

Selection Versus Demography: A Multilocus Investigation of the Domestication Process in Maize

Maud I. Tenaillon,*† Jana U'Ren,† Olivier Tenaillon,‡ and Brandon S. Gaut†

*Station de Génétique Végétale UMR C8120, Gif sur Yvette, France; †Department of Ecology and Evolutionary Biology, University of California Irvine; ‡INSERM EMI0339, Faculté Xavier Bichat, Paris, France

The domestication of maize (*Zea mays* ssp. *mays*) from its wild ancestor (*Zea mays* ssp. *parviglumis*) led to a loss of genetic diversity both through a population bottleneck and through directional selection at agronomically important genes. In order to discriminate between those effects and to investigate the nature of the domestication bottleneck, we analyzed nucleotide diversity data from 12 chromosome 1 loci in *parviglumis*. We found an average loss of nucleotide diversity of 38% across genes, but this average was skewed downward by four putatively selected loci (*tb1*, *d8*, *ts2*, and *zagl1*). To better understand the domestication process, we used the coalescent with recombination to simulate bottlenecks under various lengths and population sizes. For each locus, we determine the likelihood of the observed data using three summary statistics: the number of segregating sites, an estimate of the population recombination parameter, and Tajima's *D*. Based on the eight neutrally evolving loci, a model with a bottleneck had a significantly higher likelihood than a model without one. The four putatively selected loci had significantly different likelihood optimums than the neutral loci, and this approach confirmed that *ts2* and *d8* were selected either during domestication or breeding. Overall, the best-fitting models had a bottleneck in which the population size and the bottleneck duration had a ratio of ~4- to ~5; for example, if the initial domestication event occurred over a 500-year period, the population size was roughly 2,000 to 2,500 individuals. However, this range did vary with the summary statistic used to assess the fit of simulations to data. In this context, Tajima's *D* performed poorly as a goodness-of-fit statistic, probably because *Z. mays* ssp. *parviglumis* has a frequency spectrum that is significantly skewed toward low-frequency variants. Finally, we found that demography is unlikely to account for the previously observed positive correlation between nucleotide diversity and the population-recombination parameter in maize, leaving this observation difficult to interpret.

Introduction

The crop plant maize (*Zea mays* ssp. *mays*) was domesticated from the wild taxon *Zea mays* ssp. *parviglumis* (hereafter "parviglumis"), through a single domestication event in southern Mexico (Matsuoka et al. 2002) between 6,250 and 10,000 B.P. (Smith 1998; Piperno and Flannery 2001). Although maize is a remarkably diverse crop, previous work has shown that domestication in maize led to a loss of genetic diversity. The loss of nucleotide diversity in maize as compared with parviglumis ranges from over 80% for loci like *cl*, *ae1*, and *tb1* (Hanson et al. 1996; White and Doebley 1999) to less than 20% for loci like *hm1*, *hm2*, *glb1*, and *sh1* (White and Doebley 1999; Whitt et al. 2002; Zhang et al. 2002a).

The loss of genetic diversity appears to be the result of two processes. The first is directional selection at target genes involved either in the differentiation between the wild and the cultivated phenotype or in subsequent maize improvement. To date, there is evidence for selection on at least six genes in maize, including *tb1*, which modifies plant architecture (Wang et al. 1999); *cl*, a regulator of anthocyanin biosynthesis (Hanson et al. 1996); three genes (*bt2*, *ae1*, and *sul*) that comprise part of the starch pathway (Whitt et al. 2002); and *zagl1*, a putative transcription factor (Vigouroux et al. 2002b). In a previous study, Tenaillon et al. (2001) also provided some evidence for selection at *d8* (see also Thornsberry et al. 2001) and *ts2*, both of which are involved in maize sex determination (Harberd and Freeling 1989; Irish and Nelson 1993).

The second, and more general, process leading to a loss of diversity during domestication is a genetic bottleneck, probably resulting from sampling processes associated with domestication and breeding. This bottleneck apparently affects all genes in maize (Whitt et al. 2002; Zhang et al. 2002a), but the effect of this bottleneck on genetic diversity has been modeled carefully with data from only two loci (Hilton and Gaut 1998a; Eyre-Walker et al. 1998). The results of coalescent simulations were consistent with a founding population of a few hundred to a few thousand individuals, depending on the duration of the bottleneck. However, these simulations assumed homogeneous mutation rates across genes and also assumed that there was no recombination within genes; both of these variables undoubtedly affect simulation results. Moreover, one expects to garner greater insights into the domestication process by examining data collected at several loci simultaneously. By understanding the domestication bottleneck, it may eventually be possible to discriminate between loci that have been under directional selection and those that have been affected primarily by demography (e.g., Vigouroux et al. 2002b).

Genetic diversity can be shaped by recombination in addition to demography and selection. Several studies in plants (Dvorak, Luo, and Yang 1998; Kraft et al. 1998; Stephan and Langley 1998) and animals (Kaplan, Hudson, and Langley 1989; Begun and Aquadro 1992; Nachman 1997; Nachman et al. 1998) have established a positive correlation between genetic diversity and recombination. This correlation may be a consequence of genetic hitchhiking due to either background selection or selective sweeps (Maynard-Smith and Haigh 1974; Charlesworth 1994; Hudson and Kaplan 1995; Begun and Whitley 2000; Andolfatto and Przeworski 2001), but the relative importance of these two processes remains difficult to evaluate

Key words: nucleotide diversity, bottleneck, coalescence, recombination.

E-mail: tenaillon@moulon.inra.fr

Mol. Biol. Evol. 21(7):1214–1225. 2004

doi:10.1093/molbev/msh102

Advance Access publication March 10, 2004

Table 1
Parviglumis Sample Used in the Present Study

Parviglumis Accession	Origin
PI566686 ^a	Guererro, Mexico
PI566688 ^a	Guererro, Mexico
PI566691 ^a	Michoacan, Mexico
HHI Site 3 ^b	Guererro, Mexico
GWB Site 4 ^b	Guererro, Mexico
BFB 967 ^b	Jalisco, Mexico
PI331783 ^a	Guererro, Mexico
PI331785 ^a	Michoacan, Mexico
PI331786 ^a	Mexico, Mexico
PI331787 ^a	Mexico, Mexico
PI331788 ^a	Michoacan, Mexico
PI331789 ^a	Michoacan, Mexico
PI384061 ^a	Guererro, Mexico
PI384062 ^a	Guererro, Mexico
PI384064 ^a	Guererro, Mexico
PI384070 ^a	Guererro, Mexico

^a These accession numbers come from the USDA North Central Regional Plant Introduction Station (Ames, Iowa).

^b More information about these accessions is available at www.panzea.org (HHI = Hugh H. Iltis; GWB = George W. Beadle; BFB = Bruce F. Benz).

(e.g., Payseur and Nachman 2002). Moreover, in some cases the correlation may have a neutral explanation, i.e., that recombination is mutagenic (Hellmann et al. 2003).

The correlation between diversity and recombination is not clear cut in maize. In previous work, we investigated the evolutionary forces acting to shape maize diversity at 21 loci distributed along chromosome 1 and found a positive correlation between recombination, as estimated by the population genetic parameter $4Nc$, and nucleotide diversity (Tenaillon et al. 2001). However, this correlation was neither confirmed with another estimator of $4Nc$ nor with estimates of recombination based on physical measures of crossover frequency (Tenaillon et al. 2002). Hence, it is still unclear whether recombination and diversity are correlated in maize. One intriguing possibility is that the demographic history of the species has jointly affected $4Nc$ and nucleotide diversity (in part because they are both functions of N , the effective population size) but obscured relationships between diversity and crossover frequency. To further explore this last possibility, additional studies of the demographic history of maize and its joint effects on linkage and diversity are necessary.

In this study we analyse 12 loci on chromosome 1—including the putatively selected genes *tb1*, *d8*, *ts2*, and *zagll*—in a common set of parviglumis individuals. Our rationale for gathering and analysing parviglumis data is to gain insight into levels and patterns of genetic diversity prior to the domestication bottleneck in maize, and to use these data to further characterize the impact of domestication on nucleotide diversity. We perform coalescent simulations to explore domestication scenarios that fit the observed, multilocus sequence data in maize and parviglumis. This multilocus approach appears to be a powerful means both for identifying loci under selection, for which the loss of diversity was not due solely to demography, and for proposing estimates of bottleneck duration and founding population size of maize. The effect of demography on the correlation between recombination and genetic diversity is also discussed.

Material and Methods

Experimental Procedure

Plant Material and Sequence Data

We sequenced eight loci in a common set of 16 parviglumis individuals (*Zea mays* ssp. *parviglumis*) covering the geographical range of the species (table 1). The number of individuals sampled for each gene is reported in table 2. Sequences were submitted to GenBank (AY513897-AY513899; AY513901-AY513905; AY513907; AY513928-AY513937; AY514500-AY514501; AY514512-AY514516; AY514529-AY514534; AY514552-AY514563; AY514584-AY514591; AY514602-AY514614; AY514890-AY514899). For four additional loci (*hml*, *adh1*, *gbl1*, and *zagll*), parviglumis sequences were obtained from previous studies that used roughly the same sample of parviglumis individuals (Eyre-Walker et al. 1998; Hilton and Gaut 1998b; Zhang et al. 2002a; Vigouroux et al. 2002b).

DNA sequence data for maize (*Zea mays* ssp. *mays*) for 10 of the 12 loci were obtained from Tenaillon et al. (2001). These loci were sequenced from a collection of 25 maize individuals including 9 U.S. inbred lines and 16 landraces. Singletons in the *tb1* data were corrected against the data of Clark et al. (2003). For the remaining loci (*hml* and *zagll*), the data were taken from Zhang et al. (2002a) and Vigouroux et al. (2002b), respectively. Sampling in *zagll* included the 16 maize landraces. Sampling in *hml* was based on 10 individuals with a geographic range similar to the 16 landraces. For 8 of the 12 loci an outgroup sequence from *Tripsacum dactyloides* was available (Tenaillon et al. 2001).

We chose to study these 12 loci for three reasons. First, they are located on chromosome 1 on the UMC98 genetic map (Davis et al. 1999), for which physical estimates of recombination rates are either available or calculable (Tenaillon et al. 2002). Second, they have been sampled from a similar set of maize and parviglumis individuals. Finally, they represent a range of function. Eight of the twelve loci are characterized genes, two are cDNA clones for which ORFs were previously described (Tenaillon et al. 2001), and two are anonymous genomic RFLP clones. Their length in maize ranged from 375 to 2740 bp, with an average of 946 bp (Table 2).

Sequencing Procedure in Parviglumis

Sequencing in parviglumis samples used PCR primers and PCR conditions described previously (Tenaillon et al. 2001). PCR products were cloned into a TA cloning vector (pGem), and one clone was sequenced in both directions using BigDye chemistries and ABI automated sequencers. Sequence data were aligned in BioEdit, version 5.0.9 (Hall 1999) and Seqman (DNASTar, Madison, Wisc.). In the initial alignments of these eight loci, we identified singleton polymorphisms. Since singletons might result either from sequence variation or from *taq* polymerase error, we reamplified and recloned individuals for which a singleton was detected. Five clones per individual per locus were sequenced in order to confirm retrieval of the allele that was first cloned and also to confirm the lack of PCR recombinants. Roughly half (52%) of the singletons were confirmed as real variants by this method; the unconfirmed

Table 2
Sequence Statistics for the Loci Studied in Parviglumis and Maize

Locus	Position ^a	Parviglumis					Maize ^h					p-values ^h
		<i>n</i> ^b	Length ^c	<i>S</i> ^d	π_{silent} ^c	<i>D</i> ^f	<i>n</i> ^b	Length ^c	<i>S</i> ^d	π_{silent} ^c	<i>D</i> ^f	
zag11	12.8	15	572 (437.56)	19	0.0090	-0.662	16	560 (443.33)	0***	0.0000	NA	1.000
csu1171 ^g	30.9	9	467 (467)	18	0.0117	-0.869	25	468 (468)	11***	0.0067	0.269	1.000
ts2 ^g	65.2	10	975 (353.5)	17	0.0160	0.388	24	946 (331.99)	8***	0.0055	0.178	1.000
asg11	116.9	6	492 (492)	10	0.0073	-1.069	24	375 (375)	12 ^{ns}	0.0089	0.120	0.528
hm1	125.1	10	1016 (496.17)	96	0.0414	-0.517	10	1026 (505)	73	0.0338	0.101	
csu1132 ^g	140.2	8	531 (418)	29	0.0251	-0.646	25	544 (430)	24***	0.0118	-0.750	1.000
bz2 ^g	178.0	12	590 (209.83)	15 ^{ns}	0.0148	-0.888	25	605 (217.22)	17 ^{ns}	0.0130	0.951	0.599
glb1 ^g	189.4	8	991 (489.06)	75	0.0358	-0.903	23	968 (468.99)	90**	0.0297	-0.843	0.996
tb1	197.6	7	2702 (2044.41)	88	0.0153	-0.650	23	2740 (2082.67)	24***	0.0011	-2.140*	1.000
d8 ^g	198.5	13	791 (206.53)	46	0.0359	-1.816*	25	800 (209.17)	8***	0.0091	-0.318	1.000
adh1 ^g	205.2	8	1257 (825.33)	57	0.0252	-0.145	25	1406 (931.83)	62***	0.0188	0.263	1.000
csu381 ^g	231.4	10	917 (868.67)	42	0.0151	-0.561	25	916 (867.67)	42***	0.0182	1.617	1.000

^a Map location along the chromosome 1 inch from the Pioneer 1999 composite map.^b Number of sequences sampled.^c Number of sites aligned sequences, excluding gaps, with number of silent sites in parentheses.^d Total number of segregating sites (*S*). Significance of a loss of diversity as measured by *S* either in maize or in parviglumis is indicated, respectively, in the maize or parviglumis columns (*, $p < 0.5$; **, $p < 0.1$; ***, $p < 0.01$; ns: nonsignificant). For *hm1*, however the resampling procedure was not applied since the number of sequences in maize and in parviglumis was identical.^e Average number of pairwise differences calculated on silent sites.^f Tajima's *D* statistics based on all sites.^g Loci for which a *T. dactyloides* outgroup was available.^h *p* values correspond to the number of simulations among 1,000 simulations for which $\theta_{\text{maize}} < \theta_{\text{parviglumis}}$.

singletons were considered polymerase artefact and discarded from analyses. This rate of singleton artefact was similar to previous reports (Eyre-Walker et al. 1998).

Data Analysis

In parviglumis data, we measured the number of polymorphic sites (*S*), the average number of pairwise differences per nucleotide site (π) (Tajima 1983), and Watterson's estimator θ per base pair (Watterson 1975) for all loci at silent and all sites using DnaSP version 3.51 (Rozas and Rozas 1999). In maize, these values were available from previous publications (Tenaillon et al. 2001; Zhang et al. 2002a). The loss of diversity in maize relative to parviglumis was measured as $1.0 - \pi_{\text{maize}}/\pi_{\text{parv}}$. For each locus, we assessed whether diversity loss in maize relative to parviglumis could be ascribed to sampling with a resampling test. We resampled, without replacement, a number of maize sequences equal to the number of parviglumis sequences and determined how often $\theta_{\text{maize}} > \theta_{\text{parviglumis}}$ (or alternatively, $S_{\text{maize}} > S_{\text{parviglumis}}$) over 1,000 simulations.

Tests of neutrality—including Tajima's *D* (Tajima 1989), the McDonald-Kreitman (MK) test (McDonald and Kreitman 1991), the Hudson, Kreitman, and Aguadé (HKA) test (Hudson, Kreitman, and Aguadé 1987), Fu and Li's *D* test with and without outgroup (Fu and Li 1993), and a haplotype test (Ewens 1972)—were performed with DnaSP, version 3.51 (Rozas and Rozas 1999). The *Tripsacum dactyloides* sequence, when available, was used as an outgroup for the MK, HKA, and Fu and Li tests.

For parviglumis data, we estimated per site values of the population recombination parameter $4Nc$ using both Hudson's 1987 method (Hudson 1987), as implemented in DnaSP, and Hudson's 2001 method (Hudson 2001), as implemented with importance sampling in the LDhat

program (Fearnhead and Donnelly 2001). We denoted the estimates as $4\hat{N}c_{\text{hud87}}$ and $4\hat{N}c_{\text{hud01}}$, respectively. For maize data, these parameters and another recombination estimate, *R*, which is based on physical observations of crossing-over frequency, have been reported (Tenaillon et al. 2001; Tenaillon et al. 2002). Correlations between sequence diversity, recombination and other parameters were measured with the Pearson correlation coefficient (*r*); significance was determined by 10,000 bootstrap resamplings of observed values.

Coalescent Simulations

Model

Coalescent simulations were used to model the impact of a bottleneck on sequence diversity and recombination. For this purpose, we modified the standard method described in Hudson (1987) by adding a bottleneck and simulating the evolution of sequences along the branches of coalescent trees. We used a bottleneck model previously described in Eyre-Walker et al. (1998), in which a single ancestral population of size *Na* experienced an instantaneous shift in size at time *t2* generations ago to the bottlenecked population size (*Nb*). Going forward in time, the bottleneck population expanded instantaneously at time *t1* generations ago to the present population size (*Np*). The bottleneck was characterized by two parameters: *d*, which is the duration of the bottleneck in generations (measured as *t2-t1*), and *Nb*, the population size during the bottleneck. Our simulations included intragenic recombination.

Simulation Parameter

The purpose of the coalescent model is to simulate bottleneck events and compare bottleneck scenarios. To achieve this purpose, we used parviglumis data as the basis for simulations and maize data for goodness-of-fit statistics

(see Eyre-Walker et al. 1998). For each locus, the bottleneck model has 10 parameters, not all of which are independent but are included here for completeness. Six parameters— N_a , N_b , N_p , t_1 , t_2 , and d —were described earlier. The remaining parameters necessary to perform simulations for each locus are sample size n , sequence length L , the mutation rate μ , and the recombination rate ($4Nc$). The following assumptions were used for simulation parameters:

1. n and L for each locus were taken from maize data (table 2).
2. t_2 was assumed to be constant across loci and set to 7,500 generations (Iltis 1983; White and Doebley 1999).
3. N_p is a nuisance parameter that has been shown to have little effect on the outcome of simulations with this model Eyre-Walker et al. (1998), reflecting the short time (on an evolutionary scale) after domestication. Furthermore, different attempts to estimate N in maize either based on nucleotide (Gaut and Clegg 1993; Eyre-Walker et al. 1998; Remington et al. 2001) or microsatellite polymorphism Vigouroux et al. (2002a) result in estimates differing by an order of magnitude. We therefore set N_p to 1×10^6 individuals, which is similar to previous estimates (Gaut and Clegg 1993; Eyre-Walker et al. 1998), for all simulations and all loci but investigated two other values ($N_p = 500,000$ and $N_p = 5,000,000$) to ensure that N_p had little influence on the results.
4. The two bottleneck parameters d ($= t_2 - t_1$) and N_b were varied over a series of values. Based on the values reported in Eyre-Walker et al. (1998) and Vigouroux et al. (2002b), we explored a range of values, with $d = 100, 300, 500, 1,000, 1,500, 2,000$ and $2,800$; $2,800$ generations (or years) is the maximum duration of a bottleneck associated with domestication, based on fossil evidence (Long et al. 1980; Eyre-Walker et al. 1998). The range of N_b values varied with d because d and N_b are positively correlated. For a given d value, we varied the ratio of N_b/d from ~ 0.5 to ≥ 6.5 . Altogether, we investigated 200 combinations (or scenarios) of N_b and d for each locus. We performed 10,000 simulations for each combination.
5. The mutation rate μ was determined for each locus i in the following manner. For the eight loci with an outgroup sequence from *T. dactyloides*, the mutation rate μ at each locus was estimated from synonymous divergence (K_s), calculated as the average synonymous distance (as measured in DNAsp) among all pairwise comparisons between each parviglumis sequence and the outgroup sequence. Given K_s for locus i , we calculated $\mu_i = K_{s_i} * K_{s_{adh1}} / \mu_{adh1}$, where μ_{adh1} is the substitution rate per synonymous site per year at grass *adh* loci, estimated to 6.5×10^{-9} substitutions per site per year (Gaut et al. 1996). For the four loci for which no outgroup sequence was available, μ was calculated as the average K_s among the other eight loci.
6. Given μ_i , N_{a_i} was calculated for each locus from the relationship $\theta_{parv} = 4 * N_{a_i} * \mu_i$

For each locus, $4\hat{N}c_{hud87}$ and $4\hat{N}c_{hud01}$ were estimated from parviglumis data (see above) and used as the population-recombination parameter in two independent sets of simulations. For comparison, we also performed a third set of simulations without recombination.

Output and Goodness-of-Fit

For each simulation at each locus, we calculated values of θ (θ_{simul}), S (S_{simul}), $4Nc_{hud87}$ ($= 4Nc_{simul}$), and Tajima's D (D_{simul}). We used $4Nc_{hud87}$ as a summary statistic because it captures aspects of both the frequency spectrum and linkage disequilibrium (LD) (Hudson 1987). Although it is likely not as good an estimator as $4Nc_{hud01}$ (Hudson 2001), it is computationally convenient. For each set of simulations we also calculated the Pearson correlation coefficient (r_{simul}) between θ_{simul} and $4Nc_{simul}$.

To find the best-fitting parameter values of d and N_b relative to observed maize data, we assessed the fit of simulated data to observed data. To assess fit, we defined levels of acceptance corresponding to a range of $\pm 20\%$ for each of three summary statistics (S_{maize} , D_{maize} , and $4Nc_{hud87\ maize}$), as well as for three joint summary statistics ($S-D$, $S-4Nc$, and $S-D-4Nc$). For each of the 200 explored bottleneck scenarios, we calculated the approximate likelihood for locus i as the proportion of simulations among 10,000 that fit the data (Weiss and von Haeseler 1998). The approximate likelihood for each of the 200 scenarios was calculated for each locus and for each of the six summary statistics.

After calculating likelihoods on a per-locus basis, we calculated the multilocus likelihood by multiplying across loci. This approach made the implicit assumption that loci are independent, which is consistent with the absence of LD among loci (Tenaillon et al. 2001).

Results

Levels of Diversity and Tests of Neutral Equilibrium

A total of 11,301 bp was aligned in parviglumis, representing the 12 sequenced loci. We identified a total of 512 SNPs and estimated π at silent sites (table 2) and θ (table 3) for each locus. For both π and θ at silent sites, minimum diversity in parviglumis was found in the gene *asg11* ($\pi = 0.0073$ and $\theta = 0.0089$), which has a level of diversity comparable to that of *te1*, a gene that may evolve slowly (White and Doebley 1999). Maximum diversity values in parviglumis were found in *hml* ($\pi = 0.041$; $\theta = 0.034$), which has been previously noted to be a high diversity gene (Zhang et al. 2002a). Figure 1 integrates these data into a chromosomal context and clearly indicates that diversity is generally higher in parviglumis than maize. Over all 12 loci, the loss of diversity in maize averaged $\sim 38\%$ (table 2). As expected, the loss of diversity is much more pronounced for putatively selected genes; for the four loci (*tb1*, *ts2*, *d8*, and *zag11*) the reduction in diversity ranges from 66% to 100%. The average loss in diversity for the eight remaining genes was $\sim 20\%$. We tested for a significant reduction in diversity with a resampling procedure (see *Materials and Methods*). Nine of eleven loci examined exhibited a significant loss of diversity in maize compared with parviglumis (table 2). The loss was not statistically significant for *asg11* and *bz2*.

Table 3
Per-Site Estimates of Nucleotide Diversity (θ) and the Population-Recombination Parameter $4Nc$ in Parviglumis and Maize for 11 Loci Located on Chromosome 1. We Estimate μ at Each Locus Based on Ks Values and Assuming a Mutation Rate $\mu = 6.5 \times 10^{-9}$ at the *adh1* Locus (cf. *Material and Methods*)

Locus	Parviglumis					Maize		
	θ	$4Nc_{hud87}$	$4Nc_{hud01}$	Ks	μ	θ	$4Nc_{hud87}$	R^a
<i>zag11</i>	0.01022	0.3300	0.0318	NA ^b	5.53×10^{-9} ^c	0.00000	NA ^c	0.1061
<i>csu1171</i>	0.01418	NA ^b	0.0061	0.03087	2.56×10^{-9}	0.00622	0.0207	0.1151
<i>ts2</i>	0.00616	0.1575	0.0666	0.06423	5.33×10^{-9}	0.00226	0.1337	0.0784
<i>asg11</i>	0.00890	0.0109	0.0000	NA ^b	5.53×10^{-9} ^c	0.00857	0.0114	0.0228
<i>hm1</i>	0.03410	0.0653	0.0155	NA ^b	5.53×10^{-9} ^c	0.02515	0.0091	0.0473
<i>csu1132</i>	0.02252	0.0237	0.0067	0.07112	5.90×10^{-9}	0.01171	0.0001	0.0481
<i>bz2</i>	0.00842	0.1688	0.0051	0.06827	5.66×10^{-9}	0.00744	0.0200	0.0305
<i>glb1</i>	0.02958	0.2048	0.0182	0.08739	7.25×10^{-9}	0.02519	0.0897	0.0615
<i>tb1</i>	0.01375	0.0053	0.0047	NA ^b	5.53×10^{-9} ^c	0.00237	0.0041	0.0965
<i>d8</i>	0.01874	0.0000	0.0000	0.09722	8.06×10^{-9}	0.00265	0.0210	0.0965
<i>adh1</i>	0.01749	0.0881	0.0108	0.07835	6.5×10^{-9}	0.01168	0.0017	0.1151
<i>csu381</i>	0.01619	0.0320	0.0110	0.03582	2.97×10^{-9}	0.01214	0.0009	0.1271

^a R in maize was estimated following the method and data reported in Tenaillon et al. (2002).

^b NA, not available.

^c Because no outgroup sequence was available for those loci, Ks was estimated from an average value of divergence (cf. *Material and Methods*).

Tajima's D was calculated on maize and parviglumis data, except for *zag11* in maize, which contains no polymorphisms (table 2). One expects D to be higher in populations that have experienced a recent bottleneck because of the loss of low-frequency variants. This appears to be true for maize where D is higher than parviglumis for 8 of 11 loci. Of the remaining three loci (*ts2*, *csu1132*, and *tb1*), two exhibit evidence of directional selection (Wang et al. 1999; Tenaillon et al. 2001), which is expected to decrease D (Tajima 1989). Another interesting feature of D for parviglumis data was that 11 of 12 loci had negative D values (Table 2). We compared the mean value of D among loci ($\bar{D} = -0.695$ for all 12 loci) to a simulated distribution (using the program of J. Hey, available at <http://lifesci.rutgers.edu/~hey/lab/DistributedProgramsAndData.htm#HKA>) and found a significant departure from neutral expectation over all loci ($p = 0.01$). This result confirms that the parviglumis single nucleotide polymorphism (SNP) frequency spectrum is skewed significantly from neutral expectations, perhaps as a consequence of population expansion or subdivision. In contrast, mean D for the maize data was -0.050 over 11 loci (table 2).

Like D , one expects LD—which is inversely proportional to $4Nc$ —to increase after a population bottleneck. For these loci, $4\hat{N}c_{hud87}$ is generally higher in parviglumis than maize. Although $4\hat{N}c_{hud87}$ is correlated between parviglumis and maize ($r = 0.82$, $P = 0.002$), 8 of 10 loci had higher $4\hat{N}c_{hud87}$ values in parviglumis (table 3). However, one of the two loci for which $4\hat{N}c_{hud87}$ is higher in maize is a putatively selected loci (*d8*), which is somewhat surprising given that selection could increase LD (Przeworski 2002) even above and beyond that generated by a bottleneck. The trend of increased LD in maize is less obvious when $4Nc$ estimates are based on Hudson's 2001 method. The $4\hat{N}c_{hud01}$ estimates are available for both taxa for nine of the 12 loci. These estimates are not significantly correlated across taxa ($r = 0.52$, $P = 0.29$), but five of seven neutral loci have higher $4\hat{N}c_{hud01}$ values in parviglumis, again suggesting a trend toward decreasing $4Nc$ (increasing LD) in maize.

Several neutrality tests were applied to the parviglumis data. For individual loci, Tajima's D statistic was significant only for the *d8* locus (table 2). Three other tests—MK, HKA, and Fu and Li's D test with outgroup—required an outgroup and were applied to eight loci. No significant departure from neutral expectation was detected for any test, except, again, *d8* (Fu and Li's D test; $p < 0.05$). Haplotype tests can be useful for detecting incomplete hitchhiking associated with selective sweeps (Depaulis, Mousset, and Veuille in press), although they may be too liberal when there has been recombination. We found a significant excess of haplotypes for *ts2* ($p = 0.008$), *zag11* ($p = 0.006$), and *d8* ($p = 0.002$), for which 13 haplotypes were detected among 13 individuals. One explanation for an excess of haplotypes is the accrual of singletons after a severe selective sweep. Under this scenario, one also expects an excess of rare variants, as measured by D . This simple sweep scenario is unlikely for *ts2* in parviglumis, because it was the only locus with $D > 0$ (table 2) and to a lesser extent for *zag11*, because its D value is not aberrantly low among parviglumis loci. However, the sweep scenario seems more likely for *d8*, in which 39 out of 46 polymorphic sites were singletons (table 2). In conclusion, *d8* may have a history of adaptive selection in parviglumis, but this conclusion must be tempered by the fact that species-wide samples from parviglumis are skewed toward low-frequency variants.

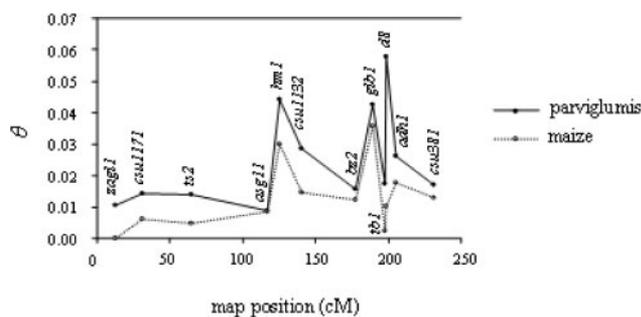


FIG. 1.—Estimates of θ at silent sites in maize and parviglumis for 12 loci located in cM along the genetic map of maize chromosome 1.

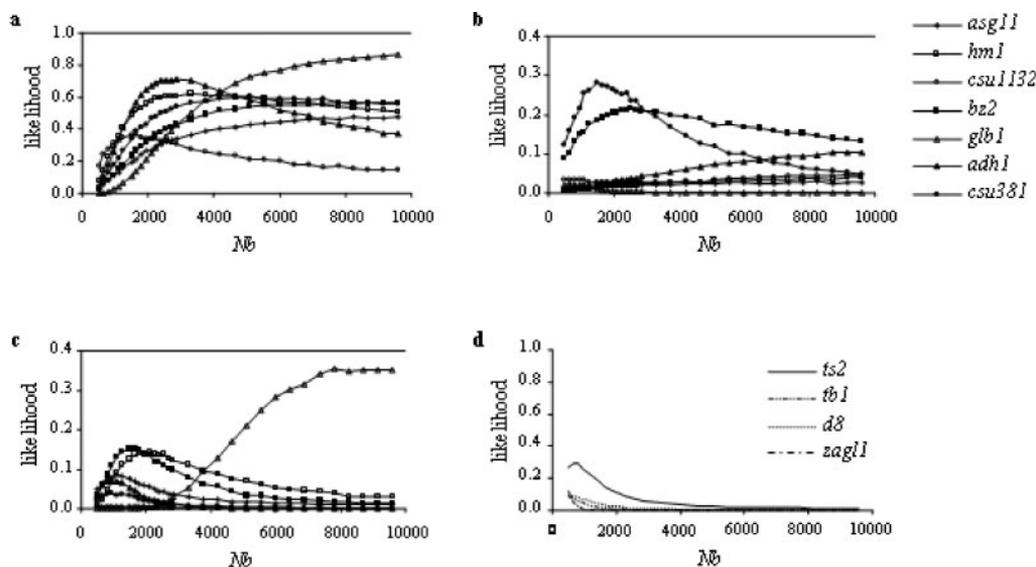


FIG. 2.—Per-locus likelihood values as a function of Nb for a bottleneck duration $d = 1,000$ generations. All simulations were based on recombination rates equivalent to $4Nc_{hud87}$. (a) likelihoods for seven neutral loci based on a fitting criterion of $\pm 20\%$ of S_{maize} . (b) Likelihoods for seven neutral loci based on a fitting criterion $\pm 20\%$ of $4Nc_{hud87maize}$. (c) Likelihoods for seven neutral loci based on a fitting criterion of $\pm 20\%$ of D_{maize} . (d) Likelihoods for four putatively selected loci, based on a fitting criterion of $\pm 20\%$ S_{maize} .

Exploring Bottleneck Scenarios

Previous work, patterns of diversity in parviglumis, and/or tests for selection continue to suggest that *tb1*, *d8*, *ts2*, and *zag1* may have been under selection. For this reason, the four genes were discarded from the multilocus likelihood for fitting the demographic model (see below). The simulation input parameters for each locus are summarized in table 3. Note that higher estimates of recombination were obtained with $4Nc_{hud87}$ compared with $4Nc_{hud01}$ for all loci.

Summary Statistics and Multilocus Likelihoods

Approximate likelihood curves for individual loci were determined for six summary statistics (S , D , $4\hat{N}c_{hud87}$, $S-D$, $S-4Nc$, and $S-D-4Nc$ combined). Figure 2a–c, generated with $d = 1,000$ and with a simulation recombination rate based on $4\hat{N}c_{hud87}$ in parviglumis, are illustrative of general trends regarding goodness-of-fit to summary statistics. When either S or $4\hat{N}c_{hud87}$ were used as summary statistics, most loci had likelihood curves with noticeable peaks; for the case illustrated in figure 2, the peak was with $Nb < 4,000$ (fig. 2a, c). In contrast, curves generated with D as a summary statistic often failed to reach a maximum within the examined parameter space (fig. 2b). Joint summary statistics that included D ($S-D$ and $S-D-4Nc$) also produced shallow curves (data not shown). Poor results with D as a summary statistic may have a simple explanation—that the skewed frequency distribution of SNPs in parviglumis is not adequately mimicked in these simulations (see *Discussion*).

We examined the effect of recombination on our inferences using S as the summary statistic (fig. 3). Multilocus likelihood (ML) curves for different values of d are reported in figure 3 for each of the three different recombination rates used in simulations ($4Nc_{hud87}$, $4Nc_{hud01}$, and no

recombination). In general, ML likelihood estimates were comparable when $4Nc_{hud87}$ and $4Nc_{hud01}$ were used in simulations. For example, with $d = 2,800$, the ML estimates of Nb for $4Nc_{hud87}$ and $4Nc_{hud01}$ was 11,800 and 10,000, respectively. Estimates of Nb were inflated slightly when simulations did not include recombination (fig. 3).

An alternative depiction of the likelihood surface reveals three features of the simulation results (fig. 4). First, confidence intervals are not tight. Many combinations of d and Nb examined here are within 1.0 likelihood ratio (LR) units of the ML estimate. LRs within 2.0 units correspond roughly to a 95% confidence interval, so LR statistics provide little statistical support for any particular joint estimate of d and Nb . Nonetheless, reasonable values of d and Nb can be inferred. Second, as noted earlier, the ML estimates of d and Nb are similar with different recombination rates, but are inflated when simulations do not include recombination. The key feature of simulations is to include recombination. Third, there is an obvious (and expected) linear relationship between d and Nb . Hence, it may be more fitting to report estimated ratios of Nb over d (Nb/d). Our estimates of Nb/d varied little over the seven values of d examined. With simulations based on $4Nc_{hud87}$ the ratio of Nb/d varied from 3.83 to 5.00 for the seven values of d ; with simulations based on $4Nc_{hud01}$, the ratio varied from 3.57 to 4.65; without recombination, this ratio varied from 5.50 to 8.30. When $4Nc_{hud87}$ was used as the summary statistic, optimal Nb/d ratios were approximately halved, with the ratio ranging from 1.8 to 2.5 (data not shown).

For comparison, we also calculated the likelihood under a model in which there was no bottleneck ($Nb = Np$; $d = 0$), with an instantaneous change in population size from N_a to N_p at $t_2 = 7,500$ generations ago. The bottleneck model is much more likely than this nonbottleneck model (fig. 3). For example, the LR between the ML of the bottleneck model and the nonbottleneck is 8.01, 6.34, and

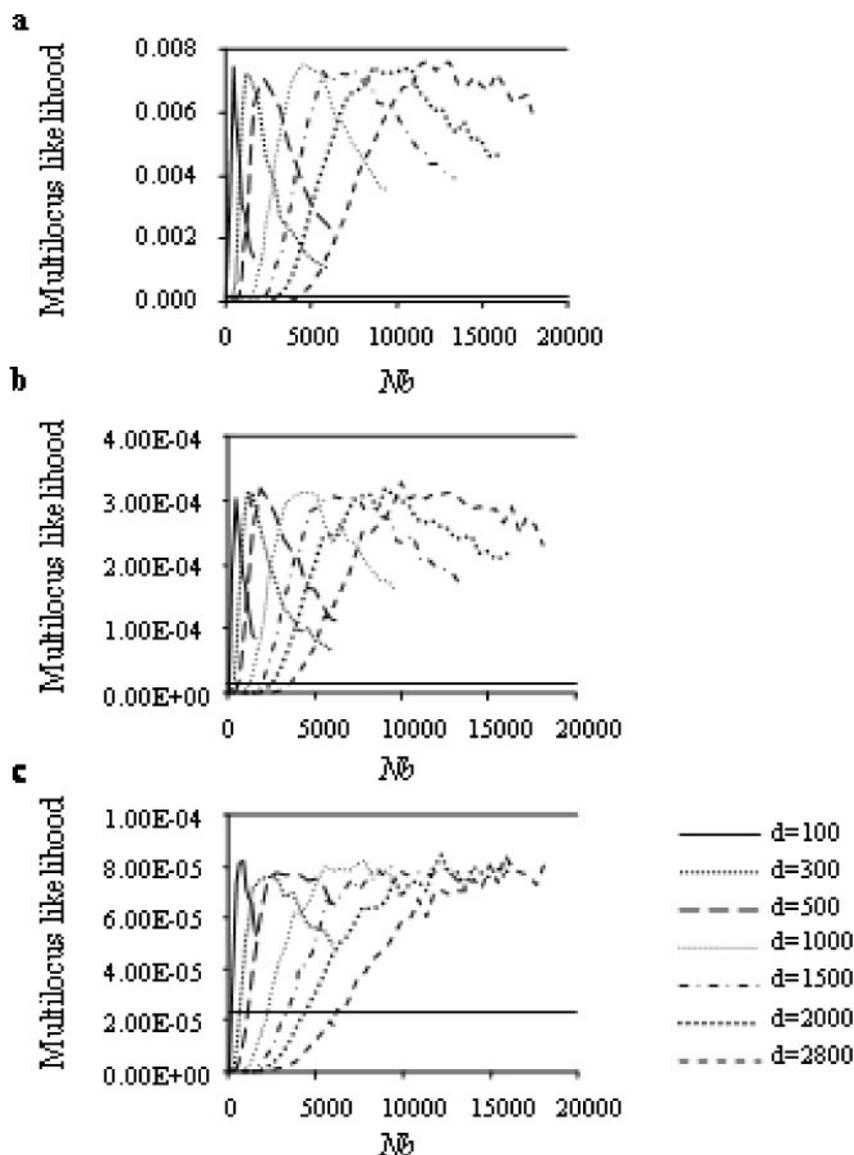


FIG. 3.—Multilocus likelihood of the observed data as calculated by the product of likelihood estimates among seven (a) and eight loci (b and c), respectively, as a function of population size during the bottleneck (N_b). The fitting criterion was $\pm 20\%$ of S_{maize} . Seven bottleneck durations (d) were explored as well as three recombination conditions (a) $4Nc_{hud87}$, (b) $4Nc_{hud01}$, and (c) no recombination. The horizontal lines indicate the multilocus likelihood value obtained without a bottleneck.

2.75 for the three recombination rates ($4Nc_{hud87}$, $4Nc_{hud01}$, and no recombination, respectively), indicating that the nonbottleneck model fits substantially more poorly than the bottleneck model for all three recombination rates. We also investigated the impact of N_p . As expected, varying N_p had very little influence on the results; similar ML estimates were obtained for $N_p = 5 \times 10^5$ and $N_p = 5 \times 10^6$. For example, with $d = 1,000$ and recombination estimates based on $4Nc_{hud01}$, the ML estimate of N_b was the same regardless of N_p , and the LR among the three N_p sizes differed by a maximum of 0.17.

Using the Demographic Model to Test for Selection

We studied four genes suspected to have been the target of selection. Physiological and molecular evidence

for selection on *tb1* and *zag11* are strong (Doebley, Stec, and Hubbard 1997; Wang et al. 1999; Vigouroux et al. 2002b), but evidence is more equivocal for *ts2* and *d8* (Tenaillon et al. 2001). The demographic model provides yet another opportunity to assess whether these loci differ substantially in their history relative to putatively neutral loci. Figure 2 provides likelihood curves for individual loci, with $d = 1,000$ as an example. The four loci under selection (fig. 2d) can easily be identified from the shape of their likelihood curves. We formalized the difference between the multilocus model, based on putatively neutral genes, and the four “selected” genes in two ways. First, we compared parameter estimates and likelihood values to the multilocus model. For example, with $d = 1,000$, N_b is estimated to be 4,650 individuals under the multilocus model (fig. 3a). For *tb1*, the ML estimate of N_b with $d = 1,000$ is 500

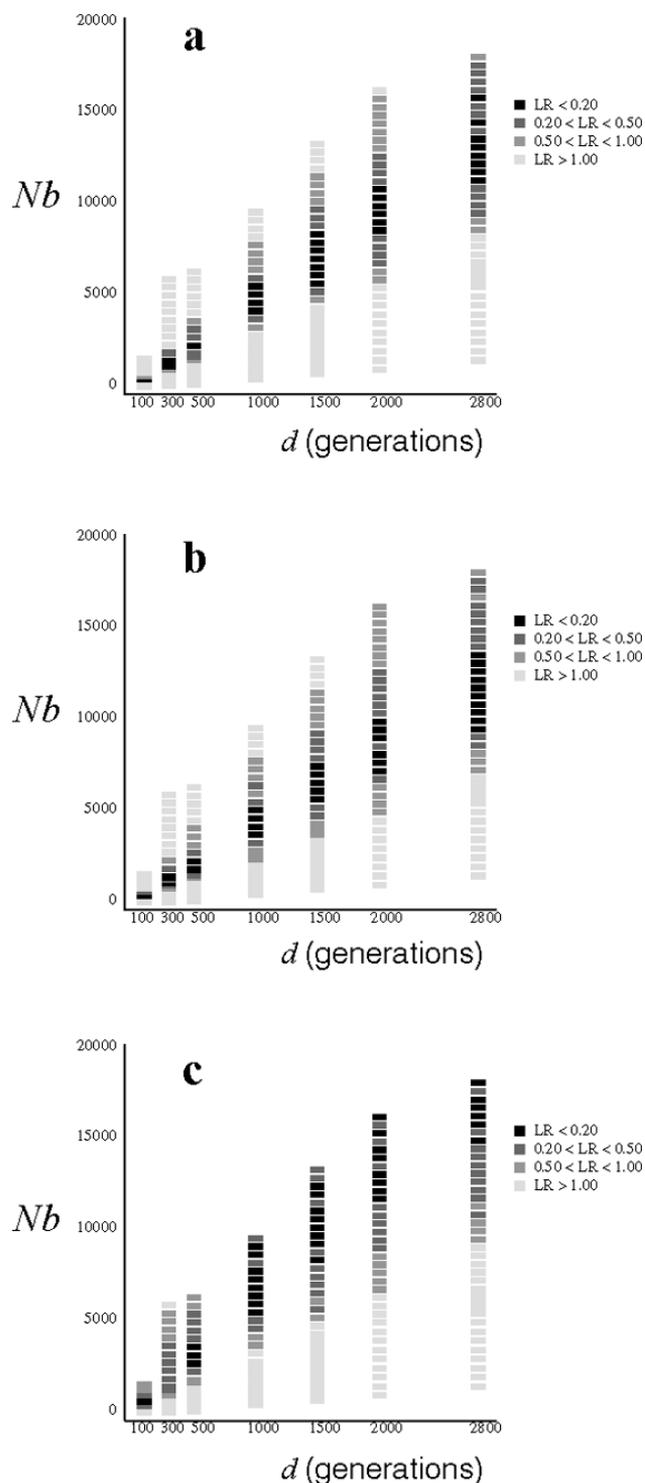


FIG. 4.—Likelihood ratio values based on the multilocus likelihood for seven loci (a) or eight loci (b, c) as a function of both bottleneck duration in generations (d) and population size during the bottleneck (N_b). Three recombination conditions were explored (a) $4Nc_{hud87}$, (b) $4Nc_{hud01}$, and (c) no recombination. The fitting criterion was $\pm 20\%$ of S_{maize} .

individuals. The *tb1* likelihoods under these two values of N_b are 0.0004 and 0.085, respectively (fig. 2d), and an LR test indicates that these values are significantly different (LR = 10.72; $P < 0.01$). Hence, not surprisingly, the *tb1*

data do not fit the “neutral” bottleneck model, but neither do *ts2* (LR = 4.97; $P < 0.05$), *d8* (LR = 7.42; $P < 0.01$), and *zagll* (LR = 22.55; $P < 0.001$). No LR tests were significant with data from putatively neutral loci (fig. 2a; data not shown).

Second, we formalized differences between the multilocus model and selected genes by simulation. We simulated each selected locus under the multilocus model and then determined the probability that the observed S falls into the 95% confident interval of the simulated distribution. Under both recombination estimates, we rejected the multilocus scenario for *tb1* ($p = 0.0002$ and $p = 0.0003$, with $4Nc_{hud87}$ and $4Nc_{hud01}$, respectively), *ts2* ($p = 0.008$ and $p = 0.0196$), *d8* ($p = 0.0021$ and $p = 0.0032$), and *zagll* ($p = 0.00$ and $p = 0.00$), strongly suggesting that the bottleneck model alone does not account for the evolutionary history of these genes in maize. In contrast, none of the eight putatively neutral genes could be differentiated from the multilocus model by this method (data not shown).

Genetic Diversity and Recombination

Previous studies reported a positive and significant correlation between recombination rate, measured by either $4Nc_{hud87}$ or Wall’s estimator (Wall 2000), and nucleotide diversity (θ) in maize ($r = 0.65$, $P = 0.007$), based on 18 putatively neutral loci. However, this same correlation was not significant when recombination was measured either by $4Nc_{hud01}$ or by a physical measure of recombination (R) (Tenaillon et al. 2002).

One of the questions we wanted ask was whether the positive correlation observed in maize was also evident in parviglumis. Among seven neutrally evolving loci for which a $4\hat{N}c_{hud87}$ value could be determined (table 3), we found no significant correlation between $4\hat{N}c_{hud87}$ and θ in parviglumis ($r = -0.07$, $p = 0.56$). Similarly, θ in parviglumis is correlated with neither R ($r = -0.116$; $p = 0.37$) nor $4\hat{N}c_{hud01}$ ($r = -0.25$, $p = 0.27$). By contrast, the correlation between $4\hat{N}c_{hud87}$ and θ in this subset of seven loci was still high in maize ($r = 0.58$), but not significant ($p = 0.32$), probably reflecting a lack of power with a small sample.

Using simulation, we explored whether the population bottleneck could generate the positive correlation between $4\hat{N}c_{hud87}$ and θ in maize. For this purpose, we performed 10,000 coalescent simulations under the best conditions defined for each of the seven values of d (fig. 3a). For each condition, the correlation between $4Nc_{simul}$ and θ_{simul} was determined among the seven neutral loci for which $4Nc_{hud87}$ could be estimated in parviglumis. We then compared the distribution of r based on simulation to the observed r ($= 0.58$) in maize. At best, only 2% of simulations (236 of 10,000) produced a correlation coefficient higher than the observed correlation. It is thus possible, but quite improbable, that bottleneck effects created the observed correlation between $4\hat{N}c_{hud87}$ and θ in maize.

Discussion

Genetic Diversity in Maize and *Z. mays* ssp. *parviglumis*

We examined the impact of domestication on sequence diversity at 12 loci, including coding and noncoding regions, in the progenitor of maize, *Zea mays* ssp. *parviglumis*. Although sequence diversity has been contrasted between parviglumis and maize previously, this study substantially increases the amount of publicly available parviglumis SNP data. In these data, the average loss of SNP diversity in maize, measured by π at silent sites, averaged $\sim 20\%$ in putatively neutral genes. However, the loss was much more severe ($>65\%$) for four putatively selected genes (*tb1*, *ts2*, *d8*, and *zag11*). These values are similar to previous estimates from neutral and selected genes; for example, three selected and three neutral genes in the starch biosynthesis pathway averaged 38.9 and 78.5% loss of diversity, respectively (Whitt et al. 2002).

The data provide other insights into genetic diversity in parviglumis. For example, LD, which is proportional to the inverse of $4Nc$, is generally lower than that in maize, which is already noted to have low LD levels (reviewed in Gaut and Long 2003). Although $4Nc_{hud87}$ values are correlated between parviglumis and maize, 8 of 10 loci had higher $4Nc$ estimates in parviglumis (table 3), consistent with a bottleneck increasing LD in maize. Our data also extend previous observations about the SNP frequency spectrum in parviglumis (Zhang et al. 2002a). The significant skew toward low-frequency polymorphisms is consistent with a history of either population subdivision or population expansion. If, however, population subdivision is driving the frequency skew, LD levels should be elevated in our admixed sample. Such an effect is not immediately apparent.

At present, then, it is difficult to pinpoint the cause(s) of the skew in SNP frequency spectrum, but it does complicate identification of selection events. Of the eight loci sequenced here, only *d8* provides reasonably strong evidence for selection in parviglumis, based on a significant Tajima's D , a significant Fu and Li test (with outgroup), and a significant haplotype excess (table 2). However, all three results are in the same direction as the frequency bias, so it is difficult to conclude if the results reflect selection or the frequency skew. Nonetheless, two arguments favor a selective interpretation. First, the D value in *d8* is more negative ($D = -1.82$) than that of *hm2* ($D = -1.63$), a plant defense gene thought to have been subject to positive selection in parviglumis (Zhang et al. 2002a) and subsequently shown to be under positive selection in other *Zea* taxa (P. Tiffin, R. Hacker, and B.S. Gaut, unpublished data). Second, *d8* is a strong candidate for adaptive selection because it is associated with variation in flowering time in maize (Thornsberry et al. 2001) and could therefore confer adaptative advantages in natural populations.

Recombination and Diversity

A previous study of chromosome 1 genes reported a positive and significant correlation between $4Nc_{hud87}$ and θ (and π) in maize (Tenailon et al. 2001) but did not detect a correlation between θ (and π) and R in maize (Tenailon et

al. 2002). One possible explanation for these observations was that demography influenced correlations between diversity and $4Nc_{hud87}$ in maize. An important test of this conjecture is to examine patterns of diversity in prebottleneck populations. We did not detect a correlation between sequence diversity and any measure of recombination—whether based on $4Nc$ or R —in parviglumis. Furthermore, our simulations indicate that the domestication bottleneck was unlikely to create the observed correlation between $4Nc_{hud87}$ and diversity in maize. Thus, the correlation between $4Nc_{hud87}$ and θ in maize remains difficult to interpret.

However, our data do reveal an interesting correlation between silent nucleotide diversity and divergence, as measured by Ks for eight parviglumis loci for which we have an outgroup sequence (table 3). The correlation was positive and significant in parviglumis (θ : $r = 0.79$, $p = 0.0015$; π : $r = 0.87$, $p < 0.001$); these correlations remain significant after removal of *d8*, which may be evolving nonneutrally, and also positive in maize (θ : $r = 0.48$, $p = 0.061$; π : $r = 0.29$, $p = 0.23$). Based on this relatively small sample, these correlations suggest that variation in SNP diversity among loci is a function of the neutral mutation rate. Other studies have noted correlations between diversity and divergence (Payseur and Nachman 2002; Hellmann et al. 2003) but differ in that divergence rates appear to be correlated with recombination (e.g., Lercher and Hurst 2002; Marais, Mouchiroud, and Duret 2003). Here there is no detectable correlation with recombination. Hence, recombination does not appear to be the proximal cause of variation in neutral SNP mutation rates.

The Demographic Model and Selection

A primary goal of this study is to explore a demographic model of domestication, using multilocus data. Similar approaches have proven useful in analysis of human and drosophilid population genetic data (Pluzhnikov, Di Rienzo, and Hudson 2002; Reich et al. 2001), but our simulations are unique in at least two respects. First, we have incorporated a wide array of locus-specific information, including $4Nc$ estimates and mutation rates. Mutation rates were estimated from outgroup data and calibrated with a rate estimated for grass *adh* loci (see *Materials and Methods*). The *adh* calibration point has some inherent weaknesses (see White and Doebley 1999 for detailed discussion), but remains one of the few point estimates of nuclear gene substitution rates in the grass family. Nonetheless, relative rates of mutation among loci should be reasonably accurate; they indicate that mutation rates vary ~ 2.8 -fold among *Zea* loci (table 3). This range corresponds with recent estimates in other plant systems. For example, Zhang, Vision, and Gaut (2002b) estimate that 90% of Arabidopsis genes fall within a relatively narrow range of 2.6-fold synonymous substitution rate variation among loci.

The second unique aspect is that, to our knowledge, these and previous simulations of maize domestication are the only simulations that incorporate empirical information about “prebottleneck” populations. The tacit assumption of this approach is that extant populations of parviglumis

provide reasonable insight into levels of diversity $\sim 7,500$ to 9,000 years ago when maize was domesticated (Iltis 1983; Matsuoka et al. 2002). Although domestication occurred very recently on an evolutionary time scale, there are some dangers to this assumption. For example, as mentioned earlier, it has now become clear from species-wide D values that parviglumis may have historically subdivided populations or has experienced recent population expansion. If the former, our “species-wide” parviglumis sample may overestimate the amount of variation that was present in the “primary source” population exploited for domestication. Ideally, one would sample exclusively from this “primary source” population, but this population may be difficult to identify and may no longer even exist (Matsuoka et al. 2002).

Given this limitation, we used coalescent simulations to investigate domestication scenarios. We used three metrics to fit simulated data to observed data—i.e., S , $4Nc_{hud87}$, and D —as well as combinations of these metrics. When S was used as a fitting criterion, the maximum likelihoods dictated that the population size N_b was about four to five times greater than the bottleneck duration d (fig. 4), but when $4Nc_{hud87}$ was used for goodness-of-fit, optimal N_b/d ratios were approximately halved. The difference between $4Nc_{hud87}$ and S as goodness-of-fit measures may be attributable, in part, to the high variance of $4Nc_{hud87}$ (Hudson 1987; Wall 2000) and to the fact that $4Nc$ depends on the frequency spectrum (Hudson 1987). In contrast to S and $4Nc_{hud87}$, D fit a far smaller proportion of simulations. The poor fit with D is likely explained both by the skewed frequency spectrum in parviglumis, which is not adequately mimicked by the coalescent process, and by its high variance. One future possibility to partially account for the skewed SNP frequency of parviglumis is incorporate population structure or growth into simulations of the ancestral population.

Our exploration of bottlenecks has an applied purpose: to differentiate the effects of directional selection from demography. A similar approach has been applied to maize microsatellite loci. Vigouroux et al. (2002b) screened 75 microsatellites for evidence of directional selection and found 15 SSRs (Simple Sequence Repeat) markers that deviate from expected diversity values, based on a bottleneck model. Six of these were localized to candidate genes, including *zagll*, that may prove to be agronomically important.

Here we applied the demographic model to data from four loci—*tb1*, *ts2*, *d8* and *zagll*—that served as test cases. None of the four loci fit the demographic model (fig. 2d; see Results), indicating that they differ significantly in patterns of diversity compared with the “neutral” genes that were used to tune the model. Given the bottleneck model, these four loci best fit a scenario in which N_b is very small (fig. 2d), consistent with the theoretical perspective that hitchhiking at a single locus can be approximated by a severe bottleneck (Depaulis, Mousset, and Veuille, in press). It is reasonable to ascribe the differences between *ts2*, *d8*, *tb1*, and *zagll* and the other eight loci to directional selection, but the phenotypic effects of such selection are not clear for *ts2* and *d8*. Both genes affect sex determination: *ts2* is involved in the feminization of the

maize tassel (Irish and Nelson 1993), and *d8* is involved in ear masculinization (Harberd and Freeling 1989). Because *tb1* has a pleiotropic effect on sexual fate, and because its phenotypic effect clearly is dependent on epistatic interactions with other genes (Doebley, Stec, and Gustus 1995; Lukens, and Doebley 1999), it is possible that selection on *ts2* and *d8* resulted from such interactions (Irish, Langdale, and Nelson 1994).

We should note the circularity of our statistical tests on *ts2* and *d8*. For these genes, we used existing neutrality tests to identify them as “different”, excluded them from model tuning on the basis of this difference, and then used the model to verify that they were, in fact, different. Our reasoning behind the circularity is simple: it is important to avoid potentially selected genes during the process of model tuning. Despite this circularity, these two genes are interesting test cases, precisely because previous neutrality tests were equivocal with maize data. For example, neither Tajima’s D nor MK tests identified these two genes as potentially selected in maize (Tenaillon et al. 2001). They were identified as potentially selected in maize on the basis of HKA tests, yet data from *ts2* and *d8* produced significant ($p < 0.05$) results in only three and two tests, respectively, out of 12 pairwise comparisons for each locus. Thus, the addition of parviglumis data has improved significantly the ability to infer selection, as has consideration of the demographic model. More importantly, application of this demographic model will help reveal deviations from neutrality in future large-scale genomic studies of maize polymorphism.

Acknowledgments

We thank Steven Wright and Frantz Depaulis for discussion and Wolfgang Stephan for helpful comments on an earlier version of the manuscript. We are grateful to John Doebley for contributing parviglumis seeds, unpublished sequences, and valuable comments. This study was supported by National Science Foundation grant number DBI-0096033 to B.S.G.

Literature Cited

- Andolfatto, P., and M. Przeworski. 2001. Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. *Genetics* **158**:657–665.
- Begun, D. J., and C. F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**:519–520.
- Begun, D. J., and P. Whitley. 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. *Proc. Natl. Acad. Sci., U.S.A.* **97**:5960–5965.
- Charlesworth, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet. Res.* **63**:213–227.
- Clark, R.M., E. Linton, J. Messing, and J. Doebley. in press. Patterns of diversity in the genomic region near the maize domestication gene, *tb1*. *Proc. Natl. Acad. Sci., U.S.A.*
- Davis, G. L., M. D. McMullen, C. Baysdorfer et al. (13 coauthors). 1999. A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. *Genetics* **152**:1137–1172.

- Depaulis, F., S. Mousset, and M. Veuille. 2003. Power of neutrality tests to detect bottlenecks and hitchhiking. *J Mol Biol Evol.* **57**:Supl 1:S190–S200.
- Doebley, J., A. Stec, and L. Hubbard. 1997. The evolution of apical dominance in maize. *Nature* **386**:485–488.
- Doebley, J., A. Stec, and C. Gustus. 1995. Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**:333–346.
- Dvorak, J., M.-C. Luo, and Z.-L. Yang. 1998. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. *Genetics* **148**:423–434.
- Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**:87–112.
- Eyre-Walker, A., R. L. Gaut, H. Hilton, D. L. Feldman, and B. S. Gaut. 1998. Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl. Acad. Sci., U.S.A.* **95**:4441–4446.
- Fearnhead, P., and P. Donnelly. 2001. Estimating recombination from population genetic data. *Genetics* **159**:1299–1318.
- Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693–709.
- Gaut, B. S., and M. T. Clegg. 1993. Molecular evolution of the *Adh1* locus in the genus *Zea*. *Proc. Natl. Acad. Sci. U.S.A.* **90**:5095–5099.
- Gaut, B. S., and A. D. Long. 2003. The lowdown on linkage disequilibrium. *Plant Cell* **15**:1502–1506.
- Gaut, B. S., B. R. Morton, B. M. McCaig, and M. T. Clegg. 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. U.S.A.* **93**:10274–10279.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**:95–98.
- Hanson, M. A., B. S. Gaut, A. O. Stec, S. I. Fuerstenberg, M. M. Goodman, E. H. Coe, and J. Doebley. 1996. Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics* **143**:1395–1407.
- Harberd, N. P., and M. Freeling. 1989. Genetics of dominant gibberellin-insensitive dwarfism in maize. *Genetics* **121**:827–838.
- Hellmann, I., I. Ebersberger, S. E. Ptak, S. Paabo, and M. Przeworski. 2003. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* **72**:1527–1535.
- Hilton, H., and B. S. Gaut. 1998a. Speciation and Domestication in maize and its wild relatives: evidence from the *Globulin-1* gene. *Genetics* **150**:863–872.
- Hilton, H., and B. S. Gaut. 1998b. Speciation and domestication in maize and its wild relatives: Evidence from the globulin-1 gene. *Genetics* **150**:863–872.
- Hudson, R. R. 1987. Estimating the recombination parameter of a finite population model without selection. *Genet. Res. Camb.* **50**:245–250.
- Hudson, R. R. 2001. Two-locus sampling distributions and their application. *Genetics* **159**:1805–1817.
- Hudson, R. R., and N. L. Kaplan. 1995. Deleterious background selection with recombination. *Genetics* **141**:1605–1617.
- Hudson, R. R., M. Kreitman, and M. Aguadé. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**:153–159.
- Itlis, H. H. 1983. From teosinte to maize: the catastrophic sexual transmutation. *Science* **222**:886–894.
- Irish, E. E., J. A. Langdale, and T. M. Nelson. 1994. Interactions between tassel seed genes and other sex determining genes in maize. *Dev. Genet.* **15**:155–171.
- Irish, E. E., and T. M. Nelson. 1993. Development of tassel seed 2 inflorescences in maize. *Am. J. Bot.* **80**:292–299.
- Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. The “Hitchhiking Effect” revisited. *Genetics* **123**:887–899.
- Kraft, T., T. Sall, I. MagnussonRading, N. O. Nilsson, and C. Hallden. 1998. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* **150**:1239–1244.
- Lercher, M. J., and L. D. Hurst. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* **18**:337–340.
- Long, A., B. F. Benz, D. J. Donahue, A. J. T. Jull, and T. J. Toolin. 1980. First Direct AMS Dates on Early Maize from Tehuacan, Mexico. *Radiocarbon* **31**:1035–1040.
- Lukens, L. N., and J. Doebley. 1999. Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genetical Research* **74**:291–302.
- Marais, G., D. Mouchiroud, and L. Duret. 2003. Neutral effect of recombination on base composition in *Drosophila*. *Genet. Res.* **81**:79–87.
- Matsuoka, Y., Y. Vigouroux, M. M. Goodman, J. Sanchez G., E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. U.S.A.* **99**:6080–6084.
- Maynard-Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. *Genet. Res.* **23**:23–35.
- McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**:652–654.
- Nachman, M. W. 1997. Patterns of DNA variability at X-linked loci in *Mus domesticus*. *Genetics* **147**:1303–1316.
- Nachman, M. W., V. L. Bauer, S. L. Crowell, and C. F. Aquadro. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**:1133–1141.
- Payseur, B. A., and M. W. Nachman. 2002. Natural selection at linked sites in humans. *Gene* **300**:31–42.
- Piperno, D. R., and K. V. Flannery. 2001. The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proc. Natl. Acad. Sci. U.S.A.* **98**:2101–2103.
- Pluzhnikov, A., A. Di Rienzo, and R. R. Hudson. 2002. Inferences about human demography based on multilocus analyses of non coding sequences. *Genetics* **161**:1209–1218.
- Przeworski, M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* **160**:1179–1189.
- Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti, D. J. Richter, T. Lavery, R. Kouyoumjian, S. F. Farhadian, R. Ward, and E. S. Lander. 2001. Linkage disequilibrium in the human genome. *Nature* **411**:199–204.
- Remington, D. L., J. M. Thornsberry, Y. Matsuoka, L. M. Wilson, S. R. Whitt, J. Doebley, S. Kresovich, M. M. Goodman, and E. S. Buckler. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. U.S.A.* **98**:11479–11484.
- Rozas, J., and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**:174–175.
- Smith, B. D. 1998. *The emergence of agriculture*. W. H. Freeman, New York.
- Stephan, W., and C. H. Langley. 1998. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics* **150**:1585–1593.

- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**:437–460.
- Tenaillon, M. I., M. C. Sawkins, L. K. Anderson, S. M. Stack, J. Doebley, and B. S. Gaut. 2002. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics* **162**:1401–1413.
- Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley, and B. S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl. Acad. Sci. U.S.A.* **98**:9161–9166.
- Thornsberry, J. M., M. M. Goodman, J. Doebley, S. Kresovich, D. Nielsen, and E. S. Buckler. 2001. Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* **28**:286–289.
- Vigouroux, Y., J. S. Jaqueth, Y. Matsuoka, O. S. Smith, W. D. Beavis, J. S. C. Smith, and J. Doebley. 2002a. Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution* **19**:1251–1260.
- Vigouroux, Y., M. McMullen, C. T. Hittinger, K. Houchins, L. Schulz, S. Kresovich, Y. Matsuoka, and J. Doebley. 2002b. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc. Natl. Acad. Sci. U.S.A.* **99**:9650–9655.
- Wall, J. D. 2000. A comparison of estimators of the population recombination rate. *Mol. Biol. Evol.* **17**:156–163.
- Wang, R. L., A. Stec, J. Hey, L. Lukens, and J. Doebley. 1999. The limits of selection during maize domestication. *Nature* **398**:236–239.
- Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**:188–193.
- Weiss, G., and A. von Haeseler. 1998. Inference of population history using a likelihood approach. *Genetics* **149**:1539–1546.
- White, S. E., and F. D. Doebley. 1999. The molecular evolution of terminal ear1, a regulatory gene in the genus *Zea*. *Genetics* **153**:1455–1462.
- Whitt, S. R., L. M. Wilson, M. I. Tenaillon, B. S. Gaut, and E. S. Buckler IV. 2002. Genetic diversity and selection in the maize starch pathway. *Proc. Natl. Acad. Sci. U.S.A.* **99**:12959–12962.
- Zhang, L., A. S. Peek, D. Dunams, and B. S. Gaut. 2002a. Population genetics of duplicated disease defense genes, hm1 and hm2, in maize (*Zea mays* ssp. *mays* L.) and its wild ancestor (*Zea mays* ssp. *parviglumis*). *Genetics* **162**:851–860.
- Zhang, L., T. J. Vision, and B. S. Gaut. 2002b. Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **19**:1464–1473.

William Martin, Associate Editor

Accepted January 12, 2004