Maydica 54 (2009): 375-386

MAIZE AMINO ACID PATHWAYS MAINTAIN HIGH LEVELS OF GENETIC DIVERSITY

S.A. Flint-Garcia[1,2,*], K.E. Guill[1], H. Sanchez-Villeda[2,3], S.G. Schroeder[2,4], M.D. McMullen[1,2]

[1] *US Department of Agriculture - Agricultural Research Service, Plant Genetics Research Unit, Columbia, MO 65211, USA*

[2] *Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA*

Received July 18, 2009

ABSTRACT - The methods commonly used to determine the genetic bases of agronomic traits rely on genetic variation to detect marker-phenotype associations. If variation at key loci is reduced due to selection during domestication, the analyses may fail to identify these genes as important loci in controlling trait expression. It is of practical importance to determine the level of genetic diversity for key genes controlling agronomic traits. Results from recent large scale selection screens in maize have revealed that a number of genes involved in amino acid synthesis exhibit signs of selection. Here we asked whether other genes in amino acid pathways were targets of selection, and to what extent amino acid metabolism has experienced artificial selection. We surveyed amino acid pathways by sequencing 15 candidate genes involved in amino acid metabolism maize inbreds and teosinte, and conducted tests of selection. We demonstrated that five of 15 additional genes involved in amino acid metabolism exhibit weak evidence for selection. Amino acid synthesis pathways, as a whole, do not show evidence of greatly reduced genetic diversity. Our results suggest that adequate genetic diversity remains for improvement of maize nutritional quality, and that limited genetic diversity may be impacting only a few specific amino acids.

KEY WORDS: Artificial selection; Domestication; Kernel composition; Teosinte.

INTRODUCTION

Maize (*Zea mays* ssp. *mays*) is the most valuable crop in the US and one of the most important crops world-wide. Pedigree analysis indicates that the majority of the inbred lines that constitute modern maize hybrids were derived from only a small subset of the available maize germplasm (GOODMAN, 1990). Many corn breeders and geneticists are concerned that recurrent recycling of breeding germplasm may diminish future breeding progress. Should we reach a yield plateau, we will need readily available diverse germplasm to resume yield increases. Equally important is the need for novel alleles for stress (e.g. disease, insect, and abiotic) resistance and improved quality traits. Therefore, it is of practical importance to determine the level of genetic diversity present in maize germplasm for key genes controlling agronomic traits.

Genetic methods including quantitative trait locus (QTL) analysis and association analysis have been used to determine the genes underlying variation for important agronomic traits in maize. Both methods require variation in the phenotype and marker data in order for phenotype-marker associations to be identified. If variation at key loci is limited, however, due to selection during domestication and/or crop improvement, the analyses would fail to identify these genes as important loci in controlling trait expression (YAMASAKI *et al.*, 2005). Therefore, the identification of genes with low diversity in modern crop varieties relative to their progenitors may aid in the discovery of genes important to crop improvement (WRIGHT *et al.*, 2005; HYTEN *et al.*, 2006; TANG *et al.*, 2006).

Maize was domesticated from teosinte (*Zea mays* ssp. *parviglumis*) in southern Mexico about 7500 years ago (ILTIS, 1983; DOEBLEY, 2004). It is believed that a founding population of teosinte individuals was isolated from the progenitor population via human selection to form an ancestral maize population. This ancestral maize population underwent a population bottleneck, resulting in genome-wide loss of diversity (WRIGHT *et al.*, 2005). Genetic varia-

[3] Current address: Genetic Resources and Enhancement Unit, CIMMYT, Mexico, DF 06600, Mexico.

[4] Current address: US Department of Agriculture - Agricultural Research Unit, Beltsville, MD USA.

* For correspondence (fax: +1 573-884-7850; e.mail: Sherry.Flint-Garcia@ars.usda.gov).

tion was further reduced in genes contributing to traits of human interest, i.e. in selected genes (WRIGHT *et al.*, 2005).

Over the last two decades, only a small number of studies have identified genes that have been targets of selection in maize. Many of these genes are responsible for the gross morphological differences between teosinte and maize such as *teosinte branched1* and *barren stalk1* involved in lateral branch inflorescence architecture (WANG *et al.*, 1999; GALLAVOTTI *et al.*, 2004), and *teosinte glume architecture1* involved in the stony fruitcase (the hardened glumes surrounding the kernel) that is characteristic of the teosinte kernel (WANG *et al.*, 2005). Other selected genes include: *colorless1* (HANSON *et al.*, 1996), *zagl1* (VIGOUROUX *et al.*, 2002), *tassel seed2*, and *dwarf8* (TENAILLON *et al.*, 2001). WHITT *et al.* (2002) found the maize starch pathway has been a target of selection, where three of six genes examined showed strong evidence of selection. This study strongly supports the hypothesis that the maize kernel was a primary target of selection during domestication. Results from recent large scale selection studies (WRIGHT *et al.*, 2005; YAMASAKI *et al.*, 2005) have revealed that a number of genes involved in amino acid synthesis exhibit signs of selection: *cysteine synthase*, *dihydrodipicolinate synthase*, *chorismate mutase*, and *hexokinase*.

We have recently demonstrated that the profiles of zein proteins, the predominant class of seed storage proteins, have been modified during the domestication process (FLINT-GARCIA *et al.*, 2009). Several studies suggest that progenitors of maize contain diversity in zein genes that is lacking in modern inbreds (WILSON and LARKINS, 1984; SWARUP *et al.*, 1995). In this study, we asked whether other genes in amino acid pathways were targets of selection, and to what extent has amino acid metabolism in maize experienced artificial selection. To investigate these questions, we sequenced 15 candidate genes encoding structural enzymes in amino acid metabolism in diverse samples of maize inbred lines and teosinte accessions. We then conducted multiple tests of selection to identify targets of selection.

MATERIALS AND METHODS

Phenotypic analysis

We analyzed the total kernel amino acid content of kernels from 11 teosinte accessions, 16 landrace accessions, and 26 inbred lines of maize. We obtained the following 11 geographically diverse teosinte accessions from the North Central Regional Plant Introduction Station (NCRPIS): PI 384063, PI 384065, PI 384066, PI 384071, Ames 21889, Ames 21785, Ames 21786, Ames 21789, Ames 21809, Ames 21812, and Ames 21814. We obtained the following 16 landrace accessions from NCRPIS or M. Goodman at North Carolina State University: Bolita (OAX68), Cateto Sulino (URG II), Chalqueno (MEX48), Chapalote (SIN2), Conico (PUE32), Costeno (VEN453), Cristalino Norteno (CHI349), Dzit Bacal (GUA131), Gordo (CHH160), Guirua (MAG450), Nal-tel (YUC7), Pisccorunto (APC13), Sabanero (SAN329), Serrano (GUA14), Tuson (CUB57), and Zapalote Chico (OAX70). The 26 maize inbred lines included B73, Mo17, and the 25 diverse parents of the Nested Association Mapping (NAM) population (YU *et al.*, 2008) except that Tzi8 was omitted: B97, CML103, CML228, CML247, CML277, CML322, CML333, CML52, CML69, Hp301, IL14H, Ki11, Ki3, Ky21, M162W, M37W, Mo17, Mo18W, MS71, NC350, NC358, Oh43, Oh7B, P39, and Tx303.

Two replicates were planted in Puerto Rico in winter 2005-2006. Entries were randomized within groups (inbred lines, landraces, and teosintes) and groups were randomized within replicates. All entries were allowed to open pollinate in order to obtain an adequate amount of seed for analysis. Ears were harvested, and balanced bulks of seed were created for each plot.

For the teosinte samples, the stony casings were removed, and the kernels were ground into a fine meal. Whole kernels of the inbred lines and landrace accessions were similarly ground into a fine meal. The samples were submitted to the University of Missouri Experiment Station Chemical Laboratories for analysis. Complete amino acid profiles were determined following the Official Methods of Analysis of AOAC International by method 982.30E (AOAC, 1995).

Analysis of variance was conducted using SAS PROC MIXED with entries fixed and replicates random. The phenotypic variance of entries was partitioned into "among groups", and specific contrasts of teosinte vs. landraces, teosinte vs. inbreds, and landraces vs. inbreds were tested.

Sequence analysis

Three sets of diverse germplasm were used for sequence analysis: 28 maize inbred lines, 16 partially-inbred teosinte plants, and two *Tripsacum dactyloides* accessions. The twenty-eight maize inbred lines included the 25 diverse parents of NAM (described above), B73, Mo17, and W22 R-r:standard. The 16 teosinte accessions were selected based on geographical distribution of *Z. mays* ssp *parviglumis* and were self-pollinated two to three generations, yielding partially-inbred lines: TIL-1 (JSGyLOS 130), TIL-2 (JSGyLOS 119), TIL-3 (JSGyMAS 401), TIL-4 (CIMMYT 8783), TIL-5 (JSG 197), TIL-6 (JSGyLOS 109), TIL-7 (JSG 378), TIL-8 (JSG 374), TIL-9 (JSGyLOS 161), TIL-10 (CIMMYT 11355), TIL-11 (USDA PI566686), TIL-12 (JSGyMAS 264), TIL-14 (Benz 967), TIL-15 (Kato Site 4), TIL-16 (Beadle & Kato Site 4), and TIL-17 (Wilkes Site 6). The teosinte partial-inbreds were derived by John Doebley at the University of Wisconsin from collections that are generally not publicly available (exceptions of TIL-11, TIL-14, and TIL-16 which were derived from accessions PI 566686, Ames 21889, and PI 384062, respectively). Currently, 11 of the 16 partial inbreds are available from NCRPIS. Only TIL-14 is directly related to the teosinte accessions phenotyped in this study. Standard protocols were used to extract DNA (SAGHAI-MAROOF *et al.*, 1984).

We chose 15 candidate genes for sequence analysis based on their positions at pathway branch-points or their regulatory roles in amino acid synthesis and nitrogen metabolism (Table 1;

TABLE 1 - *Amino acid candidate genes examined in this study.*

Gene name	Original GenBank/ TIGR/MAGI accession	Alignment ID [a]	Gene function
chorismate mutase	AY107907 [b]	PZA00406	Tyrosine and phenylalanine synthesis
cysteine synthase	AY108388 [b]	PZA01782	Cysteine synthesis
dihydrodipicolinate synthase	AY107228 [b]	PZA00514	Lysine Synthesis
hexokinase	AY106111 [b]	PZA00093	Proposed carbon:nitrogen sensing
2-isopropylmalate synthase	MAGI_34045	PZA03249	Leucine synthesis
acetohydroxy acid synthase	X63554	PZA03231	Branched chain amino acid synthesis
anthranilate synthase β	TC249514	PZA03233	Aromatic amino acid synthesis
asparagine synthetase	X82849	PZA03253	Asparagine synthesis
aspartate amino transferase	TC269156	PZA03255	Aspartate-derived amino acid synthesis
aspartate kinase - homoserine dehydrogenase			Methionine, lysine, isoleucine and threonine synthesis
akh1 - transit peptide & AK domain	L33912	PZA03254	
akh1 - AK domain	APTA52033	PZA03705	
akh1 - HSDH domain	MAGI4_124888	PZA03706	
akh2 - AK domain	MAGI4_68123	PZA03707	
akh2 - HSDH domain	MAGI4_68157	PZA03708	
cystathionine γ-synthase	AF007786	PZA03235	Methionine degradation
glutamate dehydrogenase	D49475	PZA03256	Glutamate-derived amino acid synthesis
phenylalanine ammonia lyase	L77912	PZA03239	Phenylpropanoid pathway from phenylalanine
proline dehydrogenase	TC251771	PZA03240	Proline degradation
s-adenosylmethionine (sam) synthetase I	TC268733	PZA03241	Methionine degradation
s-adenosylmethionine (sam) synthetase II	TC268809	PZA03242	Methionine degradation
sucrose-nonfermenting1 (snf1)-related protein kinase	AY486125	PZA03243	Catabolic repressor involved in nutritional stress
threonine deaminase	TC260741	PZA03244	Isoleucine synthesis
tryptophan synthase β-1	M76684	PZA03247	Tryptophan synthesis

[a] Sequence alignments are available from www.panzea.org.
[b] Original GenBank unigene sequence from WRIGHT *et al.* (2005) and YAMASAKI *et al.* (2005).

CORUZZI and LAST, 2000). Several sequence databases including the TIGR maize gene index (http://www.tigr.org) and NCBI (http://www.ncbi.nlm.nih.gov) were searched by BLASTn to obtain maize sequence for each candidate gene. No published maize sequence was available for *2-isopropylmalate synthase*. In this case, the rice sequence (TIGR rice gene index TC249453) was used to conduct a BLASTn search against the Maize Assembled Genomic Island (MAGI) database (version 3.1; http://www.plantgenomics.iastate.edu/maize), resulting in an alignment of 447 bp of maize genomic sequence that was 86% identical to the rice sequence.

Polymerase chain reaction (PCR) primers were designed to amplify approximately 600 bp of each candidate gene using the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Primers were targeted to a single region in 13 of the 15 candidate genes. For the two paralogs of *aspartate kinase-homoserine dehydrogenase* (*akh1* and *akh2*), multiple regions were targeted. Each region was PCR amplified for all maize inbred, teosinte, and tripsacum accessions. Duplicate samples of Mo17 and B73 were included to monitor sequence quality. The PCR was conducted with PCR 2X Master Mix Taq (Promega, Madison, WI) as previously described (YAMASAKI *et al.*, 2005).

The PCR products were sequenced with both forward and reverse primers using BigDye version 3.1 terminators (Applied Biosystems, Foster City, CA) and analyzed on ABI 3100 sequencers (Applied Biosystems). PHRED software (EWING and GREEN, 1998; EWING *et al.*, 1998) was used to call bases, assign quality scores, and trim trace files, and sequence reads from the same amplicon were assembled using PHRAP. Sequences from each accession were aligned with CLUSTAL W (THOMPSON *et al.*, 1994) and the alignments were manually edited using DNA Align Editor (SANCHEZ-VILLEDA *et al.*, 2008). Rare single nucleotide polymorphisms (SNPs) and insertion/deletions (indels) were visually confirmed by examining trace files. When a heterozygote was encountered, one allele was randomly chosen. A master alignment containing all three species was used to create sub-alignments containing (1) maize and tripsacum, (2) teosinte and tripsacum, and (3) maize and teosinte. Duplicate alignments were created to assess the diversity contributed by indel polymorphisms. In the duplicated alignment, indel haplotypes were converted to an arbitrary nucleotide designation (A, T, C, or G) to allow their inclusion into subsequent diversity/selection analyses with each indel allele considered equivalent to a SNP polymorphism. Sequences are available from GenBank (accession numbers BV680562-BV681145 and BV693818-BV693972), and sequence alignments are available from www.panzea.org.

Selection and diversity

The maize-tripsacum and teosinte-tripsacum alignments were used to obtain the average number of pair-wise differences per nucleotide site (π; TAJIMA, 1983) and the number of segregating sites (S) for both maize and teosinte, as well as divergence estimates between maize and tripsacum, and teosinte and tripsacum using DnaSP (ROZAS *et al.*, 2003). DnaSP was also used to calculate the Tajima's D test statistic (TAJIMA, 1989). In order to determine an empirical significance threshold for the diversity ratio (π_{maize}:$\pi_{teosinte}$), we calculated the mean of the π ratios for ten genes reported as selected in the literature (*tb1*, CLARK *et al.*, 2004; ba1, GALLAVOTTI *et al.*, 2004; c1, HANSON *et al.*, 1996; ts2, TENAILLON *et al.*, 2001; d8, THORNSBERRY *et al.*, 2001; zagl1, VIGOUROUX *et al.*, 2002; tga1, WANG *et al.*, 2005; ae1, bt2, and su1, WHITT *et al.*, 2002).

The Hudson-Kreitman-Aguadé (HKA) test of selection (HUDSON *et al.*, 1987) was conducted for each candidate gene versus 15 documented neutral genes (EYRE-WALKER *et al.*, 1998; HILTON and GAUT, 1998; TENAILLON *et al.*, 2001; HUFFORD *et al.*, 2007). In this test, diversity within maize is compared to the divergence between maize and tripsacum, the outgroup species, for the candidate gene and the 15 neutral genes simultaneously using the HKA software package (http://lifesci.rutgers.edu/~heylab). We employed the test of maximum cell value to determine the significance level of the HKA test (WANG and HEY, 1996). The same comparison was made between teosinte and tripsacum. A significant HKA test for maize versus tripsacum and a non-significant test for teosinte versus tripsacum constituted an overall significant test of selection in maize. When HKA tests were significant for teosinte versus tripsacum, the gene was considered selected in teosinte and could not be tested for selection in maize.

The maize-teosinte alignment subset was used to determine the population recombination parameter (ρ; HUDSON, 1987) and Watterson's estimator of the population mutation parameter (θ; WATTERSON, 1975) for teosinte using DnaSP. When DnaSP was unable to calculate ρ due to low polymorphism levels within the alignment, ρ was conservatively assigned a zero value. The teosinte estimates ρ and θ were used in coalescent simulations to generate 5000 sequence alignments (INNAN and KIM, 2004) simulated to have gone through the bottleneck that maize experienced (WRIGHT *et al.*, 2005). The following parameters were used in the simulations: 940,000, 6850, and 1,000,000 for the size of the ancestral, bottleneck, and present populations, respectively; and 2800 years as the duration of the bottleneck that began 7500 years ago. The primary parameter of interest is the bottleneck severity (k) which is the ratio of the bottleneck population size to the duration of the bottleneck in generations. The estimate of $k = 2.45$ is robust to variation in model parameters such as time since domestication, and population sizes before and after the domestication event (WRIGHT *et al.*, 2005). The number of segregating sites in the observed maize alignment was compared to the null distribution of S from the simulations to determine the probability that the diversity remaining in the maize lines could be explained by general bottleneck effects.

RESULTS

Amino acid profiles

We analyzed the total kernel amino acid content of maize inbred lines, teosinte accessions, and maize landraces, the historical intermediate between inbred lines and teosinte. When amino acid content was examined on a kernel weight basis, all amino

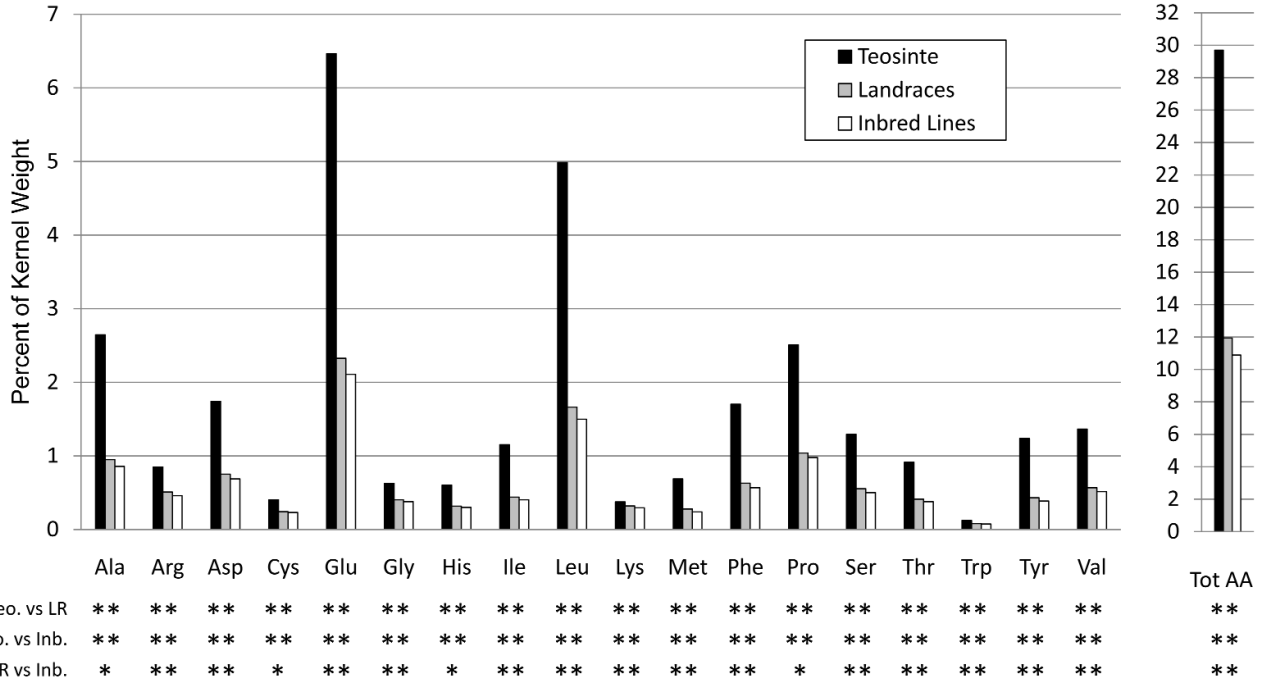

FIGURE 1 - Amino acid content of maize inbred lines (N=26), landraces (N=16), and teosinte (N=11). **, * = germplasm group means were significantly different at $P < 0.01$ and $P < 0.05$, respectively; ns = not significant at $P = 0.05$.

acids were present in significantly higher quantities in teosinte than landraces and inbreds (Fig. 1). Some amino acids such as glutamic acid, leucine, phenylalanine, tyrosine, and valine were nearly three times as abundant in teosinte as the other groups. Of the nutritionally limiting amino acids in maize, methionine was far more abundant in teosinte than landraces or inbreds. Lysine and tryptophan levels were statistically higher in teosinte, although differences of these magnitudes aren't sufficient to have a major impact on human and animal nutrition. The difference in total kernel amino acid content between teosinte and both landraces and inbred lines was highly significant, with teosinte kernels containing approximately twice the total kernel amino acid content of either the landraces or inbred lines. This is consistent with results of prior studies where teosintes contained twice the protein as maize (WANG *et al.*, 2008; FLINT-GARCIA *et al.*, 2009).

Sequence analysis

Eighteen sequence alignments from 15 candidate genes were created from up to 28 maize inbred lines, 16 teosinte inbred plants, and two Tripsacum accessions. The mean alignment length was 511 bp and ranged from 352 bp to 811 bp (Table 2). The minimum number of inbred lines and teosinte accessions in the alignments was 17 and 6, respectively, with means of 25 and 13, respectively. Previous studies have demonstrated that these sample sizes are sufficient to determine whether a gene has been selected (WRIGHT *et al.*, 2005; YAMASAKI *et al.*, 2005).

We analyzed two parallel alignments for each candidate gene: one with the unedited indel polymorphisms left in the alignment (thus excluded from further analyses by DnaSP which only considers alignment positions with a nucleotide in every sample), and one with the indel haplotypes converted to nucleotide polymorphisms. Conversion of the indels into nucleotide polymorphisms allowed their

TABLE 2 - *Population statistics and selection results.*

	Inbred Lines				Teosinte				π_{Maize} : $\pi_{Teosinte}$	Tajima's D	HKA test	CS [d]
	N [a]	L [b]	S [c]	π	N	L	S	π				
SELECTED GENES [e]												
chorismate mutase	11	222	0	0.0000	12	215	8	0.0120	0.000		– [f]	***
cysteine synthase	14	441	0	0.0000	16	476	22	0.0084	0.000		t.s.[g]	***
dihydrodipicolinate synthase	13	331	0	0.0000	15	319	7	0.0040	0.000		– [f]	***
hexokinase	11	222	0	0.0000	12	215	8	0.0120	0.000		– [f]	***
NEUTRAL GENES [h]												
adh1	14	1217	42	0.0124	12	1386	88	0.0193	0.642			ns
bz2	14	519	11	0.0092	14	544	18	0.0069	1.333			ns
fus6	14	237	7	0.0073	10	226	10	0.0125	0.584			ns
CL10432	12	650	13	0.0087	16	616	18	0.0087	0.991		ns	ns
CL11572	12	649	17	0.0098	12	591	20	0.0121	0.816		ns	ns
CL26230	13	548	8	0.0046	11	544	6	0.0027	1.675		ns	ns
CL38764	13	634	21	0.0130	10	648	28	0.0160	0.812		ns	ns
CL40819	14	551	14	0.0073	13	555	26	0.0113	0.644		ns	ns
CL5367	9	643	5	0.0034	13	617	9	0.0032	1.040		ns	ns
CL6312	10	656	18	0.0095	14	671	32	0.0154	0.617		ns	ns
PCO061840	12	543	4	0.0024	15	502	11	0.0061	0.400		ns	ns
PCO095943	8	571	13	0.0093	10	666	28	0.0136	0.685		ns	ns
PCO103663	14	528	21	0.0120	15	541	33	0.0153	0.785		ns	ns
PCO117716	11	673	12	0.0059	5	681	14	0.0091	0.647		ns	ns
PCO136491	11	650	12	0.0068	16	599	25	0.0101	0.676		ns	ns
GENES IN CURRENT STUDY												
2-isopropylmalate synthase [i]	28	420	2	0.0003	16	420	7	0.0035	0.086	ns	ns	ns
	–	–	–	–	–	–	–	–	–	–	–	–

TABLE 2 - *Continued.*

	Inbred Lines				Teosinte				π_{Maize} : $\pi_{Teosinte}$	Tajima's D	HKA test	CS [d]
	N [a]	L [b]	S [c]	π	N	L	S	π				
acetohydroxy acid synthase	27	499	3	0.0015	16	499	6	0.0029	0.517	ns	ns	ns
	–	–	–	–	–	–	–	–	–	–	–	–
anthranilate synthase	23	688	22	0.0079	10	688	30	0.0145	0.545	ns	ns	ns
	23	678	25	0.0094	10	680	35	0.0169	0.556	ns	ns	ns
asparagine synthetase	27	542	12	0.0074	16	543	17	0.0072	1.028	ns	ns	ns
	27	545	13	0.0081	16	545	18	0.0077	1.052	ns	ns	ns
aspartate amino transferase	26	502	7	0.0051	16	501	16	0.0068	0.750	ns	ns	ns
	26	504	9	0.0064	16	504	22	0.0104	0.615	ns	ns	ns
akh1 - transit peptide & AK domain	27	505	6	0.0021	15	498	11	0.0037	0.568	ns	ns	ns
	27	508	7	0.0024	15	501	13	0.0054	0.444	ns	ns	ns
akh1 - AK domain	27	449	3	0.0010	13	448	12	0.0050	0.191	ns	ns	**
	27	450	5	0.0016	13	451	15	0.0068	0.233	ns	ns	**
akh1 - HSDH domain	28	412	8	0.0060	16	412	13	0.0057	1.040	ns	ns	ns
	28	416	8	0.0059	16	416	14	0.0067	0.882	ns	ns	ns
akh2 - AK domain	23	352	8	0.0075	12	352	13	0.0096	0.781	ns	– [f]	ns
	23	355	11	0.0098	12	355	16	0.0129	0.759	ns	– [f]	ns
akh2 - HSDH domain	21	487	10	0.0046	11	487	12	0.0101	0.452	ns	ns	ns
	21	484	12	0.0079	11	484	16	0.0137	0.577	ns	n	ns
cystathionine -synthase	27	477	7	0.0044	6	476	8	0.0060	0.733	ns	ns	ns
	27	476	8	0.0046	6	476	9	0.0067	0.687	ns	ns	ns
glutamate dehydrogenase	24	423	8	0.0041	13	418	10	0.0064	0.641	ns	*	ns
	24	427	12	0.0076	13	422	14	0.0120	0.633	ns	**	ns
phenylalanine ammonia lyase	27	498	5	0.0024	16	498	5	0.0029	0.828	ns	ns	ns
	–	–	–	–	–	–	–	–	–	–	–	–
proline dehydrogenase	27	519	14	0.0094	14	516	34	0.0201	0.468	ns	*	*
	27	525	20	0.0134	14	528	46	0.0293	0.457	ns	**	*
sam synthetaseI	24	433	7	0.0045	11	433	6	0.0046	0.978	ns	ns	ns
	–	–	–	–	–	–	–	–	–	–	–	–
sam synthetaseII	28	501	5	0.0032	16	501	14	0.0072	0.444	ns	– [f]	*
	–	–	–	–	–	–	–	–	–	–	–	–
threonine deaminase	17	804	11	0.0025	7	803	8	0.0042	0.595	ns	ns	ns
	17	811	15	0.0037	7	807	12	0.0058	0.638	ns	ns	ns
tryptophan synthase β-1	27	591	5	0.0019	16	589	10	0.0028	0.679	ns	ns	ns
	27	546	5	0.0021	16	546	12	0.0037	0.568	ns	ns	ns
MEAN (CURRENT STUDY)	25	511	9	0.0051	13	510	15	0.0084	0.627			

[a] Number of individuals sampled.
[b] Length, including gaps.
[c] Number of segregating sites.
[d] Significance of coalescent simulations after 1000 simulations.
[e] Published previously in YAMASAKI *et al.* (2005) and WRIGHT *et al.* (2005). These data are not included in the mean calculations.
[f] HKA analysis not conducted because no tripsacum sequence was available.
[g] t.s. = The HKA test was significant for teosinte versus tripsacum. The gene was considered selected in teosinte and could not be tested for selection in maize.
[h] Neutral genes published previously in EYRE-WALKER *et al.* (1998), TENAILLON *et al.* (2001), and HUFFORD *et al.* (2007).
[i] For each gene analyzed in the current study, the top line represents the unedited alignment (i.e. indel polymorphisms were excluded from analysis) and the bottom line represents the edited alignment with indel polymorphisms converted to nucleotide polymorphisms to allow their inclusion into selection analyses.
*, **, ***: significant at $P = 0.05$, 0.01, and 0.001, respectively.

inclusion into analyses of diversity by population genetics software. For five of our candidate genes, there were no indels; thus only the original alignment was analyzed. The results were generally the same regardless of which alignment was used.

There was tremendous variation in *S* across the candidate genes (Table 2). The number of segregating sites ranged from 2 to 25 (mean = 9) in maize, and ranged from 5 to 46 (mean = 15) in teosinte. In total, we identified 143 SNPs among the inbred lines and 132 SNPs among the teosinte accessions. Diversity, as measured by the average number of nucleotide differences, π, ranged from 0.0003 to 0.0134 (mean = 0.0051) for maize, and ranged from 0.0028 to 0.0293 (mean = 0.0084) for teosinte. For comparison sake, the average diversity (π) of a set of more than 700 randomly chosen genes was 0.0065 in maize inbred lines and 0.0098 in teosinte accessions (WRIGHT *et al.*, 2005).

Tests of selection

We conducted four statistical tests to determine whether our candidate genes were targets of selection: relative loss in diversity from teosinte to inbred lines, Tajima's D, the HKA test, and coalescent simulations.

Although selection results in a loss of diversity, absolute diversity levels vary across genes depending on the functional constraints placed on polymorphism in coding regions. We examined the diversity in maize (π_{maize}) relative to teosinte ($\pi_{teosinte}$) with the expectation that a low π-ratio reflects artificial selection during domestication or crop improvement. We determined an empirical significance threshold of 0.16 based on the mean of the π-ratios of previously reported selected genes (π-ratios *ae1* = 0.161; *ba1* = 0.030; *bt2* = 0.256; *c1* = 0.267; *d8* = 0.253; *su1* = 0.278; *tb1* 5' region = 0.009; *tga1* promoter = 0.050; *ts1* = 0.344; *zagl1* = 0.000). The π-ratios in the set of 15 candidate genes ranged from 0.086 to 1.052. Only *2-isopropylmalate synthase* had a π-ratio below the empirical threshold (Table 2). However, this may have been an artifact of the very low levels of nucleotide diversity for this gene (π_{maize} = 0.0003 and $\pi_{teosinte}$ = 0.0035). It should be noted that this empirical threshold is highly conservative, as many of the selected genes reported to date are estimated to be among the most strongly selected genes in maize. In contrast, the π-ratio for neutral genes is typically around 0.75 (EYRE-WALKER *et al.*, 1998; ZHANG *et al.*, 2002).

The Tajima's *D* test determines whether mutations are selectively neutral by comparing the π and *S* sequence diversity metrics (TAJIMA, 1989). In the case of a significant Tajima's *D* test, the frequency distribution of polymorphisms would be shifted toward higher frequency alleles as increased rates of genetic drift during a bottleneck would tend to preferentially remove rare alleles from the population. The Tajima's *D* test was not significant for any of the sequence alignments analyzed in this study (Table 2). Thus based on polymorphism distribution data, none of the 15 candidate genes were targets of selection.

The HKA test (HUDSON *et al.*, 1987) is based on the neutral theory of molecular evolution (KIMURA, 1983) such that a neutral gene that evolves rapidly between species will also display high levels of intraspecific sequence variation. However, selection affects inter- and intra-specific variability differently, thus violating neutral theory. The comparison of the divergence between two species (interspecific variability) and polymorphism within a species (intraspecific variability) will indicate a history of selection. We conducted HKA tests in teosinte and maize for our candidate genes using fifteen neutral genes and tripsacum as an outgroup. We were unable to obtain a tripsacum sequence for *sam synthetase II* and the AK domain of *akh2*, and consequently were unable to conduct the HKA test for these genes. A gene was considered selected if the HKA test was non-significant for teosinte (neutral) and significant for maize (non-neutral). In this study, *glutamate dehydrogenase* and *proline dehydrogenase* both had significant HKA tests and selection (Table 2).

Finally, we performed coalescent simulations using the demographics of the maize domestication bottleneck (WRIGHT *et al.*, 2005). For each candidate gene region, we generated 5000 random maize sequence alignments based on teosinte population metrics (population recombination parameter, ρ; and diversity, θ), and generated a null distribution of *S* from the simulations to determine the significance of *S* from the observed maize alignment. Using this approach, we determined that the AK domain of *akh1*, *proline dehydrogenase*, and *sam synthase II* had significantly ($P \leq 0.05$) fewer segregating sites than teosinte, consistent with a history of selection for these loci (Table 2).

Selection and paralogous sequences

In order to investigate the impact of selection on paralogous gene regions and on multiple domains

within the same gene, we targeted multiple regions of the two paralogs of *aspartate kinase-homoserine dehydrogenase* (*akh1* and *akh2*). One region in each of the aspartate kinase (AK) and homoserine dehydrogenase (HSDH) domains was sequenced in *akh1* and *akh2*, and a region of the transit peptide-AK domain boundary was sequenced in *akh1*. The π-ratio varied greatly between the two enzymatic domains of *akh1* (Table 2), demonstrating that selection on the AK domain (as supported by coalescent simulations and a low π-ratio) had little effect on the diversity levels of the HSDH domain. This result is supported by the rapid decay of linkage disequilibrium in teosinte populations (REMINGTON *et al.*, 2001; WEBER *et al.*, 2007). Likewise, a comparison of the AK domain from *akh1* and *akh2* showed differential selection during maize domestication/improvement.

DISCUSSION

Like most other cereals, maize is deficient in the essential amino acids tryptophan, lysine, and methionine. This is because these amino acids are rare in zeins, the primary class of storage proteins. Attempts have been made to increase lysine and tryptophan content in maize with the *opaque2* (*o2*) mutant. The *o2* gene alters gene expression of zein proteins (SCHMIDT *et al.*, 1990). Several studies suggest that progenitors of maize contain diversity in zein genes that is lacking in modern inbreds (WILSON and LARKINS, 1984; SWARUP *et al.*, 1995). Recently, we have demonstrated that zein profiles have been modified during the domestication process (FLINT-GARCIA *et al.*, 2009). Can such changes in zein profiles occur without selection in amino acid genes?

Amino acid content of diverse Zea

At the heart of this study is the question of whether selection in the amino acid pathway is associated with amino acid content in maize and teosinte. Previously, a single study reported amino acid composition of teosinte based on a single teosinte sample (BRESSANI and MERTZ, 1958). We found that teosinte contained twice as much total amino acid (i.e. total kernel protein) as landraces and inbred lines, and therefore had significantly higher concentrations of all amino acids than the other germplasm groups. However, examination of amino acid profiles that were adjusted for protein content revealed that these high levels of amino acids are primarily driven by overall higher protein content.

An important concept in improving kernel quality is amino acid imbalance, where one or more amino acids are limiting in a food or feedstock. In maize, the limiting amino acids are lysine, tryptophan, and methionine. Both human and animal diets based heavily on maize will manifest several consequences including protein-energy malnutrition, higher costs as amino acid supplements are added, and increased nitrogenous waste as unutilized nitrogen is excreted (DARRIGUES *et al.*, 2006). Maize has slightly higher levels of lysine, tryptophan, and methionine, and therefore a better amino acid balance than teosinte, despite having greatly decreased kernel protein content as compared to teosinte. It is possible that selection, direct or indirect, has improved the amino acid balance of maize compared to its wild ancestor.

Tests of selection for amino acid genes

We chose candidates based on their positions at pathway branch-points or their regulatory roles in amino acid synthesis and nitrogen metabolism (Fig. 2). Many enzymes involved in amino acid pathways occur in small gene families and, preferably, all members would be resequenced in this study. However, due to a lack of genomic sequence for these gene families in maize, a single representative was chosen for our survey of the pathway for most candidates. In order to investigate selection in paralogous sequences and multiple regions of the same gene, we resequenced regions of two paralogs of *aspartate kinase-homoserine dehydrogenase* (MUEHLBAUER *et al.*, 1994). Expression of each candidate gene in developing ear tissue was bioinformatically demonstrated by conducting BLASTn searches against non-normalized EST databases, using an expect value (e-value) of e-50 (HUFFORD *et al.*, 2007). The resulting hits were examined for descriptions of the tissue used to create the EST library, with the keywords “endosperm”, “embryo”, “kernel”, and “ear” providing evidence for expression in developing ears (data not shown).

Interestingly, no candidate gene was classified as selected by all four selection tests. Three genes - the aspartate domain of *akh1*, *proline dehydrogenase*, and *sam synthase II* - were found to be selected by coalescent simulation, but diversity was only marginally reduced. *Proline dehydrogenase* was the most diverse gene in teosinte; therefore, even

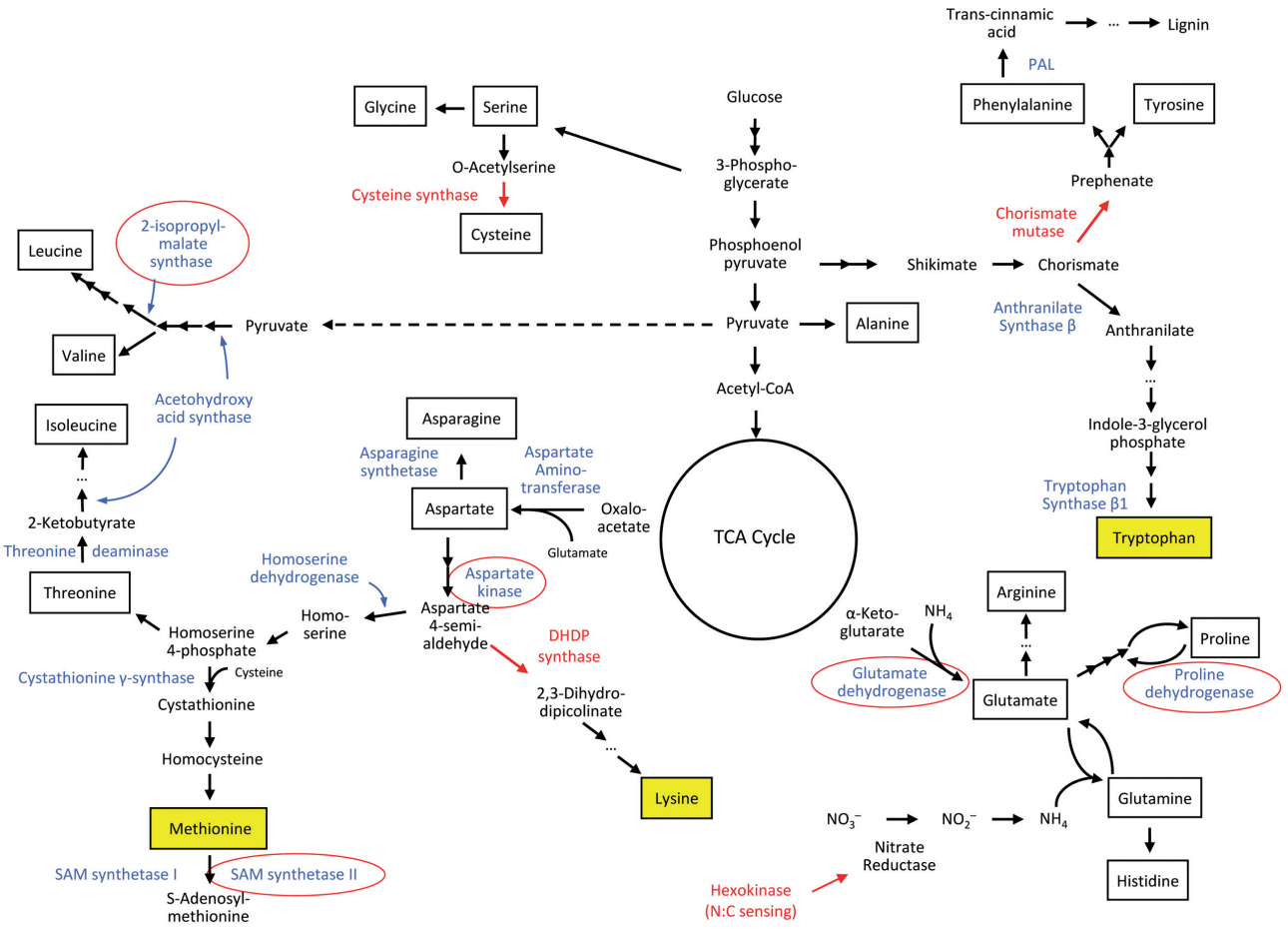

FIGURE 2 - Generalized amino acid pathway indicating the positions of the candidate genes. Amino acids are in boxes, and the amino acids particularly deficient in maize are in yellow boxes. Genes exhibiting signs of selection in randomly sequenced genes (WRIGHT *et al.*, 2005; YAMASAKI *et al.*, 2005) are colored in red, genes chosen for candidate sequencing in this study are colored in blue, and genes showing signs of selection in this study are circled in red. Figure adapted from CORUZZI and LAST (2000).

though the loss of segregating sites in this gene was significantly greater than expected by the bottleneck alone (significant coalescent simulation result), substantial diversity remains in this gene in maize. The opposite situation occurred for *2-isopropylmalate synthase*. Although the π-ratio suggests that this gene may be selected, the gene has low diversity in teosinte ($S = 7$); therefore, one has limited power to test this gene for selection as all diversity may have been lost by the bottleneck effects alone. Each of the selection tests focuses on a different demographic aspect of selection; therefore, conflicting assessments should be viewed as suggestive of selection, and that further experimentation such as extended sequencing is required to confirm or refute selection status.

Effect of selection on paralogous sequences

In order to investigate the impact of selection on paralogous gene regions, and on multiple domains within the same gene, we targeted multiple regions of two *aspartate kinase - homoserine dehydrogenase* paralogs for sequencing and tests of selection. We chose the bifunctional polypeptide AKHSDH as the subject of this analysis primarily because two paralogs have been fully sequenced in maize and the domains have clearly defined functions in the pathway. Secondly, AKHSDH has been intensively studied due to its role as a major regulatory step of the aspartate-derived pathway, where its activity is sensitive to feedback inhibition by the end products threonine and lysine.

The AKHSDH enzyme is composed of three do-

mains: a transit peptide directing localization to plastids, and the aspartate kinase (AK) and homoserine dehydrogenase (HSDH) domains. In the inbred line A188, AKHSDH1 and AKHSDH2 are 38% and 88% identical at the amino acid level in the transit peptide and mature peptide coding regions, respectively (MUEHLBAUER *et al.*, 1994). We sequenced and analyzed the transit peptide-AK domain junction, a region of the AK domain, and a region of the HSDH domain from *akh1*, and a region of the AK domain and HSDH domains from *akh2*. Of the five total regions, only the AK domain of *akh1* had reduced diversity in inbred lines relative to teosinte and a significant coalescent simulation test result. Our data support the theory of gene duplication and subsequent functional divergence where one paralogous gene is freed from selective pressure, allowing it to acquire mutations that result in a novel function, or temporal or tissue specificity (OHNO, 1970). Differences in intragenic selective pressure have been previously demonstrated in the maize gene, *teosinte branched1*. In this case, the 60-90 kb promoter region was selected during domestication, while diversity levels in the coding region were consistent with neutral evolution (CLARK *et al.*, 2004). This is not unlikely in maize since LD typically decays rapidly within individual maize loci (REMINGTON *et al.*, 2001; TENAILLON *et al.*, 2001; FLINT-GARCIA *et al.*, 2003).

Selection and crop improvement

While we have identified several amino acid genes as new candidates for selection, it is apparent that the amino acid pathway, as a whole, has not been a strong target of selection as appears to be the case for maize starch accumulation (2002). Yet, for most amino acids, there are small but significant differences in content among the germplasm groups. These differences may be a result of either the free amino acids or the ratio of the storage protein classes in the kernel. If the differences are due to free amino acid content, then modification of these selected genes would have a direct impact on the amino acid content of the kernel. However, if the differences in individual amino acids are due to the storage protein profile, then selection may have acted on genes controlling the seed storage protein as previously hypothesized (WILSON and LARKINS, 1984; SWARUP *et al.*, 1995; FLINT-GARCIA *et al.*, 2009). Our results indicate that the dramatic change in total amino acid content (i.e. total protein content) and the smaller differences in individual amino acids have been uncoupled from selection for amino acid synthesis genes themselves.

Upon examining the selection results of this study and previous studies, we observed that several of the selected amino acid candidates were clustered in only small number of pathways. These results indicate that for most amino acid pathways, enough variation probably exists in maize inbred lines to enable crop improvement as has been accomplished in many plants (GALILI *et al.*, 2005).

Our results indicate a clustering of selected genes exists for methionine and lysine. The AK domain of AK-HSDH is responsible for the flow of aspartate into the methionine and lysine pathways. From there, *dihydrodipicolinate synthase* is involved in the initial step in lysine synthesis. In the methionine pathway, *sam synthase II* has a negative effect on methionine content as it converts free methionine to the methyl donor S-adenosylmethionine, while *cysteine synthase* has a direct effect in methionine synthesis as cysteine is a precursor in methionine synthesis. It may be possible to reintroduce genetic variation from landraces or teosinte for these loci to alter the kernel for these key amino acids. Indeed, a recent study showed that introgression of *Zea mays* ssp *mexicana* (a teosinte more distantly related to maize than *Z. mays* ssp. *parviglumis*) into maize yielded lines with higher protein content, as well as higher lysine, methionine, and/or phenylalanine content (WANG *et al.*, 2008).

ACKNOWLEDGMENTS - We would like to express our appreciation to Major Goodman and John Doebley for providing seed of the landrace and teosinte accessions, respectively; Joan Zhao for her insight into coalescent simulations; Christie Hovis for her assistance in candidate gene sequencing and selection analysis; and Brandon Gaut for his constructive comments and bioinformatic validation of gene expression. This research was supported by a National Science Foundation Plant Genome Award DBI0321467, by research funds provided by USDA-ARS, and by a USDA Edminster Postdoctoral Fellowship.

Names of products are necessary to report factually on available data; however, neither the USDA nor any other participating institution guarantees or warrants the standard of the product, and the use of the name does not imply approval of the product to the exclusion of others that may also be suitable.

Sequence data from this article have been deposited with the GenBank Data Libraries under accession nos. BV680562-BV681145 and BV693818-BV693972.

REFERENCES

AOAC, 1995 Official Methods of Analysis of Analysis, 15 ed. Association of Official Analytical Chemists, Washington, DC.

BRESSANI R., E.T. MERTZ, 1958 Studies on corn protein. IV. Protein and amino acid content of different corn varieties. Cereal Chem. **35:** 227-235.

CLARK R., E. LINTON, J. MESSING, J. DOEBLEY, 2004 Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proc. Natl. Acad. Sci. USA **101:** 700-707.

CORUZZI G., R. LAST, 2000 Amino acids. pp. 358-410. *In:* B. Buchanan, W. Gruissem, R. Jones (Eds.), Biochemistry and Molecular Biology of Plants. Am. Society of Plant Physiologists, Rockville, Maryland.

DARRIGUES A., K.R. LAMKEY, M.P. SCOTT, 2006 Breeding for grain amino acid composition in maize. pp. 335-344. *In:* K.R. Lamkey, M. Lee (Eds.), Plant Breeding. Arnel R. Hallauer Intl. Symposium. Blackwell Publishing, Ames, IA.

DOEBLEY J., 2004 The genetics of maize evolution. Annu. Rev. Genet. **38:** 37-59.

EWING B., P. GREEN, 1998 Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:** 186-194.

EWING B., L. HILLIER, M.C. WENDL, P. GREEN, 1998 Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:** 175-185.

EYRE-WALKER A., R.L. GAUT, H. HILTON, D.L. FELDMAN, B.S. GAUT, 1998 Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. USA **95:** 4441-4446.

FLINT-GARCIA S.A., J.M. THORNSBERRY, E.S. BUCKLER, 2003 Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54:** 357-374.

FLINT-GARCIA S.A., A.L. BODNAR, M.P. SCOTT, 2009 Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. Theor. Appl. Genet. **119:** 1129-1142.

GALILI G., R. AMIR, R. HOEFGEN, H. HESSE, 2005 Improving the levels of essential amino acids and sulfur metabolites in plants. Biol. Chem. **386:** 817-831.

GALLAVOTTI A., Q. ZHAO, J. KYOZUKA, R.B. MEELEY, M.K. RITTER, J.F. DOEBLEY, M.E. PE, R.J. SCHMIDT, 2004 The role of *barren stalk1* in the architecture of maize. Nature **432:** 630-635.

GOODMAN M., 1990 Genetic and germplasm stocks worth conserving. J Hered. **81:** 11-16.

HANSON M.A., B.S. GAUT, A.O. STEC, S.I. FUERSTENBERG, M.M. GOODMAN, E.H. COE, J.F. DOEBLEY, 1996 Evolution of anthocyanin biosynthesis in maize kernels: The role of regulatory and enzymatic loci. Genetics **143:** 1395-1407.

HILTON H., 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., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245-250.

HUDSON R.R., M. KREITMAN, M. AGUADÉ, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153-159.

HUFFORD K.M., P. CANARAN, D.H. WARE, M.D. MCMULLEN, B.S. GAUT, 2007 Patterns of selection and tissue-specific expression among maize domestication and crop improvement loci. Plant Physiol. **144:** 1642-1653.

HYTEN D.L., Q. SONG, Y. ZHU, I.-Y. CHOI, R.L. NELSON, J.M. COSTA, J.E. SPECHT, R.C. SHOEMAKER, P.B. CREGAN, 2006 Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. USA **103:** 16666-16671.

ILTIS H.H., 1983 From teosinte to maize: the catastrophic sexual transmutation. Science **222:** 886-894.

INNAN H., Y. KIM, 2004 Pattern of polymorphism after strong artificial selection in a domestication event. Proc. Natl. Acad. Sci. USA **101:** 10667-10672.

KIMURA M., 1983 The neutral theory of Molecular Evolution. Cambridge University Press, Cambridge, Massachusetts.

MUEHLBAUER G.J., D.A. SOMERS, B.F. MATTHEWS, B.G. GENGENBACH, 1994 Molecular genetics of the maize (*Zea mays* L.) aspartate kinase-homoserine dehydrogenase gene family. Plant Physiol. **106:** 1303-1312.

OHNO S., 1970 Evolution by Gene Duplication. Springer-Verlag, Berlin, Germany.

REMINGTON D.L., J.M. THORNSBERRY, Y. MATSUOKA, L.M. WILSON, S.R. WHITT, J. DOEBLEY, S. KRESOVICH, M.M. GOODMAN, E.S. BUCKLER, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA **98:** 11479-11484.

ROZAS J., J.C. SÁNCHEZ-DELBARRIO, X. MESSEGUER, R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496-2497.

SAGHAI-MAROOF M.A., K.M. SOLIMAN, R.A. JORGENSEN, R.W. ALLARD, 1984 Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA **81:** 8014-8018.

SANCHEZ-VILLEDA H., S.G. SCHROEDER, S.A. FLINT-GARCIA, K.E. GUILL, M. YAMASAKI, M.D. MCMULLEN, 2008 DNAAlignEditor: DNA Alignment editor tool. BMC Bioinformatics **9:** 154.

SCHMIDT R.J., F.A. BURR, M.J. AUKERMAN, B. BURR, 1990 Maize regulatory gene *opaque-2* encodes a protein with a "leucine-zipper" motif that binds to zein DNA. Proc. Natl. Acad. Sci. USA **87:** 46-50.

SWARUP S., M.C. TIMMERMANS, S. CHAUDHURI, J. MESSING, 1995 Determinants of the high-methionine trait in wild and exotic germplasm may have escaped selection during early cultivation of maize. Plant J. **8:** 359-368.

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.

TANG T., J. LU, J. HUANG, J. HE, S.R. MCCOUCH, Y. SHEN, Z. KAI, M.D. PURUGGANAN, S. SHI, C.-I. WU, 2006 Genomic variation in rice: genesis of highly polymorphic linkage blocks during domestication. PLoS Genet. **2:** e199.

TENAILLON M.I., M.C. SAWKINS, A.D. LONG, R.L. GAUT, J.F. DOEBLEY, B.S. GAUT, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA **98:** 9161-9166.

THOMPSON J.D., D.G. HIGGINS, T.J. GIBSON, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673-4680.

THORNSBERRY J.M., M.M. GOODMAN, J. DOEBLEY, S. KRESOVICH, D. NIELSEN, E.S. BUCKLER, 2001 *Dwarf*8 polymorphisms associate with variation in flowering time. Nature Genetics **28:** 286-289.

VIGOUROUX Y., M. MCMULLEN, C.T. HITTINGER, K. HOUCHINS, L. SCHULZ, S. KRESOVICH, Y. MATSUOKA, J. DOEBLEY, 2002 Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc. Natl. Acad. Sci. USA **99:** 9650-9655.

WANG H., T. NUSSBAUM-WAGLER, B. LI, Q. ZHAO, Y. VIGOUROUX, M. FALLER, K. BOMBLIES, L. LUKENS, J.F. DOEBLEY, 2005 The origin of the naked grains of maize. Nature **436:** 714-719.

WANG L., C. XU, M. QU, J. ZHANG, 2008 Kernel amino acid composition and protein content of introgression lines from *Zea mays* ssp. *mexicana* into cultivated maize. J. Cereal Sci. **48:** 387-393.

WANG R.L., J. HEY, 1996 The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the *period* locus. Genetics **144:** 1113-1126.

WANG R.L., A. STEC, J. HEY, L. LUKENS, 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.

WEBER A., R.M. CLARK, L. VAUGHN, J. DE JESUS SANCHEZ-GONZALEZ, J. YU, B.S. YANDELL, P. BRADBURY, J. DOEBLEY, 2007 Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). Genetics **177:** 2349-2359.

WHITT S.R., L.M. WILSON, M.I. TENAILLON, B.S. GAUT, E.S. BUCKLER, 2002 Genetic diversity and selection in the maize starch pathway. Proc. Natl. Acad. Sci. USA **20:** 12959-12962.

WILSON D.R., B.A. LARKINS, 1984 Zein gene organization in maize and related grasses. J. Mol. Evol. **20:** 330-340.

WRIGHT S.I., I. VROH BI, S.G. SCHROEDER, M. YAMASAKI, J.F. DOEBLEY, M.D. MCMULLEN, B.S. GAUT, 2005 The effects of artificial selection on the maize genome. Science **308:** 1310-1314.

YAMASAKI M., M.I. TENAILLON, I. VROH BI, S.G. SCHROEDER, H. SANCHEZ-VILLEDA, J.F. DOEBLEY, B.S. GAUT, M.D. MCMULLEN, 2005 A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell **17:** 2859-2872.

YU J., J.B. HOLLAND, M.D. MCMULLEN, E.S. BUCKLER, 2008 Genetic design and statistical power of nested association mapping in maize. Genetics **178:** 539-551.

ZHANG L., A.S. PEEK, D. DUNAMS, B.S. GAUT, 2002 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.