Recombination, Genetic Diversity and Plant Domestication

by

Jeffrey Ross-Ibarra

(Under the direction of James L. Hamrick)

Abstract

The following chapters are the result of my work investigating the roles of two potential genetic preadaptations to domestication: recombination and diversity. Two chapters directly address this possibility, testing hypotheses about the preadaptive value of recombination or diversity. The remaining two chapters extend these results, investigating further aspects of the roles of diversity and recombination in domestication.

In chapter 1, I use data on chiasma frequencies available from almost a century of plant cytogenetical literature to test two hypotheses regarding the role of recombination in plant domestication. I show that recombination does not function as a preadaptation, but is instead selected indirectly by the process of domestication itself.

In my second chapter I investigate genetic diversity and effective population size as potential preadaptations to domestication. I develop a forward population genetic simulation model to test the preadaptive role of effective population size, and compare these results to patterns of genetic diversity in the genus *Zea*. The results suggest that effective population size may well function as a preadaptation to domestication for many crop plants.

In chapter 3 I address the effect of genome size on recombination. I analyze the relationship between genome size and recombination rate in a phylogenetic context, and though I

find a significant positive correlation, I am also able to show that domestication still explains meaningful differences in recombination even after genome size is taken into account.

In my final chapter I look at the effects of domestication on patterns of allozyme diversity and quantitative genetic variation for fruit and leaf size in tomatillo, *Physalis philadelphica*, I find that domestication has had little effect on overall levels of tomatillo diversity but that wild and weedy accessions nonetheless harbor diversity not found in cultivated types. I also show that directional selection on fruit size during tomatillo domestication has had a dramatic effect on patterns of quantitative genetic diversity in cultivated populations, and that this pattern of variation is unexpectedly found for leaf size but not leaf shape characters as well.

INDEX WORDS: recombination, preadaptation, domestication, diversity, Physalis, genome size

Recombination, Genetic Diversity and Plant Domestication

by

Jeffrey Ross-Ibarra

B.A., University of California, Riverside, 1998

M.S., University of California, Riverside, 2000

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2006

Recombination, Genetic Diversity and Plant Domestication

by

Jeffrey Ross-Ibarra

Approved:

Major Professor: James L. Hamrick

Committee: Rodney Mauricio
Kathleen Parker
Andrew Paterson
Daniel Promislow

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2006

Dedication

I dedicate this dissertation to Priscilla Ross, without whom I would not be a scientist, to Stuart Ross, without whom I would not be able to look it up, and to Claudia Ross-Ibarra, without whom I would not be finished.

Acknowledgments

This work was funded by fellowships from the Graduate School of the University of Georgia and a Training Grant from the National Institutes of Health to the Department of Genetics. Funding was also provided by Sigma Xi, The Center for Latin American and Caribbean Studies, and a collection grant from the United States Department of Agriculture.

I would like to thank my parents, Priscilla and Stuart Ross, for proofreading early versions of much of this dissertation, my advisor, James L. Hamrick, for giving me the freedom to explore unusual projects and pursue my own interests, and above all my wife, Claudia, for her endless patience.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature Review

Researchers have long been aware of the dramatic discrepancy between the number of known plant species and the number of species successfully domesticated by humans (Hawkes, 1983; Darwin, 1899). Of the nearly 300,000 estimated plant species, only a few hundred have ever been domesticated (Harlan, 1992). Many workers have explained this discrepancy by arguing that some plant species are easier to domesticate than others, possessing characterisitcs that preadapt them to domestication (Hawkes, 1983; Hammer, 1988; Diamond, 2002; Rindos, 1984; Gornall, 1983). And though the idea of preadaptation to domestication is not unique to plants (e.g. Cameron-Beaumont et al., 2002), it is also not universally accepted (Darwin, 1899).

Most of the characters suggested as potential preadaptations relate to plant life history. Self-compatibility, an annual habit, and weediness have all been considered traits likely to function as preadaptations (Hawkes, 1983; Hammer, 1988; Gornall, 1983). It is argued that these characters adapt plants to initial survival in disturbed habitats associated with human habitation (Rindos, 1984; Hammer, 1988), but to date there are no comparisons which adequately test the predicted distributions of these characters among wild congeners of domesticated taxa. Similarly, arguments for a preadaptive role characters associated with animal-mediated dispersal via ingestion (Hammer, 1988) have also remained untested.

In addition to life history characters, genetic factors have been proposed as potential preadaptations. Diamond (2002) highlights the importance of the genetic architecture of domestication traits, suggesting that traits with a simple genetic basis are easier to select. And while there is some evidence to support this claim (Ross-Ibarra, 2005), it cannot explain

why only one of a closely related group of species with shared characteristics would be domesticated. Another potential genetic preadaptation, polyploidy, is commonly argued to be an important feature in plant domestication, but has nonetheless been shown to have no statistical association with domestication (Hilu, 1993).

The following chapters are the result of my work investigating the roles of two potential genetic preadaptations to domestication: recombination and diversity. Two chapters directly address this possibility, testing hypotheses about the preadaptive value of recombination or diversity. The remaining two chapters extend these results, investigating further aspects of the roles of diversity and recombination in domestication.

In chapter 2, I use data on chiasma frequencies available from almost a century of plant cytogenetical literature to test two hypotheses regarding the role of recombination in plant domestication. The first, initially proposed by Rees and Dale (1974) and later echoed by Burt and Bell (1987) and Otto and Barton (2001), predicts an increase in recombination rate through domestication. The second hypothesis, elaborated by Gornall (1983), argues that high recombination rate should serve as a preadaptation to domestication: because high recombination rate increases response to strong selection, he argued, plants without this advantage would be less likely to be successfully domesticated. The two hypotheses are not mutually exclusive, but, while based on very similar theoretical underpinnings, make distinct and readily testable predictions about patterns of recombination rate in domesticated plants and their wild relatives. Rees and Dale's (1974) hypothesis predicts a higher recombination rate in domesticated plants relative to their wild progenitors, while Gornall's (1983) hypothesis predicts a higher recombination rate in the wild progenitors of domesticated plants relative to their other wild congeners. This study uses data on chiasma frequencies available from almost a century of plant cytogenetical literature to test these two hypotheses. The results not only support the predictions of Rees and Dale and others, but in rejecting Gornall's hypothesis, suggest directions for future research into the possibility of preadaptation to domestication.

After discarding a possible preadaptive role for recombination, in chapter 3 I investigate genetic diversity as a potential preadaptation. The central role of genetic diversity in the process of adaptation has long been well-established, but the idea that a lack of variation may have limited adaptation to domestication has nonetheless been rarely addressed. Though data in the literature provide evidence suggesting a preadaptive role for diversity, I argue that the important causative factor behind observed patterns of diversity is differences in effective population size. In this chapter I develop a forward population genetic simulation model to test the preadaptive role of effective population size in domestication success. The model is explicitly framed in terms relevant to maize domestication, allowing us to create a more realistic model and compare the results directly with data from maize and its congeners. I argue that the results of these simulations suggest that the inferred differences in effective population size among species in the genus *Zea* are sufficient to explain domestication success. Relating these findings with patterns of genetic diversity found across many plant species, I suggest that the effective population size may well function as a preadaptation to domestication for most crop plants.

In chapter 4 I return to recombination rate comparisons, addressing a confounding factor that could have implications for my findings in chapter 2. Burt and Bell's (1987) study of chiasma frequencies not only suggests a relationship between domestication and recombination rate, but also hints at a possible relationship between genome size and recombination rate. A review of the literature finds a general consensus that recombination rate is likely uncorrelated to genome size, but these arguments are based on scarce data. I rework some of the data from chapter 4, combined with information on plant genome sizes, to analyze the relationship between genome size and recombination rate in a phylogenetic context. Though I find a significant positive correlation, I am also able to show that domestication and other life history characteristics still explain meaningful differences in recombination even after genome size is taken into account.

In my final chapter I look at the effects of domestication on patterns of allozyme diversity and quantitative genetic variation for fruit and leaf size in tomatillo, *Physalis philadelphica*. This work is motivated by the opportunity that the genus *Physalis* offers for understanding the possible preadaptive significance of effective population size. Darwin (1899) sees little value in arguments of preadaptation, and argues that virtually any species could be domesticated if subjected to sufficient selection. This idea makes testing the hypothesis of preadaptation problematic for most crop taxa (such as maize), for which it is impossible to know if the wild congeners could have been domesticated had they been subjected to selection. The genus *Physalis* provides a study system that includes domesticated plants, wild relatives, and several species that function as negative controls: though having been exposed to artificial selection for hundreds of generations, these species have never been successfully domesticated. In this study I take the first steps towards characterizing domestication in tomatillo, building the foundation for future work that can explicitly test both the hypothesis that effective population size is a preadaptation to domestication and Darwin's argument that selection alone is required for domestication success. I investigate levels and patterns of genetic diversity at multiple polymorphic allozyme loci and several morphological traits in a common garden study of wild and domesticated tomatillo. I find that domestication has had little effect on overall levels of tomatillo diversity but that wild and weedy accessions nonetheless harbor diversity not found in cultivated types. I also show that directional selection on fruit size during tomatillo domestication has had a dramatic effect on patterns of quantitative genetic diversity in cultivated populations, and that this pattern of variation is unexpectedly found for leaf size but not leaf shape characters as well.

Chapter 2

The Evolution of Recombination under Domestication: A Test of Two Hypotheses[1][2]

[1] Ross-Ibarra, J. 2004. *American Naturalist* 163: 105-112. Reprinted here with permission of publisher.

Abstract

The successful domestication of wild plants has been one of the most important human accomplishments of the last 10,000 yr. Though our empirical knowledge of the genetic mechanisms of plant domestication is still relatively limited, there exists a large body of theory that offers a host of hypotheses on the genetics of domestication. Two of these that have not been addressed concern the role of recombination in the process of domestication. The first predicts an increase in recombination rate through domestication, while the second argues that recombination rate should serve as a preadaptation to domestication. This study makes use of data on chiasma frequencies available from almost a century of plant cytogenetical literature to test these two hypotheses. The results support the hypothesis that domestication selects for an increase in recombination, and in rejecting the preadaptation hypothesis, they suggest directions for future research into the possibility of preadaptation to domestication.

2.1 Introduction

The successful domestication of wild plants has been one of the most important human accomplishments of the last 10,000 years. Changes wrought by domestication enabled human populations to harness and control a food supply tremendously greater than was previously possible. The magnitude of these changes and the rapidity with which they were effected are convincing evidence of the strong directional selection to which these plants were subjected. Though clearly each species is a product of its own unique history, several patterns of morphological and genetic change can nonetheless be discerned among many domesticated plant species (Harlan, 1992; Paterson, 2002). Morphological and physiological changes including gigantism, loss of natural dispersal and defense mechanisms, and loss of seed dormancy have been known for decades and are well documented for a wide variety of crop plants (Smartt and Simmonds, 1995). Genetic evidence of strong selective pressure (e.g. Wang et al., 1999), loss of genetic diversity (Doebley, 1989), and polyploidy (Hilu, 1993) have been shown for

some of the more important crop plants, but our empirical knowledge of the genetics behind the domestication process is still rather limited.

Though the empirical literature on domestication is less developed than could be desired, there exists a large body of theory on the genetics of plant domestication. Both quantitative and population genetics provide a host of hypotheses regarding domestication. Two of these that have not been addressed are concerned with the role of recombination in the process of domestication.

The first, initially proposed by Rees and Dale (1974) and later echoed by Burt and Bell (1987) and Otto and Barton (2001), predicts an increase in recombination rate through domestication. Both theory and simulations show that selection generally favors an increased recombination rate during periods of rapid evolutionary change (Otto and Barton, 1997). High recombination is of most value when selection is strong and genetic variability is limited by negative linkage disequilibrium (Feldman et al., 1996). In domesticated plants this disequilibrium might have been generated by population bottlenecks and genetic drift (Otto and Barton, 1997, 2001; Felsenstein and Yokoyama, 1976), or by negative epistasis among beneficial alleles (Barton, 1995; Charlesworth, 1993).

Gornall (1983) elaborated the second hypothesis, arguing that high recombination rate should serve as a preadaptation to domestication: because high recombination rate increases response to strong selection, he argued, plants without this advantage would be less likely to be successfully domesticated.

The two hypotheses are not mutually exclusive, but, while based on very similar theoretical underpinnings, make distinct and readily testable predictions about patterns of recombination rate in domesticated plants and their wild relatives. Rees and Dale's (1974) hypothesis predicts a higher recombination rate in domesticated plants relative to their wild progenitors, while Gornall's (1983) hypothesis predicts a higher recombination rate in the wild progenitors of domesticated plants relative to their other wild congeners.

This study makes use of data on chiasma frequencies available from almost a century of plant cytogenetical literature to test these two hypotheses. The results not only support the predictions of Rees and Dale and others, but in rejecting Gornall's hypothesis, suggest directions for future research into the possibility of preadaptation to domestication.

2.2 Methods

Recombination data were collected from published studies citing chiasmata frequencies for species of vascular plants. Though it has been suggested that chiasma frequency is not always an accurate measure of recombination in plants (Nilsson et al., 1993), recent studies have shown the direct one-to-one relation between chiasmata and recombination events (King et al., 2002; Knox and Ellis, 2002). Data on other characteristics were obtained from electronic databases, floras, and a variety of other sources. A text file of the recombination data used is available online, and a complete list of the sources consulted is available upon request to the author.

Five parameters were recorded for the recombination data: haploid chromosome number (N), total number of chiasmata per nucleus (XMA), excess chiasmata per nucleus (XS = XMA-N), chiasmata per bivalent (II = XMA/N), and recombination index (RI = XMA + N). Missing data were calculated using the above relationships, or, in the case of haploid number, taken from the Index to Plant Chromosome Numbers (IPCN) database (Goldblatt and Johnson, 2003) or from Darlington (1956). Data from each study were entered into the database as a species average, weighted by the number of plants or cells of each species examined in the study. Species data from multiple publications were averaged to give a single value in the final data set. If entries for the same species differed in one of the other characteristics recorded, however, those entries were not averaged but instead kept separate in the data set. Data were recorded separately for male and female meioses; only data from male meiosis were included in this study. Whenever possible, chiasma frequencies at metaphase were used. Likewise, data from plants with supernumerary chromosomes were excluded whenever

possible. Nonetheless, sources that included only data from other phases of meiosis or from plants with supernumerary chromosomes were included in the database.

Each species was classified for five additional characteristics: ploidy level, life form (annual or perennial), mating system (selfing, mixed, or outcrossing), weediness (weed or not), and domestication status (wild, cultivated, or domesticated). Data on ploidy level, life form, and mating system were occasionally available from the original source, but were otherwise gleaned from a wide range of sources. Plants were considered weeds if named as such by three or more sources as cited in the Global Compendium of Weeds (Randall et al., 2002). The domestication status provided by the original source was used whenever available. If not available, various sources were consulted to determine the appropriate status; the majority of the determinations, however, were made using Smartt and Simmonds (1995). Plants grown horticulturally, and plants for which no clear evidence of domestication was available (mostly forage grasses and relatives of domesticated taxa) were considered cultivated but not domesticated for purposes of this study.

Statistical analyses took the form of an across-species analysis of variance of the entire data set followed by pairwise comparison of specific groups. The raw data for haploid number and chiasma frequency were significantly non-normally distributed and were transformed using a Box-Cox transformation for use in the analysis of variance and pairwise tests.

To test the importance of domestication in determining recombination rate, a stepwise regression analysis was performed on the entire data set. Life form, mating system, weediness, domestication, and all two-way interactions were included as variables in the initial model, but all three- and four-way interactions were ignored. More direct tests of the effect of domestication were carried out by making pairwise comparisons of domesticated species to their wild progenitors, and of the progenitors to their nearest congeners. These were compared using both a standard paired t-test as well as a Wilcoxon matched pairs test of the untransformed data. To test for possible correlations between these pairwise differences and the other species characteristics, the sets of (transformed) pairwise differences were then

analyzed by two stepwise regressions, again ignoring all three- and four-way interaction terms. The first regression analyzed the effect of the other species characteristics on the size of the difference in recombination rate between paired taxa (e.g. whether selfing taxa have smaller differences than outcrossing taxa). The second looked for correlations between a change in these species characteristics as a result of domestication and the difference in recombination rate (e.g. whether a change from selfing to outcrossing is correlated with larger differences in recombination than no change in mating system). Statistical calculations were carried out using Statistica 5.5 (Statsoft Inc. 2000).

2.3 Results

Data were collected for 601 species of vascular plants from 124 genera and 37 families. After elimination of incomplete cases, the across-species analyses of chiasma frequency were limited to a sample size of 196, including 46 domesticated species. Species used for pairwise comparisons are listed in table 2.1.

In order to separate the effect of chromosome number in determining chiasma frequency, it was desirable to use a measure of chiasma frequency independent of haploid number and ploidy level. In spite of claims that excess chiasmata is independent of chromosome number Burt and Bell (1987); Koella (1993), all measures of chiasma frequency in this data set are significantly correlated with haploid number (Pearson product-moment correlation of >0.5 for all measures besides chiasmata per bivalent, significant at the $p<0.05$ level), and all but chiasmata per bivalent are significantly influenced by ploidy level ($p<0.0001$ in all cases). Chiasmata per bivalent is least correlated with haploid number (Pearson product-moment correlation of -0.20) and is independent of ploidy level ($p>0.27$). Though only data using chiasmata per bivalent are reported here, analyses performed using the other measures do not differ qualitatively from those presented.

Results from an initial across-species regression analysis are shown in table 2.2. Though domestication is not the sole determinant of recombination rate in the final model, it is

Table 2.1: Species used for pairwise analyses. Numbers refer to data points in Figures 2.2-2.3 and taxa in table 2.3.

No.	Domesticate	Progenitor	Congener	Family
1	*Allium schoenoprasum*	*A. schoenoprasum*	*A. touricola*	Liliaceae
2	*Avena sativa*	*A. insularis*	*A. murpheyi*	Poaceae
3	*Capsicum baccatum*	ssp. *baccatum*	*C. chacoense*	Solanaceae
4	*Cicer arietinum*	*C. reticulatum*	*C. echinospermum*	Fabaceae
5	*Cucumis melo*	ssp. *agrestris*	*C. sagitattus*	Cucurbitaceae
6	*Cucumis sativus*	ssp. *hardwickii*	*C. pustulatus*	Cucurbitaceae
7	*Daucus carota*	ssp. *carota*	–	Apiaceae
8	*Helianthus annuus*	*H. annuus*	*H. argophyllus*	Asteraceae
9	*Hordeum vulgare*	*H. spontaneum*	*H. bulbosum*	Poaceae
10	*Lactuca sativa*	*L. serriola*	–	Asteraceae
11	*Lens culinaris*	ssp. *orientalis*	*L. ervoides*	Fabaceae
12	*Linum usitatissimum*	*L. bienne*	*L. africanum*	Linaceae
13	*Lycopersicon esculentum*	ssp. *cerasiforme*	*L. hirsutum*	Solanaceae
14	*Oryza glaberrima*	*O. barthii*	*O. longistaminata*	Poaceae
15	*Oryza sativa*	*O. nivara*	*O. rufipogon*	Poaceae
16	*Pennisetum glaucum*	*P. violaceum*	*P. squamulatum*	Poaceae
17	*Phaseolus vulgaris*	var. *aborigineus*	–	Fabaceae
18	*Pisum sativum*	ssp. *eliatus*	–	Fabaceae
19	*Secale cereale*	*S. cereale*	*S. afghanicum*	Poaceae
20	*Setaria italica*	*S. viridis*	*S. verticillata*	Poaceae
21	*Solanum melongena*	*S. incanum*	*S. virginianum*	Solanaceae
22	*Sorghum bicolor*	ssp. *arundinaceum*	*S. sudanense*	Poaceae
23	*Triticum turgidum*	ssp. *dicoccoides*	*T. urartu*	Poaceae
24	*Vicia narbonensis*	*V. narbonensis*	*V. serratifolia*	Fabaceae
25	*Vigna radiata*	ssp. *sublobatus*	*V. riukiuensis*	Fabaceae
26	*Zea mays*	ssp. *parviglumis*	*Z. diploperennis*	Poaceae

Table 2.2: Regression analysis of entire data set for chiasmata per bivalent. Factors showing no data were removed from the model because their corresponding p-values were greater than 0.1.

	DF	SS	MS	F	P
Intercept	1	28.35	28.35	322.96	0
Domestication	2	0.55	0.28	3.14	0.05
Mating	2	0.64	0.32	3.63	0.03
Weediness	1	0.03	0.03	0.32	0.57
Life-Form	1	0.06	0.06	0.63	0.43
Weediness X Life-Form	1	0.43	0.43	4.94	0.03
Domestication X Weediness					
Domestication X Life-Form					
Domestication X Mating					
Weediness X Mating					
Life-Form X Mating					
Error	191	16.77	0.09		

clearly significant. A post-hoc analysis reveals that although cultivated and wild plants are not discernibly different from each other, domesticated plants have a significantly higher recombination rate than either of the former (Scheffe's test, $p<0.03$ for wild and $p<0.01$ for cultivated, Fig. 2.1), providing support for the hypothesis that domestication increases recombination rate. It is worth noting that none of the other characteristics (mating system, life form, etc.) interact significantly with domestication, suggesting that the role these have played in the effect of domestication on recombination rate is relatively insignificant. Including the progenitors of crop plants in the regression model as a category distinct from other wild plants adds no meaningful information; the relative sample size of the category is small and the standard error such that it cannot be distinguished from either other wild or domesticated taxa.

By performing an across-taxa regression for genera and families, the data make possible a test for preadaptation at these taxonomic levels. Were preadaptation acting at these higher levels, genera or families that gave rise to domesticated taxa would be expected to have

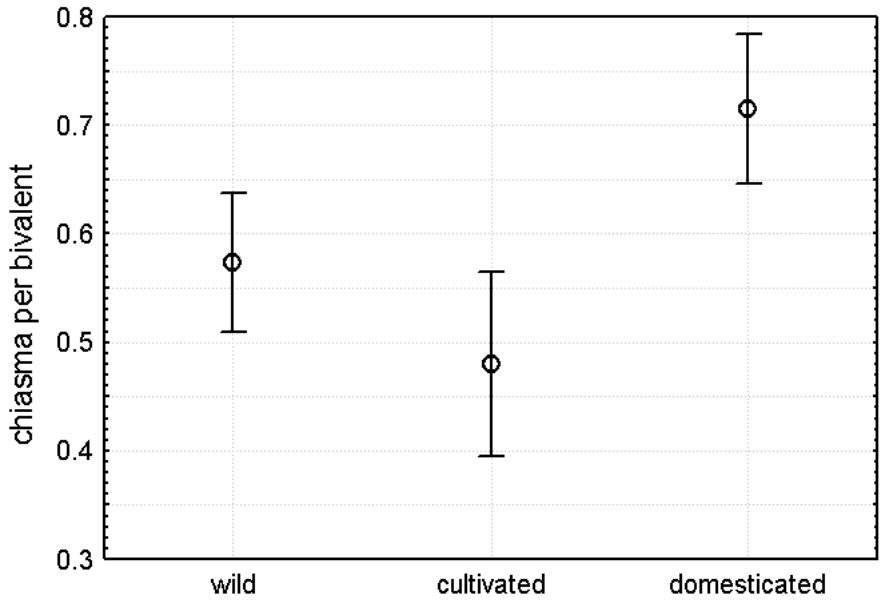


Figure 2.1: Observed means and 95% confidence intervals of the 3 categories of domestication. Transformed data shown.

higher recombination rates than genera or families that did not. Such an analysis reveals no significant effect ($p>0.7$ for both families and genera; results do not differ if data from domesticated species are included in the analysis), suggesting that preadaptation is not an important factor at the genus or family level.

A more direct test of the role of recombination is a pairwise test of the actual taxa involved. These are shown graphically in Figures 2.2 and 2.3, and p-values for these tests are reported in table 2.3. Though the overall data conform reasonably well to normality after transformation, the set of transformed pairwise comparisons (both the raw data and the set of differences themselves) is still markedly non-normal. While the t-test is generally robust to non-normality, a nonparametric test is perhaps more appropriate for these analyses; results for both a standard paired t-test and the Wilcoxon's matched pairs test are reported in table 2.3.

Table 2.3: Number of species pairs, taxa excluded, average difference in chiasma per bivalent and its standard error, and p-values associated with domesticate-progenitor (D-P) and progenitor-congener (P-C) pairwise comparisons. A paired t-test was performed on the transformed data, and a Wilcoxon matched pairs test was calculated using non-transformed data. The numbers of taxa excluded in each comparison refer to table 2.1, and the rationalization for each comparison is explained in the text. The mean standard deviation of replicate measurements of chiasma per bivalent within a species was 0.140.

	D-P	D-P	D-P	P-C	P-C
Species pairs	26	22	20	22	20
Taxa excluded	none	7,10 17,18	7,10,14 17,18,24	7,10 17,18	7,10,14 17,18,24
Avg. difference (s.e)	0.095 (0.044)	0.097 (0.046)	0.117 (0.045)	0.048 (0.083)	0.031 (0.037)
Paired T	0.02	0.02	0.01	0.28	0.21
Wilcoxon	0.03	0.03	0.02	0.19	0.16

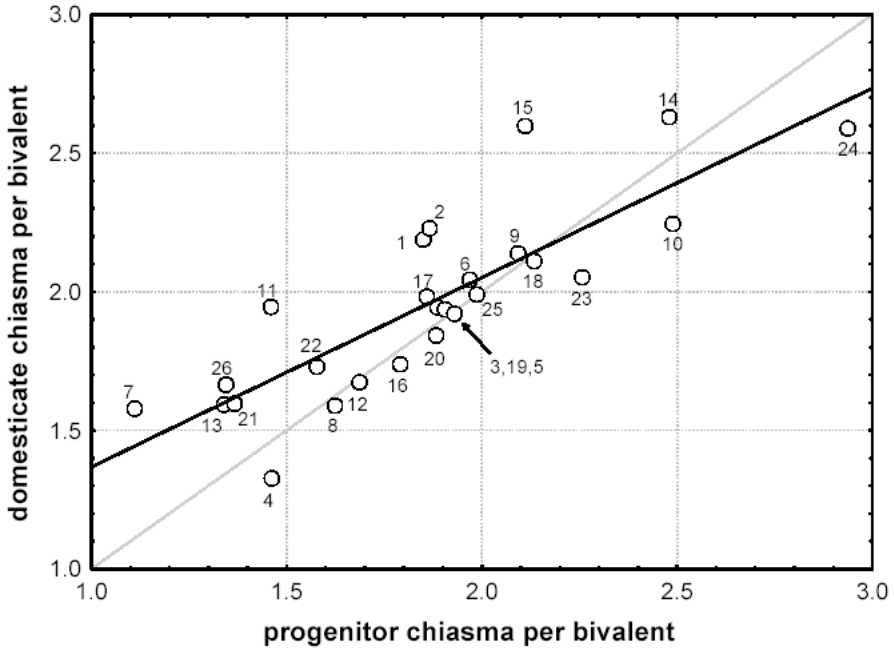


Figure 2.2: Pairwise comparison of domesticated taxa to their nearest congener. In each case the chiasmata per bivalent of the domesticate plotted against the chiasmata per bivalent of the progenitor. Numbers refer to species in table 2.1.

The comparison of domesticates to their wild progenitors is shown in Figure 2.2 . Domesticated taxa in general show a significant increase in recombination when compared to their wild relatives (table 2.3). Hexaploid domesticated oats (*Avena sativa*) differ in ploidy from their most direct wild progenitor, the tetraploid *Avena insularis*. Given that chiasmata per bivalent is independent of ploidy level, this difference should not affect the results; however, even if *Avena* is removed from the list the results do not meaningfully change.

Of the 26 species pairs from table 2.1 compared above, recombination data for both a progenitor and a congener were available in only 22 cases. For three of the progenitor taxa (*Cucumis*, *Pennisetum*, and *Triticum*), the closest available congener was of different ploidy level; removal of these taxa once again does not change the results. Two cases (*Oryza* and *Vicia*) were found to be statistical outliers (Grubb's test for outliers, $p<0.05$). In both cases, the congeners used, though the closest relative of the progenitor, are dramatically different from other species in their genera and probably should be excluded as misrepresentative. Nonetheless, results from both comparisons (the original 22 and the 20 non-outliers) are shown in table 2.3 and are plotted on Figure 2.3. The comparison does not detect a significant difference between progenitors and their congeners, thus failing to support Gornall's hypothesis of preadaptation. Though the number of taxa analyzed is small, it is unlikely that any large difference was missed by this test. Using the same reduced set of 22 or 20 taxa to make the paired domesticate-progenitor comparison still produces a statistically significant result (table 2.3), indicating that even if recombination rate is of preadaptive value, its effect is less than that of domestication in increasing recombination rate via selection. Moreover, a post-hoc power analysis (Hintze, 2001) based on the 20 progenitor-congener pairs does not reveal any real lack of power (80% power to detect a difference of 0.061 in the transformed data, a difference less than that found in any of the comparisons of domesticates and progenitors).

Finally, data on mating system (outcrossing, mixed, inbreeding), weediness, and life form (annual vs. perennial) were used to determine if these characters, or changes in these char-

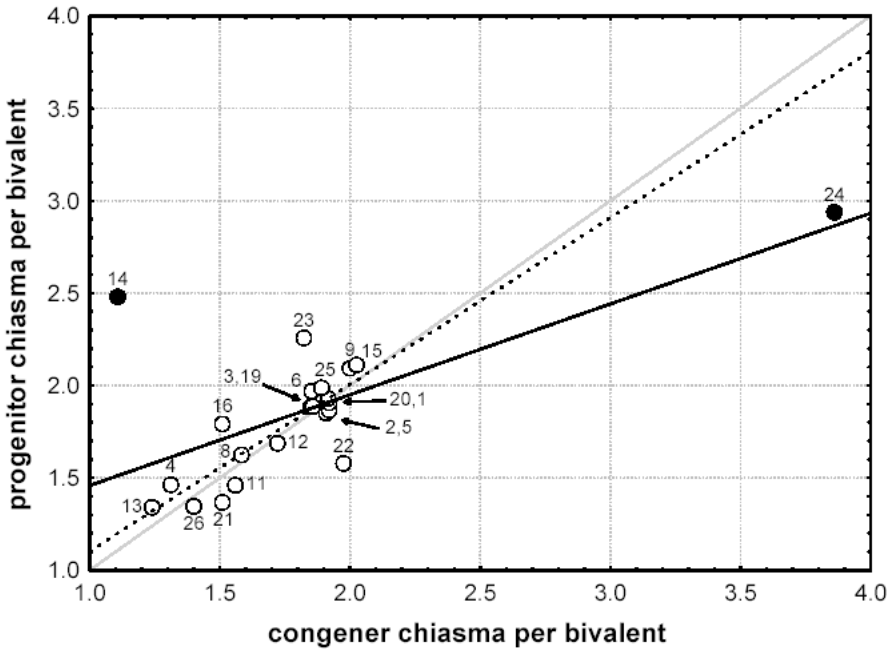


Figure 2.3: Pairwise comparison of progenitor taxa to their nearest congener. In each case the chiasmata per bivalent of the progenitor plotted against the chiasmata per bivalent of the congener. Numbers refer to species in table 2.1. The solid line represents the regression when all data points are included. The dotted line is the regression when the two outlier taxa (shown as solid circles) are removed from the analysis. Numbers refer to species in table 2.1.

acters, are correlated with differences found in the pairwise tests. None of the ANOVAs of pairwise differences in recombination rate with regard to these species characteristics or changes in these characteristics was significant ($p>0.10$ in all cases) for either of the comparisons (domesticate-progenitor or progenitor-congener). The increase in recombination rate due to domestication does not seem to be affected by any of the other traits studied or changes in these traits, nor does the lack of preadaptation seem to be explained by these other characters.

2.4 Discussion

The results from both an across-species analysis and a paired comparison of domesticates to their progenitor taxa strongly suggest that the domestication process generally increases the recombination rate of a species. Domesticated taxa show a higher overall recombination rate than non-domesticated taxa (Fig. 2.1), and pairwise comparison to their progenitors reaffirms the result. Though chiasmata per bivalent is undoubtedly the most appropriate measure of recombination for this analysis, it is comforting to note that all measures tested showed similar results.

In spite of the data available, the literature includes few observations that domestication might affect chiasma frequencies. Though several studies have published data on both a domesticate and its progenitor, very few authors have noted the difference in recombination rate (but see Koul and Hamal (1989)), and none have offered a plausible explanation. In addition to these few empirical papers, a study by Burt and Bell (1987) on recombination rate and lifespan in mammals has occasionally been cited as evidence of the effect of domestication on recombination rate (Koul and Hamal, 1989; Otto and Barton, 2001). While their data do show a statistically significant increase in excess chiasma among several domesticated mammal species, their study does not include any of the progenitors of these species, and the effect disappears if the analysis is done using measures of recombination shown to be less strongly correlated with chromosome number (data not shown). With the exception of

the limited work mentioned above, then, the present study is the first to conclusively show the general effect of domestication on recombination rate.

In contrast to the lack of empirical work, there is a large body of theory that would predict a correlation between domestication and a change in recombination rate. Much effort has been devoted to determining the potential sources of the negative linkage disequilibria and the conditions under which recombination is favored. It is widely agreed, however, that strong directional selection (Feldman et al., 1996), especially at multiple loci (Otto and Barton, 1997) or in concert with genetic drift (Felsenstein and Yokoyama, 1976; Otto and Barton, 2001) can generate negative disequilibria sufficient to select for increased recombination. And while it has been suggested that introgression from wild relatives could potentially select for a decreased recombination rate in domesticates (Lenormand and Otto, 2000), the majority of the conditions provided by the process of domestication (new environment, strong directional selection at multiple loci, and at least in some cases small population sizes) concur with those thought to select for recombination.

Theoretical analysis offers many scenarios for the evolution of recombination; it also cautions of the negative consequences of recombination rate. Not only does recombination break down negative disequilibria among beneficial alleles, but it disrupts positive disequilibria among them as well, potentially splitting up adaptive gene complexes (Barton, 1995). This 'recombination load' is thought to limit selection for increased recombination and effectively set a threshold above which higher recombination is selected against. Selection against excessive recombination would predict a regression slope of less than one between the recombination rate of a domesticate and that of its progenitor. Such a relation is clearly seen in Figure 2.2 ($p<0.05$). Interestingly, the regression line crosses the 1:1 line of equality at just above 2 chiasmata per bivalent, a number suggested by Kondrashov (1988) as a potential limit beyond which the effects of recombination become detrimental. Perhaps most significantly, a slope of less than one is not in keeping with an alternative explanation for

the association between domestication and high recombination rate: the idea that increased homozygosity, and not selection, is the primary cause of increased recombination.

While the results obtained strongly support the hypothesis that domestication selects for increased recombination rate, they offer no support for the hypothesis that recombination rate is an important preadaptation to domestication. Pairwise comparisons reveal that not only do the wild progenitors of crop plants generally possess lower recombination rates than their domesticated descendants, but also that they are not discernibly different from their wild congeners. Furthermore, if recombination functions as a preadaptation to domestication, constraints due to recombination load should be evident as recombination increases in the congeners of domesticated species. As recombination increases in congeneric taxa, selection in favor of progenitors with higher recombination should decrease, resulting in a slope of less than one as seen in Fig. 2.2. Though this relation is in fact seen in Figure 2.3, the removal of the two outlier taxa reveals that it is probably artefactual; the remaining tightly-clustered comparisons do not reveal any such relation (the regression line is not different from one). Likewise, across-taxa comparisons at the genus and family levels find no effect of preadaptation, and there are no correlations with mating system, life form, or weediness that explain the lack of an effect.

Though we are beginning to form an idea of the effect that domestication has had on the genetics of plant species, we are still far from understanding the role genetics has played in determining which species were successfully domesticated. It seems implausible that ecological factors alone (mating system, life form) could explain the impressive discrepancy between the 250,000 species of flowering plants and the few hundred species of domesticates (Hawkes, 1983; Diamond, 2002). Yet the very extent to which humans have modified or utilized the botanical diversity of their environments almost requires a mechanistic explanation for this discrepancy, and it seems quite plausible that genetic factors could have played a central role in the success or failure of the domestication process. While the genetic bases of many important domestication traits have been identified (Paterson, 2002), we will not know how

these genes have influenced the success of domestication until they have been studied in many wild species as well. Other hypothesized genetic preadaptations, such as polyploidy (Hilu, 1993), have been shown to be unimportant in determining the successful domestication of plant species, and the present analysis suggests that recombination rate is likewise of little importance. Clearly, much work is still needed before we can approach a more complete understanding of the genetic mechanisms involved in the domestication process.

2.5 Acknowledgments

I would like to thank the patience, advice and guidance of D. Promislow, S. Otto, and J. Hamrick, as well as the helpful comments of many reviewers, including R.S. Cornman, M. Eubanks, M.J. Godt, K. Parker, A. Paterson, M.C.Q. Provance, P. Ross, S. Ross, and two anonymous reviewers. Thanks are also due to A. Burt for kindly sharing his data. This work was funded in part by a fellowship from the University of Georgia.

Chapter 3

Diversity and effective population size as a preadaptation to domestication[1]

[1]Ross-Ibarra, J. and J.L. Hamrick. To be submitted to *Journal of Heredity*

Abstract

One of the most pervasive mysteries in the study of crop origins is why so few species have been successfully domesticated, and researchers have long questioned what characteristics, if any, may have allowed some plants to successfully respond to domestication. This idea, that some species may have been preadapted to domestication, continues to gain following, and numerous characters have been identified as potential preadaptations. In this paper, we investigate the potential for genetic diversity to function as a preadaptation to domestication. We first show that patterns of sequence diversity in the genus *Zea* conform to predictions of the preadaptation hypothesis, and then use *Zea* as a framework to build simulations of the domestication process. Our simulation results reveal that differences in effective population size — the factor responsible for differences in genetic diversity — are sufficient to explain differential domestication success of species in the genus. We argue that these results, combined with broad-scale patterns in the literature, suggest that genetic diversity and effective population size may generally serve as preadaptations to domestication in plant species.

3.1 Introduction

One of the most pervasive unknowns in the study of crop origins is why so few species have been successfully domesticated. Of the approximately 300,000 species of flowering plants, only a few hundred can be considered as domesticates (Harlan, 1992). Since Darwin, researchers have long questioned why more plants were not domesticated and what characteristics, if any, may have allowed some plants to successfully respond to domestication Darwin (1899); Hawkes (1983); Gornall (1983); Gepts (2004). This idea — that some plants may have been preadapted to domestication — continues to gain following, and numerous characters have been identified as potential preadaptations Diamond (2002); Gornall (1983); Hammer (1984); Gepts (2004). And while convincing arguments have been made for a preadaptive role for some of these characters (Hawkes, 1983; Hammer, 1988), others have been shown to have

no association with domestication success (Hilu, 1993; Ross-Ibarra, 2004), and our overall ability to explain why so few species have been domesticated has advanced little.

An intriguing possibility that has received relatively little attention is the idea that genetic diversity itself could function as a preadaptation to domestication. The central role of genetic diversity in the process of adaptation has long been well-established. Darwin (1899) recognized the fundamental importance of genetic diversity for the process of adaptation, and early population geneticists formalized the relationship between diversity and adaption for relatively simple scenarios Fisher (1930); Wright (1932). The fields of conservation and quantitative genetics are based on the implicit idea that variation is necessary for future adaptation (Avise, J.C. and J.L. Hamrick, 1996; Lynch and Walsh, 1998) and some authors have even argued that a lack of variation may generally be a limiting factor for adaptation (Bradshaw, 1991). Nonetheless, the idea that a lack of variation may have limited adaptation to domestication has been mentioned by only a few authors. Buckler et al. (2001) briefly cite the potential importance of genetic diversity, Gepts (2004) mentions diversity along with numerous other potential preadaptations, and Ross-Ibarra (2005) touches on the idea in the context of standing variation. The only real discussion of diversity as a preadaptation, however, is by Darwin himself. While accepting the importance of variation as the raw material of selection, Darwin (1899) explicitly rejects the possibility that plant (or animal) species might have failed as domesticates because they lack sufficient variation. Instead, he argues that failure is due primarily to the lack of sufficient human-mediated selection.

Moreover, there is little theoretical support for the idea that overall genetic diversity would have much effect on domestication. Variation would only be preadaptive if it occurred at loci important to domestication — seed dormancy, dispersal, fruit size, etc. Furthermore, population genetic theory argues that the presence of adaptive (domesticate-type) alleles in wild populations and their eventual fate after artificial selection begins should be largely determined by effective population size. The chance of loss and average expected frequency of future adaptive alleles segregating as neutral or nearly neutral variants in the wild population

is entirely determined by the mutation rate and effective population size (Ewens, 1972). Likewise, once artificial selection begins, the effectiveness and efficiency of selection at fixing adaptive variants is predominantly determined by the strength of selection and effective population size, a result which holds regardless of whether an adaptive allele originates as a novel mutation or comes from standing variation (Hermisson and Pennings, 2005). It thus stands to reason that, barring large differences in mutation rate, species with larger effective population sizes should be more likely to maintain adaptive variants and more likely to fix adaptive variants during domestication. And while species with higher overall levels of diversity would be more likely to be diverse at a given locus important for domestication, higher overall levels of diversity are a result of a higher effective population size. Rather than overall genetic diversity, then, it seems that effective population size might have the potential to function as a preadaptation to domestication.

In this paper we develop a simulation model to test the role of effective population size on domestication success. Because of the wealth of available genetic data available for maize and its congeners, we explicitly frame our model in terms relevant to maize domestication. The genus *Zea* is one of the few taxa for which extensive species-wide sequence data is available for a domesticated species and multiple wild relatives, thus allowing estimation of relative effective population sizes. Additionally, the domestication process in maize has been relatively well characterized: the timing, demography, strength of selection and genomic architecture (Eyre-Walker et al., 1998; Wang et al., 1999; Doebley, 2004; Tenaillon et al., 2004; Wright et al., 2005) involved in the maize domestication have all been investigated, allowing for a more realistic formulation of a domestication model. We utilize this wealth of information to construct a forward population genetic simulation model that tests the importance of effective population size for several possible parameters of the domestication process.

3.2 Methods

3.2.1 Diversity in *Zea*

We collected sequences from nuclear loci for which data were available for the wild progenitor of domesticated maize, *Z. m. parviglumis,* and at least one congener. The genus *Zea* consists of the annual *Z. mays*, the wild annual diploid *Z. luxurians*, and two wild perennials: the diploid *Z. diploperennis* and the autotetraploid *Z. perennis*. In addition to the domesticated subspecies *mays*, *Z. mays* itself consists of three wild annual subspecies: *Z. m. parviglumis*, *Z. m. mexicana*, and *Z. m. huehuetenangensis*. Samples and sequence data are scarce for this last taxon, which is not included in these analyses. Though some authors have recently relegated populations of *Z. luxurians* from Nicaragua to a novel species (Iltis and Benz, 2000), we consider these populations as part of *Z. luxurians*. The list of loci, taxa, and number of individuals sampled is shown in Table 3.4. Descriptions of the individual loci can be found in Tenaillon et al. (2004) and Moeller and Tiffin (2005). Additional sequence data not publicly available were generously provided by B. Gaut.

For each locus, we aligned the combined data set of all taxa using ClustalW software (Thompson et al., 1994) and then edited the alignment by hand. We then estimated the number of segregating sites (S), the average number of pairwise differences between sequences θ_π, and the value of the Watterson (1975) estimator θ_w for each taxon. Analyses were performed separately for all sites, nonsynonymous sites and silent sites (noncoding and synonymous sites combined). All polymorphism analyses were performed using the software SITES (Hey and Wakeley, 1997). We also conducted a suite of common tests of neutrality for all loci. We used the program DNASP (Rozas et al., 2003) to calculate Tajima's D (Tajima, 1989), Fu and Li's D* and F* (Fu and Li, 1993), the ratio of synonymous to nonsynonymous substitutions (ka/ks), and applied the McDonald-Kreitman test (McDonald and Kreitman, 1991) to each locus for each taxon. Following Wright et al. (2003), a multilocus estimate of θ_w was also calculated for each taxon. This method utilizes the recursion equations of Hudson

(1990) to numerically estimate the likelihood curve for θ_w assuming a constant mutation rate among loci and no intragenic recombination. Significant differences among these multilocus estimates can be tested using the χ^2 approximation of the relative log likelihood curves.

3.2.2 Domestication model

3.2.2.1 Genome and genetic architecture

We simulated a maize population during its domestication bottleneck. We assumed that effective population size during the bottleneck is constant and proportional to the effective population size previous to the bottleneck; we can thus model among-species differences in effective population size by changing the effective population size in the domestication bottleneck and monitoring domestication success.

We modeled a maize genome of 10 chromosomes, each of 100 cM length. To cluster genes within the genome, we placed a random normal $N(1, 2)$ genes on each chromosome for each simulation. This process was repeated for each chromosome until there were at least 10 domestication genes in the genome. A location L along the chromosome was then drawn randomly and each gene placed randomly at a position determined by $N(L, 0.0100)$. Also placed randomly along each chromosome was a single recombination locus. Each chromosome thus had one recombination gene and 0-1 clusters of domestication genes.

3.2.2.2 Mutation

All genes were modeled as bi-allelic loci; domestication genes were either wild-type or domesticate-type, and recombination genes either added to recombination rate or did not. We assumed a mutation rate of 6.5×10^{-9} mutations per base pair per generation, equivalent to values estimated at the Adh gene in maize (Gaut et al., 1996). Genes were assumed to be 4kb long, and mutations at one of every three base pairs gave rise to a domesticate-type allele. This is roughly equivalent to assuming that domesticate-type alleles are loss of function alleles and any amino acid change can cause the domesticate phenotype, or that

domesticate-type alleles are formed from regulatory changes and that the regulatory region is approximately 1/3 the length of the entire locus. Assuming that back mutations can only happen if the original mutation is reversed, back mutations were modeled at a frequency of 1/3 the forward mutation rate. Mutation rate was constant across loci, and identical for domestication and recombination loci.

3.2.2.3 Recombination and selection

Each cM along a chromosome was further divided into 100 possible recombination sites, providing for 10,000 potentially recombining units along each chromosome. We assumed no intragenic recombination. The number of chiasmata placed on a chromosome during meiosis was drawn separately for each chromosome from a truncated Poisson with a genome wide mean of $\lambda = 0.5$ and a minimum of 1 crossover per generation (producing an approximate mean of 1.27 chiasmata per chromosome). This distribution changed as a result of changes in allele frequency at recombination loci; each mutant allele added 0.05 to the genome-wide mean for a given individual. Recombination loci were assumed to be completely codominant.

We assumed soft reproductive selection in which individuals with no domestication alleles have a fitness of 1, and domestication alleles add to reproductive fitness; all individuals have equivalent viability. Fitness within a locus was additive, with genotypes 00, 01, and 11 having fitness of 1, $1+hs$, and $1+s$, respectively. Fitness across loci was also considered to be additive.

3.2.2.4 Demographics

We modeled an individual-based Wright-Fisher population in which, each generation, N_e diploid individuals mutate, recombine, reproduce, and perish. Reproduction was random with each individual represented in the pool of potential mates in proportion to its fitness. The simulation ran for g generations, where $g = \frac{2Ne}{k}$ and the constant k — representing the strength of the domestication bottleneck — has been estimated for maize at approximately

2.45 Wright et al. (2005). Simulations ended at g generations or if the domesticate-type allele became fixed at five loci.

3.2.2.5 Runs

Diversity-based estimates of the bottleneck size and duration, combined with archaeological information on the timing of maize domestication, suggest a maximum size for the bottlenecked population of approximately 3500 diploid individuals (Wright et al., 2005; Matsuoka et al., 2002; Piperno and Flannery, 2001). This limit was used as an upper bound of the population sizes for simulations: we modeled effective population sizes of N_e= 1000,1500,2000,2500,3000, and 3500). In line with the current understanding of the genetic architecture of maize domestication (Doebley, 2004), we assumed five loci are required for domestication success. We tested two selection coefficients, one similar to ($s = 0.03$) and one stronger than ($s = 0.07$) the magnitude of selection previously estimated at domestication loci in maize (Wang et al., 1999, 2005). We simulated both recessive ($h = 0.05$) and codominant ($h = 0.5$) fitness for domestication loci. Finally, we tested two different starting frequencies for domestication (and recombination) alleles. Populations were initiated with an allele frequency of 0.025 or a frequency determined by mutation-selection balance. Mutation-selection balance was calculated for the pre-domestication populations assuming that dominance and the selection coefficient of the allele remain unchanged, but that the sign of the selection coefficient is reversed. Ten simulations were run for each possible combination of population size, selection coefficient, dominance, and starting frequency.

3.2.2.6 Analysis

For each simulation, we recorded time to domestication, relative mean fitness (defined as a percentage of the difference between the maximum and minimum possible fitness), number of genes fixed, and the average chiasmata per chromosome. Each variable was analyzed separately in a multiple regression analysis. For each variable, a full regression model was

built including each independent variable *(s*, *h*, N_e, starting frequency, and the total number of loci). For determination of bottleneck success (fitness/fixed loci/time to domestication) average chiasmata per bivalent was also included in the model. For the analysis of chiasmata, a separate analysis was performed for each of the three measures of domestication success. Only second order interaction terms were included in all models. For each analysis, a stepwise analysis was performed, utilizing the Akaike Information Criteria to determine the best subset of variables for inclusion in the final model (Hastie and Pregibon, 1992; Venables and Ripley, 2002).

3.3 Results

Analysis of polymorphism data from 42 nuclear sequence loci showed a strong trend towards higher diversity in the progenitor of maize, *Z. mays ssp. parviglumis* (Table 3.4). *Z. m.* ssp. *parviglumis* had a higher estimated θ_w than all of the other wild species at 39 of 42 loci (sign test, $p<0.001$), a pattern which held true for nonsynonymous sites as well (28 of 42, $p<0.03$). Measures of diversity for *Z. m. parviglumis* did not differ statistically from diversity estimates for the conspecific *Z. m. mexicana*, however. Multilocus estimates at both silent and nonsynonymous sites showed a similar pattern to the per locus data — *Z. m. parviglumis* had a significantly higher multilocus θ_w than any of the three wild species, and did not differ significantly from *Z. m. mexicana* (Fig. 3.1). Evidence of non-neutral evolution was detected at several of the loci. To account for the possibility that departure from the neutral model could explain lower levels of diversity for some loci, we compared diversity estimates after removing these loci. Including only loci not showing evidence of non-neutral evolution, both single and multilocus results remained significant ($p<0.05$, results not shown).

Estimates of θ_w can be converted into effective population size with knowledge of the mutation rate. The most widely used estimate of the mutation rate in grasses is currently 6.5×10^{-9} (Gaut et al., 1996), though some recent work has challenged the applicability of this value (Clark et al., 2005). Using this value, we calculate an approximate N_e for *Z.*

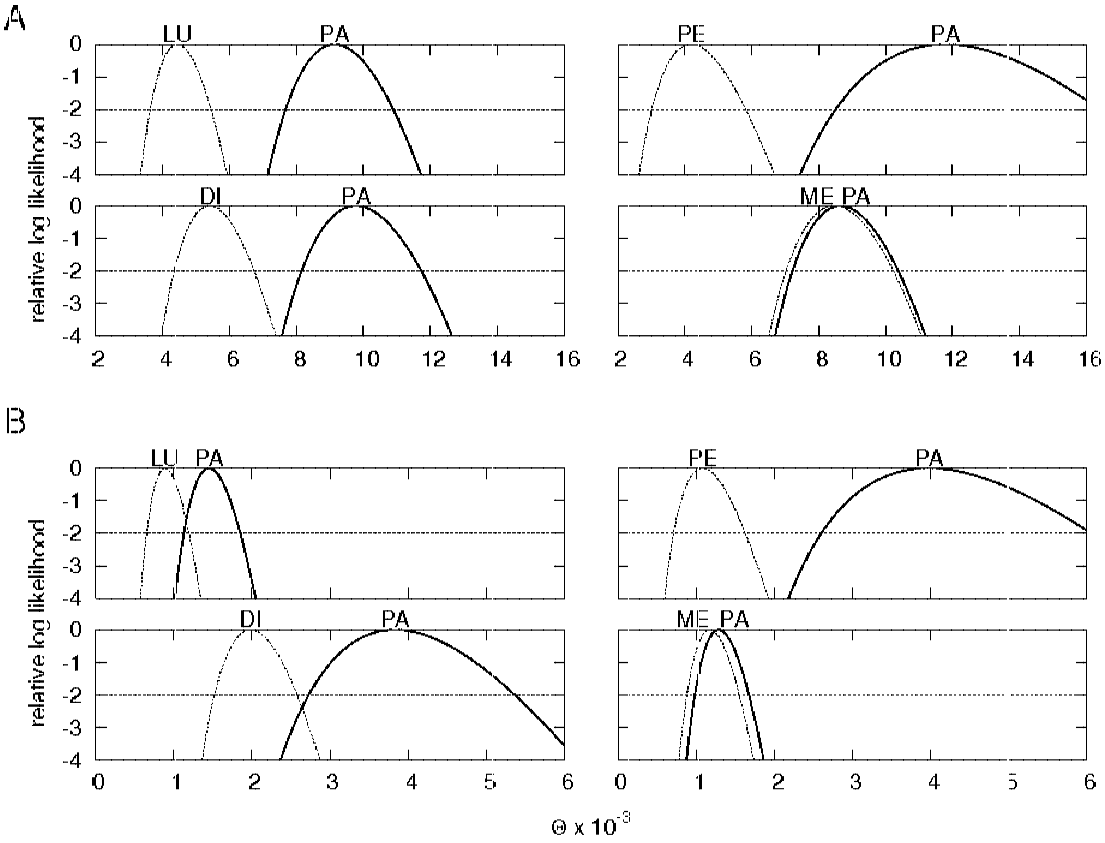


Figure 3.1: Multilocus estimates of diversity for wild taxa of the genus Zea. Each graph shows the relative log likelihood curve of a multilocus estimate of theta per base pair for one wild taxon (dotted lines) and the corresponding estimate for *Z. m. parviglumis*. Estimates for *Z. m. parviglumis* differ among graphs due to the different subsets of loci used in each comparison. A. Silent sites B. Nonsynonymous sites.

m. parviglumis (and *Z. m. mexicana*) of 350,000, and 150-200,000 for the other wild taxa. However, unless there are dramatic differences in mutation rates across species, the ratio of the effective population sizes should be congruent with the ratio of the estimated θ_w. Thus, these results suggest that the wild subspecies of *Z. mays* have an evolutionary effective population size 2-3 times that of the other wild taxa in *Zea.*

Data were generated for 480 simulations of the domestication bottleneck, including 142 successful domestication events. Regression results for mean fitness and fixed loci are shown in tables 3.1-3.2. Results from the regression of successful domestication time were nearly identical to those from the regression of fitness, differing only in the marginal significance of two interaction effects (N_e x s and N_e x starting frequency, results not shown). All of the factors investigated had significant effects on relative fitness (Fig. 3.2), and several interaction effects were significant as well (Fig. 3.3). Importantly, though effective population size interacted significantly with starting frequency, selection coefficient, and dominance, none of these interactions changed the basic result that fitness increased substantially with change in effective population size.

Though the main result of a significant effect of N_e was similar for the analysis of fixed loci, other results differed substantially from the data on fitness (Table 3.2 and Fig. 3.4). These differences can probably be attributed to the upper limit placed on fitness during the simulations.

Thus, if enough loci were at high frequency, such that the maximum fitness had been reached, selection would no longer favor fixation of an allele. This is evidenced by the fact that 103 of the 245 simulations with relative fitness > 0.999 have fewer than five fixed loci. This effect has implications for dominance and starting frequencies as well: maximum fitness will be more quickly reached when heterozygotes contribute to fitness, and when starting frequency is high more loci will contribute to fitness from the beginning of the simulations so fewer loci are likely to reach fixation. Additionally, the fact that starting frequency alone was not significantly related to the number of loci fixed suggests that new mutations were

Table 3.1: ANOVA table of mean relative fitness.

	Df	Sum Sq.	Mean Sq.	F	Pr(>F)
N_e	1	5.8	5.8	114.95	< 0.01
h	1	6.69	6.69	132.78	< 0.01
s	1	1.01	1.01	19.95	< 0.01
start freq.	1	1.73	1.73	34.29	< 0.01
total loci	1	0.64	0.64	12.63	< 0.01
recombination	1	0.24	0.24	4.7	0.03
N_ex h	1	2.98	2.98	59.15	< 0.01
N_ex s	1	0.35	0.35	7	0.01
N_ex start freq.	1	1.96	1.96	38.9	< 0.01
h x start freq.	1	0.91	0.91	17.97	< 0.01
h x loci	1	0.17	0.17	3.28	0.07
s x loci	1	0.14	0.14	2.8	0.09
start freq. x loci	1	0.26	0.26	5.17	0.02
Residuals	466			23.49	6.69

Table 3.2: ANOVA table of mean number of fixed genes.

	Df	Sum Sq.	Mean Sq.	F	Pr(>F)
N_e	1	147.23	147.23	83.02	<0.01
h	1	0.07	0.07	0.04	0.84
s	1	46.88	46.88	26.43	<0.01
start freq.	1	0.83	0.83	0.47	0.49
loci	1	1.82	1.82	1.03	0.31
recombination	1	16.04	16.04	9.05	<0.01
N_ex h	1	59.56	59.56	33.59	<0.01
N_e x s	1	7.64	7.64	4.31	0.04
N_e x start freq.	1	55.76	55.76	31.44	<0.01
h x start freq.	1	135.86	135.86	76.61	<0.01
h x total loci	1	10.68	10.68	6.02	0.01
Residuals	468	829.95	1.77		

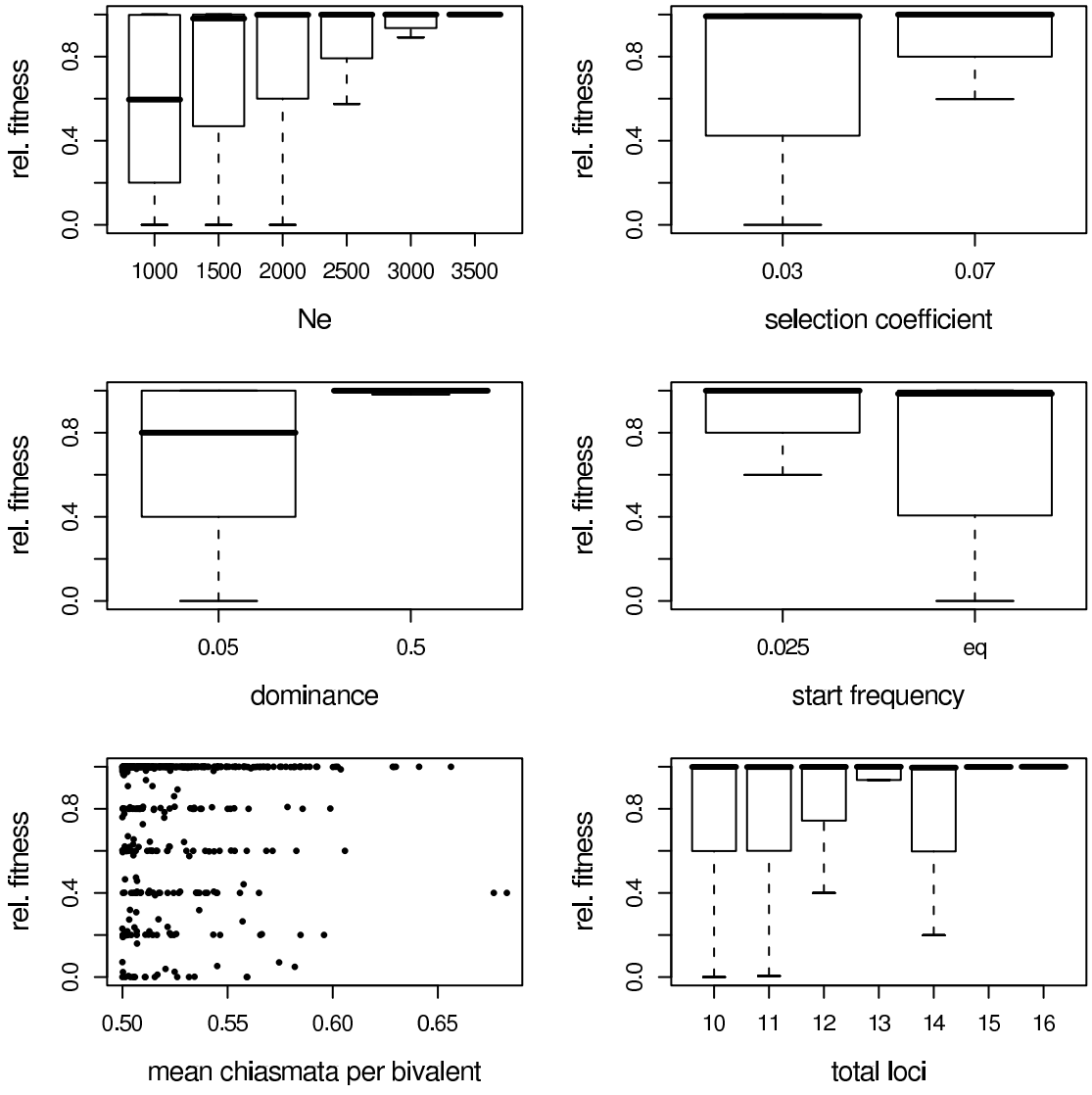


Figure 3.2: Significant main effects on relative fitness. Horizontal bars represent the median, boxes the inter-quartile range, and whiskers extend out twice the inter-quartile range.

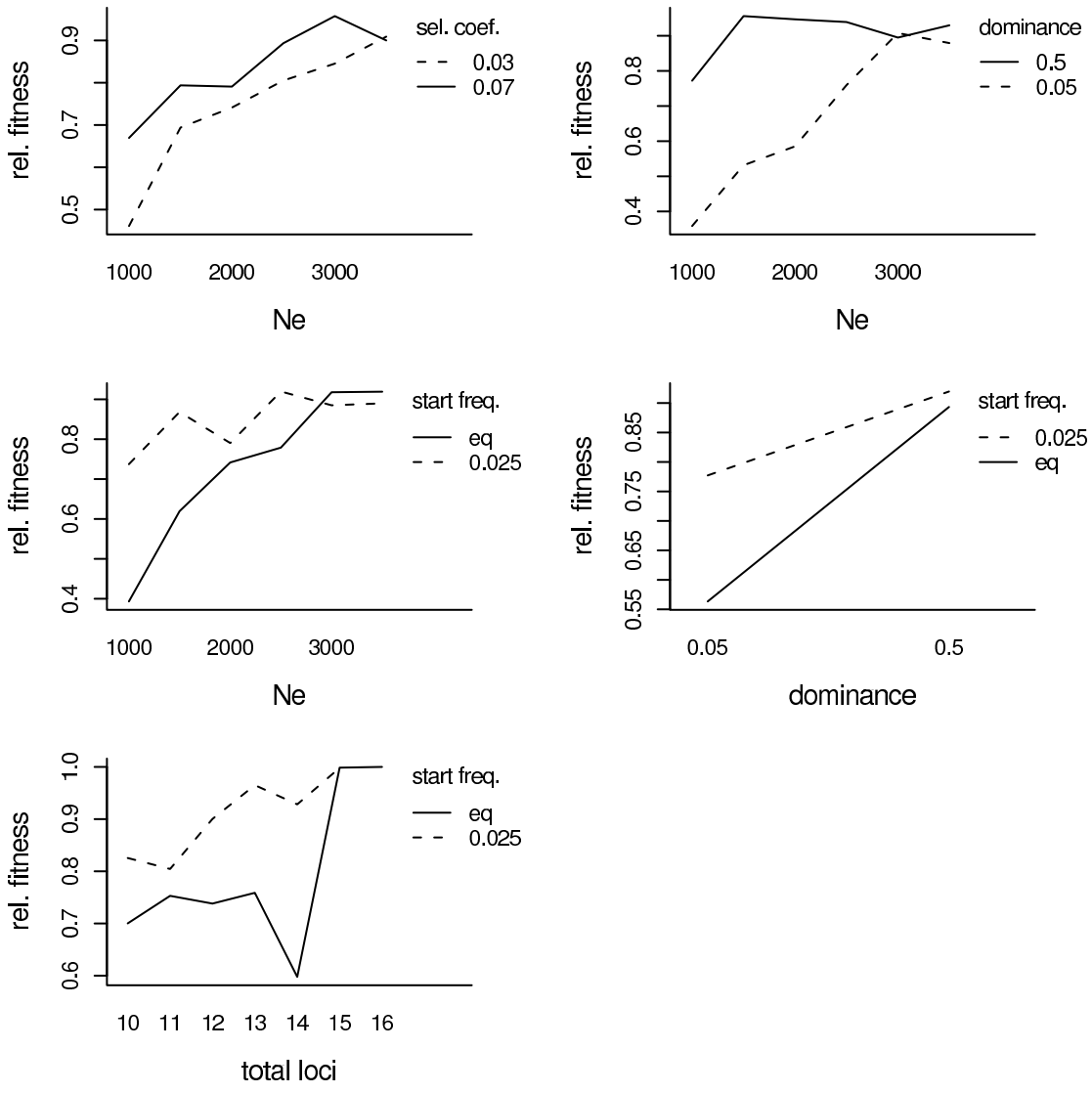


Figure 3.3: Significant two-way interaction effects on relative fitness.

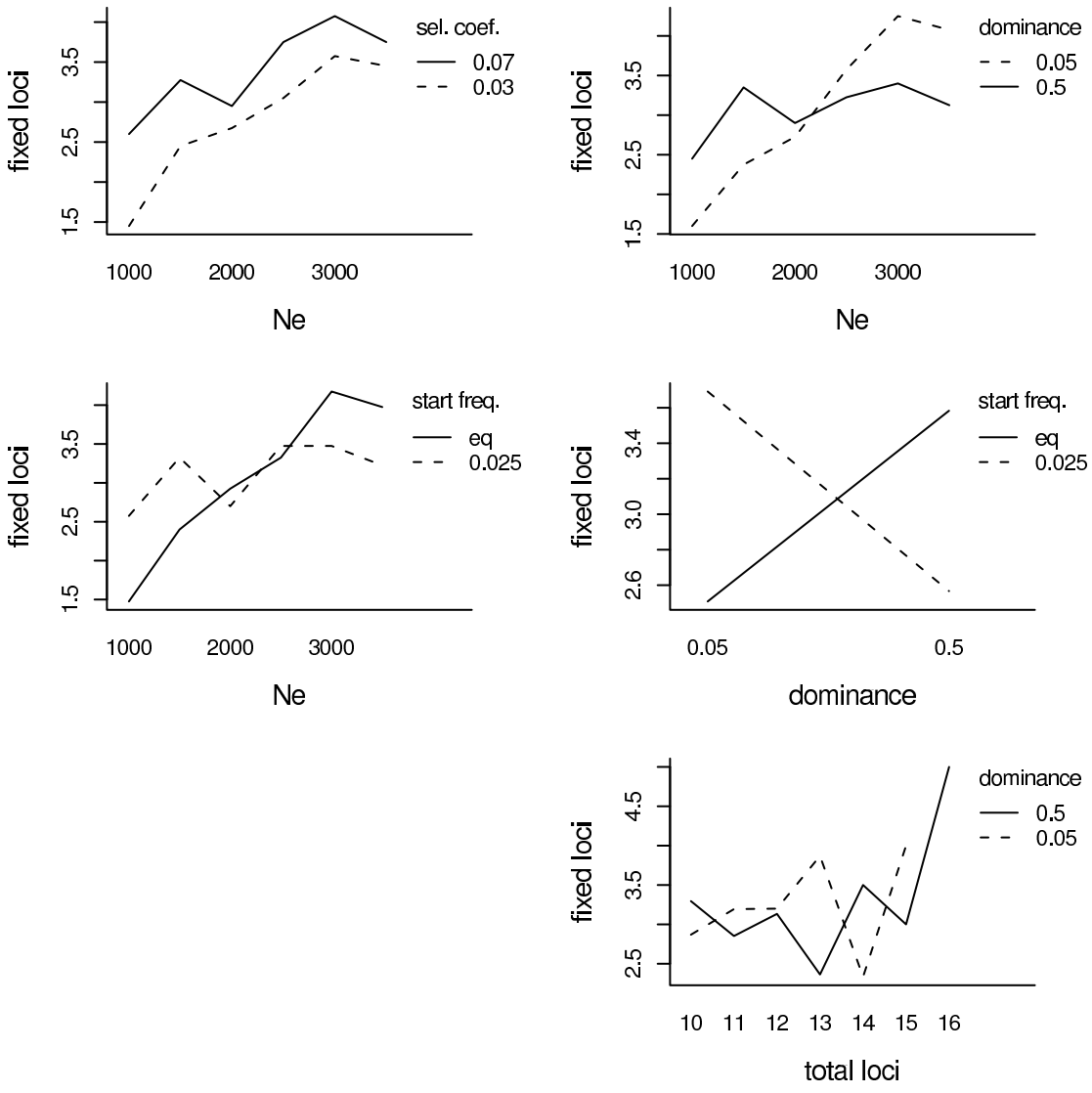


Figure 3.4: Significant two-way interaction effects on number fixed loci.

Table 3.3: ANOVA table of mean chiasmata per chromosome

	Df	Sum Sq.	Mean Sq.	F	p
N_e	1	0.0015	0.0015	2.3	0.13
h	1	0.0007	0.0007	1.08	0.3
s	1	0.0003	0.0003	0.5	0.48
start freq.	1	0.0579	0.0579	88.68	<0.01
total loci	1	0.0009	0.0009	1.34	0.25
fixed loci	1	0.0046	0.0046	7.03	<0.01
h x s	1	0.0012	0.0012	1.88	0.17
h x start freq.	1	0.0017	0.0017	2.63	0.11
s x start freq.	1	0.0006	0.0006	0.98	0.32
start freq. x loci	1	0.0039	0.0039	5.91	0.02
start freq. x fixed	1	0.0017	0.0017	2.63	0.11
Residuals	468	0.3053	0.0007		

more important in eventual domestication success than mutations in the population at the beginning of the simulations. Closely tracking allele frequencies in a number of simulations corroborated these results — for those simulations followed, nearly all of the loci that became fixed reached 0 frequency at some time during the simulation, such that the fixed allele always arose as a new mutation.

Results from an analysis of recombination were equivocal. Though recombination was found as a significant predictor of both fixed loci and relative fitness, domestication success was a significant predictor of recombination rate only when fixed loci were used as an independent variable (Table 3.3).

The initial frequency of recombination alleles (simulated for convenience as equal to the initial frequency of domestication alleles) was also significant. These data, however, are based on the final average recombination rate in the population. A closer look the allele frequency of recombination loci during a single simulation run suggests that, in spite of the significant correlations found, analysis of the final average recombination rate is a poor method for investigating the evolution of recombination (Fig. 3.5).

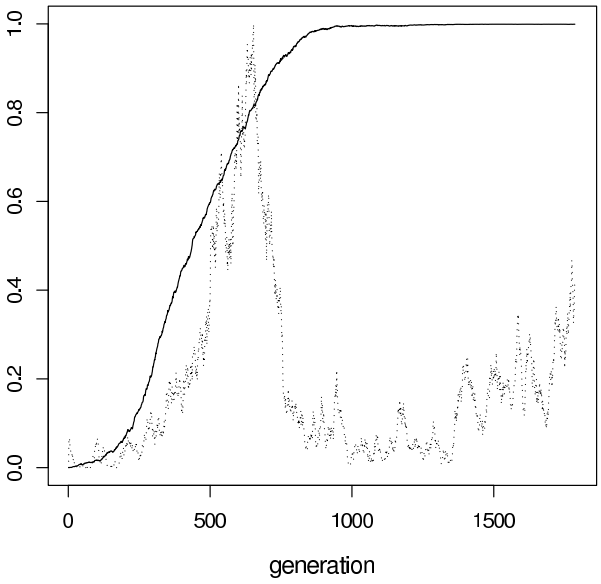


Figure 3.5: Generational change in fitness and recombination rate. Data shown is from a single simulation with $h = 0.5$, starting frequency = 0.025, $N_e = 3000$, and $s = 0.03$. Solid lines represent relative fitness, dotted lines average chiasma frequency.

3.4 Discussion

The central prediction of the hypothesis that effective population size functions as a preadaptation to domestication is that the wild progenitor of a successful domesticate should have a larger effective population size than its wild congeners. While this is impossible to investigate directly, evolutionary effective population size should be reflected in patterns of genetic diversity. Using patterns of diversity at more than 40 sequence loci, we have shown that this prediction holds for *Zea*: the wild progenitor of maize, *Zea mays* ssp. *parviglumis*, is more diverse than any of its wild congeners at nearly all loci examined. Previous allozyme analysis of these species provided similar results — *Z. m. parviglumis* revealed a more polymorphic loci and more alleles per polymorphic locus than any of its congeners (Doebley et al., 1984). Interestingly, the conspecific *Z. mays mexicana* showed similar levels of diversity to *Z. mays parviglumis*; the fact that this former taxon has been implicated in the evolution of modern

maize (Fukunaga et al., 2005) is thus not surprising in light of the arguments made here. The possibility that this observed difference in diversity is a general feature of domesticated plant taxa is further suggested by two reviews of allozyme diversity in crop plants. Doebley (1989) compared levels of allozyme diversity between domesticates and their wild progenitors for a variety of crops, and convincingly showed that the process of domestication has involved a loss of genetic diversity. Yet a later review comparing crop plants as a group to other wild plants found that crop plants harbored significantly more genetic diversity at allozyme loci (Hamrick and Godt, 1997). If crop plants in general have greater diversity than most wild plant species, but have lost diversity in the process of domestication, it follows that their wild progenitors should be more diverse (and thus have higher effective population sizes) than other wild species, a conclusion consistent with the major prediction of the preadaptation hypothesis.

Our simulation results suggest that effective population size can function as a preadaptation. Larger effective population size substantially increased the likelihood of domestication success in our models, and differences in population size equivalent to observed differences in diversity among species (2-3 fold) are more than sufficient to prevent successful domestication. In the scenarios that most closely approximate what is known about maize domestication ($N_e = 3500$, $h = 0.05$, $s = 0.03$, initial frequency = 0.025), 7 of 10 simulations successfully fixed all 5 alleles, while the corresponding simulation at either $N_e = 1000$ or $N_e = 1500$ was successful only 1 of 10 times.

In addition to supporting the role of effective population size in domestication success, these results corroborate previous findings that domestication itself selects for increased recombination Ross-Ibarra (2004). It is evident, however, that analysis of the final recombination rate is an inappropriate measure for studying the evolution of recombination during domestication (Fig. 3.5). The limited within-simulation data available suggest an even closer relationship between recombination and fitness during the spread of selected alleles, followed by a relaxation of selection for recombination or perhaps even selection against recombina-

tion after the population approaches a fitness peak. A more detailed study of the change in frequency of recombination alleles during the domestication process seems warranted if we are to understand the role of recombination in domestication.

Though these results confirm the possibility that effective population size alone could determine domestication success for a realistic set of parameter values, it is important to recognize the many limitations of these analyses. In particular, these results are very much dependent on assumptions of the simulation models. For example, additional simulations suggest that the mode of intergenic selection (multiplicative vs. additive) and the number of loci required for successful domestication both strongly effect the possibility of domestication success (not shown). It is also clearly an oversimplification to think that any 5 of 10 possible loci can achieve a domesticated phenotype — while the genetic pathways leading to some domestication traits may have multiple loci that could theoretically achieve the same result (e.g. Whitt et al., 2002), other changes associated with domestication may only be achievable via mutations in a single gene. Similarly, these simulations assume a single Wright-Fisher population, while it is highly unlikely that domestication involved a panmictic population.

In summary, the results presented here suggest that differences in diversity and effective population size among species in the genus *Zea* are sufficient to explain the observation that only the wild subspecies of *Z. mays* was successfully domesticated. Combined with general findings on diversity levels in cultivated crops and their wild ancestors, we speculate that effective population size may well function as a preadaptation to domestication for most crop plants. We argue that these data are sufficient to warrant further empirical investigation — is it generally the case that the wild ancestor of successful domesticates are more diverse than their congeners, as implied from reviews of allozyme diversity? However, since congeneric wild taxa have not been exposed to artificial selection, it will be difficult to refute the possibility raised by Darwin that any species could be domesticated given sufficient selection. Of special interest, then, would be genetic analysis of a genus in which some congeners have been exposed to selection — cultivation, gathering, etc. — but have never been successfully

domesticated. Showing that these 'negative controls' also have lower effective population size than the wild ancestor of a successful domesticate would provide strong empirical corroboration for the results presented here in favor of the hypothesis of preadaptation.

3.5 Acknowledgements

We wish to thank D. Brown and G. Derda for help in procuring and using computer resources, M. Arnold and J. Wares for helpful discussion, R. Cartwright for useful ideas, and programming advice, and S. Cornman and M. Poelchau for a great deal of patience and coffee. B. Gaut and P. Tiffin kindly provided unpublished sequence data.

Table 3.4: Loci, sample sizes, and basic diversity statistics for the loci used in the analysis. Shown are locus names, number of haplotypes sampled (N), average base pair length of locus (bp), number of segregating sites (S), Watterson's (1975) estimator of θ (θ_W), and Tajima's (1983) estimator of θ (θ_π)

locus	**taxon**	**N**	**bp**	**S**	θ_w	θ_π
adh1	dip	10	1278	42	14.85	15.09
adh1	lux	8	1374.8	27	10.41	11.29
adh1	mays	31	1308	59	14.77	17.37
adh1	mex	8	1366	65	25.07	22.29
adh1	parv	8	1262.8	57	21.98	21.39
adh1	peren	16	1382	52	15.67	17.44
asg11	lux	5	441.6	10	4.8	4.6
asg11	mays	24	349	11	2.95	3.17
asg11	mex	7	494	13	5.31	4.1
asg11	parv	6	491	10	4.38	3.6
asg35	lux	10	497	5	1.77	2.22
asg35	mays	22	496	19	5.21	6.99
asg35	mex	12	499	12	3.97	5.18
asg35	parv	12	468	31	10.27	7.24
asg64	lux	11	586	21	7.17	9.56
asg64	mays	20	556	22	6.2	5.04
asg64	mex	11	565	14	4.78	4.04
asg64	parv	16	509.9	29	8.74	6.36
asg65	dip	8	933	12	4.63	4.32
asg65	lux	11	798	9	3.07	3.38
asg65	mays	17	798.9	19	5.62	3.6

locus	**taxon**	**N**	**bp**	**S**	θ_w	θ_π
asg65	mex	12	795	12	3.97	2.86
asg65	parv	14	800	22	6.92	7.59
bnl7-13	lux	11	895.8	7	2.39	1.27
bnl7-13	mex	11	894	40	13.66	11.82
bnl7-13	parv	12	893.7	27	8.94	8.89
bz2	lux	11	590	13	4.44	3.71
bz2	mays	25	594	17	4.5	5.71
bz2	mex	7	592	15	6.12	6.38
bz2	parv	12	590	15	4.97	3.94
c1	dip	14	638	15	4.72	4.08
c1	lux	7	716	21	8.57	10.29
c1	parv	6	734	21	9.2	10
c1	peren	15	650	14	4.31	4.15
chiA	dip	8	859	0	0	0
chiA	parv	13	848	23	7.41	5.38
chiB	dip	7	862	19	7.76	7.71
chiB	parv	14	862	27	8.49	8.29
chiI	dip	9	1074	12	4.42	4.28
chiI	parv	15	1078	25	7.69	7.01
csu1132	lux	10	537	18	6.36	5.27
csu1132	mays	25	543	24	6.36	5.01
csu1132	mex	10	531	35	12.37	9.71
csu1132	parv	8	531	29	11.19	10.14
csu1138	lux	10	351	7	2.47	1.56
csu1138	mays	25	324	16	4.24	5.07

locus	**taxon**	**N**	**bp**	**S**	θ_w	θ_π
csu1138	mex	10	351	16	5.66	5.31
csu1138	parv	14	351	6	1.89	2.42
csu1171	lux	8	463	10	3.86	3.68
csu1171	mays	25	468	11	2.91	3.15
csu1171	mex	10	432	13	4.6	4.42
csu1171	parv	9	467	18	6.62	5.44
csu381	lux	9	921.8	7	2.58	2.22
csu381	mays	25	916	42	11.12	15.9
csu381	mex	9	917	58	21.34	20.39
csu381	parv	10	917	42	14.85	13.13
csu838	lux	11	757	22	7.51	7.27
csu838	mays	21	697	15	4.17	3.83
csu838	mex	12	762	13	4.31	3.97
csu838	parv	13	759	25	8.06	5.58
d8	lux	8	812	8	3.09	3.5
d8	mays	25	800	8	2.12	1.91
d8	mex	10	812	12	4.24	3.78
d8	parv	13	803	47	15.15	9.08
fus6	lux	12	254	2	0.66	0.33
fus6	mays	23	248	8	2.17	2.5
fus6	mex	12	250	4	1.33	1.12
fus6	parv	10	250	5	1.77	1.67
gl15010	lux	12	721	5	1.66	1.83
gl15010	mays	22	691	7	1.92	1.53
gl15010	mex	13	721	20	6.45	5.41

locus	**taxon**	**N**	**bp**	**S**	θ_w	θ_π
gl15010	parv	15	714	13	4	3.41
glb1	peren	15	1005	47	14.46	15.78
glb1	mays	30	957.4	94	23.73	17.28
glb1	dip	9	1034	36	13.25	10.17
glb1	parv	8	1033.3	75	28.93	24.93
glb1	lux	6	1082.7	24	10.51	10.53
hag	dip	8	651	9	3.47	3.32
hag	parv	15	613	37	11.38	11.31
hm2	dip	11	1495	3	1.02	0.69
hm2	mays	10	1144	25	8.84	8.73
hm2	parv	11	1109	41	14	9.35
hm2	peren	12	1418	38	12.58	15.89
mgs3020	lux	12	787	6	1.99	2.2
mgs3020	mays	22	963	9	2.47	2.85
mgs3020	mex	13	789	12	3.87	3.51
mgs3020	parv	14	789	14	4.4	3.56
mir1	dip	9	971	20	7.36	7.5
mir1	parv	13	931	68	21.91	19.87
mpi	dip	13	526	11	3.55	2.13
mpi	parv	17	523	10	2.96	2.79
mpi	peren	14	558	11	3.46	3.68
pepc1070	lux	10	757	6	2.12	2.02
pepc1070	mays	18	741	52	15.12	17.3
pepc1070	mex	8	714	39	15.04	15.46
pepc1070	parv	13	752	64	20.62	21.15

locus	**taxon**	**N**	**bp**	**S**	θ_w	θ_π
pepc1150	lux	11	675	8	2.73	2.62
pepc1150	mays	18	671	21	6.11	6
pepc1150	mex	12	671	24	7.95	9.05
pepc1150	parv	16	671	18	5.43	4.38
plt1	dip	9	482	17	6.26	6.39
plt1	parv	13	491	42	13.53	9.64
plt2	dip	11	609	6	2.05	1.82
plt2	parv	14	813	24	7.55	5.55
pr1	dip	9	582	10	3.68	2.83
pr1	parv	16	507	15	4.52	4.08
pr5	dip	10	479	11	3.89	4.56
pr5	parv	13	479	19	6.12	5.58
pr6	dip	8	1170	14	5.4	5.18
pr6	parv	14	1110	26	8.18	7.69
prms	dip	8	699	11	4.24	2.75
prms	parv	16	699	21	6.33	4.57
rip1	dip	8	1135	35	13.5	12.82
rip1	parv	14	1147	55	17.3	15.16
rip2	dip	8	416	421	162.37	310.75
rip2	parv	13	415.5	421	135.67	309.29
tb1	lux	11	2423	54	18.44	11.89
tb1	mays	23	2737	22	5.96	2.61
tb1	mex	7	2679.7	118	48.16	44.24
tb1	parv	7	2701	87	35.51	32.33

locus	**taxon**	**N**	**bp**	**S**	θ_w	θ_π
ts2	lux	8	973	13	5.01	5.32
ts2	mays	24	958	11	2.95	3.24
ts2	mex	9	907	13	4.78	4.67
ts2	parv	10	972	15	5.3	5.4
vp1010	lux	6	726	4	1.75	1.93
vp1010	mays	22	1295.8	25	6.86	5.39
vp1010	mex	9	726	15	5.52	5.56
vp1010	parv	12	726	17	5.63	4.56
wx1	dip	10	884	12	4.24	3.53
wx1	lux	21	885.9	13	3.61	3.63
wx1	mays	37	850	34	8.15	6.83
wx1	mex	20	864	40	11.28	8.99
wx1	parv	19	868.8	34	9.73	8.2
wx1	peren	17	864	19	5.62	6.82
wip1	dip	10	617	26	9.19	10.96
wip1	lux	9	541	20	7.36	7.22
wip1	mays	12	541	28	9.27	10.29
wip1	mex	10	548	28	9.9	10.47
wip1	parv	15	475	29	8.92	8.31
wip1	peren	10	583	14	4.95	5
zlp	dip	8	722	5	1.93	1.86
zlp	parv	14	727	14	4.4	3.51

Chapter 4

Genome size and recombination in Angiosperms: a second look[1]

[1] J. Ross-Ibarra. Submitted to *Journal of Evolutionary Biology* on 07/12/2006

Abstract

Despite dramatic differences in genome size — and thus space for recombination to occur — previous workers found no correlation between recombination rate and genome size in flowering plants. Here I re-investigate these claims using phylogenetic comparative methods to test a large dataset of recombination data in Angiosperms. I show that genome size is significantly correlated with recombination rate across a wide sampling of Angiosperm species and that change in genome size explains a meaningful proportion (~20%) of observed variation in recombination rate. Though the magnitude of the observed effect of genome size on recombination rate is comparable to the effects of several life history characters previously linked to evolutionary change in recombination rate, consideration of processes of genome size change in Angiosperms suggests that the observed correlation is likely conservative. And finally, while recombination rate was found to increase less than proportionally to change in genome size, several mechanistic and theoretical arguments suggest that this result is not unexpected.

4.1 Introduction

Recombination and genome size are highly labile characteristics of plant genomes. Variation in recombination rate is seen among genes within plant genomes (Tenaillon et al., 2002) as well as among individuals within a population and populations within a species (Rees and Dale, 1974). Moreover, recombination rate has been shown to respond to direct or indirect selection over short periods of time (Harinarayana and Murty, 1971; Ross-Ibarra, 2004), and putative adaptive correlations have been documented with several life history characteristics (Koella, 1993). Variation in genome size among angiosperms is even more impressive: flowering plant genome sizes span a range of more than three orders of magnitude (Bennett and Leitch, 2005), and dramatic variation in genome size is evident even among congeners (Rees and Durrant, 1986; Narayan and McIntyre, 1989; Jakob et al., 2004).

Given the dramatic differences in genome size — and thus space in which recombination could occur — and the labile nature of recombination, a positive correlation between genome size and recombination rate would seem natural. Abundant evidence for a strong within-genome correlation between chromosome size and recombination rate strengthens this hypothesis (Anderson et al., 2001, 2003). It is thus somewhat surprising that all investigations of this relationship among species have found either no correlation (Cavalier-Smith, 1985; Rees and Durrant, 1986; Anderson et al., 2001) or a negative one (Narayan and McIntyre, 1989). These findings, however, have been based on limited datasets from a few genera or widely unrelated taxa, and have failed to incorporate phylogenetic information into their analyses. Yet there is a strong phylogenetic component to the observed variation in genome size (Soltis et al., 2003), and clear differences among genera and families for average levels of recombination (Rees and Dale, 1974).

Here I re-investigate findings that recombination rate is independent of genome size. Utilizing a dataset of more than 270 Angiosperm species, I find that both standard linear regression and phylogenetic comparative methods reveal a significant positive correlation between genome size and recombination rate both across all species as well as within individual families and genera. I also show that the magnitude of the effect of genome size on recombination rate is similar to or greater than the effect of several life history characteristics commonly linked to the evolution of recombination rates. Finally, however, I find that increases in recombination are not proportional to increases in genome size, as might be predicted from purely spatial or mechanistic considerations, and that this relationship is robust to comparisons using only euchromatic or nonrepetitive fractions of the genome.

4.2 Materials and Methods

I utilized a previously published dataset on chiasmata counts and other plant characteristics (Ross-Ibarra, 2004), supplementing these data with additional life history data from a range of sources. Haploid genome sizes (pg DNA) were extracted from the Kew Plant C

value Database (Bennett and Leitch, 2005). Both genome size and recombination data were available for a subset of 279 species from 63 genera in 22 families; the complete dataset is available upon request. To correct for the strong correlation between total chiasmata and chromosome number, I used chiasmata per bivalent as a measure of recombination; analyses using the residuals of a regression of total chiasmata on chromosome number gave qualitatively similar results, but are not presented. All analyses were performed using both total genome size and average bivalent size; only data on total genome size are reported for most analyses. Chiasmata and genome size data were natural log-transformed for all analyses.

I initially explored the relationship between genome size and recombination rate with standard linear regressions of chiasmata per bivalent on genome size for the entire dataset as well as within each of the two largest families (Poaceae, Fabaceae) and four largest genera (*Senecio, Vicia, Allium, Lathyrus*). To evaluate the correlation between genome size and recombination rate in a phylogenetic context, I mapped haploid genome size and chiasmata per bivalent onto four different phylogenies (Fig. 1.3). I constructed the first phylogeny using the PHYLIP software package (Felsenstein, 1993) to generate a neighbor-joining tree from hand-aligned nucleotide sequence data of the chloroplast gene rbcL. I used a single GenBank rbcL sequence to represent each of 55 genera from the total dataset (not all sequences used were from species represented in the data; a list of sequences is available upon request). Given the paucity of infrageneric data available for most taxa, I coded species' relationships within each genus as polytomies, assigned terminal branch lengths of 0.005, and then arbitrarily resolved polytomies into bifurcations by inserting branches of length 10^{-13}.

I constructed a second phylogeny using independent topological information from the literature and constraining all branches to a length of 1. Deeper branches of the tree (family and above) came from Stevens (2005) and the Angiosperm Phylogeny Group (2003), while information for the shallower clades was taken from individual studies (Hsiao et al., 1995; Olmstead et al., 1999; Downie et al., 2000; Kajita et al., 2001; Hu et al., 2000; Mason-Gamer et al., 2002; Salamin et al., 2002; Steele and Wojciechowski, 2003). I again assumed

infrageneric polytomies which I arbitrarily divided into bifurcations by inserting branches of length 10^{-5}.

Finally, I constructed species-level trees for the two largest families in the dataset (Poaceae and Fabaceae). Sequences from the 5.8S rDNA subunit and both internal transcribed spacers were available in GenBank for 29 species from each family. PHYLIP neighbor-joining trees were constructed from sequence alignments generated in ClustalW (Thompson et al., 1994) and subsequently checked by hand.

Character evolution was analyzed on each of the four phylogenies using the phylogenetic least squares model (PGLS; Martins and Hansen, 1997) implemented in the software COMPARE (Martins, 2004). In addition to correlations between characters, the model includes a parameter α which evaluates the strength of evolutionary constraint acting throughout the phylogeny. Under the assumption of $\alpha = 0$ the PGLS becomes a standard model of independent contrasts (Felsenstein, 1985), and with an arbitrarily large α the model is equivalent to a nonphylogenetic linear regression. COMPARE provides regression results and a log likelihood value for comparison of each of the three models.

To compare the effects of genome size on recombination to the effects of other factors linked to changes in recombination rate, I re-analyzed the data using both of the two complete trees, including data on several other characters thought to be correlated with recombination: weediness (weedy, not weedy), domestication status (domesticated, cultivated, wild), mating system (selfing, mixed, outcrossing), and perenniality (annual/biennial, perennial). Characters with more than two character states were recoded as a series of binary dummy variables (e.g. selfing vs. other, outcrossing vs. other). Data were not available for all species for these additional characters, and these analyses were performed on trees pruned to only include taxa with complete data (not shown). Comparisons using only weediness, domestication status, and bivalent size, for which complete data were available, were also performed using the full tree.

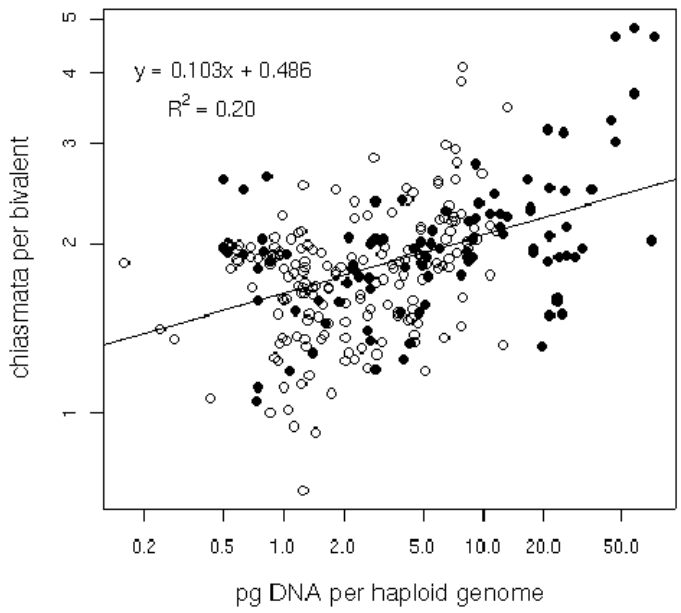


Figure 4.1: Standard linear regression of chiasmata per bivalent on total genome size; solid circles represent monocotyledons and open circles dicotyledons. The regression equation displayed is for the combined data.

4.3 Results and Discussion

Consistent with a well-established pattern in Angiosperms (Levin and Funderburg, 1979), no significant correlation was found between genome size and chromosome number (r=-0.026, p>0.1) in these data. As expected, total genome size is thus tightly correlated (r=0.92) to average bivalent size, and comparisons of total genome size to recombination rate are qualitatively identical and quantitatively similar to comparisons made with average bivalent size. For most additional analyses, only results from comparisons to total genome size are therefore reported.

A significant positive correlation between recombination rate and genome size was observed in standard regression analysis of all 279 species in the dataset (Fig. 4.1). While recombination rates were higher among monocotyledons (t-test, p<0.001), the slope of the

Table 4.1: Nonphylogenetic linear regression of chiasmata per bivalent on genome size and for all available species in each taxon.

Fabaceae (n = 85)	
Percent r^2	24.71
Slope (s.e.)	0.14 (0.03)
Poaceae (n = 63)	
Percent r^2	7.74
Slope (s.e.)	0.06 (0.03)
Senecio (n = 30)	
Percent r^2	27.02
Slope (s.e.)	0.14 (0.04)
Vicia (n = 19)	
Percent r^2	28.7
Slope (s.e.)	0.24 (0.09)
Lathyrus (n = 16)	
Percent r^2	5.51
Slope (s.e.)	-0.09 (0.09)
Allium (n = 16)	
Percent r^2	0.17
Slope (s.e.)	-0.01 (0.09)

regression of recombination rate and genome size was nearly identical in monocots and dicots (not shown). This regression remained significant when performed within either of the two largest families (Poaceae and Fabaceae) or the two most speciose genera (*Senecio* and *Allium*), but no such relationship was observed in two other well-sampled genera (*Vigna* and *Lathyrus*; Table 4.1). Variation in genome size explained more than 20% of the observed variance in recombination rate in all of the significant regressions except within the Poaceae, where genome size variation explained only ~8% of the variation in recombination rate.

Phylogenetic estimation is an integral part of nearly all comparative methods (Harvey and Pagel, 1991), but errors in branch length or topology can lead to misleading results (Ackerly, 2000; Symonds, 2002). Parallel comparative analysis on two independently derived

— and incongruent — topologies differing in the treatment of branch lengths provided nearly identical results, suggesting that phylogenetic inaccuracy or error are unlikely to be a serious concern for these data. To further account for the observation that substantial change in both genome size and recombination rate is often evident among species within a genus, comparisons were also made on species-level phylogenies of the two largest families in the dataset. All four analyses revealed a similar significant positive correlation between genome size and chiasmata per bivalent (Table 4.2). Moreover, genome size alone explained a substantial portion of the observed variation in recombination — from more than 10% across all species to nearly 60% of the observed variation in recombination rate within the Fabaceae. For the two largest trees, likelihood ratio tests show that the PGLS model was a significantly better fit than either a nonphylogenetic regression or Felsenstein's (1985) independent contrasts, and the estimated α values suggest that although there is considerable phylogenetic signal in the correlation, other factors also contribute substantially to the observed variation.

Phylogenetic regression analysis utilizing either of the family-level trees (Fabaceae or Poaceae) revealed a much stronger relationship than that observed for the larger phylogenies. For both of these families, the correlation between genome size and recombination rate was twice that observed in the complete phylogeny, and variation in genome size explained a large portion of the observed variation in recombination rate ($r^2 > 0.35$, Table 4.2). Estimated values of α for both families were at the extreme high end of the range estimated by the COMPARE software, and log-likelihood values reveal little difference between the PGLS and nonphylogenetic regression models, suggesting little evolutionary constraint acting on the genome size - recombination correlation. Comparison to standard regression results that incorporate data from all taxa in each family suggests that the strength of the above regressions is due at least in part to sampling effects. In the Poaceae, for example, it is likely that exclusion of the genus *Oryza* — with the lowest average genome size in the family, but above average recombination rate — artificially inflated the correlation in the phylogenetic analysis.

Table 4.2: Correlation coefficients (r) and slope of regression of chiasmata per bivalent on genome size. The three models shown are the phylogenetic generalized least squares (PGLS), Felsenstein's independent contrasts (FIC), and a nonphylogenetic regression (NR).

	PGLS	FIC	NR
ITS: Fabaceae (n = 29)			
α	15.5	0	NA-large
log likelihood	30.42	23.27	30.42
r	0.60	0.08	0.74
slope (s.e.)	0.23 (0.06)	0.03 (0.08)	0.28 (0.05)
ITS: Poaceae (n = 29)			
α	15.5	0	NA-large
Ln likelihood	40.47	31.31	38.58
r	0.70	0.78	0.62
Slope (s.e.)	0.19 (0.04)	0.19 (0.03)	0.17 (0.04)
rbcL (n = 260)			
α	8.27	0	NA-large
Ln likelihood	276.53	256.93	240.15
r	0.31	0.16	0.46
Slope (s.e.)	0.10 (0.02)	0.07 (0.03)	0.11 (0.01)
Equal (n = 279)			
α	7.5	0	NA-large
Ln likelihood	294.88	281.93	259.48
r	0.3	0.21	0.45
Slope (s.e.)	0.10 (0.02)	0.08 (0.02)	0.10 (0.01)

To gauge the relative significance of the genome size-recombination rate correlation, I compared the effect of genome size on recombination to the effects of several factors that have previously been linked to differences in recombination. With a reduced dataset that included information on perenniality, domestication status, weediness, and mating system, I performed a multiple regression of these factors and average bivalent size on recombination rate within the phylogenetic context of the rbcL tree used above (Table 4.3). Average bivalent size was used instead of total genome size to avoid spurious negative correlations with polyploidy. Results did not differ when total genome size was used and polyploidy included in the analysis or when analyses were performed using the equal-branch length phylogeny. Partial regression results for each character show that genome size has an effect at least as large as the effect of any of the life history characters surveyed.

Results for the other life history characters were not entirely congruent with previous results, and should probably be interpreted with caution. Outcrossing taxa were found here to have lower recombination rates than taxa with mixed or selfing mating systems, in agreement both with previous findings (Ross-Ibarra, 2004; Gibbs et al., 1975, and references therein) and theoretical predictions (Roze and Lenormand, 2005). Perennial plants showed lower rates of recombination than annuals or biennials; this contrasts with previous nonphylogenetic results (Ross-Ibarra, 2004), but is not surprising given previous evidence of the relevance of phylogenetic depth to this comparison (Koella, 1993). Weediness showed a significant effect only when analyzed on the full tree lacking information on perenniality and mating system; previous workers have presented conflicting evidence on the role of weediness in determining recombination rates (Gornall, 1983; Evans and Weir, 1981) and nonphylogenetic multiple regression has suggested the importance of the interaction between weediness and perenniality (Ross-Ibarra, 2004). Domestication status similarly only shows a significant effect on the complete tree, though both across-species regression and a more appropriate sister-taxon comparison of domesticates and their wild progenitors using the same dataset showed significant results (Ross-Ibarra, 2004).

Table 4.3: Multiple regression of chiasmata per bivalent on bivalent size and several plant life history characteristics. The partial regression slope and standard error are shown for each characteristic. The three models shown are the phylogenetic generalized least squares (PGLS), Felsenstein's independent contrasts (FIC), and a nonphylogenetic regression (NR).

	PGLS	FIC	NR
rbcL (n = 142)			
α	3.64	0	NA-Large
log likelihood	189.73	184.04	167.66
Model r^2	34.1	34.87	38.66
Domesticate	0.06 (0.04)	0.06 (0.04)	0.06 (0.05)
Wild	0.01 (0.04)	0.01 (0.03)	0.06 (0.05)
Weedy	-0.03 (0.04)	-0.04 (0.03)	0.07 (0.04)
Perennial	-0.06 (0.03)	-0.06 (0.03)	-0.07 (0.04)
Outcrossing	-0.12 (0.05)	-0.14 (0.05)	-0.04 (0.06)
Selfing	0.03 (0.05)	0.00 (0.05)	0.08 (0.06)
Bivalent size	0.12 (0.02)	0.11 (0.03)	0.11 (0.01)
rbcL (n = 279)			
α	7.98	0	NA-Large
log likelihood	301.43	287.83	274.46
Model r^2	13.91	8.27	28.13
Domesticate	0.03 (0.04)	0.03 (0.04)	0.06 (0.04)
Wild	-0.07 (0.03)	-0.07 (0.03)	-0.04 (0.04)
Weedy	-0.07 (0.03)	-0.08 (0.03)	-0.02 (0.03)
Bivalent size	0.10 (0.02)	0.09 (0.02)	0.11 (0.01)

The general conclusion from the above analyses is that genome size is a significant determinant of recombination rate in Angiosperm species, with an effect as large as that observed for several life history characteristics. A closer look at genome size and recombination rate data in the genus *Lathyrus*, however, highlights a potentially important caveat in these results: rather than being compared to overall genome size, recombination should be compared to the size of the euchromatic or nonrepetitive fraction of the genome. Heterochromatic or highly repetitive portions of the genome are much less recombinationally active, and empirical work has shown that euchromatin amount is a more important determinant of recombination than total DNA amount among chromosomes within a cell (Sherman and Stack, 1995). Furthermore, amplification of repeat-rich or heterochromatic regions have been implicated as a predominant cause of genome expansion in plants (Bennetzen, 2002). A clear example of the potential problem this causes is evident in chiasma and genome size data from 8 species of the genus *Lathyrus* (Narayan and McIntyre, 1989). Reanalysis of this data shows that while recombination is negatively correlated with total genome size ($r = -0.25$, $p > 0.10$), comparison of recombination to euchromatic genome size reveals a positive correlation ($r = 0.40$, $p > 0.10$). These arguments suggest that the large-scale analyses conducted above could underestimate to a potentially large extent the correlation between euchromatic genome size and recombination. Nonetheless, comparison of a regression of recombination rate onto size of the nonrepetitive fraction of several plant genomes (Flavell et al., 1974; Wenzel and Hemleben, 1982) to a regression on total genome size revealed no difference in slope (Fig. 4.2). The dataset is small ($n=18$) and regression results are not corrected for phylogeny, but the analysis suggests that estimates of the effect of genome size on recombination are not likely to be extremely biased.

The analyses presented here show that clear positive correlation between recombination rate and genome size in plant species, and issues of repetitive DNA and genome expansion probably contrive to make the relationship found here an underestimate of the true correlation between genome size and euchromatic DNA. Nonetheless, the slope of the observed

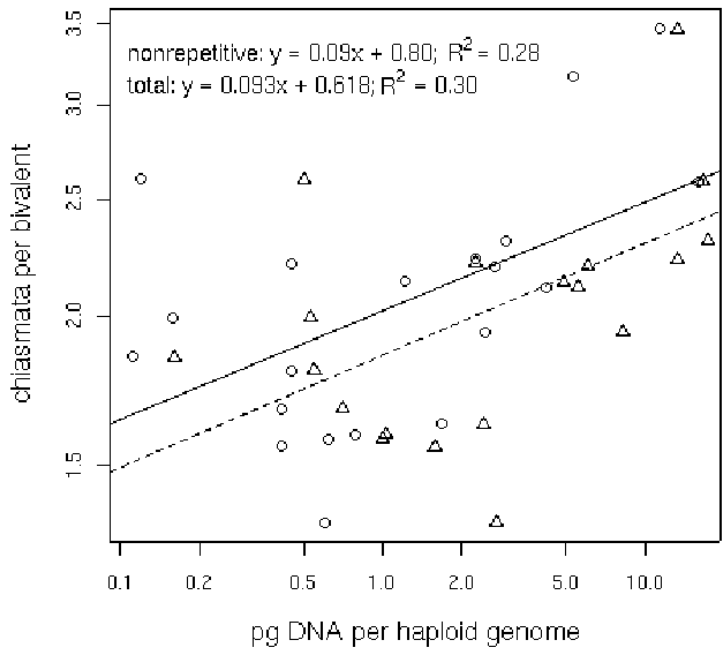


Figure 4.2: Linear regression of chiasmata per bivalent on nonrepetitive and total genome size. Circles and the solid line represent the regression of chiasmata per bivalent on nonrepetitive genome size, while triangles and the dashed line represent the regression on total genome size.

log-log regression never exceeded 0.25 and the above survey of the available data on euchromatic genome size provide little reason to believe that the true regression slope would be substantially higher. Recombination events require physical space along a chromosome, and a purely mechanistic view would predict change in recombination proportional to change in genome size and thus the space available for recombination events. There are several reasons to believe that this expectation is unrealistic, however. All bivalents, regardless of size, must have at least one chiasmata to ensure normal meiotic pairing and assortment (Dawe, 1998), and this effect will be exacerbated if crossover interference extends over a sizable length of the chromosome (Egel, 1995). Moreover, theoretical considerations suggest that even when recombination is selectively favored a small amount of recombination is usually sufficient to break up correlations between loci (Burt, 2000) and excessive recombination can actu-

ally be detrimental (Barton, 1995), such that selection alone could constrain increases in recombination to be less than proportional to change in genome size.

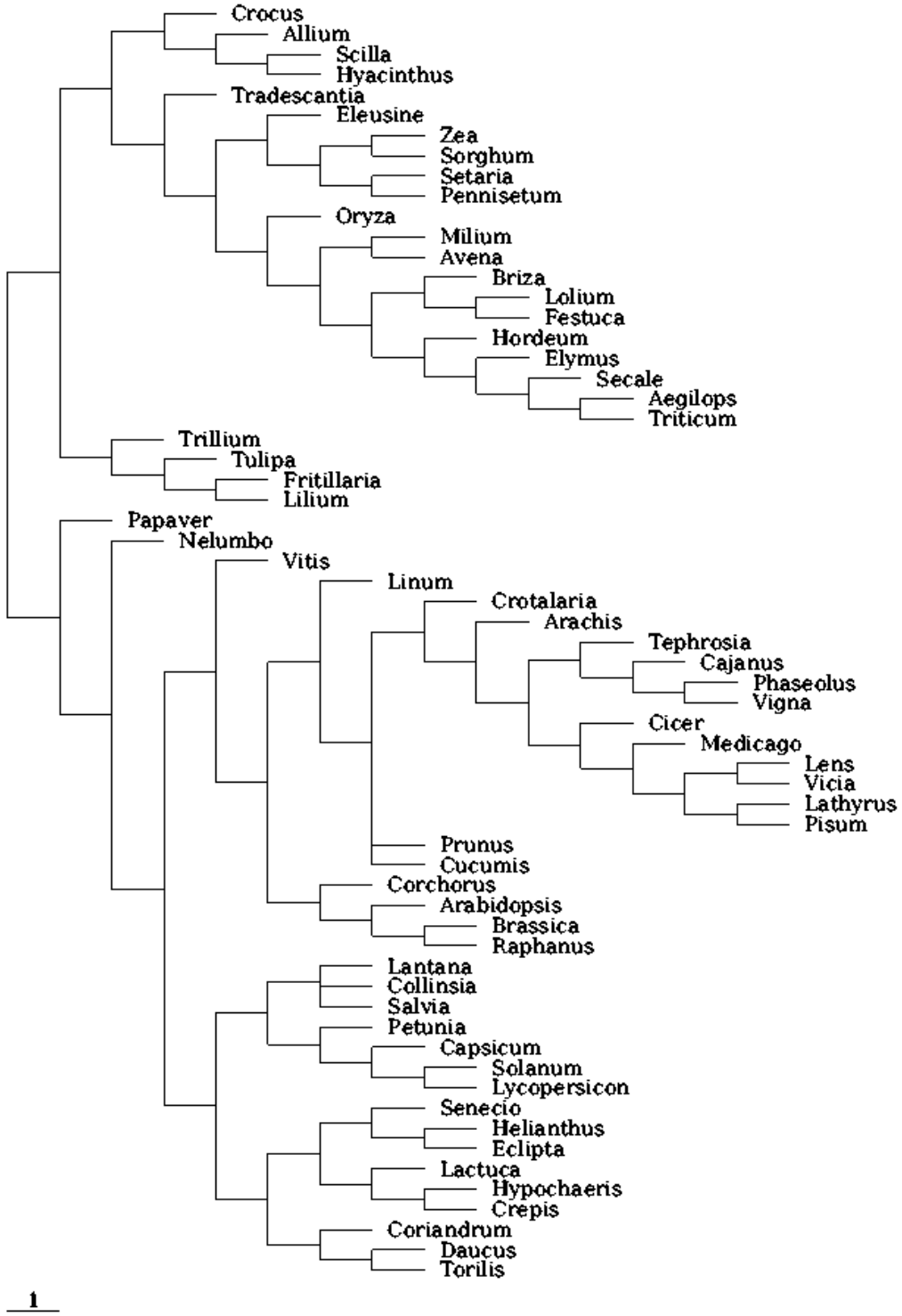
I have shown in this paper that genome size is significantly correlated to recombination rate across a wide sampling of Angiosperm species even after correction for phylogenetic effects, and that change in genome size explains a meaningful proportion of the observed variation in recombination rate. Though the magnitude of the observed effect of genome size on recombination rate is comparable to the effects of several life history characters previously linked to evolutionary change in recombination rate, consideration of processes of genome size change in Angiosperms suggests that the observed correlation is likely conservative. And finally, while recombination rate was found to increase less than proportionally to change in genome size, the theoretical and mechanistic arguments mentioned above suggest that this result is not unexpected.

4.4 Acknowledgements

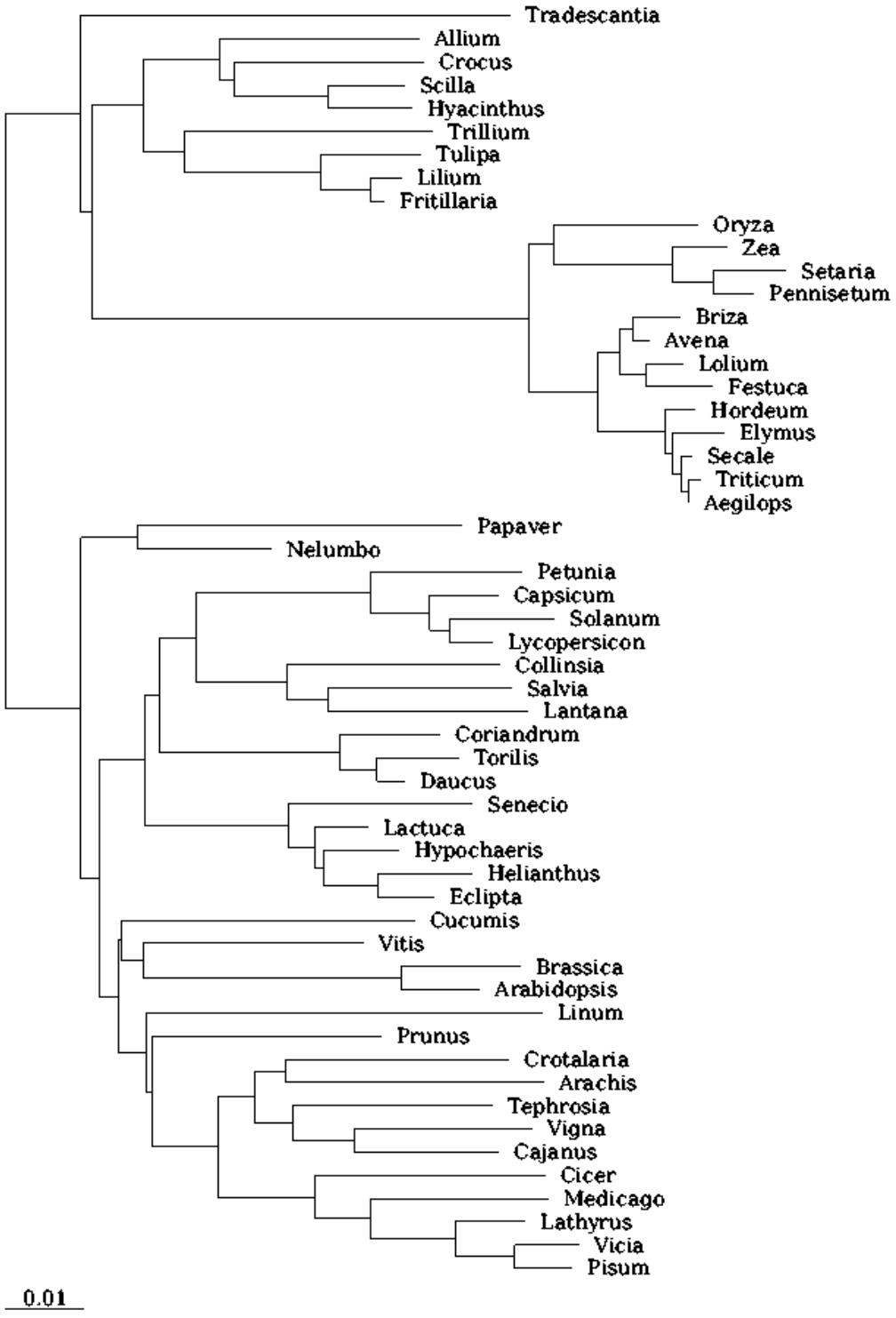
I would like to thank E. Martins and E. Housworth for help with the COMPARE software, R. Mauricio and D. Promislow for statistical help, M. Arnold, K. Dawe, W. Parrott, and S. Ross for helpful discussion, and J. Mank for advice and the motivation to write this paper. M. Arnold, S. Cornman, J. Hamrick, E. Kuntz, N. Martin, W. Parrott, S. Small, and an anonymous reviewer helped improve previous drafts of this manuscript. Finally, both the manuscript and author have greatly benefited from the considerable guidance provided by S. Otto.

Fig. 1.3. Phylogenies used for analysis. A. Equal branch lengths. B. Chloroplast rbcL neighbor-joining tree. C. Fabaceae rDNA neighbor-joining tree. D. Poaceae rDNA neighbor-joining tree. In A and B species within a genus are coded as polytomies (not shown).

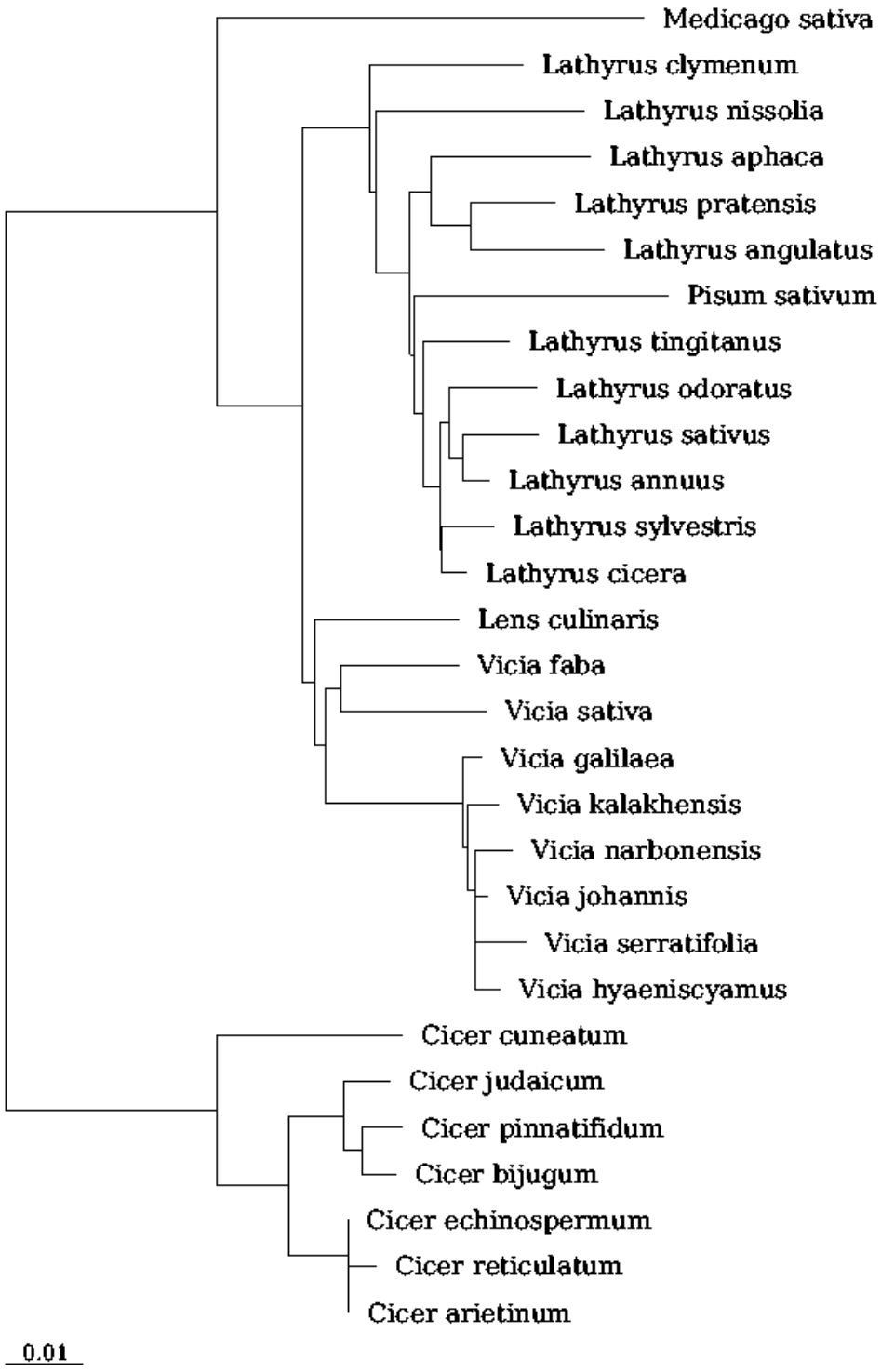
A.


Crocus
Allium
Scilla
Hyacinthus
Tradescantia
Eleusine
Zea
Sorghum
Setaria
Pennisetum
Oryza
Milium
Avena
Briza
Lolium
Festuca
Hordeum
Elymus
Secale
Aegilops
Triticum
Trillium
Tulipa
Fritillaria
Lilium
Papaver
Nelumbo
Vitis
Linum
Crotalaria
Arachis
Tephrosia
Cajanus
Phaseolus
Vigna
Cicer
Medicago
Lens
Vicia
Lathyrus
Pisum
Prunus
Cucumis
Corchorus
Arabidopsis
Brassica
Raphanus
Lantana
Collinsia
Salvia
Petunia
Capsicum
Solanum
Lycopersicon
Senecio
Helianthus
Eclipta
Lactuca
Hypochaeris
Crepis
Coriandrum
Daucus
Torilis
1

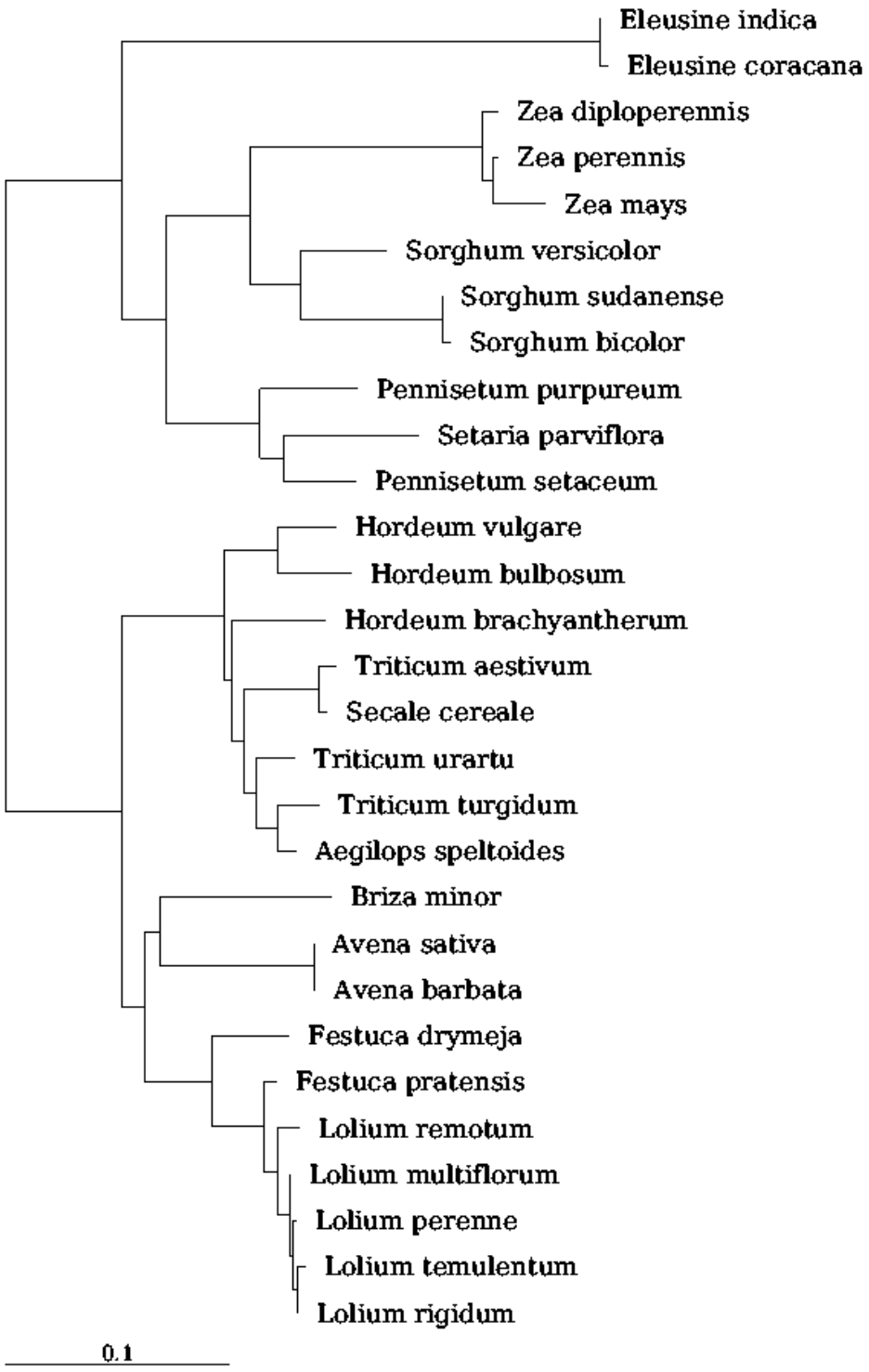
B.


Tradescantia
Allium
Crocus
Scilla
Hyacinthus
Trillium
Tulipa
Lilium
Fritillaria
Oryza
Zea
Setaria
Pennisetum
Briza
Avena
Lolium
Festuca
Hordeum
Elymus
Secale
Triticum
Aegilops
Papaver
Nelumbo
Petunia
Capsicum
Solanum
Lycopersicon
Collinsia
Salvia
Lantana
Coriandrum
Torilis
Daucus
Senecio
Lactuca
Hypochaeris
Helianthus
Eclipta
Cucumis
Vitis
Brassica
Arabidopsis
Linum
Prunus
Crotalaria
Arachis
Tephrosia
Vigna
Cajanus
Cicer
Medicago
Lathyrus
Vicia
Pisum
0.01

C.


Medicago sativa
Lathyrus clymenum
Lathyrus nissolia
Lathyrus aphaca
Lathyrus pratensis
Lathyrus angulatus
Pisum sativum
Lathyrus tingitanus
Lathyrus odoratus
Lathyrus sativus
Lathyrus annuus
Lathyrus sylvestris
Lathyrus cicera
Lens culinaris
Vicia faba
Vicia sativa
Vicia galilaea
Vicia kalakhensis
Vicia narbonensis
Vicia johannis
Vicia serratifolia
Vicia hyaeniscyamus
Cicer cuneatum
Cicer judaicum
Cicer pinnatifidum
Cicer bijugum
Cicer echinospermum
Cicer reticulatum
Cicer arietinum
0.01

D.


Eleusine indica
Eleusine coracana
Zea diploperennis
Zea perennis
Zea mays
Sorghum versicolor
Sorghum sudanense
Sorghum bicolor
Pennisetum purpureum
Setaria parviflora
Pennisetum setaceum
Hordeum vulgare
Hordeum bulbosum
Hordeum brachyantherum
Triticum aestivum
Secale cereale
Triticum urartu
Triticum turgidum
Aegilops speltoides
Briza minor
Avena sativa
Avena barbata
Festuca drymeja
Festuca pratensis
Lolium remotum
Lolium multiflorum
Lolium perenne
Lolium temulentum
Lolium rigidum
0.1

Chapter 5

Diversity and selection in tomatillo, *Physalis philadelphica*[1]

[1]Ross-Ibarra, J., J.L. Hamrick, and R. Mauricio. To be submitted to *American Journal of Botany*.

Abstract

In this study we investigate the effects of domestication on genetic diversity in tomatillo (*Physalis philadelphica*, Solanaceae), an economically important pre-Columbian domesticate from central Mexico. Cultivated tomatillos are morphologically highly diverse, but little is known about the genetics of tomatillo or the process of its domestication beyond the fact that selection for fruit size has played an important role. We investigate how genetic diversity has changed under the process of domestication and compare the distribution of neutral genetic diversity with patterns of quantitative diversity for traits known to be under selection and traits not thought to have been important for artificial selection. We planted a number of accessions of wild, landrace, and cultivated populations of tomatillo from throughout its range in a common garden and measured levels and patterns of genetic diversity at multiple polymorphic allozyme loci and several morphological traits. We find that domestication has had little effect on overall levels of tomatillo diversity but that wild and weedy accessions nonetheless harbor diversity not found in cultivated types. We also show that directional selection on fruit size during tomatillo domestication has had a dramatic effect on patterns of quantitative genetic diversity in cultivated populations, and that this pattern of variation is found for leaf size but not leaf shape characters as well.

5.1 Introduction

Domestication has had dramatic effects on the evolution of numerous plant species. Strong artificial selection and population bottlenecks usually associated with domestication are responsible for large changes in the morphology and genetic diversity of these species. Nonetheless, the process of domestication is far from homogeneous, and the effects of domestication differ greatly depending on the biology of a species as well as its history of anthropogenic use.

In this study we investigate several hypotheses concerning the effects of domestication on genetic diversity in tomatillo (*Physalis philadelphica*, Solanaceae), an economically im-

portant pre-Colombian domesticate from central Mexico. Though not well known outside of Mexico, the tomatillo has long been an important agricultural product within Mexico, probably preceding the tomato in domestication Jenkins (1948). The tomatillo is an annual, self-incompatible herb that is widely distributed throughout Mexico and Central America, often cultivated in close proximity to weedy or wild populations. Cultivated tomatillos are morphologically highly diverse, but little is known about the genetics of tomatillo or the process of its domestication beyond the fact that selection for fruit size has played an important role — fruit in domesticated varieties are up to 15 times larger than wild fruit. In this study, we investigate how genetic diversity has changed under the process of domestication and compare the distribution of neutral genetic diversity with patterns of quantitative diversity for traits known to be under selection and traits not thought to have been important for artificial selection. Comparison of the structure of quantitative genetic variation with that at neutral marker loci can be used as evidence to support hypotheses of selection acting on a trait Wright (1951); Spitze (1993), and should allow us to distinguish between traits under directional artificial selection and those unrelated to domestication.

Given what is known about the biology of the species, we predict little loss of genetic diversity in cultivated tomatillo compared to that seen in other crops. The increased opportunity for gene flow among cultivated varieties and weedy populations that is associated with domestication should also result in a decrease in observed levels of genetic structure. Quantitative traits not associated with domestication should show a similar pattern, and genetic structure statistics for these traits should follow genetic structure at neutral loci closely. Traits important for domestication, on the other hand, should show a greater loss of diversity than neutral markers, and a decrease in structure among populations beyond that observed at neutral markers.

To test these hypotheses in tomatillo, we collected a number of accessions of wild, landrace, and cultivated populations of tomatillo from throughout its range. We planted these accessions in a common garden and measured levels and patterns of genetic diversity at mul-

tiple polymorphic allozyme loci and several morphological traits. We measure fruit weight, a trait known to be under directional artificial selection, as well as leaf size and shape, thought unlikely to have any relation to tomatillo domestication Hudson (1983) and shown to be uncorrelated to fruit and floral traits in other Solanaceous genera Frary et al. (2003, 2004).

5.2 Materials and Methods

Seeds were collected from a variety of sources including germplasm collections, cultivated fields, and wild populations, representing a broad geographic sampling of cultivated and wild populations (Table 5.6). Seed from cultivated fields or wild populations was combined in a bulk sample similar to germplasm accessions. Wild collections were defined as those that had not been planted or cultivated by humans in recent memory. As such, our definition of wild includes weedy and potentially feral populations. Landraces were defined as populations found in cultivated fields but described by farmers as self-sowing or naturally occurring — "milpero" or "criollo" varieties. Cultivated populations include both local varieties and commercially developed lines.

In June 2005 seeds were germinated in soil flats under greenhouse conditions, and then transplanted to the field at the University of Georgia Plant Sciences Farm four weeks after sowing. Plant height was measured when each individual was planted in the field. Plants were fertilized once during the growing season, and watered 2-3 times weekly until all leaves and fruit were harvested. The field was divided into 10 rows, with a single individual of each accession randomly planted at 1m intervals in each row. Plants that died within the first three weeks of the field experiment were replaced with individuals from the greenhouse when possible. Because of limited germination and mortality in some accessions, we grouped the 10 rows into two blocks of five rows each to ensure replication within blocks.

5.3 Genetic Analysis

We sampled leaf tissue from every surviving individual in the common garden. Leaf samples were also taken from additional individuals grown in the greenhouse that were not included in the common garden (Table 5.6), allowing us to increase the sample size for several accessions and include accessions that were not planted in the common garden. Leaf samples were packed on ice and taken to the laboratory, where they were extracted onto paper using the extraction buffer of Wendel and Parks (1982) and stored at -70^oC until analysis. We used starch gel electrophoresis to resolve allozyme genotypes for the following enzyme systems: diaphorase (DIA, one locus), isocitrate dehydrogenase (IDH, one locus), menadione reductase (MNR, one locus), 6-phosphogluconate dehydrogenase (6-PGDH, one locus), phosphoglucoisomerase (PGI, three loci), phoshoglucomutase (PGM, two loci) and triose-phosphate isomerase (TPI, two loci). Gel and electrode buffer systems were adapted from Soltis et al. (1983). For PGM and TPI we used a buffer 34, while buffer 10 was used for PGI, a modified buffer 8 used for DIA and MNR, and buffer system 11 was used for all other loci. Staining recipes and procedures followed Cheliak and Pitel (1984) for DIA and MNR, and Soltis et al. (1983) for all others.

We calculated standard genetic diversity statistics for cultivated, landrace, and wild types as well as for individual accessions. An analysis of molecular variance (AMOVA, Excoffier et al., 1992) was carried out first to partition variation among populations and types and then to investigate variation among geographic regions. Diversity statistics and the AMOVA among geographical regions made use of all available individuals; the reported genetic structure among populations within a cultivation type only includes individuals sampled from the common garden plot. All genetic analyses were performed using the GenAlEx software Peakall and Smouse (2006).

5.3.1 Morphological Traits

Fruit were harvested from plants when a minimum of five mature fruit could be simultaneously harvested from an individual plant; all available mature fruit were harvested from each individual. In cases in which this was not possible because of low fruit production, fruit were harvested as they matured. Fruit were separated from the calyx and weighed. Fruit badly damaged from herbivory were discarded.

To minimize the potential effect of early leaf harvesting on fruit production, leaves were harvested after fruit collection. One leaf was taken from the fifth node basal to the apical meristem on each of five shoots from each individual. When this was not possible, leaves were occasionally taken from either the fourth or sixth node. In a small number of cases, it was not possible to sample five leaves from a plant. Leaves were pressed in newspaper and dried in a drying oven at 45^oC. Dried leaves were digitally photographed and measured using the ImageJ image analysis software (http://rsb.info.nih.gov/ij/). The software measures the perimeter and area of each leaf, and then fits an ellipse to the shape of the leaf and measures the major and minor axes of the ellipse (hereafter referred to as leaf length and width for simplicity). Damaged leaves were excluded from area and perimeter measurements, but included in the linear measurements. Additionally, we calculated circularity — a dimensionless statistic calculated as $\frac{4\pi(area)}{(perimeter)^2}$ and varying from 1 (a perfect circle) to 0 — as a measure of leaf shape.

Statistical analysis of the morphological data was performed using the R statistical language (http://www.r-project.org/). A mixed effect model was fitted to the data for each trait, with block as a fixed effect and type, accession, and individual as nested random effects, allowing calculation of the variance components of each level. All data except circularity were log-transformed before analysis. To compare the distribution of quantitative genetic variation with neutral variation at allozyme loci, we calculated the Q_{ST} statistic as $\frac{V_{type}}{V_{type}+V_{pop}+2V_{ind}}$ among types or $\frac{V_{pop}}{V_{pop}+2V_{ind}}$ among accessions within types or for comparison across all accessions.

Table 5.1: Total genetic diversity H_T, average population diversity H_S, percent polymorphism $\%P$, and alleles per polymorphic locus $A_P$calculated from entire data set. Genetic structure statistics, $F_{ST}$are calculated from an AMOVA of the common garden collections only. Shown are pooled values for all populations of a given type.

	H_T	H_S	$\%P$	A_P	F_{ST}
Types					0.106
Cultivated	0.310	0.263	0.7545	2.105	0.059
Landrace	0.326	0.298	0.8182	2.141	0.087
Wild Pops	0.334	0.266	0.7706	2.191	0.139

5.4 Results

5.4.1 Genetic Diversity

All of the 11 loci analyzed were polymorphic in at least one population. Total genetic diversity was high across all three types of cultivation (mean He = 0.323), but the cultivated class had slightly lower diversity than landraces or wild types (Table 5.1), and five alleles not found in cultivated types were present in landrace and wild populations (three alleles were unique to wild types and one unique to landraces). This trend was not evident for diversity statistics within populations, in which landrace populations showed higher diversity than either of the other categories.

Analysis of molecular variance of individuals from the common garden revealed that the majority of genetic variation (89%) occurs within populations, with a relatively small amount (11%) due to differences among populations within a type and virtually zero variation due to differences among types. Nonetheless, genetic structure within types was not identical, evidencing a distinctive trend towards lower structure in the landrace and cultivated populations (Table 5.1). No evidence of clustering due to geographic locality was observed: less than 1% of the total variation was due to differences among geographical regions, and the within-region structure among populations was nearly identical to structure for all popula-

Table 5.2: Variance component analysis and Q_{ST} for fruit size data. Shown are variances and standard deviation of the variances.

	types	accessions	individuals	fruit	Q_{ST}
All – types	1.272 (1.128)	0.050 (0.223)	0.237 (0.487)	0.183 (0.428)	0.708
All – pops		1.024 (1.012)	0.235 (0.485)	0.183 (0.428)	0.685
Cultivated		< 0.001	0.282 (0.531)	0.210 (0.459)	0.001
Landrace		0.131 (0.362)	0.129 (0.359)	0.167 (0.408)	0.338
Wild		0.074 (0.271)	0.234 (0.484)	0.177 (0.420)	0.136

Table 5.3: $Q_{ST}$for leaf traits.

	All Populations	Type	Cultivated	Landrace	Wild
length	0.192	0.089	0.031	0.061	0.197
width	0.126	0.019	0.027	0.128	0.173
area	0.159	0.048	0.024	0.106	0.193
perimeter	0.157	0.072	< 0.001	0.113	0.179
circularity	0.135	0.034	0.067	0.159	0.159

tions ($F_{ST}$of 0.122 vs. 0.119). Patterns of structure were not qualitatively different when all populations and individuals were included.

5.4.2 Quantitative Traits

High levels of variation within plants were evident for both fruit (Table 5.2) and leaf traits (data not shown), though within-plant variation never exceeded among-plant variation for any trait. None of the leaf traits or fruit weight were correlated with initial height at planting ($p>0.1$) and were accordingly left uncorrected.

Fruit weights closely matched expectations based on cultivation type: cultivated varieties had larger fruits than either landrace or wild types ($p<0.001$), and landraces were significantly larger than wild types ($p<0.05$). Nonetheless, the magnitude of differences between landrace and wild types was relatively small (Fig. 5.1). Fruit weight $Q_{ST}$was extremely high

and much larger than F_{ST} both among types and among all accessions, reflecting the large observed size differences (Table 5.2). Within wild accessions Q_{ST} for fruit weight was very similar to F_{ST}, while it was close to 0 among cultivated accessions and high among landraces (Table 5.2).

The four leaf size measurements were all strongly correlated with each other ($p<0.001$, Pearson correlation coefficient ≥ 0.9 for all pairwise comparisons). Circularity was not significantly correlated to leaf area or leaf width, but was significantly negatively correlated with leaf length and leaf perimeter ($p<0.05$, Pearson correlation coefficient -0.24 and -0.12, respectively). Cultivated individuals had significantly larger leaves than either wild plants or landraces for all four leaf size measurements ($p<0.01$) and significantly lower circularity ($p<0.05$, Fig. 5.1). No significant differences were observed between landrace and wild individuals for any leaf measurements. Not surprisingly, Q_{ST} for the four size measurements behaved similarly, decreasing from wild to landrace and cultivated populations. Comparison of Q_{ST} to F_{ST} for leaf size traits revealed greater quantitative structure among quantitative traits than neutral markers for wild accessions, similar levels of structure in landraces, and noticeably lower Q_{ST} than F_{ST} among cultivated accessions (Table 5.3). The single exception to this pattern was leaf length among landrace accessions, in which Q_{ST} was slightly lower than the observed F_{ST}. For leaf circularity, $Q_{ST}$was similar to $F_{ST}$for both cultivated and wild types, but somewhat higher among landraces. All five leaf measurements showed a Q_{ST} larger than F_{ST} across all pooled accessions.

5.5 Discussion

Consistent with our predictions, the allozyme data revealed a negligible loss of diversity in cultivated tomatillo. This contrasts strongly with the pattern for most domesticates Doebley (1989), in which the domesticate exhibits 75% or less of the diversity of its wild progenitor. Nonetheless, this observation is in good accord with the biology and history of tomatillo: it is not surprising that a weedy, annual, obligately outcrossing plant frequently subject

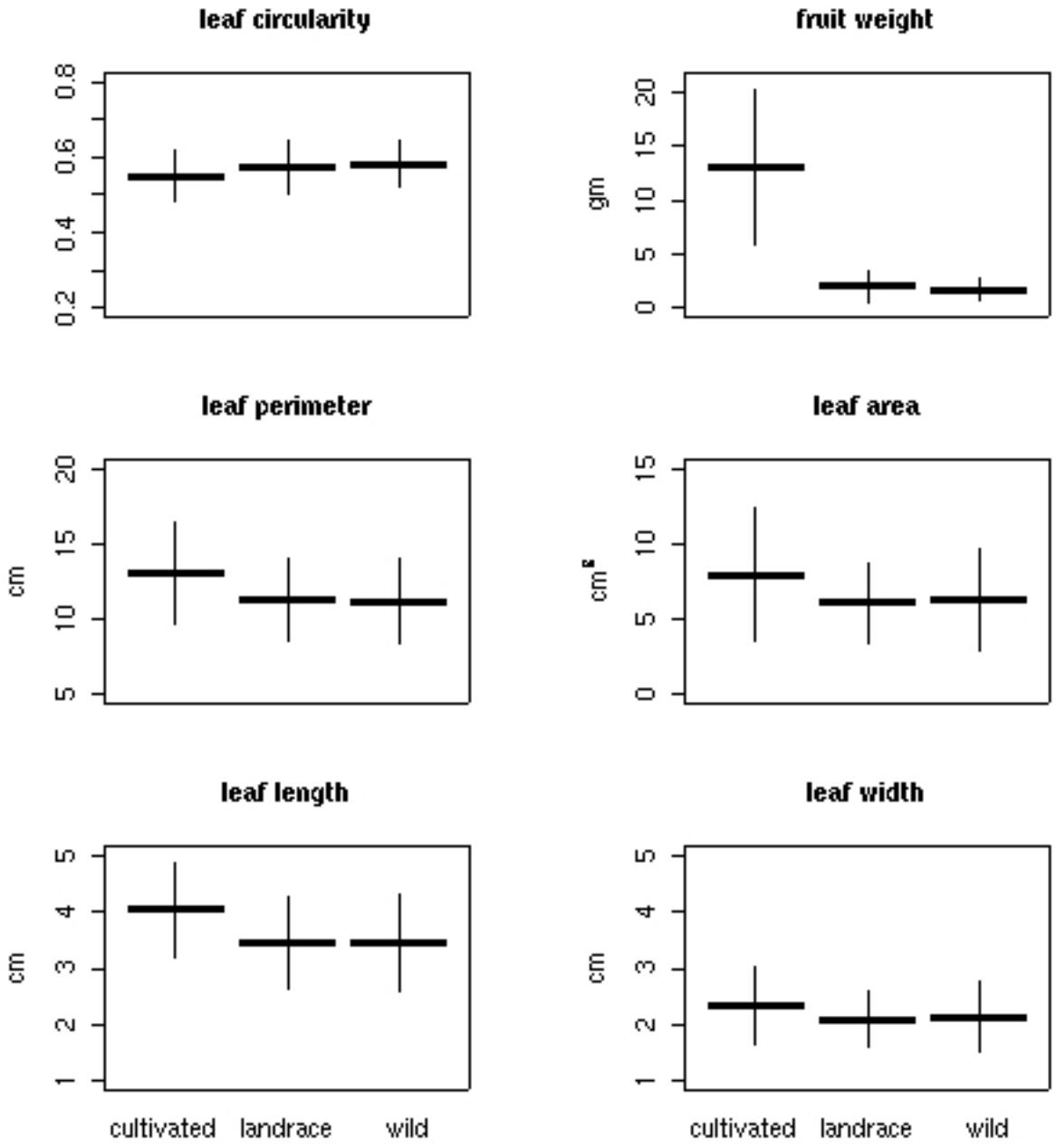


Figure 5.1: Fruit weight and leaf morphology across cultivation types. Shown are individual plant means and standard deviations. Cultivated taxa are significantly different from wild taxa for all characters, and from landraces for all characters except circularity. Landrace and wild taxa are only significantly different for fruit weight.

to anthropogenic dispersal would have high levels of genetic diversity Hamrick and Godt (1996). Few advanced cultivars have been developed for tomatillo, and even those have been created over generations of mass selection, minimizing the bottleneck effect evident in advanced lines of most crops. Moreover, it is common practice for farmers to plant seed from multiple cultivated varieties and utilize seed from their own field for future growing seasons (J. Ross-Ibarra, pers. obs.), providing a mechanism for high levels of gene flow among populations of this predominantly outcrossing species. Additionally, weedy or feral tomatillo can be found growing in virtually all of the areas in which tomatillo is cultivated, and weedy and domesticated types are fully compatible Hudson (1983). All of these factors contribute to high expected levels of genetic diversity within individual populations and low levels of differentiation among cultivated populations.

In spite of the negligible loss of diversity in cultivated populations and low levels of differentiation among types, the allozyme data lends support to the idea that our sampled populations include truly "wild" tomatillo. All of the wild collections utilized were collected in disturbed habitat, raising the concern that they represent weedy or feral forms instead of truly wild tomatillo. However, the observation of unique alleles in wild populations at three loci support the idea that the wild collections are not simply feral or weedy samples of cultivated material but, in fact, include diversity not found in cultivated types.

Analyses of molecular variance allowed us to partition the observed genetic diversity among populations and types. Genetic differentiation among types is similar to that among populations within types, consistent with the mating system and the possibility of interbreeding and anthropogenic dispersal between types. The observed decrease in F_{ST}in cultivated and landrace populations is undoubtedly due to anthropogenic effects. While any bottleneck or founder effect stemming from varietal choices has had minimal impact on diversity, increased mixing among cultivated populations and the potential for gene flow with wild or weedy populations has decreased differentiation among populations in an anthropogenic context.

Variation in fruit data closely approximated our predictions for wild and cultivated populations. Differentiation in fruit size among types was substantially greater than differentiation at allozyme loci, consistent with directional selection for fruit size in cultivated but not wild types. Patterns of fruit size variation among wild populations closely matched the differentiation observed in allozyme loci, providing no evidence for selection on fruit size in the wild. Fruit size differentiation among cultivated accessions contrasted starkly with the pattern observed in wild populations, with less than 1% of the observed variation due to differences among accessions — again consistent with directional selection operating similarly in all cultivated populations. Population bottlenecks Lopez-Fanjul et al. (2003) are unlikely to be responsible for the observed difference judging from the high levels of allozyme diversity in cultivated tomatillo.

Contrasting the relatively good fit of allozyme and fruit weight data to expectations for wild and cultivated species, landrace accessions showed unexpected results. Average population level genetic diversity was higher than in either wild or cultivated species, due in part to a single population (accession 49) which was the only population to show variability at locus TPI-2. And though landrace fruit sizes were intermediate, patterns of quantitative genetic variation were far from intermediate. The observed $Q_{ST}$for fruit size was much higher than F_{ST} and more than twice the Q_{ST}in wild populations. Much of this effect is due to accession 55, which has a significantly larger mean fruit size than any of the other landrace accessions (student's t-test, $p<0.5$). Additionally, our sample size of landrace accessions is small, and the inflated Q_{ST} may be an artifact of this sampling.

Nonetheless, our initial assumption that landrace accessions would be intermediate to the two other categories seems poorly founded. Landrace populations are generally self-sowing, and some farmers claim that they have never planted seed, suggesting that the harvested plants could be, in effect, weedy or wild plants allowed to grow in cultivated sites. In other cases farmers clearly cultivate tomatillo, weeding plots, spraying with insecticide, etc., differentiating these populations little from other cultivated types. And while some landraces

may be harvested weedy individuals with small fruit, many farmers actually select for small-fruited types either because of their superior taste or the increased market price they can garner by claiming the tomatillos were wild-collected (J. Ross-Ibarra, pers. obs.). It is thus likely that "landrace" includes a mixture of populations at various stages of domestication, which would explain both the presence of a novel allele (at locus TPI-2) and the high Q_{ST} suggestive of directional selection varying across populations.

Previous studies led us to hypothesize that leaf characters would show patterns different from fruit weight, a trait known to be under selection. There is no evidence for direct artificial selection on leaf size during tomatillo domestication, little differentiation between domesticated and wild varieties in the distribution of leaf morphologies (Fig. 5.1, Hudson (1983)), and evidence for differential genetic architecture for leaf and floral or fruit traits in other related Solanaceous crops Frary et al. (2004, 2003). In spite of this previous evidence, Q_{ST} for leaf size traits behaved similarly to those for fruit weight. In wild populations, Q_{ST} values suggest the presence of selection on leaf size. Among landrace populations Q_{ST} is only slightly higher than F_{ST}, due in part to the significantly smaller leaves of at least one population (49), and potentially to differences in cultivation technique or population type (weedy vs. cultivated landraces).

It is not entirely clear why leaf size traits behaved similarly to fruit size, though it is clear that the two traits are correlated ($p<0.01$ for all individuals, $p<0.05$ among cultivated or wild individuals, Fig. 5.2). There is no evidence that artificial selection on tomatillo has targeted leaf size, and it thus seems that change in leaf size is a pleiotropic effect of selection on another trait. We suggest that increased leaf size likely resulted from selection for increased plant vigor. Though we do not have measurements of final plant size in the field, cultivated plants were significantly larger than wild plants when initially planted (mean height = 39.8 cm vs. 32.2 cm, $p<0.05$).

Furthermore, 33 out of 39 populations show a positive correlation between fruit size and leaf size among individuals within populations, suggesting a genetic correlation between the

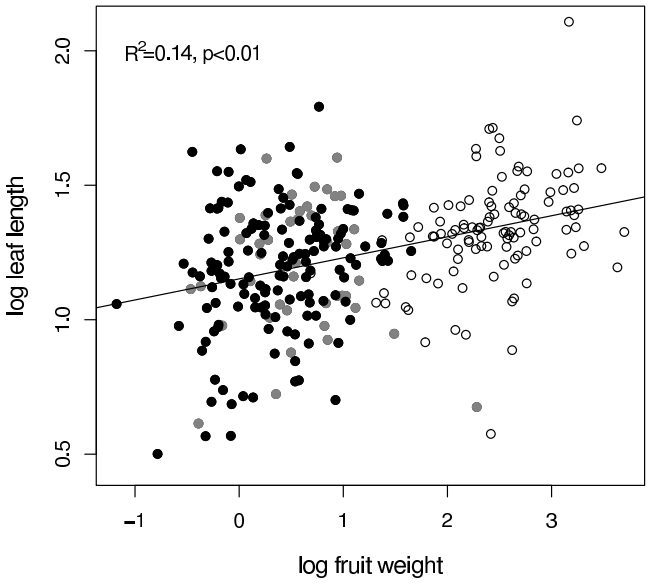


Figure 5.2: Correlation between fruit weight and leaf length among individual plants. Black circles represent wild, grey circles landraces, and open circles are cultivated plants. Shown is the regression line for all individuals.

two traits. Previous work in tomatillo showed no correlation between population means for leaf size and fruit diameter Hudson (1983), which is also seen in our data ($p>0.1$), but does not report any test for a correlation among individuals. And while QTL studies suggest little overlap in the genetic architecture of both traits Frary et al. (2003, 2004), the QTL found by these studies do not explain all of the variance for either trait and these results are limited to individual crosses and may thus miss some of the variation present in natural populations. It is also worth noting that correlations between leaf size and fruit size are found in several other plant species Herrera (2002).

Leaf shape, measured here as circularity, only differed from neutral expectation in landrace populations. Thus, although both Q_{ST} and F_{ST} are lower in cultivated populations, there is no change in the relationship of the two measures. Unlike leaf size traits, there is no correlation in our data between leaf shape and fruit size, and leaf shape is only weakly correlated with measurements of leaf size (data not shown). Leaf shape thus serves as as

a useful contrast to fruit size, evidencing the pattern expected of a neutral trait and thus supporting our interpretation of the genetic data as evidence of selection on fruit size.

Finally, it is worth commenting on the significance of Q_{ST} comparisons. We have refrained from estimating confidence intervals or making claims about the statistical significance of Q_{ST} - F_{ST} comparisons for several reasons. There are a number of statistical issues bedeviling accurate calculation of confidence intervals for Q_{ST} O'Hara and Merila (2005), and our relatively small sample size only exacerbates this problem. Similarly, there are a number of assumptions underlying calculation of Q_{ST}, including mutation rates, Hardy-Weinberg equilibrium, and the neutrality of allozymes, and the additivity of a quantitative trait. Violation of these can affect the estimation of Q_{ST} and its confidence interval. O'Hara and Merila (2005) have argued that a large percentage of the confidence intervals calculated for Q_{ST} are likely to not include the true value; we have thus avoided the situation altogether.

Our investigation of genetic diversity in cultivated and wild tomatillo has shown that domestication has had little effect on overall levels of tomatillo diversity but that wild and weedy accessions nonetheless harbor diversity not found in cultivated types. We have additionally shown that directional selection on fruit size during tomatillo domestication has had a dramatic effect on patterns of quantitative genetic diversity in cultivated populations. The simple observation of decreased Q_{ST} due to artificial selection is of some import. This is the first study of which we are aware that tests the expected pattern of variation from artificial selection. Other studies comparing quantitative genetic variation to neutral marker F_{ST} utilize only wild populations, and the majority find values of Q_{ST} exceeding the observed F_{ST} Merila and Crnokrak (2001). While this implies a dichotomy between artificial selection and natural selection, we argue that numerous scenarios in the wild — including climate change, disease, changes in interspecific interactions, or simple stabilizing selection — could create selection regimes that would lead to the observed pattern of Q_{ST} and F_{ST}. Looking for such examples is not as simple as picking a random morphological trait and testing for differential selection across populations, but we suspect that as more authors make a concerted effort

to look for stabilizing or species-wide directional selection, we will discover that it is much more common than current reviews would suggest.

5.6 Acknowledgements

We would like to thank the G. Baucom, S. Cornman, M. Poelchau, C. Ross-Ibarra, and S. Small for help with field and laboratory work, S. Cornman and M. Poelchau for useful discussion, and the Plant Biology Greenhouse staff and the University of Georgia Plant Sciences Farm staff for their invaluable assistance. K. Parker, A. Paterson, D. Promislow, and P. Ross helped revise an earlier version of this manuscript.

Table 5.6 Accessions, type, geographic origin, and number of individuals used in field and total genetic diversity analyses.

Accession	Type	Origin	N-field	N-genetic
1	wild	Nayarit	7	13
2	wild	San Luis Potosi	8	9
3	wild	Jalisco	10	18
4	wild	Jalisco	9	11
5	wild	Jalisco	9	22
8	wild	Puebla	8	21
11	wild	Guanajuato	3	5
15	wild	Jalisco	2	6
21	wild	Jalisco	0	6
23	wild	Guerrero	9	13
24	wild	Guerrero	9	9
27	wild	Oaxaca	9	22
29	wild	Oaxaca	3	11
35	wild	Guerrero	8	25
38	wild	Chihuahua	9	16
39	wild	Chihuahua	10	16
40	wild	Chihuahua	10	11
41	wild	Chihuahua	8	8
50	wild	Chihuahua	9	25
51	wild	Chihuahua	8	22
52	wild	New Mexico, U.S.	10	22
7	cultivated	Jalisco	9	16
12	cultivated	Chihuahua	9	14

Accession	Type	Origin	N-field	N-genetic
13	cultivated	Morelos	10	12
14	cultivated	Guerrero	8	8
17	cultivated	Guanajuato	10	18
25	cultivated	Puebla	0	8
32	cultivated	Oaxaca	8	9
34	cultivated	Michoacan	10	12
36	cultivated	Guatemala	9	23
42	cultivated	Chihuahua	9	22
43	cultivated	Chihuahua	10	14
44	cultivated	Michoacan	0	5
45	cultivated	Sinaloa	9	16
46	cultivated	Puebla	8	9
47	cultivated	Guanajuato	9	9
9	landrace	San Luis Potosi	0	5
22	landrace	Chiapas	10	18
26	landrace	Oaxaca	9	9
30	landrace	Chiapas	10	12
33	landrace	Oaxaca	10	24
49	landrace	Guerrero	5	6
55	landrace	Hidalgo	3	3

Chapter 6

Conclusions

The work presented here advances our understanding of the roles of recombination and diversity in plant domestication. The central focus of this work has been the idea of preadaptation to domestication (Hammer, 1988), the possibility that some plant species are more readily domesticated than others.

In chapter 2 I use data on chiasma frequencies from more than 600 plant species to show that patterns of recombination rates do not support a preadaptive role for recombination, but do provide evidence for selection on recombination during the process of domestication.

Chapter 4 takes up the question of recombination rate and genome size. This question is of independent interest for our understanding of the relationship between genome size change and the evolution of recombination, but also allows me to address the concern that genome size differences may overshadow observed differences in recombination due to domestication. I find a significant positive correlation between genome size and chiasma frequencies, but also show that life history characteristics — including domestication status — remain significant predictors of recombination rate.

After addressing the importance of recombination, I investigated the possible preadaptive role of genetic diversity. I show that patterns of genetic diversity among species of the genus *Zea* are consistent with the predictions of preadaptation. Nonetheless, I argue that the important causative factor determining domestication success is more likely to be effective population size than overall levels of genetic diversity. I build a forward population genetic simulation model, explicitly framed in terms of maize evolution, and show that the estimated differences in effective population size among wild species of the genus *Zea* are sufficient to

explain the differential domestication success of these taxa. I note that evidence from broad surveys of allozyme diversity in cultivated and wild plants provides some support to the idea that effective population size generally serves as a preadaptation to domestication.

One of the most serious problems with any argument for preadaptation to domestication is that made by Darwin (1899), that any species could be domesticated if subjected to sufficient selection. In the case of maize, for example, there is little evidence that any of the wild taxa other than the progenitor of maize were ever subjected to human-mediated selection. While differences in estimated effective population size are consistent with the prediction of a preadaptation hypothesis, one cannot say that the other species could not have been domesticated if given the opportunity. I thus looked for a study system that would allow me to address this concern. The tomatillo, *Physalis philadelphica*, is a domesticated species with several wild relatives which, though exposed to human-mediated selection for many generations, have never been successfully domesticated. However, virtually no genetic work had been done in this genus. I thus undertook the first steps towards describing the domestication of tomatillo, investigating how patterns of diversity have changed over the course of domestication. This is the first study of which I am aware that tests the expected pattern of quantitative and neutral genetic variation resulting from artificial selection. I showed here that domestication has had little effect on overall levels of tomatillo diversity but that wild and weedy accessions nonetheless harbor diversity not found in cultivated types. Additionally, I found that directional selection on fruit size during tomatillo domestication has had a dramatic effect on patterns of quantitative genetic diversity in cultivated populations, and that this effect is unexpectedly seen in some leaf traits not related to domestication. Importantly, these findings will allow me to return in future work to look at patterns of diversity in these same traits in other species in the genus, thus providing a test of the preadaptive role of effective population size that circumvents Darwin's argument against preadaptation.

In addressing the roles of recombination and diversity in plant domestication, I hope to have taken evolutionary biology a few small steps closer to an understanding of the

population genetics of plant domestication and shed some light on why so few of the many known species of plants were successfully domesticated.

Bibliography

Ackerly, D. D. (2000). Taxon sampling, correlated evolution, and independent contrasts. *Evolution*, 54:1480–1492.

Anderson, L., Doyle, G., Brigham, B., Carter, J., Hooker, K. D., Lai, A., Rice, M., and Stack, S. M. (2003). High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. *Genetics*, 165:849–865.

Anderson, L. K., Hooker, K. D., and Stack, S. M. (2001). The distribution of early recombination nodules on zygotene bivalents from plants. *Genetics*, 159:1259–1269.

Angiosperm Phylogeny Group (2003). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.*, 141:399–436.

Avise, J.C. and J.L. Hamrick, editor (1996). *Conservation genetics: case histories from nature*. Chapman & Hall.

Barton, N. H. (1995). A general model for the evolution of recombination. *Genet. Res.*, 65:123–145.

Bennett, M. and Leitch, I. (2005). Plant DNA C-values database. http://www.kew.org/cval/homepage.html. V. 4.0.

Bennetzen, J. L. (2002). Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica*, 115:29–36.

Bradshaw, A. D. (1991). The Croonian Lecture, 1991 - Genostasis and the Limits to Evolution. *Phil Tran. Roy. Soc. B.*, 333:289–305.

Buckler, E. S. I. V., Thornsberry, J. M., and Kresovich, S. (2001). Molecular diversity, structure and domestication of grasses. *Genet. Res.*, 77:213–218.

Burt, A. (2000). Perspective: sex, recombination, and the efficacy of selection–was Weismann right? *Evolution*, 54:337–351.

Burt, A. and Bell, G. (1987). Mammalian chiasma frequencies as a test of two theories of recombination. *Nature*, 326:803–805.

Cameron-Beaumont, C., Low, S., and Bradshaw, J. (2002). Evidence suggesting preadaptation to domestication throughout the small Felidae. *Biol. J. Linn. Soc.*, 75:361–366.

Cavalier-Smith, T. (1985). Eukaryote gene numbers, non-coding DNA and genome size. In Cavalier-Smith, T., editor, *The Evolution of Genome Size*, pages 69–104. John Wiley and Sons Ltd.

Charlesworth, B. (1993). Directional selection and the evolution of sex and recombination. *Genet. Res.*, 61:205–224.

Cheliak, W. and Pitel, J. A. (1984). Techniques for starch gel electrophoresis of enzymes from forest tree species. Technical report, Petawana National Forestry Institute, Canadian Forestry Service.

Clark, R. M., Tavare, S., and Doebley, J. (2005). Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Mol. Biol. Evol.*, 22:2304–2312.

Darlington, C. (1956). *Chromosome Atlas of Flowering Plants*. The MacMillian Company.

Darwin, C. (1899). *The Origin of Species*. H.M. Caldwell Co., New York, NY.

Dawe, R. K. (1998). Meiotic chromosome organization and segregation in plants. *Annu. Rev. Plant. Physiol. Plant Mol. Biol.*, 49:371–395.

Diamond, J. (2002). Evolution, consequences, and future of plant and animal domestication. *Nature*, 418:700–707.

Doebley, J. (1989). Isozymes in plant biology. In Soltis, D. and Soltis, P., editors, *Isozymes in Plant Biology*, pages 165–191. Chapman and Hall, London.

Doebley, J. (2004). The genetics of maize evolution. *Annu. Rev. Genet.*, 38:37–59.

Doebley, J., Goodman, M., and Stuber, C. (1984). Isoenzymatic variation in *Zea* (Gramineae). *Syst. Bot.*, 9:203–218.

Downie, S., Katz-Downie, D., and Watson, M. (2000). A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA rpl16 and rpoC1 intron sequences: towards a suprageneric classification of subfamily Apioideae. *Am. J. Bot.*, 87:273–292.

Egel, R. (1995). The synaptonemal complex and the distribution of meiotic recombination events. *Trends Genet.*, 11:206–208.

Evans, A. and Weir, J. (1981). The evolution of weed beet in sugar beet crops. *Genet. Resour. Crop Evol.*, 29:301–310.

Ewens, W. J. (1972). Sampling theory of selectively neutral alleles. *Theor. Popul. Biol.*, 3:87–112.

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131:479–491.

Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L., and Gaut, B. S. (1998). Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci. U. S. A.*, 95:4441–4446.

Feldman, M. W., Otto, S. P., and Christiansen, F. B. (1996). Population genetic perspectives on the evolution of recombination. *Annu. Rev. Genet.*, 30:261–295.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *Am. Nat.*, 125:1–15.

Felsenstein, J. (1993). Phylip (phylogeny inference package) version 3.6. distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Felsenstein, J. and Yokoyama, S. (1976). The evolutionary advantage of recombination. ii. individual selection for recombination. *Genetics*, 83:845–859.

Fisher, R. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press.

Flavell, R., Bennet, M., Smith, J., and Smith, D. (1974). Genome size and proportion of repeated nucleotide sequence dna in plants. *Biochem. Genet.*, 12:257–259.

Frary, A., Doganlar, S., Daunay, M. C., and Tanksley, S. D. (2003). QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species. *Theor. Appl. Genet.*, 107:359–370.

Frary, A., Fritz, L. A., and Tanksley, S. D. (2004). A comparative study of the genetic bases of natural variation in tomato leaf, sepal, and petal morphology. *Theor. Appl. Genet.*, 109:523–533.

Fu, Y. X. and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133:693–709.

Fukunaga, K., Hill, J., Vigouroux, Y., Matsuoka, Y., Sanchez, G. J., Liu, K., Buckler, E. S., and Doebley, J. (2005). Genetic diversity and population structure of teosinte. *Genetics*, 169:2241–54.

Gaut, B. S., Morton, B. R., McCaig, B. C., and Clegg, M. T. (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proc. Natl. Acad. Sci. U. S. A.*, 93:10274–10279.

Gepts, P. (2004). Crop domestication as a long-term selection experiment. In Janick, J., editor, *Plant Breeding Reviews*, volume 24, pages 1–23. John Wiley and Sons.

Gibbs, P., Milne, C., and Carillo, M. (1975). Correlation between the breeding system and recombination index in five species of *Senecio*. *New Phytol.*, 75:619–626.

Goldblatt, P. and Johnson, D. (2003). *Index to Plant Chromosome Numbers 1998-2000*, volume 94 of *Monogrpahs in Systematic Botany*.

Gornall, R. J. (1983). Recombination systems and plant domestication. *Biol. J. Linn. Soc.*, 20:375–383.

Hammer, K. (1984). Das domestikationssyndrom. *Kulturpflanze*, 32:11–34.

Hammer, K. (1988). Preadaptation and domestication of crops and weeds. *Biologisches Zentralblatt*, 107:631–636.

Hamrick, J. L. and Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 351:1291–1298.

Hamrick, J. L. and Godt, M. J. W. (1997). Allozyme diversity in cultivated crops. *Crop Sci.*, 37(1):26–30.

Harinarayana, G. and Murty, B. R. (1971). Cytological regulation of recombination in *Pennisetum* and *Brassica*. *Cytologia*, 36:435–448.

Harlan, J. R. (1992). *Crops & Man*. American Society of Agronomy : Crop Science Society of America, Madison, Wis., 2nd edition.

Harvey, P. H. and Pagel, M. D. P. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press.

Hastie, T. and Pregibon, D. (1992). *Statistical Models in S*, chapter Generalized linear models. Wadsworth & Brooks/Cole.

Hawkes, J. (1983). *The Diversity of Crop Plants*. Harvard University Press.

Hermisson, J. and Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4):2335–52.

Herrera, C. M. (2002). Correlated evolution of fruit and leaf size in bird-dispersed plants: species-level variance in fruit traits explained a bit further? *Oikos*, 97:426–432.

Hey, J. and Wakeley, J. (1997). A coalescent estimator of the population recombination rate. *Genetics*, 145:833–46.

Hilu, K. W. (1993). Polyploidy and the evolution of domesticated plants. *Am. J. Bot.*, 80:1494–1499.

Hintze, J. (2001). NCSS and PASS. Number Cruncher Statistical Systems, Kaysville, Utah.

Hsiao, C., Chatterton, N. J., Asay, K. H., and Jensen, K. B. (1995). Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. *Genome*, 38:211–223.

Hu, Lavin, Wojciechowski, and Sanderson (2000). Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast trnK/matK sequences and its implications for evolutionary patterns in Papilionoideae. *Am. J. Bot.*, 87(3):418–430.

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

Hudson, W. (1983). *The relationship of wild and domesticated tomate, /textitPhysalis philadelphica Lamarck (Solanaceae)*. PhD thesis, University of Indiana.

Iltis, H. and Benz, B. F. (2000). *Zea nicaraguensis* (Poaceae), a new teosinte from pacific coastal nicaragua. *Novon*, 10:382–390.

Jakob, S. S., Meister, A., and Blattner, F. R. (2004). The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. *Mol. Biol. Evol.*, 21:860–9.

Jenkins, J. (1948). The origin of the cultivated tomato. *Econ. Bot.*, 2:379–398.

Kajita, T., Ohashi, H., Tateishi, Y., Bailey, C. D., and Doyle, J. J. (2001). rbcL and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies. *Syst. Bot.*, 26:515–536.

King, J., Roberts, L. A., Kearsey, M. J., Thomas, H. M., Jones, R. N., Huang, L., Armstead, I. P., Morgan, W. G., and King, I. P. (2002). A demonstration of a 1:1 correspondence between chiasma frequency and recombination using a /textitLolium perenne/Festuca pratensis substitution. *Genetics*, 161:307–314.

Knox, M. R. and Ellis, T. H. N. (2002). Excess heterozygosity contributes to genetic map expansion in pea recombinant inbred populations. *Genetics*, 162:861–873.

Koella, J. (1993). Ecological correlates of chiasma frequency and recombination index of plants. *Biol. J. Linn. Soc.*, 48:227–238.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*, 336:435–440.

Koul, P., A. K. and Hamal, I. (1989). Reproductive biology of wild and cultivated carrot *Daucus carota* l. *New Phytol.*, 112:437–444.

Lenormand, T. and Otto, S. P. (2000). The evolution of recombination in a heterogeneous environment. *Genetics*, 156:423–428.

Levin, D. and Funderburg, S. (1979). Genome size in angiosperms: temperate versus tropical species. *Am. Nat.*, 114:784–795.

Lopez-Fanjul, C., Fernandez, A., and Toro, M. A. (2003). The effect of neutral nonadditive gene action on the quantitative index of population divergence. *Genetics*, 164:1627–1633.

Lynch, M. and Walsh, J. (1998). *Genetics and Analysis of Quantitative Traits*. Sinaur, Associates, Inc.

Martins, E. (2004). Compare. computer programs for the statistical analysis of comparative data.

Martins, E. P. and Hansen, T. (1997). Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into analysis of interspecific data. *Am. Nat.*, 149:646–667.

Mason-Gamer, R. J., Orme, N. L., and Anderson, C. M. (2002). Phylogenetic analysis of North American *Elymus* and the monogenomic Triticeae (Poaceae) using three chloroplast DNA data sets. *Genome*, 45:991–1002.

Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez G., J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *P Natl Acad Sci USA*, 99:6080–6084.

McDonald, J. H. and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in /textitDrosophila. *Nature*, 351:652–654.

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

Moeller, D. A. and Tiffin, P. (2005). Genetic diversity and the evolutionary history of plant immunity genes in two species of /textitZea. *Mol. Biol. Evol.*, 22:2480–2490.

Narayan, R. and McIntyre, F. K. (1989). Chromosomal DNA variation, genomic constraints and recombination in *Lathyrus*. *Genetica*, 79:45–52.

Nilsson, N. O., Sall, T., and Bengtsson, B. O. (1993). Chiasma and recombination data in plants: are they compatible? *Trends Genet.*, 9:344–348.

O'Hara, R. B. and Merila, J. (2005). Bias and precision in QST estimates: problems and some solutions. *Genetics*, 171:1331–1339.

Olmstead, R., Sweere, J., Spangler, R., and JD, J. P. (1999). Phylogeny and provisional classification of the Solanaceae based on chloroplast dna. In Nee, M., Lester, R., and Hawkes, J., editors, *Solanaceae IV*, pages 111–137. Royal Botanic Gardens, Kew.

Otto, S. P. and Barton, N. H. (1997). The evolution of recombination: removing the limits to natural selection. *Genetics*, 147:879–906.

Otto, S. P. and Barton, N. H. (2001). Selection for recombination in small populations. *Evolution*, 55:1921–1931.

Paterson, A. H. (2002). What has QTL mapping taught us about plant domestication? *New Phytol.*, 154:591–608.

Peakall, R. and Smouse, P. (2006). GENALEX 6: genetic analysis in Excel. population genetic software for teaching and research. *Molecular Ecology Notes*, 6:288–295.

Piperno, D. R. and Flannery, K. V. (2001). The earliest archaeological maize (Zea mays L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proc. Natl. Acad. Sci. U. S. A.*, 98(4):2101–2103.

Randall, R., Richardson, R., and Richardson, F. (2002). *A global compendium of weeds*. Missourri Botanical Garden Press.

Rees, H. and Dale, P. J. (1974). Chiasmata and variability in *Lolium* and *Festuca* populations. *Chromosoma*, 47:335–351.

Rees, H. and Durrant, A. (1986). Recombination and genome size. *Theor. Appl. Genet.*, 73:72–76.

Rindos, D. (1984). *The origins of agriculture: an evolutionary perspective.* Academic Press.

Ross-Ibarra, J. (2004). The evolution of recombination under domestication: a test of two hypotheses. *Am. Nat.*, 163:105–112.

Ross-Ibarra, J. (2005). Quantitative trait loci and the study of plant domestication. *Genetica*, 123:197–204.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19:2496–2497.

Roze, D. and Lenormand, T. (2005). Self-fertilization and the evolution of recombination. *Genetics*, 170:841–857.

Salamin, N., Hodkinson, T. R., and Savolainen, V. (2002). Building supertrees: an empirical assessment using the grass family (Poaceae). *Syst. Biol.*, 51:136–150.

Sherman, J. D. and Stack, S. M. (1995). Two-dimensional spreads of synaptonemal complexes from Solanaceous plants. VI. High-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). *Genetics*, 141:683–708.

Smartt, J. and Simmonds, N., editors (1995). *Evolution of Crop Plants.* Longman Scientific and Technical, 2nd edition.

Soltis, D., Haufler, C., Darrow, D., and Gastony, G. (1983). Starch gel electrophoresis of ferns: a compilation of grinding buffers, gel and electrode buffers, and staining schedules. *American Fern Journal*, 73:9–27.

Soltis, D. E., Soltis, P. S., Bennett, M. D., and Leitch, I. J. (2003). Evolution of genome size in the angiosperms. *Am. J. Bot.*, 90:1596–1603.

Spitze, K. (1993). Population structure in daphnia obtusa: quantitative genetic and allozymic variation. *Genetics*, 135:367–374.

Steele, K. and Wojciechowski, M. (2003). Phylogenetic analyses of tribes Trifolieae and Vicieae based on sequences of the plastid gene matK (Papilionoideae: Leguminosae). In Klitgaard, B. and Bruneau, A., editors, *Advances in Legume Systematics, Part 10*, pages 355–370. Royal Botanic Gardens, Kew.

Stevens, P. (2005). Angiosperm phylogeny website. http://www.mobot.org/MOBOT/research/APweb/.

Symonds, M. R. E. (2002). The effects of topological inaccuracy in evolutionary trees on the phylogenetic comparative method of independent contrasts. *Syst. Biol.*, 51:541–553.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105:437–460.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595.

Tenaillon, M. I., Sawkins, M. C., Anderson, L. K., Stack, S. M., Doebley, J., and Gaut, B. S. (2002). Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays ssp. mays* L.). *Genetics*, 162:1401–1413.

Tenaillon, M. I., U'Ren, J., Tenaillon, O., and Gaut, B. S. (2004). Selection versus demography: A multilocus investigation of the domestication process in maize. *Mol. Biol. Evol.*, 21:1214–1225.

Thompson, J., Higgins, D., and Gibson, T. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22:4673–4680.

Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S*. Springer.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J. F. (2005). The origin of the naked grains of maize. *Nature*, 436:714–719.

Wang, R.-L., Stec, A., Hey, J., Lukens, L., and Doebley, J. (1999). The limits of selection during maize domestication. *Nature*, 398(6724):236–239.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.*, 7(2):256–276.

Wendel, J. F. and Parks, C. R. (1982). Genetic control of isozyme variation in /textifCamillia japonica l. *Heredity*, 73:197–204.

Wenzel, W. and Hemleben, V. (1982). A comparative study of genomes in angiosperms. *Plant Sys. Evol.*, 139:209–227.

Whitt, S. R., Wilson, L. M., Tenaillon, M. I., Gaut, B. S., and Buckler, Edward S., I. (2002). Genetic diversity and selection in the maize starch pathway. *PNAS*, 99(20):12959–12962.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. Sixth Intern. Congr. Genet.*, 1:356–366.

Wright, S. (1951). The genetical structure of populations. *Annual Eugenics*, 15:323–354.

Wright, S., Bi, I., Schroeder, S., M., Y., Doebley, J., McMullen, M., and Gaut, B. (2005). The effects of artificial selection on the maize genome. *Science*, 308:1310–1314.

Wright, S. I., Lauga, B., and Charlesworth, D. (2003). Subdivision and haplotype structure in natural populations of Arabidopsis lyrata. *Mol. Ecol.*, 12:1247–1263.