassen wordt wereldwijd in de velden van boeren in stand gehouden. De populatiegenetica van op traditionele wijze beheerde landrassen is daarom van belang voor het behoud van genetische hulpbronnen. Door de toenemende modernisering van de landbouw en de mogelijke introductie van genetisch gemodificeerde variëteiten in oorsprongsgebieden van gewassen is de behoefte toegenomen de determinanten van de genetische structuur in de landrassen van onze belangrijkste cultuurplanten te begrijpen. Dat patronen van genetische diversiteit worden beïnvloed door omgevingsfactoren en geografie was al bekend; tegenwoordig is er echter ook toenemende belangstelling voor de rol van boeren in deze. De laatste jaren is werk gepubliceerd over zowel genetische differentiatie tussen gewaspopulaties als over de landbouwmethoden die genetische differentiatie beinvloeden. Helaas bestaan er vrijwel geen studies die de waargenomen patronen van differentiatie met landbouwmethodes verbinden. Het ontbreken van een geëigend analytisch kader heeft waarschijnlijk bijgedragen aan deze omissie. De populatiegenetica van landrassen is complex: de distributie van genetische variatie wordt door vele menselijke en omgevingsfactoren beïnvloed. In dit proefschrift beogen wij een beter begrip te krijgen van de processen die ten grondslag liggen aan de genetische structuur van de maïs-landrassen in hun centrum van oorsprong, Mexico. Wij combineren een verscheidenheid aan theoretische en empirische methoden om geobserveerde patronen van genetische structuur te verklaren. Bovendien gebruiken wij deze gereedschappen voor het voorspellen van enkele recente en toekomstige gevolgen van het management van zaad door boeren voor de genetische identiteit van populaties van landrassen.

In hoofdstuk II presenteren wij een metapopulatiemodel dat rekening houdt met de verschillende kenmerken die uniek zijn voor traditioneel beheerde maispopulaties. Wij ontwikkelden een op coalescentie gebaseerd model van een metapopulatie, waarin zowel zaaduitwisseling en vervanging als pollenverspreiding worden gesimuleerd. In tegenstelling tot bestaande modellen, behandelt ons model zaadmigratie als episodische, partiële vervanging uit een enkele bron in plaats van constante immigratie vanuit de gehele metapopulatie. Wij toonden aan dat deze vorm van migratie tot nieuwe resultaten leidt. De coalescentietijd binnen individuele populaties was niet invariant met betrekking tot de hoeveelheid migrerend zaad, hetgeen afwijkt van wat klassieke modellen voorspellen. Genetische structuur verhield zich parabolisch tot de hoeveelheid migrerend zaad in plaats van de verwachte exponentiële afname te vertonen.

Daarentegen waren de effecten van de frequentie van zaadmigratie op diversiteit en structuur wel in overeenstemming met de klassieke voorspellingen. Wij concludeerden dat het onmogelijk is zaadmigratie door een enkele parameter te beschrijven. We toonden aan dat genetische structuur afhangt van de omvang van de individuele populaties, indien de hoeveelheid migrerend zaad groot is. Zaadvervanging kan de genetische structuur zowel vergroten of verkleinen, afhankelijk van het niveau van migratie en het aantal afzonderlijke populaties. Tenslotte lieten wij zien dat hogere niveaus van pollenmigratie het effect van zaadmanagement kan maskeren. Dit model voorziet in een een belangrijke eerste stap in ons vermogen de effecten te begrijpen van de praktijken in de landbouw op de populatiegenetica van de landrassen van maïs.

In hoofdstuk III bestuderen wij de gezamenlijke rol van omgeving en mens als determinanten van genetische differentiatie. Wij presenteren resultaten van de hiërarchisch genetische structuur in een monster van partijen zaad uit hoog- en laagland omgevingen in centraal Mexico. F_{st} and Q_{st} waarden binnen en tussen dorpen werden gebruikt als maatstaven voor neutrale en respectievelijk agronomisch genetische differentiatie. Wij ontwikkelden en gebruikten een computermodel om F_{st} te voorspellen in de twee omgevingen op basis van data over de lokale werkwijzen met betrekking tot beheer van zaad en plantingspatronen. Sterke genetische verschillen werden gevonden tussen hoogland en laagland maïs, zowel op basis van merkers als van kenmerken. In drie hoogland dorpen werd maïs geplant van gemengde oorsprong, zoals bleek uit zowel merkers als fenologische kenmerken. Dit suggereert dat door de mens bemiddelde genverspreiding van laagland- naar hooglandomgevingen heeft plaatsgevonden. Moleculaire differentiatie was laag voor moleculaire merkers, maar aanmerkelijk hoger in de laaglanden. Door ons model werd dit verschil, gebaseerd op lagere pollen verspreiding en kleinere populatieomvang in de laaglanden, correct voorspeld. Agronomische kenmerken toonden hogere differentiatie tussen dorpen en waren waarschijnlijk onderworpen aan diversifiërende selectie. Fenolgische kenmerken toonden de sterkste differentiatie. Velddata suggereerden dat verschillen in zaaidata een verklarende factor kan zijn voor de waargenomen verschillen. Fenologische differentiatie was was het hoogst in het transect dat de gemengde partijen zaad bevatte. Dit bewijst dat genetische structuur het gevolg kan zijn van de introgressie van kenmerken die buiten het groeigebied zijn gedivergeerd.

In hoofdstuk IV richten wij ons op de kwestie van genetische erosie in gemoderniseerde kleinschalige landbouwsystemen. Genetische erosie wordt geacht plaats te vinden wanneer moderne variëteiten traditionele landrassen vervangen. Feitelijk bewijs van genetische erosie in specifieke gebieden of gewassen is echter zelden gevonden.

Een complicerende factor in de studie van het verlies van diversiteit is de dikwijls opgemerkte coëxistentie tussen traditionele en verbeterde variëteiten. Bovendien, de adoptie van moderne variëteiten in het traditionele systeem van zaadvoorziening kan het onderscheid tussen moderne en traditionele variëteiten verdoezelen. Het onvermogen tot het onderbrengen van germoplasma in discrete typen maakt het moeilijk diversiteit te meten. Wij benaderden deze problemen door middel van een case study over gemoderniseerde kleinschalige maïsteelt in zuidelijk Mexico. Wij karakteriseerden zaad, verkregen van zowel boeren als zaadhandelaren, op agronomische kenmerken en moleculaire merkers. Interviews met boeren werden gebruikt om traditionele landrassen te onderscheiden van gerecyclede moderne variëteiten. Wij calculeerden genetische diversiteit, gedefinieerd als de gemiddelde differentiatie tussen individuele partijen zaad, voor verschillende typen germoplasma. Modern germoplsma was duidelijk onderscheiden van traditionele landrassen. De grote gelijkenis tussen moderne en als landras geadopteerde moderne variëteiten, bewees dat ondanks jaren van onafhankelijke evolutie, geadopteerde moderne variëteiten niet veel van hun oorspronkelijke voorouders zijn gedivergeerd. Wij toonden aan dat verschillende kenmerken verschillende diversiteits niveaus onthullen, daarbij de inherente moeilijkheid van het taxeren van diversiteitsverlies demonstrerend. De groep van geadopteerde moderne variëteiten vertoonde de laagste diversiteit voor alle gemeten kenmerken. Wij konden hierdoor voorspellen dat volledige vervanging van landrassen door deze variëteiten de diversiteit in het traditionele zaadsysteem zal reduceren. Bij de huidige patronen van coëxistentie echter, veroorzaakt de onderscheidenheid van moderne en traditionele variëteiten slechts een beperkte vermindering van genetische diversiteit.

Hoofdstuk V handelt over de effecten van reproductieve en populatie genetische processen op de detectiekans van transgenen die onbedoeld in maïslandrassen zijn geïntroduceerd. Dit onderwerp is relevant geworden, sinds initiële onderzoeksresultaten die wezen op besmetting van Mexicaanse landrassen, werden gevolgd door daarmee tegenstrijdige resultaten in volgende jaren. Theoretische en gesimuleerde resultaten lieten zien, dat de voortplantings biologie van maïs een negatief effect kan hebben op de waarschijnlijkheid van detectie. Wij toonden aan dat de potentieel sterkste beperking van detectie werd veroorzaakt door de geaggregeerde frequentieverdeling, die het gevolg is van de introductie van transgenen door individuele boeren. Analyse van recente steekproeven laat zien dat detectiekansen veel lager kunnen zijn, dan eerder werd aangenomen. Hierdoor kunnen de recente inconsistente resultaten gedeeltelijk worden verklaard.

Resumen

Una gran cantidad de la diversidad genética de las plantas cultivadas es mantenida por los agricultores de todo el mundo. La genética de poblaciones de las razas cultivadas es, entonces un tema de enorme interés para la conservación de los recursos genéticos. La tendencia creciente hacia una modernización agrícola y la perspectiva de introducción de variedades genéticamente modificadas en los centros de origen ha aumentado la necesidad de entender las causas que determinan la estructura genética de nuestros cultivos básicos. Los patrones de diversidad genética se sabe que son afectados por factores ambientales y geográficos pero que ahora incluye un creciente interés por el papel de los agricultores. En los años recientes ha habido mucho trabajo tanto en la diferenciación genética entre los lotes de semillas como en las prácticas agrícolas que se espera afecten esta diferenciación. Desafortunadamente, pocos estudios han podido ligar los patrones observados de diferenciación con la práctica agrícola. La falta de un enfoque analítico adecuado ha sido probablemente una de las causas de este problema. La genética de poblaciones de las razas cultivadas es compleja, con muchos factores humanos y ambientales afectando la distribución de la variación genética. En esta tesis, nuestro objetivo es tener un mejor entendimiento de los procesos que estructuran la variación genética en razas criollas de maíz en el centro de origen, México. Para ello combinamos un amplio rango de métodos empíricos y teóricos para explicar los patrones observados de variación genética. Adicionalmente, usamos estas herramientas para explicar los patrones observados de variación genética de las poblaciones de las razas criollas.

En el capítulo II, presentamos un modelo metapoblacional que explica muchos de los rasgos que son únicos para las poblaciones manejadas de maíz. Desarrollamos un modelo de coalescencia de una metapoblación en la que hay tanto flujo de polen como de semillas así como extinción en forma de reemplazo de semillas. Contrario a los modelos previos, nuestro modelo considera a la migración de semillas como episódica, reemplazando semillas a partir de una fuente en lugar de tener un flujo continuo de toda la metapoblación. Nosotros encontramos que esta forma particular de migración lleva a resultados novedosos. El tiempo de coalescencia dentro de los demes no cambia con la cantidad de semillas migrantes como lo predicen los modelos clásicos. Además la estructura genética tuvo una relación parabólica con la cantidad de semilla migrante en lugar de una relación de reducción exponencial. Por otro lado, los efectos de la migración de semillas en la diversidad y estructura genéticas fueron acordes con las predicciones clásicas.

Nosotros concluimos que es imposible describir la migración de semillas con un solo parámetro. La estructura genética mostró depender en el tamaño del deme cuando la cantidad de semillas migrantes es grande. La extinción disminuyó o incrementó la estructura genética dependiendo del nivel de migración y del número de demes. Finalmente pudimos demostrar que altos niveles de flujo de polen pueden enmascarar los efectos del manejo de las semillas. Este modelo provee un primer paso en nuestra capacidad para entender los efectos de las prácticas agrícolas en la genética de poblaciones de las razas criollas de maíz.

En el capítulo III, estudiamos el efecto conjunto del ambiente y los seres humanos como factores que determinan la diferenciación. En él presentamos resultados acerca de la estructura jerárquica en una muestra de semillas tanto en regiones altas como en ecosistemas de baja altitud en el centro de México. Dentro y entre pueblos, los valores de F_{st} y Q_{st} fueron utilizados como medidas de diferenciación genética neutra y agronómica respectivamente. Asimismo, nosotros desarrollamos y usamos un modelo computacional para predecir F_{st} en los dos ambientes basándonos en datos de manejo local de semillas así como prácticas de siembra. De esta manera encontramos diferencias genéticas muy fuertes entre las tierras altas y las bajas tanto para los marcadores como para los rasgos usados. Tres pueblos de tierras altas sembraron maíz de origen mezclado y esto pudo ser inferido a través de los marcadores moleculares y los rasgos fenológicos. Esto sugiere que ha habido migración mediada por los campesinos de las tierras bajas a las tierras altas. La diferenciación fue baja para los marcadores moleculares pero claramente mayor en las tierras bajas. Nuestro modelo predijo en forma correcta esta diferencia basado en un menor flujo de polen y un menor tamaño de los lotes de semillas en las tierras bajas. Los rasgos agronómicos, por otro lado, mostraron una mayor diferenciación entre pueblos debido probablemente a selección diversificadora. Por último, los rasgos fenológicos mostraron la mayor diferenciación. Datos de campo sugieren que diferentes momentos de siembra pueden explicar las diferencias observadas. La diferenciación fenológica fue la mayor en los transectos que contenían los lotes de semillas mezcladas mostrando que la estructura genética puede ser el resultado de la introgresión de rasgos que divergieron en un ambiente diferente.

En el capítulo IV nos enfocamos al tema de la erosión genética en una agricultura moderna de subsistencia. La erosión genética se supone que ocurre cuando la variedades modernas sustituyen a las razas criollas.

Prueba fehaciente de que ha ocurrido erosión genética en un área particular o en algún cultivo ha sido encontrada muy rara vez. Un factor que complica el estudio de la pérdida de diversidad genética en la agricultura tradicional es la coexistencia de variedades tradicionales y mejoradas. Más aún, la adopción de variedades modernas en la fuente de semillas tradicionales puede obscurecer la distinción entre variedades modernas y tradicionales. Además la incapacidad para clasificar el germoplasma en tipos discretos hace muy difícil medir la diversidad. Nosotros enfrentamos estos problemas usando un estudio de caso en una pequeña propiedad con agricultura de maíz en el sur de México. Caracterizamos lotes de semillas tanto de agricultores como de comerciantes usando marcadores moleculares y rasgos agronómicos. Se hicieron entrevistas para distinguir entre razas criollas tradicionales y variedades modernas recicladas. Calculamos la diversidad genética definida como la diferenciación promedio entre lotes individuales de semillas para diferentes tipos de germoplasma. El germoplasma moderno fue claramente diferente de las razas tradicionales. La similitudes entre las variedades modernas y las modernas recicladas fueron muy claras aún después de varios años de evolución independiente demostrando que las recicladas no han divergido mucho de sus ancestros. Mostramos además que diferentes rasgos muestran distintos niveles de diversidad relativa mostrando la incapacidad inherente de evaluar la pérdida de diversidad. El grupo de variedades modernas recicladas mostraron la menor diversidad para todos los rasgos medidos. Podríamos predecir, entonces, que el reemplazo completo de las razas criollas por estas variedades reduciría la diversidad en el sistema tradicional de semillas. Bajo las condiciones actuales de coexistencia, la distinción de las variedades modernas y tradicionales ha producido una reducción limitada de la variación genética que pudo ser inferida.

El capítulo V se refiere a los efectos de los procesos reproductivos y poblacionales en la probabilidad de detectar transgenes introducidos de forma inconsciente en las razas criollas. Este aspecto se ha vuelto importante desde que se reportaron datos previos que sugieren contaminación en las razas criollas seguidos de otra evidencia contradictoria en años subsecuentes. Resultados teóricos y producto de simulaciones mostraron que la biología reproductiva del maíz puede afectar de forma negativa la probabilidad de detección. Nosotros además demostramos que la limitación potencial mayor en la detección era producida por una distribución de frecuencias agregada que es una consecuencia de la introducción, por parte de los agricultores de los transgenes. Análisis de esfuerzos recientes revelaron que la probabilidad de detección puede ser mucho menor de lo que se había supuesto, resultado que en parte explica los resultados inconsistentes obtenidos.

Acknowledgments

Let it be said, working towards a PhD is a lonely endeavor. By its very nature, thesis research takes one beyond the reach of interest and sympathy of all but a few very close friends and collaborators. It is therefore that those who have provided relief from this sense of loneliness by means of their solidarity and support deserve most credit for the successful completion of this thesis. First and foremost, I wish to thank the love of my life, my soon to be wife and biggest source of joy and understanding. Maru, I doubt if without your magical presence in my life I would have found the strength to bring this work to an end. Mi vida, gracias por tu apoyo incondicional, tu paciencia y tu amor. Te amo. Then there is my father, who has been my most faithful reader, critic and coach. With his continuous expressions of curiosity and support and his relentless (and admittedly biased) faith in my abilities he gave me the motivation and confidence to keep at it. Not to mention his generous contribution to this thesis in the form of the Dutch summary. My mother, to whom I owe much of what I am and who has stood by me silently without receiving much in return save some hurried attempts at conversation. I hope we may talk of other things than my worries now that this is done. Thanks also to my sister Meinske, who has agreed to be at my side at such an important moment even though I was absent from so many events in her life. Thank you Linus, the fact that our friendship has endured the tests of time and distance has been a great source of comfort. I also wish to mention those good friends who might have felt forgotten but were not: Igor, Emiel, Coen, Marije and Joris.

On the professional side of things many people and institutions have made important contributions. I thank my supervisors Fred and Richard for taking me on as a PhD student and bringing this process to a successful conclusion. Fred, I know it hasn't been easy to supervise a PhD student on the other side of the Atlantic. I have appreciated your bursts of attention and criticism. Cheers. Thanks are due to Mauricio Bellon and Julien Berthaud who brought me to CIMMYT. Mauricio deserves special credit, as the socio-economic data used in chapter III were collected by- and thanks to him. This thesis would have not existed without the hard work of many CIMMYT field workers. They have been key in all agronomic work including the taking of measurements. I have learnt much from them and have had the privilege to build on their skill, advice and Judgment. I am especially grateful to the following people: Juan Juárez, Germán Velazquez, Cristobal Almaraz (El Cachiris), Demetrio Soto and Don Francisco.

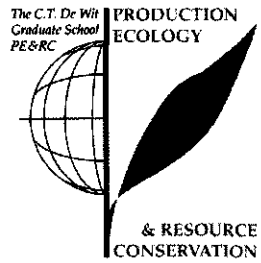
Most labwork was performed by the skilled hands of Marta Hernandez and Maria Zaharieva. Many of the recurrent logistical problems in the laboratory were solved thanks to the efforts of Susana Velazquez. I thank Dagoberto Flores, Alexandro Ramirez and Jon Hellin for their stimulating company in the field and beyond. My thanks also to Marilyn Warburton for having me in her lab, to Jose Crossa for statistical advice, Dave Hodson for GIS support and Ciro Sanchez for keeping an eye on the Tlalti planting. I am indepted to Elena Alvarez-Buylla, Alma Piñeyro and Hugo Perales for their collaboration on the issue of GMOs. By incorporating my work into their project and by sharing their data they have significantly added to the content of this thesis. Hopefully our efforts will soon be rewarded by a publication on this important topic. I also express my gratitude to Montgomery Slatkin for his constructive criticism on my theoretical model and to Daniel Piñero and all the boys and girls from his laboratory for giving me a home after my PhD. I am sincerely grateful to The International Maize and Wheat Improvement Centre (CIMMYT) for providing me with the space, tools and people to make this work a success. The Rockefeller Foundation as well as the Food and Agriculture Organization of the United Nations (FAO) are recognized for their financial support.

My last words of gratitude go towards the numerous smallholder farmers, too many to name, that did not only provide the all-important samples and information for this work but also taught a misguided biologist a thing or two about maize and agriculture. I would like to express my deepest respect to these experts of the most basic, challenging and vital of all fields of human knowledge, that of subsistence.

Curriculum vitae

Joost van Heerwaarden was born on April 12th 1975 in Eindhoven, The Netherlands. During his biology studies at the then named Wageningen Agricultural University, he tried his best to have as little to do as possible with agriculture. His MSc theses ranged from molecular evolution in exotic insects and fish to seed dispersal in tropical rainforests but steered clear of corn and cows. After graduating in the year 2000, he moved to Mexico where he had his first encounter with subsistence agriculture and its importance for significant parts of humanity. After a short flirt with the evolutionary genetics of the common frog in Finland in 2001-2002, he decided that maize landraces presented more satisfying objects of study. It was then that the idea for the present thesis was born. While waiting for sources of funding to materialize, he was given the opportunity to explore the realities of smallholder maize agriculture while acting as coordinator of a small seed relief effort in response to hurricane Isidore in the Yucatan Peninsula. Shortly before completing the latter project, news of funding finally arrived. A grant requested by Mauricio Bellon and Julien Berthaud of the International Maize and Wheat Improvement Center (CIMMYT) had been approved. The great similarity between their project and the proposed thesis research led to the collaboration that made the present work possible. Thanks to the funding provided by the Rockefeller Foundation, three years of research ensued. The experience accumulated during the Rockefeller project led to the writing of a proposal to investigate the effects of improved seed on genetic diversity in Chiapas. It was funded by the FAO in 2005 and 2006. The work performed at CIMMYT from 2004 to 2007 forms the core of this thesis. The part on transgene sampling has been the result of a collaboration with Dr. Elena Alvarez-Buylla of the Ecology Institute of the National Autonomous University of Mexico (UNAM). Academic supervision was sought and provided by Prof. Richard Visser and Prof. Fred van Eeuwijk, of the departments of plant breeding and applied statistics of Wageningen University. Joost van Heerwaarden is currently working as a postdoctoral fellow at the laboratory of genetics and evolution at the Ecology Institute of UNAM, where he hopes to delve deeper into the wonderful world of maize population genetics.

PE&RC PhD Education Certificate



With the educational activities listed below the PhD candidate has complied with the educational requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of Literature (4.2 credits)

- Understanding genetic structure in traditional farming systems (2004)

Laboratory Training and Working Visits (4.3 credits)

- Training in use SSR; CIMMYT (2004)

Post-Graduate Courses (2.6 credits)

- Summer Institute in Statistical genetics, coalescence module; Department Biostatistic University Washington (2007)
- Workshop, using markets to promote the sustainable utilization of crop genetic resources; FAO HQ, Rome (2006)

Deficiency, Refresh, Brush-up Courses (2.8 credits)

- Modern statistics for the life sciences; WUR (2004)

Competence Strengthening / Skills Courses (1.4 credits)

- Self study, programming in C++

Discussion Groups / Local Seminars and Other Scientific Meetings (4.9 credits)

- Project progress report to Rockefeller Foundation oversight committee, other CIMMYT work planning meetings (2004)
- Model presentation and work planning meeting. El Colegio de la frontera Sur (ECOSUR), San Cristobal de las Casas. Chiapas. Other project work meetings at UNAM, Mexico City (2004-2007)
- Cimmyt ABC seminars, science week (2004-2006)

International Symposia, Workshops and Conferences (6.3 credits)

- 46th Maize genetics conference March (2004)
- Taller sobre conservación de cultivos Mexicanos: amenazas y sistemas de monitoreo, workshop (2005)
- SF bio-complexity maize mini-symposium; Mexico City (2006)

Courses in which the PhD Candidate Has Worked as a Teacher

- Population genetics (5 days)