

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*)

Liqing Zhang, Andrew S. Peek,¹ Detiger Dunams and Brandon S. Gaut²

Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697

Manuscript received March 14, 2002

Accepted for publication June 10, 2002

ABSTRACT

Plant defense genes are subject to nonneutral evolutionary dynamics. Here we investigate the evolutionary dynamics of the duplicated defense genes *hm1* and *hm2* in maize and its wild ancestor *Zea mays* ssp. *parviglumis*. Both genes have been shown to confer resistance to the fungal pathogen *Cochliobolus carbonum* race 1, but the effectiveness of resistance differs between loci. The genes also display different population histories. The *hm1* locus has the highest nucleotide diversity of any gene yet sampled in the wild ancestor of maize, and it contains a large number of indel polymorphisms. There is no evidence, however, that high diversity in *hm1* is a product of nonneutral evolution. In contrast, *hm2* has very low nucleotide diversity in the wild ancestor of maize. The distribution of *hm2* polymorphic sites is consistent with nonneutral evolution, as indicated by Tajima's *D* and other neutrality tests. In addition, one *hm2* haplotype is more frequent than expected under the equilibrium neutral model, suggesting hitchhiking selection. Both defense genes retain >80% of the level of genetic variation in maize relative to the wild ancestor, and this level is similar to other maize genes that were not subject to artificial selection during domestication.

PLANT defense-related genes encode a wide range of functions, including pathogen recognition, signal transduction, and direct enzymatic defense. To date, evolutionary studies have focused primarily on pathogen recognition genes, which are usually members of large multigene families (MEYERS *et al.* 1999). Diversification among gene family members is often mediated by positive selection (PARNISKE *et al.* 1997; MEYERS *et al.* 1998; WANG *et al.* 1998), presumably in response to intense selection for new resistance specificities (MICHELMORE and MEYERS 1998). There have been comparatively few evolutionary studies of defense genes that either act in signal cascades or interact directly with pathogens to limit the extent of infection (but see BISHOP *et al.* 2000; TIFFIN and GAUT 2001).

One useful approach for characterizing the evolution of defense genes is molecular population genetics, but to date intraspecific polymorphism has been studied only in two *Arabidopsis thaliana* resistance genes (*rpm1* and *rps2*) and one maize gene (*wip1*) that may play a role in defense. Both *rpm1* and *rps2* are pathogen recognition genes, and both have evolved in response to selective pressures that lead to the long-term maintenance of allelic diversity (CAICEDO *et al.* 1999; STAHL *et al.* 1999). In contrast, *wip1*, which codes for an adenine protease inhibitor, presumably limits the severity of in-

fection by interacting with and inhibiting pathogen proteases (RYAN 1990). Polymorphism data provide little evidence of recent balancing or positive selection at *wip1*, but this gene may have experienced episodic selection events through time (TIFFIN and GAUT 2001). In short, intraspecific studies have provided valuable insights into the evolutionary forces acting on defense genes.

Here we study DNA sequence polymorphism in *hm1* and *hm2*, two resistance genes in maize (*Z. mays* ssp. *mays*). The evolution of *hm1* and *hm2* is of great interest not only because they confer resistance to a plant pathogen (JOHAL and BRIGGS 1992; MULTANI *et al.* 1998), but also because they are duplicates. The two genes are 84.5% similar at the DNA sequence level, are located on different chromosomes (1L and 9L, respectively), and appear to confer a resistance function that is conserved throughout the grass family (MULTANI *et al.* 1998).

The *hm1* and *hm2* genes encode nitrate reductases that detoxify the toxin of *Cochliobolus carbonum* race 1 (JOHAL and BRIGGS 1992). In the absence of functional *hm1* and *hm2* alleles, the *C. carbonum* fungus causes leaf spot and ear mold, one of the most damaging diseases in maize (NELSON and ULLSTRUP 1964). Adult maize plants resist *C. carbonum* infection if either gene is functional, but the two loci differ in resistance response. This difference was characterized by NELSON and ULLSTRUP (1964), who infected plants of varied *hm1* and *hm2* genotypes with *C. carbonum* race 1. They studied three *hm1* alleles and found that one of the three was dominant and conferred complete resistance to *C. carbonum* at all

¹Present address: Integrated DNA Technologies, 1710 Commercial Park, Coralville, IA 52241.

²Corresponding author: Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525. E-mail: bgaut@uci.edu

developmental stages, independently of the *hm2* background (including an *hm2*-null background). The remaining *hm1* alleles conferred intermediate resistance with incomplete dominance. These two *hm1* alleles acted largely, but perhaps not completely, independently of the *hm2* background. Finally, plants with one or two functional *hm2* alleles in an *hm1*-null background were susceptible to fungal attacks in juvenile stages but exhibited intermediate to low resistance as adults. These observations led NELSON and ULLSTRUP (1964) to two conclusions. First, they concluded that there is little selective pressure on *hm2* alleles in a functioning *hm1* background. Second, because they could find no phenotypic variation among *hm2* alleles, they conjectured that at one point in time *hm2* provided the only source of resistance to *C. carbonum*.

To compare the evolutionary history of these defense genes, we have sampled allelic diversity of the two genes in maize and its wild ancestor (*Z. mays* ssp. *parviglumis*; hereafter called "parviglumis"). We focus on these two taxa both because we are interested in the long-term evolutionary dynamics of *hm1* and *hm2*, which can be investigated properly only in a wild species like parviglumis, and because we are interested in the genetic effects of domestication on these genes. It has already been shown that the domestication process affects maize genes differentially. Most maize genes contain ~60–80% of the level of variation found in parviglumis (WHITE and DOEBLEY 1999), but others have much reduced levels of genetic variation due to artificial selection during domestication (HANSON *et al.* 1996; WANG *et al.* 1999). From an agronomic standpoint, it is important to determine whether defense-related genes fall into the latter category and are thus genetically depauperate in maize.

To study the population genetics and evolution of *hm1* and *hm2*, we sampled sequences from several maize and parviglumis individuals and addressed the following questions:

- i. Is *hm2* less variable at the DNA sequence level, as implied by the phenotypic study of NELSON and ULLSTRUP (1964)?
- ii. Given that *hm1* confers greater resistance, is there evidence of selection on this disease defense gene?
- iii. How did the process of domestication affect genetic diversity in maize *hm1* and *hm2*?
- iv. And overall, what evolutionary forces have shaped genetic diversity in these two resistance genes?

MATERIALS AND METHODS

DNA sequencing: *Hm1* and *hm2* were PCR amplified and sequenced. For most individuals, *hm1* was amplified with the F1 (HM1-5'2) primer (5' cggattctgtctgctgggtgtgc 3'), which targets the first intron, and the R1 (HM1-3'3) primer (5' gatgtcgaggtagggaac 3'), which targets the fourth exon (Figure 1). PCR reactions with F1 and R1 consisted of 30 cycles

of 95° for 1 min, 65° for 1 min, and 72° for 2 min. For some individuals, it was necessary to use a nested PCR approach to amplify *hm1*. These individuals were first amplified with the F1B (F2965) and R1B (B1091) primers, which target the 5' untranslated region and the fifth exon, respectively (Figure 1), and then reamplified with the F1/R1 primer pair. F1B and R1B sequences were 5' atttcaggggcagccatggccga 3' and 5' tgcttctgttagccgagc 3'. Amplification with F1B and R1B consisted of 30 amplification cycles of 95° for 1 min, 60° for 1 min, and 72° for 2 min.

The *hm2* gene was also amplified with nested PCR. The first amplification used the R1 primer and the *hm2*-5F (5' atgaacagcagtagcagtgaagt 3') primer, which anneals to exon 1 (Figure 1). The first amplification consisted of 30 cycles of 95° for 1 min, 60° for 1 min, and 72° for 2 min. The nested primers were F91 (5' gggttcatcggtctctggctcgtcag 3'), which also targets exon 1, and R1 (Figure 1). Nested PCR was performed with 30 cycles of 94° for 30 sec, 52° for 30 sec, and 72° for 2 min. Altogether, amplification products consisted of the gene region from exon 2 to exon 4 for *hm1* and the gene region from exons 1 to 4 for *hm2* (Figure 1).

For comparison's sake, we also sequenced population samples of two additional genes on chromosome 9: *c1* and *waxy*. The population genetics of the *c1* locus have been studied in detail (HANSON *et al.* 1996); for this study we increased the number of individuals sampled from parviglumis, using methods similar to those described previously (HANSON *et al.* 1996) and examining roughly the same gene region. Forward and reverse primers were 5' agcaccagcacagcagtgctc 3' and 5' catagg taccagctgtctgttccagtagt 3', respectively, and PCR reactions were based on 30 cycles of 94° for 1 min, 60° for 1 min, and 72° for 2 min. The second gene, *waxy*, encodes a starch synthase. Following MASON-GAMER *et al.* (1998), we amplified *waxy* in a region spanning from exons 8 to 13. Forward and reverse primers were 5' tgcgagctagacaacatcatgccc 3' and 5' agggcgcggcactgtctcc 3', respectively. PCR amplification protocols followed those of MASON-GAMER *et al.* (1998).

All PCR-amplified products were cloned into a TA cloning vector (pGem), and one clone was sequenced for each PCR product, using BigDye chemistries and an ABI377 automated sequencer. After double-stranded sequences were obtained for all individuals, the sequences for each gene were aligned and all polymorphisms were identified. Because polymorphism can be caused by Taq-polymerase misincorporation, we repeated PCR, cloning, and sequencing for alleles that contained singletons—*i.e.*, variants that appeared only once among sequences. Singletons were either verified as a true polymorphism or corrected. We did not examine variants that were found in more than one individual, because the probability that shared variants are caused by a Taq artifact is negligible (EYRE-WALKER *et al.* 1998). This method of error verification also helped ensure that our sequences were not interallelic PCR recombinants.

Individuals sampled: The number of individuals sampled for each gene is provided (Table 1). Maize and parviglumis were sampled randomly throughout their geographic ranges (see APPENDIX). For *hm1*, we did not include five GenBank sequences in our sample because their sampling was biased with respect to phenotype (MULTANI *et al.* 1998). Their inclusion does not substantially alter results, however. Similarly, the *hm2* and *waxy* samples did not contain GenBank sequences from previous studies. For *c1*, five parviglumis individuals were sampled; the remaining sequences used in this study were described previously (HANSON *et al.* 1996). Sequence samples for *adh1* and *glb1* have been described elsewhere (EYRE-WALKER *et al.* 1998; HILTON and GAUT 1998); for *glb1* we used the alignment of TIFFIN and GAUT (2001). Sequences were submitted to GenBank (*hm1*, AY101968–AY101987; *hm2*,

AY101988–AY102008; *c1*, AF292540–AF292553; *waxy*, AF292500–AF292539).

Sequence analysis: The average pairwise difference among sequences, π (TAJIMA 1983), was used as a diversity measure and was calculated for all genes in both taxa. π was based on silent sites (synonymous sites plus noncoding sites) or nonsynonymous sites. Most tests of neutrality were performed using DnaSP, version 3.53 (ROZAS and ROZAS 1999). The tests included Tajima's D test (TAJIMA 1989), Fu and Li's D test with and without outgroup (FU and LI 1993), the HKA test (HUDSON *et al.* 1987), and the McDONALD and KREITMAN (1991) (MK) test. The MK test was applied using *hm1/hm2* paralogs as outgroups and also using an outgroup sequence from rice (DDBJ E10912). The rice sequence was also used as an outgroup for Fu and Li's (1993) D test and FAY and WU's (2000) H test (<http://crimp.lbl.gov/htest.html>). We tested for homogeneity in the distribution of polymorphism and divergence between *hm1* and *hm2* with the run test of McDONALD (1998), using all sequence sites.

To determine whether domestication had homogeneous effects across maize loci, we devised a statistical test on the basis of the ratio R of genetic variation between maize and parviglumis. For locus i , R was defined as $R_i = \hat{\theta}_{i,mze}/\hat{\theta}_{i,parv}$, where $\hat{\theta}_{i,mze}$ was based on observed sequence diversity in maize, $\hat{\theta}_{i,parv}$ was based on observed sequence diversity in parviglumis, and both measures were based on WATTERSON'S (1975) estimator and silent sites. The average level of diversity in maize relative to parviglumis over n loci was $\bar{R} = \sum_{i=1}^n R_i/n$, where n was the number of genes assayed.

If domestication affected genetic diversity equally in all genes, then the null hypothesis of homogeneity ($R_1 = R_2 = R_3 = \dots = R_n \approx \bar{R}$) should hold. We tested for homogeneity in R across loci by coalescent simulation. For each of n loci, coalescent simulations of the neutral equilibrium model were performed with and without recombination for both maize and parviglumis, using the program of DEPAULIS *et al.* (2001). Coalescent simulations for locus i in parviglumis were performed with parameter values of $4N\mu_{i,parv} = \hat{\theta}_{i,parv}$ and recombination rates were estimated from the data with HUDSON'S (1987) estimator of the population-recombination parameter. Coalescent simulations for locus i in maize were performed with parameter values of $4N\mu_{i,mze} = \bar{R}\hat{\theta}_{i,parv}$ and recombination rates were estimated with HUDSON'S (1987) estimator. For each simulation over n loci, we calculated a statistic similar to that of HUDSON *et al.* (1987),

$$\chi^2 = \sum_{i=1}^n \left[\frac{(S_{i,mze} - E(S)_{i,mze})^2}{\text{Var}(S)_{i,mze}} + \frac{(S_{i,parv} - E(S)_{i,parv})^2}{\text{Var}(S)_{i,parv}} \right],$$

where $S_{i,mze}$ and $S_{i,parv}$ were the number of segregating sites determined by simulation in maize and parviglumis, respectively; $E(S)_{i,mze}$ and $E(S)_{i,parv}$ were the expected number of segregating sites given $4N\mu_{i,parv}$ and $4N\mu_{i,mze}$; and $\text{Var}(S)$ was the variance in the number of segregating sites (TAJIMA 1983) assuming no recombination. We performed 1000 simulations and compared the distribution of χ^2 based on simulated data to the χ^2 value based on the observed numbers of segregating sites in the n loci. The null hypothesis was rejected when the χ^2 statistic based on observed data was >95% of simulated χ^2 statistics.

RESULTS

Indel variation in *hm1* and *hm2*: Despite their similarity in function, *hm1* and *hm2* had different levels and types of genetic variation. The most obvious difference between genes was the frequency and size of indel poly-

morphisms. In *hm1*, for example, the 20 maize and parviglumis sequences contained 4 indels >100 bp in length (Figure 1) and another ~30 small (<20 bp) indels. As a result, the *hm1* alignment contained a gap (reflecting an indel polymorphism) in at least one individual in 1554 bases of 2308 aligned base pairs (~67% in length). Tajima's D was positive, although not significantly so, for indels in both maize ($D = 1.17$, $P > 0.10$) and parviglumis ($D = 0.31$; $P > 0.10$). None of the indels in this random sample interrupted the reading frame. In contrast to *hm1*, the sample of 22 *hm2* sequences from maize and parviglumis had a total of six indel polymorphisms, the largest 18 bp in length. Only ~3.5% of the *hm2* alignment contained indels. The six *hm2* indels in parviglumis were present in frequencies <0.15, resulting in a negative Tajima's D statistic ($D = -0.041$; $P > 0.10$).

The large indels of *hm1* were located in introns and appear to be miniature inverted-repeat transposable elements (MITEs; Figure 1). For example, intron 2 contained an ~220-bp *Tourist* MITE in one maize (landrace Conico) and one parviglumis individual (accession no. PI133783). Intron 3 had three apparently separate >100-bp indels. The first 308-bp insertion was found in one parviglumis individual (accession no. 304707). A BLAST search using the indel as a query yielded one hit (BLAST score = $4e^{-67}$) to an unannotated region of a maize genomic cosmid clone. The insertion appeared to be a MITE because it ended in 13-bp terminal inverted repeats (TIRs) and was flanked by 3-bp direct repeats. We named this MITE *Trek* (Figure 1). The second indel, which was present in one randomly sampled individual but also present in GenBank *hm1* sequences, had similarity (BLAST score = $1e^{-6}$) to the 5' portion of the maize *Zea mays* 6-phosphogluconate dehydrogenase isoenzyme *A* gene. Because this indel sequence was present in two different maize genes, it likely represents another MITE-like element that we named *Litespeed*. The third large indel in intron 3 was a 128-bp *Tourist* element that was found previously in *hm1* (MULTANI *et al.* 1998). This *Tourist* element was present in six individuals—three maize and three parviglumis.

Nucleotide variation and tests of selection: We estimated nucleotide diversity at *hm1* and *hm2* on the basis of parviglumis and maize sequences (Table 1). GenBank sequences of *hm1* were not included in these and other population genetic calculations both because they were sampled from relatively narrow U.S. inbred germplasm and because their sampling was biased with respect to phenotype. Comparing π between *hm1* and *hm2* led to two conclusions. First, *hm1* had higher levels of nucleotide diversity, whether diversity was sampled in maize or parviglumis or measured at silent sites, synonymous sites (data not shown), or nonsynonymous sites (Table 1). Second, the ratio of nonsynonymous to silent diversity was higher in *hm1*. For example, the ratio of π_{nonsyn} to π_{silent} was 0.50 for *hm1* in parviglumis, whereas it was

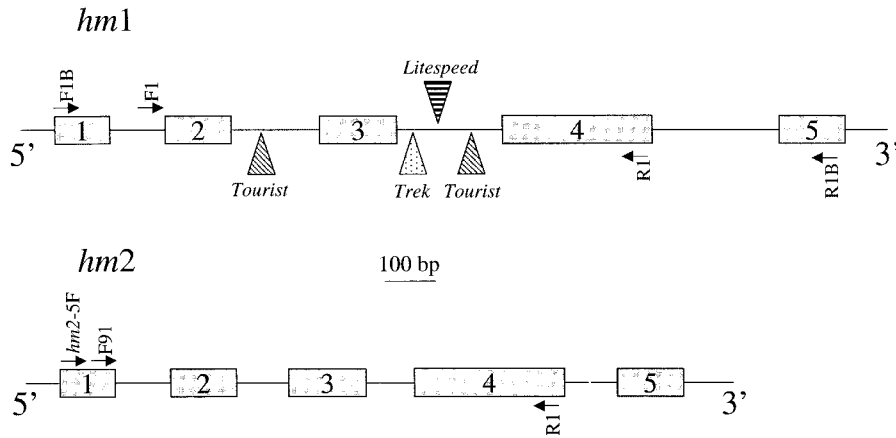


FIGURE 1.—Schematic of the *hm1* and *hm2* genes. Shaded boxes represent exons; the thin lines connecting boxes represent introns and other noncoding regions. Arrows indicate the names and locations of PCR primers. Triangles represent MITE insertions in *hm1* sequences.

0.27 for *hm2*. The high ratio of π_{nonsyn} to π_{silent} in *hm1* is not due to obvious differences in the frequency spectrum of nonsynonymous polymorphisms between genes but rather to the relatively high number of segregating nonsynonymous sites (Table 1).

Haplotype patterns also differed between *hm1* and *hm2*. *hm1* sequences are typified by substantial variation: No two sequences were identical, but common *hm1* polymorphisms were shared between parviglumis and maize (Figure 2). In contrast, the sample of 11 parviglumis *hm2* sequences contained 4 identical sequences and a fifth sequence that differed by only 1 base (Figure 3). We used coalescent simulations to determine whether four identical sequences in the parviglumis *hm2* sample were consistent with the neutral equilibrium model (see HUDSON *et al.* 1994). Four or more identical haplotypes were found only rarely (334 of 10,000 simulations; $P = 0.033$), assuming no recombination. There is thus a significant excess of one haplotype in the *hm2* parviglumis sample relative to the neutral expectation. The remaining parviglumis *hm2* haplotypes contained a high number of singletons; 33 of 40 polymorphic *hm2* sites were singletons, whereas only ~ 8 singletons are expected under the neutral-equilibrium model $n = 11$ (Tajima 1989, Equation 50). Although some polymorphic sites were shared between parviglumis and maize *hm2* samples, maize *hm2* sequences contributed an additional 9 singleton polymorphisms (Figure 3).

To further explore the distribution of variants, we applied neutrality tests. Tajima's D was not significant for *hm1* in maize, *hm1* in parviglumis, and *hm2* in maize (Table 1). However, D was marginally significant ($P < 0.10$) with parviglumis *hm2* data (Table 1). To investigate this result further in a comparative context, we examined *hm1* and *hm2* exon data. (Exon data were examined because intron data were difficult to align and therefore introns could not be compared between *hm1* and *hm2* nor among *hm1*, *hm2*, and the rice outgroup sequence.) With exon data, Tajima's test ($D = -1.883$; $P < 0.05$), Fu and Li's test ($D = -2.353$; $P < 0.01$), and Fu and Li's tests with outgroup ($D = -3.133$;

$P < 0.01$) rejected the null hypothesis for parviglumis *hm2*; all rejections were due to an excess of rare polymorphisms. These tests did not reject the null hypothesis with maize *hm2*, maize *hm1*, and parviglumis *hm1* exon data, but all statistics were less than zero (Table 1; data not shown). FAY and WU's (2000) test for selective sweeps was not significant for *hm1* or *hm2* in either taxon (data not shown).

We also applied the MK test, comparing *hm1* and *hm2* paralogs to one another to estimate divergence. MK tests did not reject neutrality whether the data were from parviglumis ($P = 1.00$), maize ($P = 0.332$), or combined between taxa ($P = 0.387$). Similarly, MK tests with the rice outgroup did not reject for *hm1* or *hm2* data (data not shown). McDONALD's (1998) run test provides an alternative means for assessing the relationship between divergence and polymorphism. With arbitrarily chosen levels of recombination, McDonald's run test provided no significant results when *hm1* polymorphism was compared to divergence, using either maize or parviglumis data (Table 2). Comparisons of *hm2* polymorphism to divergence were borderline significant ($P < 0.10$) for parviglumis data, but not significant with maize data (Table 2).

DNA sequence diversity in *hm1* and *hm2* relative to other chromosome 1 and 9 loci: Several aspects of the polymorphism data suggest that *hm1* and *hm2* have experienced different evolutionary histories. These observations, while valuable, require a broader context for interpretation. It is thus helpful to begin to formulate a genomic picture of Zea diversity. To construct a broader picture of Zea polymorphism, we compared diversity among *hm1*, *hm2*, and five additional chromosome 1 and 9 genes. Prior to this study, three chromosome 1 genes had been sampled extensively for both parviglumis and maize (*adh1*, *glb1*, and *tb1*), and we enhanced sampling for two chromosome 9 genes (*waxy* and *c1*). The chromosomal location of these genes is given (Table 1).

Comparison of nucleotide diversity among chromosome 1 and 9 genes leads to five observations. First, *hm1*

TABLE 1
Sequence statistics for seven genes in maize and parviglumis

Gene	Location ^a	Parviglumis					Maize					R		
		n^b	Length ^c	S^d	π_{silent}	π_{nonsyn}	D^e	n^b	Length ^c	S^d	π_{silent}		π_{nonsyn}	D^e
<i>hm1</i>	1 (125.1)	10	1016 (496)	96 (36)	0.0414	0.0204	-0.517	10	1026 (505)	73 (30)	0.0338	0.0180	0.101	0.82
<i>hm2</i>	9 (72.8)	11	1109 (563)	40 (10)	0.0126	0.0033	-1.630*	10	1142 (596)	24 (9)	0.0105	0.0004	-0.159	0.83
<i>adh1</i>	1 (205.2)	8	1259 (812)	59 (1)	0.0270	0.0013	-0.052	9	1368 (921)	48 (1)	0.0215	0.0005	0.452	0.80
<i>glb1</i>	1 (189.4)	8	992 (490)	75 (22)	0.0358	0.0139	-0.903	9	1033 (517)	51 (18)	0.0237	0.0108	-0.354	0.66
<i>tb1</i>	1 (197.6)	7	2599 (1977)	98 (9)	0.0165	0.0045	-0.740	12	2687 (2068)	14 (2)	0.0013	0.0009	-1.324	0.08
<i>c1</i>	9 (33.1)	11	672 (453)	29 (2)	0.0236	0.0017	0.367	15	698 (479)	7 (0)	0.0035	0.0000	-0.765	0.15
<i>waxy</i>	9 (63.7)	9	1480 (686)	38 (3)	0.0159	0.0016	-0.858	15	1243 (669)	28 (2)	0.0127	0.0008	-0.253	0.80

* $P < 0.10$.

^a Chromosomal location, with map location in centimorgans from the Pioneer 1999 Composite Map in parentheses.

^b Number of sequences sampled.

^c Number of sites in aligned sequences, excluding gaps, with number of silent sites in parentheses.

^d Total number of segregating sites, S , with nonsynonymous segregating sites in parentheses.

^e Tajima's D , based on all sites.

^f Ratio of π_{silent} in maize to π_{silent} in parviglumis.

contains high levels of silent diversity (Table 1); *hm1* is the most diverse gene sampled to date in parviglumis, but the regulatory gene *opaque2* on chromosome 7 has apparently higher nucleotide diversity within maize (HENRY and DAMERVAL 1997). Second, the ratio of π_{nonsyn} to π_{silent} for *hm1* is high compared to all other genes except *glb1* (Table 1), which is a seed storage protein that is presumably under little selective constraint (HILTON and GAUT 1998). Third, *hm2* contains the second lowest level of silent diversity among sampled parviglumis genes. Only the chromosome 3 gene *te1* has lower levels of silent site diversity, but *te1* exhibits no evidence of deviation from neutral equilibrium (WHITE and DOEBLEY 1999).

The fourth observation is that domestication has affected genes differentially. All genes demonstrate lower genetic diversity in maize than in its wild relative parviglumis, but *tb1* and *c1*, two genes putatively selected during domestication, have experienced a severe loss of genetic diversity in maize, as noted previously (HANSON *et al.* 1996; WANG *et al.* 1999). In contrast, *hm1* and *hm2* retain the highest π_{silent} proportion in maize relative to parviglumis, at 83% (Table 1; WHITE and DOEBLEY 1999; TIFFIN and GAUT 2001). We used a simulation method (see MATERIALS AND METHODS) to test whether relative levels of genetic variation between taxa were homogeneous among loci. When all seven genes in Table 1 were included for analysis, the homogeneity test was not significant without recombination ($P = 0.85$) but was significant when recombination was included in coalescent simulations ($P = 0.05$). The latter results suggest that the effects of domestication are not equivalent across loci, but this is not surprising because of known domestication effects on *c1* and *tb1*. With the five remaining genes (*waxy*, *glb1*, *hm1*, *hm2*, and *adh1*), the homogeneity test was not significant, either with ($P = 0.24$) or without recombination ($P = 0.94$). Thus, there is no strong evidence that *hm1* or *hm2* retains an aberrantly high proportion of sequence diversity in maize relative to other loci, but it is also clear that the domestication process did not preferentially remove genetic diversity from these two disease-related genes.

Finally, it is striking that Tajima's D was negative for six of the seven parviglumis loci in Table 1. Furthermore, Tajima's D increased in maize relative to parviglumis for five of the seven genes. The two genes (*tb1* and *c1*) that do not conform to this pattern are genes that were subjected to artificial selection during domestication.

DISCUSSION

This study was designed to compare evolutionary histories between two loci that confer resistance to the fungus *C. carbonum*. Previous molecular evolutionary studies have shown that disease defense genes can be subject to positive selection, which drives divergence between resistance paralogs (PARNISKE *et al.* 1997; MEY-

the neutral equilibrium model. However, it must be remembered that neutrality tests have low power (SIMONSEN *et al.* 1995) and may not be able to detect some alternative models of selection, such as frequency-dependent selection (BRAVERMAN *et al.* 1995). Given the catastrophic effect of loss-of-function *hm1* mutations in the presence of fungal pathogens (NELSON and ULLSTRUP 1964), it is possible that *hm1* is under selective pressures that cannot be detected in the species-wide sequence sample studied here.

In contrast, *hm2* has low diversity, and the distribution of diversity in parviglumis deviates from the neutral model on the basis of several measures, including Tajima's *D*, McDonald's run test, and haplotype distribution. With regard to the latter, *hm2* is atypical among parviglumis loci. For example, *hm2* contains four identical sequences and one additional sequence that differs by a single base pair (Figure 3). By comparison, *glb1*, *hm1*, *tb1*, *waxy*, and *te1* (WHITE and DOEBLEY 1999) samples contain no identical sequences, and *adh1*, *c1*, and *wip1* (TIFFIN and GAUT 2001) samples contain at most two identical sequences. We should also note that the four identical *hm2* sequences show no obvious geographic clustering (the four identical sequences are from three different Mexican states), suggesting that population subdivision is not responsible for the haplotype distribution. Altogether, significant neutrality tests, coupled with the unique features of *hm2* relative to other parviglumis loci, suggest that *hm2* has been subjected to diversity-reducing selection in parviglumis.

Previous studies of the *Drosophila SOD* (HUDSON *et al.* 1994) and *rp49* genes (ROZAS *et al.* 2001) have also detected an excess of a single haplotype, and in both cases the observation was interpreted as evidence of either a recently established balanced polymorphism or an ongoing selective sweep. If *hm2* has experienced a recent selective sweep, two aspects of the results are puzzling. First, the *H* statistic does not detect a significant excess of high-frequency variants, as expected under a hitchhiking model (FAY and WU 2000). However, statistics like *D* and *H* are critically dependent on parameters of a sweep event, such as the age of the event, the strength of selection, and the recombination rate. As a result, it is common that either *D* or *H* (but not both) detects a selective sweep (FAY and WU 2000). Furthermore, our outgroup sequence for the *H* test was evolutionarily distant and restricted the test to exon regions, thereby limiting statistical power.

The second puzzling aspect of a potential selective sweep is the excess of singletons in the remaining (non-common) *hm2* parviglumis haplotypes (Figure 3). Although an excess of singletons is expected after a selective sweep, both in the presence and absence of recombination (FAY and WU 2000), the proportion of singleton segregating sites in *hm2* is extremely high (33 singletons of 40 polymorphic sites = 82%). However, the proportion of singleton polymorphisms in the other

six parviglumis genes in Table 1 is also high, ranging from 20% (*c1*) to 70% (*tb1*) with an average of 53%. Furthermore, six of seven parviglumis genes have a negative Tajima's *D* (Table 1). Taken together, these observations suggest that parviglumis has experienced population expansion, population subdivision, or other demographic events that contribute to the accumulation of low-frequency polymorphisms. Thus, demographic effects may be responsible for a substantial (but unknown) proportion of *hm2* singleton polymorphisms.

If *hm2* has been subjected to a recent or ongoing selective sweep, as suggested by the overabundance of one haplotype, an important question is whether it has been the target of selection or has been hitchhiking with a more distant, and possibly fixed, mutation. If *hm2* is the target of selection, the selected site is probably not in the genic region we sampled. In the absence of recombination and population subdivision, a selected site should be a "fixed" difference between the common and uncommon haplotypes; such a variant is not evident (Figure 3). With recombination and population subdivision, a selected site need not be "fixed" between common and uncommon haplotypes, but the only common variants in our sample (sites 780 and 1055; Figure 3) are silent sites, which seem unlikely to be the target of selection.

Nonetheless, the "footprint" of selection may be small in *Zea* taxa and hence any selected site may be physically near *hm2*. An example of a small selective footprint comes from the *tb1* locus. Selection at *tb1* is evident by reduced diversity in the 5' regulatory region but *not* in the coding region a few hundred bases downstream (WANG *et al.* 1999, 2001). Although it is not yet clear how far upstream the footprint of selection extends, *tb1* demonstrates that recombination in maize is sufficient to decouple strongly selected regions (the regulatory region) from nonselected regions (the coding region) in a short physical distance (<500 bp) over a short time span (maize was domesticated ~7500 years ago; ILLIS 1983). Additional studies suggest that recombination in maize is generally sufficient to break down linkage disequilibrium (LD) along physical distances of a few hundred base pairs, even in centromeric regions that may have suppressed recombination (TENAILLON *et al.* 2001). LD in parviglumis decreases more rapidly than LD in maize (data not shown). Although we do not have direct information about recombination rates in regions near *hm2*, these observations suggest that recombination in maize and parviglumis may be sufficient to limit the footprint of selection to small physical distances unless selection is very strong—*i.e.*, above the 4–8% selection coefficient postulated for *tb1* (WANG *et al.* 1999).

The effect of domestication on genetic diversity: Although there is evidence of nonneutral evolution in parviglumis *hm2*, there is no corresponding evidence for nonneutrality in maize. This contradiction may be

the result of demographic differences between taxa (parviglumis typically grows in wild areas away from corn fields; DOEBLEY 1990) and domestication. Domestication affects genetic diversity in two ways. First, domestication decreases genetic diversity. All genes in this and other studies (WHITE and DOEBLEY 1999; TIFFIN and GAUT 2001) exhibit a decrease in sequence diversity in maize relative to its wild ancestor. On the basis of Table 1, genes retain from 8 to 83% of the level of π_{silent} in maize relative to parviglumis, with *hm1* and *hm2* on the high end of the range. The average ratio of π_{silent} maize relative to parviglumis for genes in Table 1 is 60%, but this number is biased downward because of *tb1* and *c1*. A more reasonable estimate, based on five genes that appear to have been affected homogeneously during domestication, is that maize retains 78% of the level of silent diversity relative to parviglumis. This estimate decreases slightly, to 77%, with the inclusion of *te1* and *wip1* (WHITE and DOEBLEY 1999; TIFFIN and GAUT 2001). In comparison, simple sequence repeat diversity in maize and parviglumis produces a slightly higher estimate of 88% (MATSUOKA *et al.* 2001). These estimates are of interest because little data on the genetic effects of domestication in crop systems are available, particularly at the DNA sequence level (BUCKLER *et al.* 2001).

The second effect of domestication is to increase Tajima's *D* in maize relative to parviglumis (Table 1), probably due to the loss of low-frequency polymorphisms during a population bottleneck (PRZEWORSKI *et al.* 2000) associated with domestication. A "domestication bottleneck" accelerates genetic drift, causing the loss of most rare variants but also increasing the frequency of some variants. The latter is evident in the *hm1* and *hm2* data, which indicate that some intermediate-to-rare variants in parviglumis are common in maize (Figures 2 and 3). In this context, we should also note that the *hm2* "hitchhiking" haplotype has a lower frequency in the maize sample (10% frequency) than in the parviglumis sample (45% frequency). This decrease may reflect demographic events, such as accelerated drift during domestication, that weaken the selective signal in maize. Alternatively, the relatively low frequency of the hitchhiking haplotype in maize may indicate that the selective coefficient on *hm2* (or a linked region) differs between the domesticate and its wild ancestor. In either case, the loss of low-frequency polymorphisms is a general property of domestication and breeding because Tajima's *D* has also increased in U.S. breeding (elite) germplasm relative to exotic (nonelite) maize germplasm (TENAILLON *et al.* 2001).

Conclusions: the evolutionary dynamics of *hm1* and *hm2*: To the extent that we can measure them accurately, the evolutionary forces on *hm1* and *hm2* have differed dramatically in their recent history. The gene that confers more complete disease resistance to the fungal pathogen (*hm1*) contains a high degree of sequence

diversity but lacks evidence of strong selection. In contrast, the gene that has "no selective advantage" (NELSON and ULLSTRUP 1964) exhibits evidence of deviation from the neutral equilibrium model, presumably due to a selective sweep. Although we cannot ascertain whether selection has targeted *hm2*, as opposed to a linked region, these results add to the general picture that disease-related genes are subject to bouts of selection, perhaps as a function of pathogen availability and specificity (STAHL *et al.* 1999; TIFFIN and GAUT 2001).

The authors thank Peter Tiffin for discussion and comment, J. Doebley for *tb1* sequence data and parviglumis seeds, and M. Goodman for maize seeds. M. Tenaillon, A. Tatarenkov, and three anonymous reviewers provided helpful comments. This study was supported by United States Department of Agriculture grant no. 98-35301-6153 and National Science Foundation (NSF) grant no. DBI-9872631 and an NSF dissertation grant to L.Z.

LITERATURE CITED

- BISHOP, J. G., A. M. DEAN and T. MITCHELL-OLDS, 2000 Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. USA* **97**: 5322–5327.
- BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–796.
- BUCKLER, E. S., J. M. THORNESBERRY and S. KRESOVICH, 2001 Molecular diversity, structure and domestication of grasses. *Genet. Res.* **77**: 213–218.
- CAICEDO, A. L., B. A. SCHAAL and B. N. KUNKEL, 1999 Diversity and molecular evolution of the RPS2 resistance gene in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **96**: 302–306.
- DEPAULIS, F., S. MOUSSET and M. VEUILLE, 2001 Haplotype tests using coalescent simulations conditional on the number of segregating sites. *Mol. Biol. Evol.* **18**: 1136–1138.
- DOEBLEY, J., 1990 Molecular evidence and the evolution of maize. *Econ. Bot.* **44**: 6–27.
- 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. USA* **95**: 4441–4446.
- FAY, J. C., and C.-I. WU, 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.
- FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GOLOUBINOFF, P., S. PAABO and A. C. WILSON, 1993 Evolution of maize inferred from sequence diversity of an *Adh2* gene segment from archaeological specimens. *Proc. Natl. Acad. Sci. USA* **90**: 1997–2001.
- HANSON, M. A., B. S. GAUT, A. O. STEC, S. I. FUERSTENBERG, M. M. GOODMAN *et al.*, 1996 Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. *Genetics* **143**: 1395–1407.
- HENRY, M., and C. DAMERRAL, 1997 High rates of polymorphism and recombination at the *Opaque-2* locus in cultivated maize. *Mol. Gen. Genet.* **256**: 147–157.
- HILTON, H., and B. S. GAUT, 1998 Speciation and domestication in maize and its wild relatives: evidence from the *Globulin-1* gene. *Genetics* **150**: 863–872.
- HUDSON, R. R., M. KREITMAN and M. AGUADE, 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIATOWSKI and F. J. AYALA, 1994 Evidence for positive selection in the superoxide dismutase (*Sod*) regions of *Drosophila melanogaster*. *Genetics* **136**: 1329–1340.
- ILTIS, H. H., 1983 From teosinte to maize: the catastrophic sexual transmutation. *Science* **222**: 886–894.
- JOHAL, G. S., and S. P. BRIGGS, 1992 Reductase activity encoded by the HM1 disease resistance gene in maize. *Science* **258**: 985–987.

- MASON-GAMER, R. J., C. F. WEIL and E. A. KELLOGG, 1998 Granule-bound starch synthase: structure, function and phylogenetic utility. *Mol. Biol. Evol.* **15**: 1658–1673.
- MATSUOKA, Y., S. E. MITCHELL, S. KRESOVICH, M. GOODMAN and J. DOEBLEY, 2001 Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. *Theor. Appl. Genet.* **104**: 436–450.
- MCDONALD, J. H., 1998 Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol. Biol. Evol.* **15**: 377–384.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh1* locus in *Drosophila*. *Nature* **351**: 652–654.
- MEYERS, B. C., K. A. SHEN, P. ROHANI, B. S. GAUT and R. W. MICHELMORE, 1998 Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell* **10**: 1833–1846.
- MEYERS, B. C., A. W. DICKERMAN, R. W. MICHELMORE, S. SIVARAMAKRISHNAN, B. W. SOBRAL *et al.*, 1999 Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* **20**: 317–332.
- MICHELMORE, R. W., and B. C. MEYERS, 1998 Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* **8**: 1113–1130.
- MULTANI, D. S., G. S. JOHAL, J. GRAY, R. B. MEELEY and S. P. BRIGGS, 1998 Plant-pathogen micro-evolution: molecular basis for the recent origin of a disease in maize. *Proc. Natl. Acad. Sci. USA* **95**: 1686–1691.
- NELSON, O. E., and A. J. ULLSTRUP, 1964 Resistance to leaf spot in maize. *J. Hered.* **55**: 195–199.
- PARNISKE, M., K. E. HAMMOND-KOSACK, C. GOLSTEIN, C. M. THOMAS, D. A. JONES *et al.*, 1997 Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf4/9 locus of tomato. *Cell* **91**: 821–832.
- PRZEWORSKI, M., R. R. HUDSON and A. DI RIENZO, 2000 Adjusting the focus on human variation. *Trends Genet.* **16**: 296–302.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- ROZAS, J., M. GULLAUD, G. BLANDIN and M. AGUADE, 2001 DNA variation at the *rp49* gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. *Genetics* **158**: 1147–1155.
- RYAN, C. A., 1990 Protease inhibitors in plants: genes for improving defenses against insects and pathogens. *Annu. Rev. Phytopathol.* **28**: 425–449.
- SIMONSEN, K. L., G. A. CHURCHILL and C. F. AQUADRO, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**: 413–429.
- STAHL, E. A., G. DWYER, R. MAURICIO, M. KREITMAN and J. BERGELSON, 1999 Dynamics of disease resistance polymorphism at the *Rpm1* locus of Arabidopsis. *Nature* **400**: 667–671.
- 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**: 585–595.
- TENAILLON, M., M. C. SAWKINS, A. D. LONG, R. L. GAUT, J. F. DOEBLEY *et al.*, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc. Natl. Acad. Sci. USA* **98**: 9161–9166.
- TIFFIN, P., and B. S. GAUT, 2001 Molecular evolution of the wound induced serine protease inhibitor *wip1* in *Zea* and related genera. *Mol. Biol. Evol.* **18**: 2091–2101.
- WANG, G.-L., D.-L. RUAN, W.-Y. SONG, S. SIDERIS, L. CHEN *et al.*, 1998 Xa21D encodes a receptor-like molecule with a leucine rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. *Plant Cell* **10**: 765–779.
- WANG, R. L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 1999 The limits of selection during maize domestication. *Nature* **398**: 236–239.
- WANG, R.-L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 2001 Correction: the limits of selection during maize domestication. *Nature* **410**: 718.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 188–193.
- WHITE, S. E., and J. F. DOEBLEY, 1999 The molecular evolution of terminal *ear1*, a regulatory gene in the genus *Zea*. *Genetics* **153**: 1455–1462.

Communicating editor: A. H. D. BROWN

APPENDIX

Individuals sampled for this study by locus

Locus	Parviglumis		Maize	
	Accession ^{a,b}	Location	Landrace ^b	Location
<i>hm1</i>	PI331783 (1)	Guerrero, Mexico	Conico (1)	Northern Mexico
	PI384070 (2)	Guerrero, Mexico	Araguito (2)	Venezuela
	PI331785 (3)	Michoacan, Mexico	Chococeno (3)	Colombia
	PI331786 (4)	Mexico, Mexico	Enano (4)	Bolivia
	PI384061 (5)	Guerrero, Mexico	Pira (5)	Colombia
	PI384062 (6)	Guerrero, Mexico	Tuxpeno (6)	Puebla, Mexico
	PI384064 (7)	Guerrero, Mexico	Coroico (7)	Bolivia (Amazon Basin)
	MO46 (8)	Jalisco, Mexico	Karapampa (8)	Bolivia (Andean Mts.)
	M063 (9)	Guerrero, Mexico	Chillo (9)	Ecuador
	M106 (10)	Guerrero, Mexico	Confite (10)	Peru
<i>hm2</i>	PI331783 (1)	Guerrero, Mexico	Bolita (1)	Southern Mexico
	PI384070 (2)	Guerrero, Mexico	Cateto Sulino (2)	Uruguay
	PI331785 (3)	Michoacan, Mexico	Assiniboine (3)	Northern United States
	PI331786 (4)	Mexico, Mexico	Chalqueno (4)	Central Mexico
	PI331788 (11)	Michoacan, Mexico	Chapalote (5)	Western Mexico
	PI384061 (5)	Guerrero, Mexico	Conico (6)	Northern Mexico
	PI384062 (6)	Guerrero, Mexico	Cristalino (7)	Chile
	PI384064 (7)	Guerrero, Mexico	Gordo (8)	Northern Mexico
	MO46 (8)	Jalisco, Mexico	Nal-tel (9)	Mexico
	M063 (9)	Guerrero, Mexico	Piscorunto (10)	Peru
<i>waxy</i>	PI384070	Guerrero, Mexico	Chococeno	Colombia
	PI331783	Guerrero, Mexico	Conico	Northern Mexico
	PI331785	Michoacan, Mexico	Coroico	Bolivia (Amazon Basin)
	PI331786	Mexico, Mexico	Karapampa	Bolivia (Andean Mts.)
	PI331787	Mexico, Mexico	Nal-tel	Mexico
	PI384061	Guerrero, Mexico	Chillo	Ecuador
	PI384062	Guerrero, Mexico	Chulpi	Ecuador
	M046	Jalisco, Mexico	Confite Puneno	Peru
	M106	Guerrero, Mexico	Enano	Bolivia
			Montana	Colombia
		Morado	Bolivia	
		Pira	Colombia	
		Araguito	Venezuela	
<i>c1</i>	PI384061	Guerrero, Mexico	—	—
	PI331783 ^c	Guerrero, Mexico		
	PI331788	Michoacan, Mexico		
	M106	Guerrero, Mexico		
	M063	Guerrero, Mexico		
	PI331783 ^c	Guerrero, Mexico		

Additional sequences for analysis were taken from GenBank, as described in MATERIALS AND METHODS.

^a Accession numbers preceded by PI were provided by the USDA germplasm center, Ames, Iowa.

^b Numbers in parentheses refer to the individuals listed in Figures 2 and 3.

^c Two alleles were sequenced from this heterozygous individual.